



**HAL**  
open science

# Contributions for the approximation and model order reduction of partial differential equations

Marie Billaud Friess

► **To cite this version:**

Marie Billaud Friess. Contributions for the approximation and model order reduction of partial differential equations: Contributions pour l'approximation et la réduction de modèle des équations aux dérivées partielles. Numerical Analysis [math.NA]. Nantes Université, 2022. tel-03826223

**HAL Id: tel-03826223**

**<https://hal.science/tel-03826223>**

Submitted on 24 Oct 2022

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

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

HABILITATION À DIRIGER DES RECHERCHES  
*Spécialité : Mathématiques Appliquées*

**Marie Billaud-Friess**

---

**Contributions for the approximation and model order reduction  
of partial differential equations**

\*\*\*\*\*

Contributions pour l'approximation et la réduction de modèle  
des équations aux dérivées partielles

---

soutenue le 21 octobre 2022, devant le jury composé de

**Albert Cohen**, *Sorbonne Université*, ..... Président  
**Virginie Ehrlicher**, *Ecole des ponts et chaussées*, ..... Examineur  
**Tony Lelièvre**, *Ecole des ponts et chaussées*, ..... Rapporteur  
**Olga Mula**, *TU Eindhoven*, ..... Examineur  
**Fabio Nobile**, *Ecole Polytechnique Fédérale de Lausanne*, ..... Rapporteur  
**Anthony Nouy**, *Centrale Nantes, Nantes Université*, ..... Examineur  
**Mario Ohlberger**, *University of Münster*, ..... Rapporteur

*Aide-toi, le Ciel t'aidera.*

LE CHARTIER EMBOURBÉ. Livre VI, fable 18.  
**Jean de la Fontaine**

# CONTENTS

<b>Introduction (version française)</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>List of publications</b>	<b>5</b>
<b>1 Low-rank approximation for parameter-dependent equations</b>	<b>7</b>
1.1 General context . . . . .	7
1.1.1 Low-rank approximation for model reduction . . . . .	7
1.1.2 Contributions . . . . .	8
1.2 Low-rank approximation in tensor format . . . . .	9
1.2.1 Classical tensor formats . . . . .	9
1.2.2 Best low-rank approximation . . . . .	10
1.2.3 Approximation of the solution of high-dimensional problems . . . . .	11
1.2.4 Numerical applications . . . . .	14
1.2.5 Summary . . . . .	16
1.3 Projection based goal oriented low-rank approximation . . . . .	16
1.3.1 Projection based low-rank approximation . . . . .	17
1.3.2 Goal oriented projections . . . . .	19
1.3.3 Numerical applications . . . . .	23
1.3.4 Summary . . . . .	25
<b>2 Dynamical low-rank approximation</b>	<b>27</b>
2.1 General context . . . . .	27
2.1.1 Low-rank approximation for parameter and time-dependent problems . . . . .	27
2.1.2 Contributions . . . . .	29
2.2 Projection based dynamical low-rank approximation . . . . .	29
2.2.1 Construction of the approximation . . . . .	30
2.2.2 Selection of sub-spaces . . . . .	30
2.2.3 Numerical applications . . . . .	31
2.2.4 Prospects for parameter-dependent conservation laws . . . . .	33
2.2.5 Summary and discussion for future research . . . . .	34
2.3 Geometry based dynamical low-rank approximation . . . . .	35
2.3.1 Geometry of the set of fixed-rank matrices . . . . .	37
2.3.2 Dynamical low-rank approximation . . . . .	38
2.3.3 Integration schemes . . . . .	39
2.3.4 Numerical applications . . . . .	41
2.3.5 Summary and discussion for future research . . . . .	43
<b>3 Approximation for PDEs with probabilistic interpretation</b>	<b>45</b>
3.1 General context . . . . .	45
3.1.1 Probabilistic approaches for approximation . . . . .	45
3.1.2 Contributions . . . . .	46
3.2 Monte Carlo methods for PDEs . . . . .	47
3.2.1 Monte Carlo methods . . . . .	47
3.2.2 Probabilistic representation of solution of PDEs . . . . .	48
3.2.3 Numerical estimation of the solution of a PDEs . . . . .	50

3.3	High dimensional PDEs . . . . .	51
3.3.1	Approximation from pointwise evaluations . . . . .	51
3.3.2	Control variate method for high dimensional PDEs . . . . .	52
3.3.3	Numerical applications . . . . .	55
3.4	Parameter-dependent PDEs . . . . .	56
3.4.1	Sample based probabilistic reduced basis method . . . . .	57
3.4.2	Probabilistic greedy algorithm . . . . .	58
3.5	Summary and discussion for future research . . . . .	60
	<b>Past research contributions</b>	<b>63</b>
	<b>Bibliography</b>	<b>67</b>

## INTRODUCTION (VERSION FRANÇAISE)

Mes travaux de recherche relèvent principalement de l'analyse numérique selon deux thématiques distinctes. Les travaux menés durant mes doctorat et post-doctorats ont porté sur le développement de méthodes numériques pour la simulation d'écoulements à plusieurs composants. A mon arrivée à Centrale Nantes, j'ai été amenée à changer de thématique pour me focaliser sur le développement et l'analyse de méthodes d'approximation et réduction de modèle pour les équations aux dérivées partielles. Dans ce manuscrit, j'ai choisi de me concentrer sur mes plus récentes contributions portant sur le second thème. Un bref aperçu de mes travaux de thèse et post-doctorats [MBF1, MBF2, MBF4, MBF3, MBF5, MBF6, MBF7] est néanmoins donné en annexe (voir page 63).

Le présent mémoire se constitue de trois chapitres. Cependant, on peut voir se dessiner deux axes suivant lesquels les contributions présentées peuvent être rangées avec d'une part des méthodes déterministes et d'autre part des approches probabilistes. Par souci de concision, les concepts utilisés seront motivés et introduits au fur et à mesure des enjeux et des objectifs soulignés.

### **Approximation de faible rang pour les problèmes dépendant de paramètres**

La première direction concerne le développement de méthodes d'approximation de faible rang pour la réduction de modèle de problèmes dépendant de paramètres.

Dans le Chapitre 1, des contributions dédiées à la résolution de tels problèmes motivés par des applications en quantification d'incertitude sont présentées. Le chapitre est divisé en deux sections selon les deux points de vue suivants: d'une part les méthodes d'approximation de faible rang sous format de tenseur et d'autre part les méthodes basées sur des projections. En particulier, une méthode d'approximation sous format de tenseur, basée sur des formulations idéales en minimum de résidu pour des problèmes en grande dimension, est d'abord proposée [MBF8, MBF9]. Puis, des méthodes basées sur des projections pour l'approximation de quantités d'intérêt à valeur fonctionnelle ou vectorielle [MBF10] sont discutées. Ces travaux ont été réalisés en collaboration avec A. Nouy et O. Zahm.

Dans le Chapitre 2, les méthodes d'approximation sous format de faible rang, discutées au chapitre précédent, sont étendues pour des problèmes paramétrés dépendant du temps. En particulier, des approches de faible rang dites dynamiques sont discutées. Premièrement, une méthode de type bases réduites dynamiques est présentée [MBF11]. La méthode proposée surpassant les approches classiques pour des problèmes de type transport, des pistes d'extension pour les systèmes de loi de conservation hyperboliques ont été explorées. Notamment, une méthode de reconstruction dans un cadre volume fini a été proposée et validée pour des problèmes de transport linéaires et dépendant de paramètres [MBF16]. Dans un second temps, une méthode de faible rang dynamique travaillant dans la variété des matrices de rang fixé est considérée. Cette approche repose sur la paramétrisation géométrique de cette variété, proposée dans [MBF13], à partir de laquelle un nouveau schéma d'intégration de type splitting a été dérivé pour l'approximation de faible rang dynamique de solutions de systèmes dynamiques matriciels [MBF14]. Ces diverses contributions ont été le fruit d'une collaboration avec T. Heuzé, A. Falcó et A. Nouy.

## Méthodes d'approximation probabilistes pour les EDP

Le deuxième axe de ce manuscrit concerne le développement de méthodes numériques pour l'approximation de la solution d'équations aux dérivées partielles (éventuellement dépendant de paramètres) en utilisant des estimations ponctuelles de la fonction recherchée.

Dans ce but, des approches probabilistes sont discutées dans le Chapitre 3. La clef de voûte des contributions présentées est le théorème de représentation de Feynman-Kac qui permet d'explicitier la solution d'une équation aux dérivées partielles, évaluée ponctuellement, comme l'espérance d'une fonctionnelle stochastique. En utilisant cet outil, une méthode probabiliste d'interpolation parcimonieuse a été proposée pour des équations en grande dimension [MBF17]. Par ailleurs, une méthode de bases réduites probabiliste basée sur des estimations ponctuelles est proposée [MBF17] pour la résolution d'EDP paramétrées. En particulier, une telle méthode s'appuie sur une procédure glouton avec un estimateur d'erreur probabiliste combiné avec un algorithme de type bandit adaptatif "probably approximately correct" [MBF15]. Ces travaux ont été réalisés en collaboration avec A. Macherey, A. Nouy et C. Prieur.

## INTRODUCTION

My research fits in the domain of numerical analysis. The work done during my PhD and post-doctorates concerned the development of numerical methods for the simulation of multi-component flows. Since I arrived in Centrale Nantes, my research work has mainly focused on the development of approximation and model order reduction methods for partial differential equations. In this manuscript, I have made the choice to not present in detail the work [MBF1, MBF2, MBF4, MBF3, MBF5, MBF6, MBF7] done before arriving in Centrale Nantes. Only a concise summary is provided in appendix (see page 63).

This document is mainly composed of three chapters, following two principal directions with deterministic and probabilistic approximation methods.

### **Low-rank approximation for parameter-dependent problems**

The first direction concerns the development of suitable low-rank approximation methods for model order reduction of parameter-dependent problems.

In Chapter 1, contributions devoted to the resolution of such problems motivated by uncertainty quantification are presented. The chapter is divided in two parts following a natural presentation of the proposed approaches within the framework of low-rank approximation mainly : low-rank approximation in tensor format and projection based low-rank approximation for parameter-dependent problems. More particularly tensor approximation methods based on ideal minimal residual formulations for the solution of high-dimensional problems [MBF8, MBF9] are first proposed. Then, goal oriented projection based model order reduction methods for the estimation of vector-valued variables of interest [MBF10] are discussed. These contributions were done in collaboration together with A. Nouy and O. Zahm.

In Chapter 2, low-rank approximation methods discussed in Chapter 1 are extended for the approximation of time and parameter-dependent problems. Here, dynamical low-rank approximation methods are discussed following two subsequent directions. First, dynamical reduced basis method for parameter-dependent dynamical systems that can be interpreted as a dynamical low-rank approximation approach with a subspace point of view is presented [MBF11]. As the proposed method is shown to perform better than usual reduced basis for transport (dominated) problems, possible extension to parameter-dependent hyperbolic conservation laws is discussed. In this direction a reconstruction method in finite volume setting has been developed and validated for linear parameter-dependent transport problems [MBF16]. Secondly, we consider a dynamical low-rank approximation method that works directly in the set of fixed rank matrices. Based on a suitable geometric parametrization of this set [MBF13], a new splitting integration scheme for the approximation of the solution of matrix dynamical systems [MBF14], arising from discretization of parameter-dependent dynamical systems, has been proposed. These works were the result of a joint work with T. Heuzé, A. Falcó and A. Nouy.

## Probabilistic approximation methods for PDEs

The second direction deals with the problem of constructing an approximation of the solution of a partial differential equation (possibly parameter-dependent) using approaches that rely on sampled estimates of the function to approximate.

To that goal, probabilistic approaches are presented in Chapter 3. The main corner stone of the proposed contributions is the Feynman-Kac representation formula of the solution of a partial differential equation. Using this key ingredient, a probabilistic sparse polynomial interpolation method has been proposed to deal with high dimensional problem [MBF12]. Then reduced basis method using pointwise estimates is discussed [MBF17] for solving parameter-dependent partial differential equations. Especially, it relies on a greedy procedure with a probabilistic error estimate gathered with a probably approximately correct bandit algorithm proposed in [MBF15]. These contributions were done in collaboration with A. Macherey, A. Nouy and C. Prieur.

## LIST OF PUBLICATIONS

Here is a summary of my publications listed in two categories, depending on whether the topics have been addressed during my PhD and post-doctorates, or since I am associate professor at Centrale Nantes.

*My publications are cited thorough the manuscript using the reference [MBF\*].*

### Work related to PhD and post-doctorates

- [MBF1] M. Billaud, *Stabilized finite element method for incompressible-compressible two-phase flows*, PhD thesis, Université Sciences et Technologies - Bordeaux I, [tel-00565815](tel:00565815), 2009.
- [MBF2] M. Billaud, G. Gallice and B. Nkonga, *Stabilized Finite Element Method for Compressible-Incompressible Diphasic Flows*. In: Kreiss G., Lötstedt P., Målqvist A., Neytcheva M. (eds) *Numerical Mathematics and Advanced Applications 2009*. Springer, Berlin, Heidelberg, 2010.
- [MBF3] M. Billaud, G. Gallice and B. Nkonga, *A simple stabilized finite element method for solving two phase compressible-incompressible interface flows*. *Comput. Methods Appl. Mech. Eng.*, 200(9), 1272–1290, 2011.
- [MBF4] M. Billaud-Friess, B. Boutin, F. Caetano, G. Faccanoni, S. Kokh, F. Lagoutière and L. Navoret, *A second order anti-diffusive Lagrange-remap scheme for two-component flows*. *ESAIM Proc.*, 32, 149–162, 2011.
- [MBF5] M. Billaud-Friess, and S. Kokh, *An anti-diffusive Lagrange-Remap scheme for multi-material compressible flows with an arbitrary number of components*. *ESAIM Proc.*, 35, 203–209, 2012.
- [MBF6] M. Billaud-Friess, J. Breil, P.-H. Maire and M. Shashkov, *A Multi-Material CCALE-MOF Approach in Cylindrical Geometry*. *Comm. Comput. Phys.*, 15(2), 330–364, 2014.
- [MBF7] M. Billaud-Friess, and S. Kokh, *Simulation of sharp interface multi-material flows involving an arbitrary number of components through an extended five-equation model*. *J. Comput. Phys.*, 273, 488–519, 2014.

### Work related to topics covered since I am at Centrale Nantes

- [MBF8] M. Billaud-Friess, A. Nouy and O. Zahm, *Méthode de réduction de modèle a priori basée sur des formulations idéales en minimum de résidu*. *Actes CSMA 2013*, 2013.
- [MBF9] M. Billaud-Friess, A. Nouy, and O. Zahm, *A tensor approximation method based on ideal minimal residual formulations for the solution of high-dimensional problems*. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 48(6), 1777–1806, 2014.

- [MBF10] Zahm, O., M. Billaud-Friess, and A. Nouy, *Projection-Based Model Order Reduction Methods for the Estimation of Vector-Valued Variables of Interest*. SIAM J. Sci. Comput., 39(4), A1647–A1674, 2017.
- [MBF11] M. Billaud-Friess, and A. Nouy, *Dynamical Model Reduction Method for Solving Parameter-Dependent Dynamical Systems*. SIAM J. Sci. Comput., 39(4), A1766–A1792, 2017.
- [MBF12] M. Billaud-Friess, A. Macherey, A. Nouy and C. Prieur, *Stochastic Methods for Solving High-Dimensional Partial Differential Equations*. Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2018. Springer Proceedings in Mathematics and Statistics, vol 324. Springer, Cham, 2020.
- [MBF13] M. Billaud-Friess, A. Falcó, and A. Nouy, *Principal bundle structure of matrix manifolds*. MDPI, Mathematics, 9(14), 1669, 2021.
- [MBF14] M. Billaud-Friess, A. Falcó, and A. Nouy, *A new splitting algorithm for dynamical low-rank approximation motivated by the fibre bundle structure of matrix manifolds*. Bit Numer. Math., 1–22, 2021.
- [MBF15] M. Billaud-Friess, A. Macherey, A. Nouy and C. Prieur, *A PAC algorithm in relative precision for bandit problem with costly sampling.*, Mathematical Methods of Operations Research, 1–25, 2022.
- [MBF16] M. Billaud-Friess, and T. Heuzé, *Reconstruction of finite volume solution for parameter-dependent linear hyperbolic conservation laws*. arXiv, 2006.1035 (preprint)
- [MBF17] M. Billaud-Friess, A. Macherey, A. Nouy and C. Prieur, *A probabilistic reduced basis method for parameter-dependent problems*. (in preparation)

# Chapter 1

## LOW-RANK APPROXIMATION FOR PARAMETER-DEPENDENT EQUATIONS

*Parameter-dependent problems are ubiquitous in large variety of applications such as uncertainty quantification, sensitivity, inverse problems, control or optimization. This chapter presents contributions devoted to the resolution of such problems with application in uncertainty quantification using low-rank approximation methods.*

*The chapter is divided in two parts following a natural presentation of the proposed approaches within the framework of low-rank approximation : low-rank approximation in tensor format and projection based low-rank approximation for parameter-dependent problems. After a brief introduction on low-rank tensor approximation methods for model order reduction of parameter-dependent problems in Section 1.1, we present in Section 1.2 a tensor approximation method based on ideal minimal residual formulation for the solution of high-dimensional problems. Then, in Section 1.3, we propose projection based methods for the estimation of vector-valued variables of interest.*

### 1.1 General context

In this chapter, we are concerned with the resolution of high-dimensional partial differential equations (PDEs) arising in physics or stochastic calculus, or the solution of parameter-dependent or stochastic equations. The following general linear parameter-dependent equation is considered

$$A(\xi)u(\xi) = b(\xi). \tag{1.1.1}$$

Here,  $A$  is a linear operator, depending on the (random) parameters  $\xi = (\xi_1, \dots, \xi_p)^T$  that take their values in some set  $\Xi \subset \mathbb{R}^p$  (with  $p$  possibly large). The parameter-dependent solution is  $u(\xi)$  which belongs to some vector space  $V$ .

#### 1.1.1 Low-rank approximation for model reduction

Computing a solution  $u$  to the problem (1.1.1) rises various difficulties, related to the complexity and cost of the computations, when one wants to access to pointwise evaluations of  $u$  over  $\Xi$  or is interested in having a global approximation of the *solution map*  $u : \Xi \rightarrow V$ , in particular in high-dimension (when  $p$  is large). As discussed in what follows, *low-rank approximation methods* are nowadays widely used for the solution of problem (1.1.1) in these contexts. We refer e.g. the reader to the surveys [91] or [134, 135].

In a multiquery context, one has to evaluate the solution  $u$  of (1.1.1) for many instances of the parameter  $\xi$ . For complex numerical model, arising for example from the standard linear approximation of a PDEs with  $\dim(V) = n$  ( $n$  large), the computation of the solution for one instance of the parameter can be very costly. To tackle this problem, *Model Order Reduction (MOR)* methods have emerged these last decades. The idea of MOR is to provide an accurate surrogate of the solution map  $u : \Xi \rightarrow V$  cheaper to compute, and allowing rapid evaluation of

the solution or quantities of interest. Such methods usually rely on *linear approximation* since there are aimed at approximating  $u$  as follows

$$u(\xi) \approx \sum_{i=1}^r \alpha_i(\xi) v_i, \quad (1.1.2)$$

where  $\{v_1, \dots, v_r\}$  are elements in  $V$  and  $\{\alpha_1, \dots, \alpha_r\}$  are elements of a some vector space  $S$  of functions defined on  $\Xi$ . Such an approximation can be interpreted as a *rank- $r$  approximation* in the tensor space  $X := V \otimes S$ . Standard MOR methods aim at finding an approximation under the form (1.1.2) with a small number of terms  $r \ll n$ . To compute an approximation under the form (1.1.2) we can distinguish two kinds of approaches depending on either they rely on constructing a basis  $\{v_i\}_{i=1}^r \subset V$ , or directly computing an approximation under the form (1.1.2) in a low-rank subset of  $X$ . The approaches of the first kind are called *projection based low-rank approximation methods*, since the approximation is obtained by a projection (e.g. Galerkin, Least-Square, interpolation ...) of  $u$  in the subspace of  $V$  spanned by a basis selected a priori. Classical projection based low-rank methods include the Proper Orthogonal Decomposition (POD) method, the Empirical Interpolation Method (EIM) or the Reduced Basis (RB) method. The approaches of the second kind include *low-rank tensor* methods that rely on direct optimization in low-rank manifold. Classical examples are the best  $r$ -term approximation that coincides with the rank- $r$  truncated Singular Value Decomposition (SVD), as well as the Proper Generalized Decomposition (PGD) methods. All, these methods differ by their selection of the reduced basis for different controls of the error over the parameter set (uniform control for EIM and RB or control in mean-square sense for POD and PGD see, e.g., [134]).

Computing a global approximation of the solution  $u$  as a function of multiple parameters  $\xi = (\xi_1, \dots, \xi_p) \in \Xi$ , is challenging in particular for high-dimensional problems. A widespread idea is to construct some suitable approximation of the solution map using a *separated representation* under the form

$$u(\xi) \approx \sum_{\nu_1=1}^{r_1} \cdots \sum_{\nu_p=1}^{r_p} a_{\nu_1, \dots, \nu_p} \phi_{\nu_1}^1(\xi_1) \cdots \phi_{\nu_p}^p(\xi_p). \quad (1.1.3)$$

Seeking an approximation in a tensorized basis is not fully satisfactory because of the so-called *curse of dimensionality* [20]. In brief, the storage or computational complexity for solving (1.1.1) grows exponentially in the parametric dimension  $p$ , i.e., it has the form  $\rho^p$  with  $\rho > 1$ . The last decades significant research effort has been done in order to develop *tensor formats* whose complexity increases slowly with the number of parameters  $p$ . In this manuscript, we focus on low-rank approximations in tensor format that aims at reducing the complexity by exploiting (high-order) low-rank structures of multivariate functions, considered as elements of tensor product spaces. Classical low-rank tensor formats include canonical tensor formats, Tucker formats, Tensor Train formats [104, 140], Hierarchical Tucker formats [102] or more general tree-based Hierarchical Tucker formats [77]. Other approaches are also available for the construction of structured approximation of the solution map. Among them, let mention sparse tensor methods that consist in searching an approximation of the form (1.1.3) with only a few non-zero terms. To this end, the idea is to properly select a small number of functions in a set of functions as e.g. multivariate polynomial [49, 50, 52].

## 1.1.2 Contributions

The contributions presented in this chapter are mainly related to the PhD of O. Zahm (INRIA, Grenoble-Rhône Alpes) [167] supervised together with A. Nouy (LMJL, ECN). In this work, we have dealt with the problem of low-rank approximation in the context of uncertainty quantification addressing the following points.

- Generally, the solution  $u$  is unknown and only given implicitly through the equation (1.1.1). As a consequence practical low-rank approximation methods often use the residual of (1.1.1) measured with a certain norm. Such methods are not able to provide optimal approximations, but only quasi-optimal approximations in the best case. This loss of accuracy can be problematic for the efficiency of the methods.
- In many applications (e.g. motivated by uncertainty quantification), we only need to compute a partial information, called a *quantity of interest* (QoI), which is a function of the solution map. In particular, the main challenge is to develop goal-oriented low-rank approximation methods which provide accurate approximation of QoI while reducing the computational complexity.

## 1.2 Low-rank approximation in tensor format

We recall briefly in Section 1.2.1 general tensor formats for approximation of a multivariate function  $u$  belonging to an Hilbert tensor space and related best low-rank approximation problem in Section 1.2.2. In Section 1.2.3 we rise difficulties of classical residual based approximation methods when  $u$  is a tensor solution of a PDEs formulated in tensor Hilbert space. Then ideal minimal residual formulation, introduced in [MBF8, MBF9] for the solution of high-dimensional problems related low-rank approximation method, is presented in Section 1.2.4. Numerical illustrations of the proposed approach are given in Section 1.2.4.

*Notations.* Let  $X_\mu$ ,  $1 \leq \mu \leq d$  be Hilbert spaces equipped with inner products  $\langle \cdot, \cdot \rangle_\mu$  and associated norms  $\| \cdot \|_\mu$ . In what follows, we denote by  $\otimes_{\mu=1}^p v^\mu = v^1 \otimes \dots \otimes v^p$ ,  $v^\mu \in X_\mu$ , an elementary tensor. A Hilbert tensor space  $X$  (see [103] for more details) equipped with the norm  $\| \cdot \|_X$  is obtained by the completion with respect to  $\| \cdot \|_X$  of the algebraic tensor space

$$X = \|\cdot\|_X \bigotimes_{\mu=1}^p X_\mu = \overline{\text{span}\{\otimes_{\mu=1}^p v^\mu : v^\mu \in X_\mu, 1 \leq \mu \leq p\}}^{\|\cdot\|_X}.$$

A natural inner product on  $X$  is induced by inner products  $\langle \cdot, \cdot \rangle_\mu$  in  $X_\mu$ ,  $1 \leq \mu \leq d$ . It is defined by  $\langle v, w \rangle_X = \prod_{\mu=1}^p \langle v^\mu, w^\mu \rangle_\mu$  for  $v, w$  two elementary tensors. Finally, we denote by  $X'$  the topological dual of  $X$  and by  $\langle \cdot, \cdot \rangle_{X', X}$  the duality pairing between  $X$  and  $X'$ . We denote by  $R_X : X \rightarrow X'$  the Riesz isomorphism.

### 1.2.1 Classical tensor formats

Here, we give some brief summary of the most classical tensor formats used in the litterature. The reader is referred to the textbook [103] and survey [91] for a more exhaustive review of the different tensor formats and their mathematical properties. In this section, for the sake of presentation, we assume that  $\dim(X_\mu) = n$ ,  $\mu = 1, \dots, p$ .

- The set of tensors in *canonical format* with rank lower than  $r \in \mathbb{N}$  is defined as the set

$$\mathcal{T}_r^{\text{can}}(X) = \left\{ v = \sum_{i=1}^r \otimes_{\mu=1}^p v_i^\mu : v_i^\mu \in X_\mu \right\}.$$

Here,  $r$  is the *canonical rank* of the tensor  $v$  in  $X$ , i.e. the smallest integer such that  $v$  can be written under the form  $\sum_{i=1}^r \otimes_{\mu=1}^p v_i^\mu$ . The canonical tensor format has a storage complexity in  $O(rnp)$ .

- The set of tensors in *Tucker format* with *multilinear rank* bounded by  $\mathbf{r} = (r_1, \dots, r_p) \in \mathbb{N}^p$  is given by

$$\mathcal{T}_{\mathbf{r}}^{tuc}(X) = \left\{ v = \sum_{i_1=1}^{r_1} \dots \sum_{i_p=1}^{r_p} \alpha_{i_1, \dots, i_p} \otimes_{\mu=1}^p v_{i_\mu}^\mu : v_{i_\mu}^\mu \in X_\mu, \alpha_{i_1, \dots, i_p} \in \mathbb{R} \right\}$$

In this format  $\alpha \in \mathbb{R}^{r_1 \times \dots \times r_p}$  is a tensor of order  $p$ . An approximation  $v$  in Tucker format, can be interpreted as an approximation in the tensor space  $U_1 \otimes \dots \otimes U_p$  where  $U_\mu = \text{span}(v_1^\mu, \dots, v_{r_p}^\mu)$  is a  $r_p$ -dimensional subspace of  $X_\mu$ . Actually, if the multilinear rank is  $\mathbf{r} = (r, \dots, r)$  (with  $r = \max_\mu(r_\mu)$ ) the storage complexity is in  $O(r^p + nrp)$ , which limits the applicability of the Tucker format for very large values of  $p$ .

- To overcome such a limitation due to storage complexity, *Tree-based (or Hierarchical) Tucker formats* [76,77,102] based on a notion of rank associated with a dimension partition tree have been introduced. As a particular example, let mention the set of tensor in *Tensor Train (TT) format* [140] with *TT-rank* bounded by  $\mathbf{r} = (r_1, \dots, r_{p-1}) \in \mathbb{N}^{p-1}$

$$\mathcal{T}_{\mathbf{r}}^{TT}(X) = \left\{ v = \sum_{i_1=1}^{r_1} \dots \sum_{i_p=1}^{r_p} v_{1,i_1}^1 \otimes v_{i_1,i_2}^2 \otimes \dots \otimes v_{i_{p-1},i_p}^{p-1} : v_{i_\mu-1,i_\mu}^\mu \in X_\mu \right\}$$

Let  $\mathbf{r} = (r, \dots, r)$  (with  $r = \max_\mu(r_\mu)$ ) be the the TT-rank. In comparison to the Tucker format, TT format has a storage of lower complexity in  $O(pr^2n)$  (which does not grow exponentially with  $p$ ).

## 1.2.2 Best low-rank approximation

Low-rank tensor approximation methods consist in searching for approximations in a subset of tensors of  $X$  noted

$$\mathcal{M}_{\leq r}(X) = \{v \in X : \text{rank}(v) \leq r\},$$

where the notion of  $\text{rank}(v)$  depends of the tensor format used for the approximation  $v$ , as previously discussed (i.e. canonical rank, multilinear rank or TT-rank). Then, the best approximation of  $u \in X$  in a given set of tensor format  $\mathcal{M}_{\leq r}$  with respect to the norm  $\|\cdot\|_X$  is

$$\min_{v \in \mathcal{M}_{\leq r}(X)} \|u - v\|_X. \quad (1.2.1)$$

### Existence of a solution

Low-rank tensor subsets  $\mathcal{M}_{\leq r}$  are neither linear subspaces nor convex sets. However, they usually satisfy topological and geometrical properties that make the best approximation problem (1.2.1) meaningful

$$\mathcal{M}_{\leq r}(X) \text{ is weakly closed (or simply closed in finite dimension),} \quad (1.2.2)$$

$$\mathcal{M}_{\leq r}(X) \subset \gamma \mathcal{M}_{\leq r}(X) \text{ for all } \gamma \in \mathbb{R}. \quad (1.2.3)$$

Property (1.2.3) is satisfied by all the classical tensor subsets (i.e.  $\mathcal{T}_{\mathbf{r}}^{can}(X)$ ,  $\mathcal{T}_{\mathbf{r}}^{tuc}(X)$  or  $\mathcal{T}_{\mathbf{r}}^{TT}(X)$ ). Property (1.2.2) ensures the existence of minimum in the set  $\mathcal{M}_{\leq r}(X)$  to problem (1.2.1). This property, under some suitable conditions on the norm  $\|\cdot\|_X$  (which is naturally satisfied in finite dimension), is verified by most tensor subsets used for approximation. For the the set  $\mathcal{M}_{\leq r}(X)$  with the notion of canonical rank for  $p \geq 3$  the property (1.2.2) is only fullfilled for the

set of elementary tensors  $\mathcal{M}_{r \leq 1}(X)$  [74]. Subsets of tensors associated with the notions of multilinear rank or TT-rank have better properties. They are closed sets and they are differentiable manifolds [76, 104, 161]. This has useful consequences in practice for optimization [161] or for the projection of dynamical systems in these manifolds (see, e.g., [121] and Chapter 2).

### Practical aspects

For the case  $p = 2$ , all classical low-rank tensor formats coincide with the canonical format so that  $\mathcal{M}_{\leq r}(X) = \mathcal{T}_r^{can}(X)$  for some rank  $r$ . When the norm  $\|\cdot\|_X$  is the canonical norm, then  $u_r$  coincides with a rank- $r$  SVD of  $u$ . Efficient algorithms for computing the SVD can therefore be applied. For  $d > 2$  and when the norm  $\|\cdot\|_X$  is the canonical norm, different algorithms based on optimization methods have been proposed for the different tensor formats (see, e.g., [72, 105] or [103] for a recent review). Very efficient algorithms based on higher order SVD have also been proposed in [61], [90] and [139], respectively for Tucker, Hierarchical Tucker and Tensor Train tensors. Note that these algorithms provide quasi-best approximations with a quasi-optimality factor bounded by a function of the dimension  $d$ . For a general norm  $\|\cdot\|_X$ , the computation of a global optimum to the best approximation problem is still an open problem for all tensor subsets, and methods based on SVD cannot be applied anymore. However, classical optimization methods can still be applied (such as Alternating Least Square (ALS)) in order to provide an approximation of the best approximation [72, 148, 160].

### 1.2.3 Approximation of the solution of high-dimensional problems

Given  $X$  (resp.  $Y$ ) a Hilbert tensor space equipped with inner product  $\langle \cdot, \cdot \rangle_X$  (resp.  $\langle \cdot, \cdot \rangle_Y$ ) and associated norm  $\|\cdot\|_X$  (resp.  $\|\cdot\|_Y$ ), we consider the following general problem

$$Au = b, \quad u \in X. \quad (1.2.4)$$

with  $A \in \mathcal{L}(X, Y')$ . We assume that  $A$  is weakly coercive operator satisfying the assumptions of the Nečas' theorem (see, e.g., [71, Chapter 2] for details). In particular, we have for all  $v \in X$ ,

$$\alpha \|v\|_X \leq \|Av\|_{Y'} \leq \beta \|v\|_X, \quad (1.2.5)$$

The constants  $\alpha, \beta$  stand for the weakly coercive and continuity constants of the operator  $A$  defined by

$$\inf_{0 \neq v \in X} \sup_{0 \neq w \in Y} \frac{\langle Av, w \rangle_{Y', Y}}{\|v\|_X \|w\|_Y} := \alpha > 0 \quad \text{and} \quad \sup_{0 \neq v \in X} \sup_{0 \neq w \in Y} \frac{\langle Av, w \rangle_{Y', Y}}{\|v\|_X \|w\|_Y} := \beta < \infty,$$

Under these assumptions the problem (1.2.4) is well-posed, and the *condition number* of the operator  $A$  is

$$\kappa(A) = \frac{\beta}{\alpha} \geq 1.$$

### Usual algorithms

Here, we want to compute an approximation in  $\mathcal{M}_{\leq r}(X)$  of the solution  $u \in X$  of the problem (1.2.4). As  $u$  is not explicitly given, the best approximation problem (1.2.1) cannot be solved directly. Low-rank approximation methods in tensor format typically rely on the definition of approximations based on the residual  $Au - b$ , which is a computable quantity. Different strategies have been proposed for the construction of low-rank approximations of the solution

of equations in tensor format.

The first family of methods rely on classical iterative algorithms for linear or nonlinear systems of equations with low-rank tensor algebra (using low-rank tensor compression) for standard algebraic operations [13, 111, 116, 126]. The second family of methods consists in directly computing an approximation of  $u$  in  $\mathcal{M}_{\leq r}(X)$  by minimizing some residual norm [24, 65, 131]

$$\min_{v \in \mathcal{M}_{\leq r}(X)} \|Av - b\|_{Y'}, \quad (1.2.6)$$

for some choice of the norm  $\|\cdot\|_{Y'}$ . To that goal, constructive approaches well known as *greedy algorithms* have been proposed (see [155] and monograph [156]). Such algorithms, often referred as the PGD [4, 117, 131] consist in computing successive corrections in a small low-rank tensor subset, typically the set of rank-1 canonical tensors. A step  $r$  of this algorithm can be summarized as follows:

$$w_r \in \arg \min_{w \in \mathcal{T}_1^{can}(X)} \|A(u_{r-1} + w) - b\|_{Y'}, \quad (1.2.7)$$

$$u_r = u_{r-1} + w_r. \quad (1.2.8)$$

where  $u_r$  stands for a rank- $r$  approximation in  $\mathcal{T}_r^{can}(X)$ . In practice, the rank-one approximation that defines the correction  $w_r$  can be solved using ALS algorithm. These greedy algorithms have been analyzed in several papers [5, 36, 37, 73, 74, 81] and a series of improved algorithms have been introduced in order to increase the quality of suboptimal greedy constructions [73, 83, 118, 132, 133].

### Good choice of the residual norm

Although minimal residual based approaches are well founded, they generally provide low-rank approximations that can be very far from optimal approximations (1.2.1), with respect to the natural norm  $\|\cdot\|_X$ , when using usual measures of the residual. Indeed, for such a choice, the operator  $A$  is often badly conditioned. If we are interested in obtaining an optimal approximation with respect to the norm  $\|\cdot\|_X$ , it implies to choose the residual norm [51, 58] such that

$$\|Av\|_{Y'} = \|v\|_X, \quad v \in X, \quad (1.2.9)$$

in (1.2.6) to recover the best approximation problem (1.2.1). This implies

$$I_X = R_X^{-1} A^* R_Y^{-1} A \Leftrightarrow R_Y = A R_X^{-1} A^* \Leftrightarrow R_X = A^* R_Y^{-1} A, \quad (1.2.10)$$

thus

$$\|w\|_Y = \|A^* w\|_{X'}, \quad (1.2.11)$$

where  $A^* : Y \rightarrow X'$  stands for the adjoint operator of  $A$ . This choice implies that  $\alpha = \beta = 1$ , and therefore

$$\kappa(A) = 1,$$

meaning that problem (1.2.4) is ideally conditioned. Imposing (1.2.9) can be interpreted as applying implicitly some preconditioning operator  $R_X^{-1} A^* R_Y^{-1}$  to  $A$ . Let us mention, that it is also possible to construct a priori an explicit preconditioner. See, e.g., [64] for stochastic problems and more recently [168] for parameter dependent algebraic problems. In the latter, the construction of an explicit parameter dependent preconditioner is obtained by interpolation of the inverse operator.

## Ideal minimal residual method

For solving (1.2.6), we have proposed in [MBF9] the following gradient-type algorithm. Letting  $u^0 = 0$ , we construct a sequence  $\{u^k\}_{k \geq 0}$  in  $\mathcal{M}_{\leq r}(X)$  and a sequence  $\{y^k\}_{k \geq 0}$  in  $Y$  defined for  $k \geq 0$  by

$$\begin{cases} y^k = R_Y^{-1}(Au^k - b) \\ u^{k+1} \in \Pi_r(u^k - \rho R_X^{-1}A^*y^k) \end{cases} \quad (1.2.12)$$

with  $\rho > 0$ . Here,  $\Pi_r : X \rightarrow \mathcal{M}_{\leq r}(X)$  denotes the set-valued map that associates to an element  $u \in X$  the set of best approximations of  $u$  in  $\mathcal{M}_{\leq r}$ :

$$\Pi_r(u) = \arg \min_{v \in \mathcal{M}_{\leq r}(X)} \|u - v\|_X. \quad (1.2.13)$$

The algorithm defined by (1.2.12) can be interpreted as a preconditioned gradient algorithm with an implicit preconditioner. Let us observe at step  $k$  we have

$$u^{k+1} \in \Pi_r(u^k - \rho R_X^{-1}A^*R_Y^{-1}(Au^k - b))$$

where the residual is ideally preconditioned. The resulting method is then referred as *Ideal Minimal Residual* (IMR) method. We have the following convergence result [MBF9, Proposition 4.3].

**Proposition 1.2.1.** *Assuming that  $\gamma(\rho) := \max\{|1 - \rho\beta^2|, |1 - \rho\alpha^2|\} < 1/2$ , the sequence  $\{u^k\}_{k \geq 1}$  defined by (1.2.12) is such that*

$$\|u^k - u\|_X \leq (2\gamma)^k \|u^0 - u\|_X + \frac{1}{1 - 2\gamma} \|u - \Pi_r(u)\|_X \quad (1.2.14)$$

and

$$\limsup_{k \rightarrow \infty} \|u^k - u\|_X \leq \frac{1}{1 - 2\gamma} \|u - \Pi_r(u)\|_X \quad (1.2.15)$$

The condition  $\gamma(\rho) < 1/2$  imposes  $\frac{\beta}{\alpha} < \sqrt{3}$  and  $\rho \in (\frac{1}{2\alpha^2}, \frac{3}{2\beta^2})$  which is a very restrictive condition, generally not satisfied without an excellent preconditioning of the operator  $A$ . Thus, for the ideal choice of norms (1.2.11), we have  $\alpha = \beta = 1$ . Then, we have convergence for all  $\rho \in [\frac{1}{2}, \frac{3}{2}]$  towards a neighborhood of  $\Pi_r(u)$  of size  $\frac{2\gamma}{1-2\gamma} \|u - \Pi_r(u)\|_X$  with  $\gamma = |1 - \rho|$ . Moreover, if  $\rho = 1$ , then  $u^1 \in \Pi_r(u)$ , the algorithm converges in one iteration for any initialization  $u^0$ .

## Approximate Ideal Minimal Residual method

The practical implementation of the IMR method rises some difficulties discussed as follows.

1. The computation of the iterate  $y^k$  requires to solve an auxiliary problem that involves the Riesz map inverse  $R_Y^{-1}$  such that  $R_Y = AR_X^{-1}A^*$ . It is equivalent to compute the initial operator inverse  $A^{-1}$ . To overcome this difficulty, following [51], we use instead  $\Lambda^\delta(r^k)$  an approximation of the residual  $r^k = R_Y^{-1}(Au^k - b) \in Y$  with a controlled relative precision  $\delta > 0$ . Here, we have introduced  $\Lambda^\delta : Y \rightarrow Y$  the approximation map such that

$$\|\Lambda^\delta(y) - y\|_Y \leq \delta \|y\|_Y, \quad y \in \{R_Y^{-1}(Av - b); v \in \mathcal{M}_{\leq r}(X)\}. \quad (1.2.16)$$

2. Moreover, we may not be able to solve the best approximation problem in  $\mathcal{M}_{\leq r}(X)$  exactly, because there is no available algorithm for computing a global optimum, or because the algorithm has been stopped at a finite precision. Here, we introduce a set of quasi-optimal approximations  $\Pi_r^\eta(u) \subset \mathcal{M}_{\leq r}(X)$  such that

$$\|u - \Pi_r^\eta(u)\|_X \leq \eta \|u - \Pi_r(u)\|_X \quad (\eta \geq 1). \quad (1.2.17)$$

Using the two approximation maps defined by (1.2.16)-(1.2.17), we have proposed a perturbed version of the ideal gradient type algorithm amenable for numerical use. Letting  $u^0 = 0$ , we construct two sequences  $\{u^k\}_{k \geq 0} \subset \mathcal{M}_{\leq r}$  and  $\{y^k\}_{k \geq 0} \subset Y$  defined for  $k \geq 0$  by

$$\begin{cases} y^k = \Lambda^\delta(R_Y^{-1}(Au^k - b)) \\ u^{k+1} \in \Pi_r^\eta(u^k - R_X^{-1}A^*y^k) \end{cases} \quad (1.2.18)$$

The resulting algorithm can be interpreted as a preconditioned gradient algorithm with an implicit preconditioner that approximates the ideal preconditioner. The algorithm works into steps. At the first step, an auxiliary problem provides an approximation of the residual  $R_Y^{-1}(Au^k - b)$  up to precision  $\delta$ . Then  $u^k$  is computed as a quasi-optimal approximation in  $\mathcal{M}_{\leq r}(X)$  of the update  $u^k - R_X^{-1}A^*y^k$ . The resulting low-rank approximation approach is called *Approximate Ideal Minimal Residual* (A-IMR) method. In practice, these two steps are performed using practical low-rank approximation algorithms discussed previously as e.g. SVD, ALS or PGD. For a detailed presentation of the method see [MBF9, Section 5]. Again we have the following convergence result [MBF9, Proposition 5.3] for the perturbed gradient type algorithm.

**Proposition 1.2.2.** *Assume (1.2.9), (1.2.16), and (1.2.17), with  $\delta(1+\eta) < 1$ . Then, the sequence  $\{u^k\}_{k \geq 1}$  defined by (1.2.18) is such that*

$$\|u^k - u\|_X \leq ((1 + \eta)\delta)^k \|u^0 - u\|_X + \frac{\eta}{1 - \delta(1 + \eta)} \|u - \Pi_r(u)\|_X. \quad (1.2.19)$$

Note that (1.2.19) implies that

$$\|u - \Pi_r(u)\|_X \leq \|u - u^k\|_X \leq (1 + \gamma_k) \|u - \Pi_r(u)\|_X, \quad (1.2.20)$$

This proposition states that the sequence  $\{u^k\}_{k \geq 0}$  converges towards a neighborhood of the best approximation of  $u$  in  $\mathcal{M}_{\leq r}(X)$ . The error depends on  $\gamma_k$  such that  $\limsup_{k \rightarrow \infty} \gamma_k \leq c(\eta, \delta)$  where  $c(\eta, \delta)$  depends on the parameters  $\delta$  and  $\eta$  that control the quality of this approximation.

**Remark 1.2.3.** *We have proposed in a [MBF9, Section 6] a (weak) greedy algorithm for the adaptive construction of an approximation of the solution of problem (1.2.4). In particular, we use the A-IMR method for the computation of the greedy corrections in rank-1 tensor subset (see Equation (1.2.7)). A convergence proof of this algorithm is provided under some conditions on the parameters  $\delta, \eta$ .*

## 1.2.4 Numerical applications

We illustrate the behavior of the A-IMR method for the numerical solution of steady reaction-advection-diffusion stochastic problem on a two-dimensional unit square domain  $D = [0, 1]^2$

$$\begin{aligned} -\nabla(\xi) \cdot (\kappa(\xi)\nabla u(\xi)) + c(\xi) \cdot \nabla u(\xi) + a(\xi)u(\xi) &= f \quad \text{in } D, \\ u &= 0 \quad \text{on } \partial D. \end{aligned} \quad (1.2.21)$$

The solution  $u(\xi) : D \rightarrow \mathbb{R}$  depends on  $\xi = (\xi_1, \dots, \xi_p)$  a random vector. Here  $a, \kappa$  are the reaction and diffusion coefficients respectively and advection field  $c$ . Combining a finite element discretization in space and stochastic Galerkin method for the parameters, the resulting Galerkin approximation still denoted by  $u$ , identified with its set of coefficients, is a tensor  $u \in X = \mathbb{R}^n \otimes (\otimes_{\mu=1}^p \mathbb{R}^s)$ . Here,  $n$  and  $s$  are the dimensions of the finite element and piecewise polynomial space respectively. The tensor  $u$  is solution of the following algebraic tensor equation

$$Au = b, \quad (1.2.22)$$

where the tensor operator  $A$  and tensor  $b$  are the algebraic representations of the bilinear and linear forms arising in the Galerkin formulation of the considered problem.

## Two dimensional case with $p = 1$

We present numerical results concerning the A-IMR method applied to equation (1.2.22). This method provides an approximation  $\tilde{u}$  of the best approximation of  $u$  in  $\mathcal{M}_{\leq r}(X)$  with respect to  $\|\cdot\|_2$  the natural canonical norm on  $X$ . This choice corresponds to an operator  $R_X = I_n \otimes I_s$ , where  $I_n$  (resp.  $I_s$ ) is the identity in  $\mathbb{R}^n$  (resp.  $\mathbb{R}^s$ ). We take  $\mathcal{M}_{\leq r}(X) = \mathcal{T}_r^{\text{can}}(X)$ . For considered numerical applications, we have  $\dim(X) \geq 10^5$ . The exact solution  $u$  of (1.2.22) is computed and used as a reference solution. For that choice, the  $\|\cdot\|_X$ -best approximation of  $u$  in  $\mathcal{M}_{\leq r}(X)$  is a rank- $r$  SVD that can be computed exactly using classical algorithms (see Section 1.2.2). For the construction of an approximation in  $\mathcal{T}_r^{\text{can}}(X)$  using A-IMR, we consider two strategies: the direct approximation in  $\mathcal{T}_r^{\text{can}}(X)$  using the gradient type algorithm with  $\mathcal{M}_{\leq r}(X) = \mathcal{T}_r^{\text{can}}(X)$ , and a greedy algorithm that consists in a series of  $r$  corrections in  $\mathcal{T}_1^{\text{can}}(X)$  computed using with the gradient type algorithm, with  $\mathcal{M}_{\leq r}(X) = \mathcal{T}_1^{\text{can}}(X)$  and with an updated residual  $b$  at each correction.

The A-IMR is compared to a standard approach, denoted CMR, which consists in minimizing the canonical norm of the residual. The latter approach has been introduced and analyzed in different papers, using either direct minimization or greedy rank-one algorithms [5, 24, 65], and is known to suffer from ill-conditioning of the operator  $A$ .

The convergence curves with respect to the rank are shown in Figure 1.2.1, where the error is measured in the  $\|\cdot\|_2$  norm. Concerning the direct approach, we observe that the different algorithms have roughly the same rate of convergence. The A-IMR convergence curves are close to the optimal SVD for a wide range of values of  $\delta$ . Concerning the greedy approach, we observe a significant difference between A-IMR and CMR. We note that A-IMR is close to the optimal SVD up to a certain rank (depending on  $\delta$ ) after which the convergence rate decreases but remains better than the one of CMR.

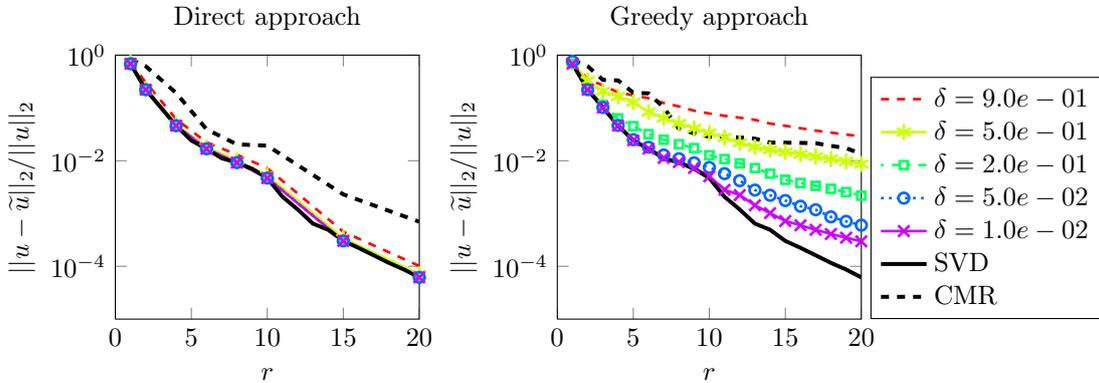


Figure 1.2.1: Comparison of minimal residual methods for  $\mathcal{M}_{\leq r}(X) = \mathcal{T}_r^{\text{can}}(X)$  and  $\|\cdot\|_X = \|\cdot\|_2$ . Convergence with the rank  $r$  of the approximations obtained with CMR and A-IMR with different precisions  $\delta$ , and with direct (left) or greedy rank-one (right) approaches.

## Higher dimensional case with $p = 9$

Here, we compute low-rank approximations of the solution of  $u$  in the canonical tensor subset  $\mathcal{T}_r^{\text{can}}(X)$  with  $r \geq 1$ . We rely on the greedy algorithm with successive corrections in  $\mathcal{M}_{\leq r} = \mathcal{T}_1^{\text{can}}(X)$  computed with the gradient type algorithm. For considered numerical applications, we have  $\dim(X) \geq 10^{11}$ . As a reference solution, we consider a low-rank approximation  $u_{\text{ref}}$  of  $u$  computed using a greedy rank-one algorithm based on a canonical minimal residual formulation.

On Figure 1.2.2, we plot the convergence with the rank  $r$  of the approximations computed by

both A-IMR and CMR algorithms and of the greedy approximations  $\tilde{u}$  of the reference solution  $u_{\text{ref}}$ . We observe (as for the lower-dimensional example) that, with different values of the parameter  $\delta$  (up to 0.9), the A-IMR method provides a better approximation of the solution in comparison to the CMR method. When decreasing  $\delta$ , the proposed algorithm seems to provide approximations that tend to present the same convergence as the greedy approximations.

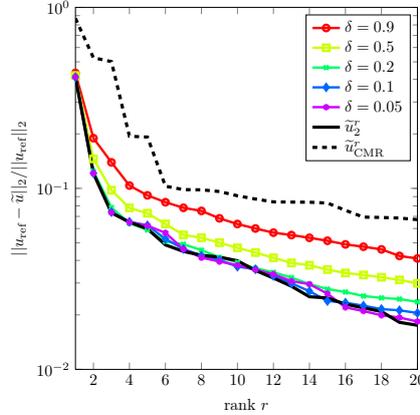


Figure 1.2.2: Convergence with the rank of approximations obtained with the greedy CMR and A-IMR algorithms for different precisions  $\delta$ . Convergence is plotted with respect to the norm  $\|\cdot\|_2$

## 1.2.5 Summary

We have proposed a new algorithm for the construction of low-rank approximations of the solutions of high-dimensional weakly coercive problems formulated in a tensor space  $X$ . This algorithm is based on the approximate minimization of a particular residual norm on given low-rank tensor subsets  $\mathcal{M}_{\leq r}(X)$ , the residual norm coinciding with some measure of the error in solution. Therefore, the algorithm is able to provide a quasi-best low-rank approximation with respect to a norm  $\|\cdot\|_X$  that can be designed for a certain objective. In view of goal oriented approximation, it is possible to design the norm  $\|\cdot\|_X$  such that it takes into account some particular quantity of interest, see first results in [MBF9]. This topic has been studied more deeply in [167, Chapter 5], but not considered for publication. However, it is discussed in Section 1.3 within the context of projection based low-rank approximation.

Let mention that the method has been applied with encouraging results for the solution of transient elastodynamics equations in space-time domain [29]. Some extension to saddle points problems [22] with possible perturbed Uzawa algorithm have also been investigated.

## 1.3 Projection based goal oriented low-rank approximation

We recall briefly in Section 1.3.1 projection based low-rank approximation methods for the solution of parameter dependent linear problem under the form (1.1.1) formulated in Hilbert space. Then, in Section 1.3.2, we discuss possible goal oriented projection methods when the quantity to compute is not the solution itself but some Quantity of Interest (QoI) expressed as  $s = Lu$ , with  $L$  a linear map. From these two ingredients we derive a goal oriented reduced basis method. This work is related to the publication [MBF10]. Some illustrative examples of the proposed approach are given in the last section.

*Notations.* For a Hilbert space  $H$ , we denote by  $R_H \in \mathcal{L}(H, H')$  the Riesz map such that  $\|v\|_H^2 = \langle R_H v, v \rangle$ , where  $\langle \cdot, \cdot \rangle$  denotes the duality pairing. The dual norm  $\|\cdot\|_{H'}$  on  $H'$  is such that  $R_{H'} = R_H^{-1}$ . Then  $\|v\|_H = \|R_H v\|_{H'}$  and  $|\langle v, w \rangle| \leq \|v\|_H \|w\|_{H'}$  hold for any  $v \in H$  and

$w \in H'$ . For any operator  $C \in \mathcal{L}(H_1, H_2)$ , with  $H_1$  and  $H_2$  two Hilbert spaces,  $C^* \in \mathcal{L}(H_2, H_1')$  denotes the adjoint of  $C$ , such that  $\langle Cv_1, v_2 \rangle = \langle v_1, C^*v_2 \rangle$  for any  $v_1 \in H_1$  and  $v_2 \in H_2$ .

### 1.3.1 Projection based low-rank approximation

Let  $V$  and  $W$  be two Hilbert spaces. We consider the linear parameter-dependent equation

$$A(\xi)u(\xi) = b(\xi) \quad (1.3.1)$$

with  $A(\xi) \in \mathcal{L}(V, W')$  and  $b(\xi) \in W'$ , where the operator  $A(\xi)$  and right-hand side  $b(\xi)$  depend on a parameter  $\xi$  which takes values in some parameter set  $\Xi \subset \mathbb{R}^p$ . We assume that  $A(\xi)$  is a norm-isomorphism such that for all  $u \in V$ ,

$$\alpha(\xi)\|u\|_V \leq \|A(\xi)u\|_{W'} \leq \beta(\xi)\|u\|_V,$$

where

$$\inf_{0 \neq v \in V} \sup_{0 \neq w \in W} \frac{\langle A(\xi)v, w \rangle}{\|v\|_V \|w\|_W} := \alpha(\xi) > 0 \text{ and } \sup_{0 \neq v \in V} \sup_{0 \neq w \in W} \frac{\langle A(\xi)v, w \rangle}{\|v\|_V \|w\|_W} := \beta(\xi) < \infty,$$

which ensures the well-posedness of (1.3.1).

We adopt here a subspace point of view for the low-rank approximation of the solution map  $u : \Xi \rightarrow V$ . It relies on two key ingredients : properly chosen (small) subspaces of  $V$ , and appropriate projections on this subspaces. Here, we will focus in particular on the well known Reduced Basis (RB) method, see monograph [144], the surveys [99], and [134, 135] for a presentation within low-rank approximation framework. RB method performs in two steps. During an *offline* step, the subspaces are constructed from snapshots of the exact solution  $u(\xi)$  for a small number of values of the parameter  $\xi$ . Then the low-rank approximation defined as some projection of the solution  $u(\xi)$  in this subspace, is computed for any new instance of the parameter  $\xi \in \Xi$  during the *online* stage.

#### Online step: Galerkin type projections

In what follows, we consider that we are given a finite-dimensional subspace  $V_r \subset V$  with  $\dim(V_r) = r$ , the *reduced space* together with the associated *reduced basis* noted  $\{v_1, \dots, v_r\} \subset V$ .

Classical methods used for problems under the form (1.3.1) are Galerkin-type projections. They provide an approximation  $u_r(\xi)$  of  $u(\xi)$ , in the dimensional subspace  $V_r$  deduced from some orthogonality condition on the residual  $r(u(\xi), \xi) = A(\xi)u(\xi) - b(\xi)$  with respect to a  $r$ -dimensional space or by minimizing some residual norm. The resulting approximation  $u_r \in V_r$  is quasi-optimal that for all  $\xi \in \Xi$  i.e. there exists  $\gamma(\xi) \geq 1$

$$\|u(\xi) - u_r(\xi)\|_V \leq \gamma(\xi)\|u(\xi) - P_{V_r}u(\xi)\|_V, \quad (1.3.2)$$

where  $P_{V_r}$  is the orthogonal projection of  $u$  in  $V_r$ , defined as

$$\|u(\xi) - P_{V_r}u(\xi)\|_V = \inf_{v \in V_r} \|u(\xi) - v\|_V.$$

We give here two typical examples of Galerkin projection methods.

**Minimal residual based projection.** The minimal residual projection  $u_r$  is defined through

$$\min_{v \in V_r} \|r(v, \xi)\|_{W_r}. \quad (1.3.3)$$

In that case (1.3.2) holds with  $\gamma(\xi) = \frac{\beta(\xi)}{\alpha(\xi)}$  the condition number of the operator  $A(\xi)$ . For problems badly conditioned, the minimal residual approximation method may lead to an approximation  $u_r$  far from the best approximation in  $V_r$  for the norm  $\|\cdot\|_V$ . In that context, it should be relevant to use parameter dependent preconditioner or ideal minimal residual based projection as discussed in Section 1.2.3.

**Petrov-Galerkin projection.** Let assume that the approximation  $u_r \in V_r$  is defined as a Petrov-Galerkin projection of  $u$  characterized by

$$\langle r(u_r(\xi), \xi), y \rangle = 0, \quad \forall y \in W_r, \quad (1.3.4)$$

where  $W_r \subset W$  is a test space of dimension  $r$ . Under the assumption that

$$\alpha_{V_r, W_r}(\xi) = \inf_{0 \neq v \in V_r} \sup_{0 \neq y \in W_r} \frac{\langle A(\xi)v, y \rangle}{\|v\|_V \|y\|_W} > 0, \quad (1.3.5)$$

Céa's lemma [71, Lemma 2.8] states that (1.3.2) holds with  $\gamma(\xi) = 1 + \frac{\beta(\xi)}{\alpha_{V_r, W_r}(\xi)}$ .

The resulting low-rank approximation is a Galerkin projection of  $u(\xi)$  in  $V_r$  given by

$$u_r(\xi) = \sum_{i=1}^r \alpha_i(\xi) v_i,$$

where  $\{\alpha_1, \dots, \alpha_r\}$  are the coefficients of  $u_r(\xi)$  in the reduced basis. If the residual  $r(u_r(\xi), \xi)$  (through the operator  $A(\xi)$  and the right-hand side  $b(\xi)$ ) admits low-rank representations, also called *affine* representation, it can be computed with a complexity only depending on  $r$ . If not, such representations can be obtained by the Empirical Interpolation Method (EIM) [14]. It allows to compute the Galerkin projection  $u_r(\xi)$  by solving a reduced system of  $r$  equations on the unknown coefficients  $\{\alpha_i(\xi)\}_{i=1}^r$  with a computational complexity only depending on  $r$ .

### Offline step: Construction of subspaces

At this step, the question is how to properly select the reduced space  $V_r$  so that the solution manifold

$$\mathcal{K} := \{u(\xi) : \xi \in \Xi\}$$

can be well approximated. To answer these question, we can use an error measure that gives a good indication on how well  $\mathcal{K}$  can be approximated by a  $r$ -linear subspace. Assuming that the manifold  $\mathcal{K}$  is compact, let us define

$$d_r^\infty(\mathcal{K})_V = \min_{\dim(V_r)=r} \sup_{\xi \in \Xi} \|u(\xi) - P_{V_r} u(\xi)\|_V$$

which is the *Kolmogorov*  $r$ -width of  $\mathcal{K}$ . This measure of error is a good indicator for uniformly accurate approximation over the parameter set  $\Xi$ . In particular, when the *Kolmogorov*  $r$ -width has a rapid decay with  $r$ , one can expect to obtain a good low-rank approximation. It is also relevant to consider subspaces optimal with respect to the  $L^2$ -norm (or more generally  $L^p$ -norm). In that case, we could consider

$$d_r^2(\mathcal{K})_V = \min_{\dim(V_r)=r} \left( \int_{\Xi} \|u(\xi) - P_{V_r} u(\xi)\|_V^2 d\mu(\xi) \right)^{1/2},$$

where  $\mu$  is the probability measure of  $\xi$ .

A simple choice is to take the subspaces  $V_r$  generated from evaluations (snapshots) of the solution  $u(\xi)$  at for some a finite number of instances (possibly randomly chosen)  $\tilde{\Xi} = \{\xi_1, \dots, \xi_K\} \subset \Xi$  of the parameter  $\xi$ . But nothing guarantees that this choice provides reduced space close to the optimal  $r$ -dimensional subspaces. Thus, in practice  $V_r$  is sought as the best  $r$ -dimensional subspace satisfying some optimality criterion related the measures  $d_r^2(\mathcal{K})_V$  or  $d_r^\infty(\mathcal{K})_V$ .

**From samples.** For the norm  $L^2$ -norm,  $V_r$  is sought such that

$$\min_{\dim(V_r)=r} \frac{1}{K} \sum_{k=1}^K \|u(\xi^k) - P_{V_r} u(\xi^k)\|_V^2$$

where  $\tilde{\Xi}$  is a  $K$ -sample according to the probability measure  $\mu$ . This choice is at the basis of POD methods for parameter-dependent equations [109, 162]. For  $L^\infty$ -norm,  $V_r$  is such that

$$\min_{\dim(V_r)=r} \max_{\xi \in \tilde{\Xi}} \|u(\xi) - P_{V_r} u(\xi)\|_V.$$

where  $\tilde{\Xi}$  is a set of  $K$  points in  $\Xi$  (called *training set*). In view of computation, a tractable definition of the subspaces relies on finding an appropriate set of parameters  $\{\xi_1, \dots, \xi_r\} \subset \tilde{\Xi}$  and impose  $V_r = \text{span}\{u(\xi_1), \dots, u(\xi_r)\}$ . Practically, *greedy algorithm* (e.g. [14, 17] for (EIM)) are prominent techniques for constructing such a nested sequence of reduced space from snapshots of the solution computed for adaptively chosen parameter values. It works as follows : given  $\{\xi_1, \dots, \xi_{r-1}\}$  and the corresponding subspace  $V_{r-1} = \text{span}\{u(\xi_1), \dots, u(\xi_{r-1})\}$ , a new parameter value  $\xi_r$  is selected such that

$$\|u(\xi_r) - P_{V_r} u(\xi_r)\|_V = \max_{\xi \in \tilde{\Xi}} \|u(\xi) - P_{V_{r-1}} u(\xi)\|_V,$$

Then the reduced space is  $V_r = \text{span}\{u(\xi_1), \dots, u(\xi_r)\}$ .

**From equations.** As in the present case,  $u(\xi)$  is solution of the parameter-dependent equation (1.3.1), the quantities  $\|u(\xi) - P_{V_r} u(\xi)\|_V$  are not longer computable. Instead, a posteriori error estimates  $\Delta(u_r(\xi), \xi)$ , computable without requiring  $u(\xi)$  and with low cost, are used. Typical choice consists in taking for  $\Delta(u_r(\xi), \xi)$  a bound of a certain norm of the residual  $r(u_r(\xi), \xi)$  which is behind the PGD and RB methods. In the particular case of RB method, when  $\Delta(u_r(\xi), \xi)$  is certified and under the quasi-optimality of the Galerkin projection (1.3.2), the resulting algorithm is the so-called *weak greedy algorithm*. Convergence results on greedy algorithms can be found in [30, 31, 62], especially they have proven that the approximation error in  $V_r$  has the same type of decay as the Kolmogorov  $r$ -width for algebraic and exponential convergence.

### 1.3.2 Goal oriented projections

In many applications one is not interested in the solution  $u(\xi)$  itself, but only in a QoI  $s(\xi)$  which is a functional of  $u(\xi)$ . Let  $Z$  be an Hilbert space, we consider

$$s(\xi) = L(\xi)u(\xi),$$

with  $L(\xi) \in \mathcal{L}(V, Z)$ . For example, for boundary value problems,  $L(\xi)$  can be defined as the trace operator providing the restriction of the solution to the boundary of the domain. In this case the QoI belongs to an infinite dimensional space or, after discretization, to a finite but possibly high dimensional space.

## MOR for evaluation of QoI

Efficient goal-oriented methods have been proposed for the estimation of a scalar-valued variable of interest  $s(\xi)$ . A standard method consists in computing an approximation of the solution of the so-called dual problem associated to (1.3.1) which is used to correct the estimation of  $s(\xi)$ . We refer to [142] for a general survey on primal-dual methods and to [46, 94, 99, 143] for the application in the context of RB methods. Difficulties arise when the variable of interest  $s(\xi)$  takes its values in a vector space of finite or infinite dimension. The standard approach, which consists in treating  $s(\xi)$  as a collection of scalar-valued variables of interest and in building one reduced dual space for each of them, has a complexity which grows proportionally to the dimension of  $s(\xi)$ . To circumvent this issue, one could construct a single reduced dual space, thus allowing to handle variables of interest with high and potentially infinite dimension. Such an approach can be found for parametric dynamical systems, see the monograph [21] for a general introduction. In this framework, projection-based model order reduction methods are used for the approximation of  $s(\xi)$  which is an output of the dynamical system. Petrov-Galerkin methods have been proposed with different ways of constructing the reduced basis for the test and trial space, such as the balanced truncation methods, (balanced) POD method, moment matching methods, etc. We refer to [23] for a recent review on these methods.

In what follows, we present some work [MBF10] concerning goal-oriented low-rank approximation for parameter dependent PDEs following the subsequent road map.

1. Develop an appropriate method for the approximation of variable of interest  $s(\xi)$  taking values in a single vector space of finite or infinite dimension.
2. Choose suitable projection principle (Petrov-Galerkin, Primal-Dual approach, Saddle point formulation) that guarantees the best approximation error on the QoI while using a single reduced space.
3. Using computable error estimate, propose greedy approaches to select the reduced spaces associated to the solution and the QoI.

### Goal oriented projections

*To alleviate the notations, we have omitted the dependence to  $\xi$  in this section. The approximation of any quantity  $a$  is noted  $\tilde{a}$ .*

**A posteriori computation of QoI.** We assume that  $\tilde{u} = u_r \in V_r$ , the Petrov-Galerkin projection of  $u$  by Equation (1.3.4) with  $W_r$  satisfying the assumption (1.3.5). Then, we consider  $\tilde{s}$  as the standard approximation of  $s$ , computed a posteriori from  $\tilde{u}$  as

$$\tilde{s} = L\tilde{u}. \quad (1.3.6)$$

The next proposition [MBF10, Proposition 2.1] provides an error bound for the standard approximation  $\tilde{s}$ .

**Proposition 1.3.1.** *Let  $\tilde{u} = u_r$  be the Petrov-Galerkin projection defined by (1.3.4), the approximation  $\tilde{s}$  of  $s$  defined by (1.3.10) satisfies*

$$\|s - \tilde{s}\|_Z \leq \frac{\delta_{W_r}^L}{\sqrt{1 - (\delta_{V_r, W_r})^2}} \min_{v \in V_r} \|u - v\|_V, \quad (1.3.7)$$

with

$$\delta_{V_r, W_r} = \max_{0 \neq v \in V_r} \min_{y \in W_r} \frac{\|v - R_V^{-1} A^* y\|_V}{\|v\|_V} < 1, \quad (1.3.8)$$

and

$$\delta_{W_r}^L = \sup_{0 \neq z' \in Z'} \min_{y \in W_r} \frac{\|L^* z' - A^* y\|_{V'}}{\|z'\|_{Z'}}. \quad (1.3.9)$$

**Primal-dual approach.** Let us introduce the *dual variable*  $Q \in \mathcal{L}(Z', W)$  defined by the *dual problem*  $A^* Q = L^*$ . For given approximations  $\tilde{u}$  and  $\tilde{Q}$ , we define the approximation  $\tilde{s}$  by

$$\tilde{s} = L\tilde{u} + \tilde{Q}^*(b - A\tilde{u}), \quad (1.3.10)$$

where  $\tilde{Q}^*(b - A\tilde{u})$  is a correction using the approximation of the dual variable. The following proposition [MBF10, Proposition 2.5] provides an error bound on the QoI, which is a generalization of the classical error bound for scalar-valued variables of interest (see [142]) to vector-valued variables of interest.

**Proposition 1.3.2.** *Let  $\tilde{u} = u_r$  be the Petrov-Galerkin projection defined by (1.3.4), the approximation  $\tilde{s}$  of  $s$  defined by (1.3.10) satisfies*

$$\|s - \tilde{s}\|_Z \leq \frac{\|L^* - A^* \tilde{Q}\|_{Z' \rightarrow V'}}{\sqrt{1 - (\delta_{V_r, W_r})^2}} \min_{v \in V_r} \|u - v\|_V, \quad (1.3.11)$$

where

$$\|L^* - A^* \tilde{Q}\|_{Z' \rightarrow V'} = \sup_{0 \neq z' \in Z'} \frac{\|(L^* - A^* \tilde{Q})z'\|_{V'}}{\|z'\|_{Z'}}. \quad (1.3.12)$$

For the approximation  $\tilde{Q}$  of  $Q \in \mathcal{L}(Z', W)$ , the bound (1.3.11) suggests that  $\|L^* - A^* \tilde{Q}\|_{Z' \rightarrow V'}$  should be small. We then propose to choose  $\tilde{Q}$  as a solution of

$$\min_{\tilde{Q} \in \mathcal{L}(Z', W_k^Q)} \|L^* - A^* \tilde{Q}\|_{Z' \rightarrow V'}, \quad (1.3.13)$$

where  $W_k^Q \subset W$  is a given approximation space (different from  $W_r$ ). The next proposition [MBF10, Proposition 2.6] shows how to construct a solution of (1.3.13).

**Proposition 1.3.3.** *The operator  $Q_k : Z' \rightarrow W_k^Q$  defined for  $z' \in Z'$  by*

$$Q_k z' = \arg \min_{y_k \in W_k^Q} \|L^* z' - A^* y_k\|_{V'} \quad (1.3.14)$$

*is linear and is a solution of (1.3.13). Moreover  $Q_k z' \in W_k^Q$  is characterized by*

$$\langle L^* z' - A^* Q_k z', R_V^{-1} A^* y_k \rangle = 0, \quad \forall y_k \in W_k^Q. \quad (1.3.15)$$

We give now a new bound [MBF10, Proposition 2.8] of the error on the QoI.

**Proposition 1.3.4.** *Let  $\tilde{u} = u_r$  be the Petrov-Galerkin projection defined by (1.3.4) and let  $\tilde{Q} = Q_k$  be defined by (1.3.15). Then the approximation  $\tilde{s}$  defined by (1.3.10) satisfies*

$$\|s - \tilde{s}\|_Z \leq \frac{\delta_{W_k^Q}^L}{\sqrt{1 - (\delta_{V_r, W_r})^2}} \min_{v \in V_r} \|u - v\|_V. \quad (1.3.16)$$

where

$$\delta_{W_k^Q}^L = \sup_{0 \neq z' \in Z'} \min_{y \in W_k^Q} \frac{\|L^* z' - A^* y\|_{V'}}{\|z'\|_{Z'}}. \quad (1.3.17)$$

For such choice of  $\tilde{Q}$ , the Proposition 1.3.4 provides a sharper bound than Proposition 1.3.2 for the error on the QoI by taking advantage of the orthogonality property (1.3.15).

**Projection based on a saddle point problem** In the line of [59] for the approximation of (vector-valued) variables of interest, we define the projection of  $u$  on the reduced space  $V_r$  by means of a saddle point problem.

Let us equip  $W$  with a norm  $\|\cdot\|_W$  such that the relation  $\|y\|_W = \|A^*y\|_{V'}$  holds for any  $y \in W$ , which is equivalent to the following relation between the Riesz maps  $R_W$  and  $R_V$ :

$$R_W = AR_V^{-1}A^*. \quad (1.3.18)$$

We introduce a subspace  $T_p \subset W$  of dimension  $p$  and we define the projection  $u_{r,p}$  in  $V_r$  as the solution of the saddle point problem

$$\min_{v \in V_r} \max_{\substack{w \in T_p \\ \|w\|_W=1}} \langle Av - b, w \rangle. \quad (1.3.19)$$

Under the condition (discrete inf-sup condition)

$$\inf_{0 \neq v \in V_r} \sup_{0 \neq y \in T_p} \frac{\langle Av, y \rangle}{\|v\|_V \|y\|_W} =: \alpha_{V_r, T_p} > 0, \quad (1.3.20)$$

we have proven [MBF10, Proposition 2.10] that there exists a unique solution  $(u_{r,p}, y_{r,p})$  in  $V_r \times T_p$  to

$$\langle R_W y_{r,p}, y \rangle + \langle Au_{r,p}, y \rangle = \langle b, y \rangle \quad \forall y \in T_p, \quad (1.3.21a)$$

$$\langle A^* y_{r,p}, v \rangle = 0 \quad \forall v \in V_r. \quad (1.3.21b)$$

Now, we define the approximation  $\tilde{s}$  as

$$\tilde{s} = L\tilde{u} + LR_V^{-1}A^*\tilde{y}, \quad (1.3.22)$$

where  $(\tilde{u}, \tilde{y}) := (u_{r,p}, y_{r,p}) \in V_r \times T_p$  is the solution of the saddle point problem (1.3.21). The following proposition [MBF10, Proposition 2.14] provides an error bound for the approximation of the QoI.

**Proposition 1.3.5.** *Let  $(\tilde{u}, \tilde{y}) := (u_{r,p}, y_{r,p}) \in V_r \times T_p$  be the solution of the saddle point problem (1.3.21). The approximation  $\tilde{s}$  defined by (1.3.22) satisfies*

$$\|s - \tilde{s}\|_Z \leq \frac{\delta_{T_p}^L}{\sqrt{1 - (\delta_{V_r, T_p})^2}} \min_{v \in V_r} \|u - v\|_V. \quad (1.3.23)$$

with

$$\delta_{T_p}^L = \sup_{0 \neq z' \in Z'} \min_{y \in T_p} \frac{\|L^*z' - A^*y\|_{V'}}{\|z'\|_{Z'}} \quad (1.3.24)$$

and

$$\delta_{V_r, T_p} = \max_{0 \neq v \in V_r} \min_{y \in T_p} \frac{\|v - R_V^{-1}A^*y\|_V}{\|v\|_V}. \quad (1.3.25)$$

In practice, we will consider for  $T_p$  spaces of the form

$$T_p = W_r + W_k^Q, \quad (1.3.26)$$

with  $\dim(W_r) = r$  which implies

$$\delta_{T_p}^L \leq \delta_{W_k^Q}^L \quad \text{and} \quad \delta_{V_r, T_p} \leq \delta_{V_r, W_r},$$

so that the error bound (1.3.23) for the QoI is better than the error bound (1.3.16) of the primal-dual method with primal approximation space  $V_r$ , primal test space  $W_r$  and dual approximation space  $W_k^Q$ . Therefore, we expect the approximation  $u_{r,p}$  to be closer to the solution  $u$  than the Petrov-Galerkin projection  $u_r$ . Also, the approximation of the QoI is expected to be improved.

## Practical aspects

The projection methods introduced for the estimation of the QoI rely on the introduction of three spaces: the primal approximation space  $V_r$ , the primal test space  $W_r$  and the dual approximation space  $W_k^Q$  and for the saddle point approach the space  $T_p = W_r + W_k^Q$ . We adopt an offline/online strategy. Reduced (low-dimensional) spaces  $V_r$ ,  $W_r$  and  $W_k^Q$  are constructed during the offline phase. Then, the projections on these reduced spaces and the evaluations of the QoI are rapidly computed for any parameter value  $\xi \in \Xi$  during the online phase.

**Reduced space  $W_r$ .** Given the primal approximation space  $V_r$ , we know that  $W_r$  should be chosen such that  $\delta_{V_r, W_r}$  is as close to zero as possible (see Propositions 1.3.1, 1.3.4 and 1.3.5). To that goal, we define the (parameter-dependent) test space as

$$W_r(\xi) = P_m(\xi)^* R_V V_r, \quad (1.3.27)$$

where  $P_m(\xi)$  is an interpolation of the inverse of  $A(\xi)$  using  $m$  interpolation points in the parameter set  $\Xi$  [168]. The interpolation points are chosen as the points where solutions (primal and dual) have already been computed, i.e. the points given by (1.3.28).

**Reduced spaces  $V_r$  and  $W_k^Q$ .** From suitable error estimates  $\Delta(\xi)$  for the vector-valued QoI (see for details [MBF10, Section 3] and Section 1.3.3), different greedy algorithms for the construction of the reduced spaces  $V_r$  and  $W_k^Q$  can be proposed. Given  $\tilde{\Xi} \subset \Xi$  a training set, they broadly consist in searching for a parameter value  $\xi^* \in \tilde{\Xi}$ , a finite set where the error estimate  $\Delta(\xi)$  is maximum, i.e.

$$\xi^* \in \arg \max_{\xi \in \tilde{\Xi}} \Delta(\xi). \quad (1.3.28)$$

A first strategy is to simultaneously enrich both the primal approximation space

$$V_{r+1} = V_r + \text{span}(u(\xi^*)) \quad (1.3.29)$$

and the dual approximation space

$$W_{k+l}^Q = W_k^Q + \text{range}(Q(\xi^*)) \quad (1.3.30)$$

at each iteration. This strategy is referred as the simultaneous construction, as opposed to the alternate construction which consists in enriching  $W_k^Q$  (resp.  $V_r$ ) if  $V_r$  (resp.  $W_k^Q$ ) were enriched at the previous greedy iteration step.

### 1.3.3 Numerical applications

We consider the benchmark problem of the cooling of electronic components [168]. The equation to solve is an advection-diffusion equation over the domain  $D \subset \mathbb{R}^2$

$$-\nabla \cdot (\kappa(\xi) \nabla T(\xi)) + \mathbf{c}(\xi) \cdot \nabla T(\xi) = f, \quad (1.3.31)$$

whose solution  $T(\xi) : D \rightarrow \mathbb{R}$  is the temperature field. Here  $\kappa$  and  $\mathbf{c}$  denote respectively the parameter-dependent diffusion coefficient and the advection field. Here, we focus on the resulting algebraic parameter-dependent equation coming from stabilized finite element discretization of (1.3.31), that is  $A(\xi)u(\xi) = b(\xi)$ , where  $u(\xi) \in \mathbb{R}^n$  are the coefficients of the finite element approximation  $T^h = \sum_{i=1}^n u_i \varphi_i$  of  $T$ , and where  $\xi = (\xi_1, \dots, \xi_4)$  is a 4-dimensional random vector. The space  $V = \mathbb{R}^n$  with  $n = 2.8 \times 10^4$  is endowed with the norm  $\|\cdot\|_V = \|\cdot\|_{V_0}$  which

corresponds to the  $H^1(D)$ -norm<sup>1</sup>. The variable of interest  $s(\xi) = (s_1(\xi), s_2(\xi))$  is the mean temperature of both electronic components, with

$$s_1(\xi) = \frac{1}{|D_{IC_1}|} \int_{D_{IC_1}} T^h(\xi) \quad , \quad s_2(\xi) = \frac{1}{|D_{IC_2}|} \int_{D_{IC_2}} T^h(\xi), \quad (1.3.32)$$

where  $D_{IC_i}$  ( $i = 1, 2$ ) are two subdomains of  $D$ . Then we can write  $s(\xi) = Lu(\xi)$  for an appropriate  $L \in \mathbb{R}^{l \times n}$ , with  $l = 2$ . Here we have  $Z = \mathbb{R}^2$ , which we equip with the canonical norm on  $\mathbb{R}^2$ .

**Comparison of the projections methods** Here the approximation spaces  $V_r$ ,  $W_k^Q$  and the test space  $W_r$  are given. We denote by  $\mathbf{V}_r$ ,  $\mathbf{W}_k^Q$  and  $\mathbf{W}_r$  the matrices containing the basis vectors of the corresponding subspaces. The matrix  $\mathbf{V}_r = (u(\xi_1), \dots, u(\xi_{50}))$  contains 50 snapshots of the primal solution ( $r = 50$ ), and  $\mathbf{W}_k^Q = (Q(\xi_1) \dots Q(\xi_{25}))$  contains 25 snapshots of the dual solution so that the dimension of  $W_k^Q$  is  $k = 25l = 50$ . The test space  $W_r$  is defined according to (1.3.27). The matrix associated to the test space is given by  $\mathbf{W}_r(\xi) = P_m^T(\xi)R_V \mathbf{V}_r$ . We consider a samples set  $\Xi_t$  of size  $t = 10^4$ . For any  $\xi \in \Xi_t$  we compute the exact quantity of interest  $s(\xi)$  and the approximation  $\tilde{s}(\xi)$  by the following methods.

- **Primal only:** solve the linear system  $(\mathbf{W}_r^T(\xi)A(\xi)\mathbf{V}_r)U_r(\xi) = \mathbf{W}_r(\xi)^T b$  of size  $r$  and compute  $\tilde{s}(\xi) = (L\mathbf{V}_r)U_r(\xi)$ .
- **Dual only:** solve the linear system

$$((\mathbf{W}_k^Q)^T A(\xi)R_V^{-1}A(\xi)^*\mathbf{W}_k^Q)Y_k(\xi) = (\mathbf{W}_k^Q)^T b$$

of size  $k$  and compute  $\tilde{s}(\xi) = (LR_V^{-1}A(\xi)^*\mathbf{W}_k^Q)Y_k(\xi)$ .

- **Primal-dual:** solve the linear system of the Primal only method, solve the linear system

$$((\mathbf{W}_k^Q)^T A(\xi)R_V^{-1}A(\xi)^*\mathbf{W}_k^Q)Y_k(\xi) = ((\mathbf{W}_k^Q)^T b) - ((\mathbf{W}_k^Q)^T A(\xi)\mathbf{V}_r)U_r(\xi)$$

of size  $k$ , and compute  $\tilde{s}(\xi) = (L\mathbf{V}_r)U_r(\xi) + (LR_V^{-1}A(\xi)^*\mathbf{W}_k^Q)Y_k(\xi)$ .

- **Saddle point:** solve the linear system of size  $p + r$

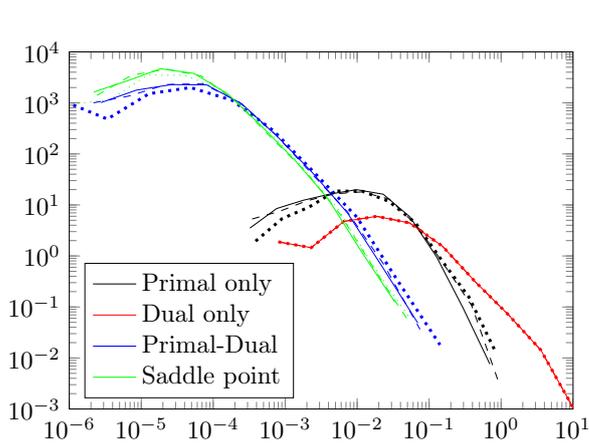
$$\begin{pmatrix} \mathbf{T}_p^T(\xi)A(\xi)R_V^{-1}A(\xi)^*\mathbf{T}_p(\xi) & \mathbf{T}_p^T(\xi)A(\xi)\mathbf{V}_r \\ (\mathbf{T}_p^T(\xi)A(\xi)\mathbf{V}_r)^T & 0 \end{pmatrix} \begin{pmatrix} Y_{r,p}(\xi) \\ U_{r,p}(\xi) \end{pmatrix} = \begin{pmatrix} \mathbf{T}_p(\xi)^T b \\ 0 \end{pmatrix}$$

with  $\mathbf{T}_p(\xi) = (\mathbf{W}_r(\xi), \mathbf{W}_k^Q)$ , and compute

$$\tilde{s}(\xi) = (L\mathbf{V}_r)U_{r,p}(\xi) + (LR_V^{-1}A(\xi)^*\mathbf{T}_p(\xi))Y_{r,p}(\xi).$$

The numerical results are given in Figure 1.3.1. The saddle point method leads to the lowest error on the variable of interest. Also, we see that a good preconditioner (for example with  $m = 30$ ) improves the accuracy for the saddle point method, the primal only method and the primal-dual method. However, this improvement is not really significant for the considered application: the errors are barely divided by 2 compared to the non preconditioned Galerkin projection ( $m = 0$ ). In fact, the preconditioner improves the quality of the test space, and the choice  $W_r = V_r$  (yielding the standard Galerkin projection) is sufficiently accurate for this example and for the chosen norm on  $V$ .

<sup>1</sup>It means that  $\|v\|_{V_0} = \|v^h\|_{H^1(D)}$  for all  $v \in V$ , where  $v^h = \sum_{i=1}^n v_i \varphi_i$ .



(a) PDF of the error. Three different preconditioners  $P_m(\xi)$  are used:  $m = 0$  (dotted lines),  $m = 10$  (dashed lines) and  $m = 30$  (continuous lines).

Primal only	$L^\infty$ -norm	$L^2$ -norm
$m = 0$	$1.284 \times 10^0$	$1.245 \times 10^{-1}$
$m = 5$	$1.203 \times 10^0$	$9.637 \times 10^{-2}$
$m = 10$	$1.458 \times 10^0$	$1.064 \times 10^{-1}$
$m = 20$	$1.068 \times 10^0$	$8.386 \times 10^{-2}$
$m = 30$	$1.066 \times 10^0$	$7.955 \times 10^{-2}$

Primal-dual	$L^\infty$ -norm	$L^2$ -norm
$m = 0$	$2.751 \times 10^{-1}$	$1.085 \times 10^{-2}$
$m = 5$	$1.308 \times 10^{-1}$	$5.708 \times 10^{-3}$
$m = 10$	$1.333 \times 10^{-1}$	$5.807 \times 10^{-3}$
$m = 20$	$1.232 \times 10^{-1}$	$5.465 \times 10^{-3}$
$m = 30$	$1.224 \times 10^{-1}$	$5.408 \times 10^{-3}$

Saddle point	$L^\infty$ -norm	$L^2$ -norm
$m = 0$	$1.023 \times 10^{-1}$	$4.347 \times 10^{-3}$
$m = 5$	$9.715 \times 10^{-2}$	$3.389 \times 10^{-3}$
$m = 10$	$9.573 \times 10^{-2}$	$3.867 \times 10^{-3}$
$m = 20$	$6.022 \times 10^{-2}$	$2.996 \times 10^{-3}$
$m = 30$	$5.705 \times 10^{-2}$	$2.896 \times 10^{-3}$

(b)  $L^\infty$  and  $L^2$  norm of the error.

Figure 1.3.1: Application 2: Probability density function,  $L^\infty$  norm and  $L^2$  norm of the error  $\|s(\xi) - \tilde{s}(\xi)\|_Z$  estimated from a samples set of size  $10^4$ .

**Quality of error estimate** We discuss now the quality of the error estimate  $\Delta(\xi)$  for the variable of interest. We consider

$$\Delta(\xi) = \|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_{V_0} \|L(\xi)^* - A(\xi)^*Q_k(\xi)\|_{Z' \rightarrow V_0'} \quad (1.3.33)$$

for the primal-dual method, and

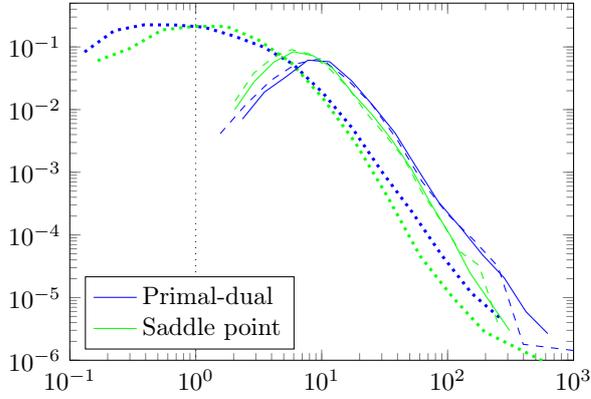
$$\Delta(\xi) = \|P_m(\xi)(A(\xi)t_{r,p}(\xi) - b(\xi))\|_{V_0} \sup_{0 \neq z' \in Z'} \inf_{y \in T_p} \frac{\|L(\xi)^*z' - A(\xi)^*y\|_{V_0'}}{\|z'\|_{Z'}} \quad (1.3.34)$$

for the saddle point method. Figure 1.3.2 shows statistics of the effectivity index  $\eta(\xi) = \Delta(\xi) / \|s(\xi) - \tilde{s}(\xi)\|_Z$  for different numbers  $m$  of interpolation points for the preconditioner. We see that the max-min ratio and the normalized standard deviation are decreasing with  $m$ : this indicates an improvement of the error estimate. Furthermore, the mean value of  $\eta(\xi)$  seems to converge (with  $m$ ) to 19.5 for the primal-dual method, and to 13.8 for the saddle point method. In fact, with a good preconditioner,  $\|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_{V_0}$  (or  $\|P_m(\xi)(A(\xi)t_{r,p}(\xi) - b(\xi))\|_{V_0}$ ) is expected to be a good approximation of the primal error  $\|u(\xi) - u_r(\xi)\|_{V_0}$  (or  $\|u(\xi) - t_{r,p}(\xi)\|_{V_0}$ ), but this does not ensure that the effectivity index  $\eta(\xi)$  will converge to 1.

### 1.3.4 Summary

We have proposed and analyzed projection based methods for the estimation of vector-valued variables of interest in the context of parameter-dependent equations. This includes a generalization of the classical primal-dual method to the case of vector-valued variables of interest, and also a Petrov-Galerkin method based on a saddle point problem.

The error estimates proposed and used during the greedy enrichment of the reduced spaces, involve the use of Cauchy-Schwarz inequalities, which are clearly not optimal. To overcome this limitation, randomized a posteriori error estimator for low-rank approximations, reliable and efficient at given high probability have been proposed recently in [150].



(a) PDF of  $\eta(\xi)$  for the primal-dual methods and the saddle point methods. Three different preconditioners  $P_m(\xi)$  are used:  $m = 0$  (dotted lines),  $m = 10$  (dashed lines) and  $m = 30$  (continuous lines)

	$\mathbb{E}(\eta(\xi))$	$\frac{\max \eta(\xi)}{\min \eta(\xi)}$	$\frac{\text{Var}(\eta(\xi))^{1/2}}{\mathbb{E}(\eta(\xi))}$
Primal-dual	$m = 0$	5.545	$3.52 \times 10^3$
	$m = 5$	16.03	$8.68 \times 10^2$
	$m = 10$	18.69	$1.01 \times 10^3$
	$m = 20$	19.20	$5.77 \times 10^2$
	$m = 30$	19.59	$3.95 \times 10^2$
Saddle point	$m = 0$	4.726	$6.93 \times 10^3$
	$m = 5$	12.61	$1.80 \times 10^2$
	$m = 10$	13.27	$1.72 \times 10^2$
	$m = 20$	13.97	$1.89 \times 10^2$
	$m = 30$	13.84	$2.17 \times 10^2$

(b) Statistics of the effectivity index  $\eta(\xi)$  for the primal-dual method and the saddle point method.

Figure 1.3.2: Application 2: PDF, mean, max-min ratio and normalized standard deviation of the effectivity index  $\eta(\xi) = \Delta(\xi) / \|s(\xi) - \tilde{s}(\xi)\|_Z$ . Here,  $\Delta(\xi)$  is defined by (1.3.33) for the primal-dual method and by (1.3.34) for the saddle point method.

# Chapter 2

## DYNAMICAL LOW-RANK APPROXIMATION

*In this chapter, model order reduction methods introduced previously are extended for the solution of parameter and time-dependent problems. Among these methods, we particularly focus on dynamical low-rank approximation approaches.*

*After a brief discussion on possible low-rank approximation methods for model order reduction of time-dependent and parameter-dependent problems in Section 2.1, we present in Section 2.2 dynamical reduced basis method that can be interpreted as a dynamical low-rank approximation method with a subspace point of view. Then, in Section 2.3, we present a new splitting integration scheme suitable for dynamical low-rank approximation in the set of fixed rank matrices.*

### 2.1 General context

Here, we focus on parameter and time-dependent problems. Typical problems could either be ordinary differential equations (ODEs), or partial differential equations (PDEs) depending on some (random) parameters. When PDEs are considered, we apply a method of lines approach, that is, we first discretize in space such that the problem is approximated by a system of ODEs. Thus, we are concerned with the solution of parameter-dependent non-autonomous dynamical systems of the form

$$\begin{cases} u'(t, \xi) = f(u(t, \xi), t, \xi), & t \in (0, T], \\ u(0, \xi) = u^0(\xi), \end{cases} \quad (2.1.1)$$

where the flux  $f$  and initial condition  $u^0$  depend on some (random) parameters  $\xi = (\xi_1, \dots, \xi_p)^T$  with values in a parameter set  $\Xi \subset \mathbb{R}^p$ . We denote  $T > 0$  the final time. We assume that the solution  $u(t, \xi)$  belongs to some vector space  $V$  for all  $t \in [0, T]$ .

#### 2.1.1 Low-rank approximation for parameter and time-dependent problems

In Chapter 1, we have seen that low-rank approximation methods are effective approaches for solving time-independent parameter-dependent problems (1.1.1). Seeking a solution  $u$  to the problem (2.1.1) leads to similar difficulties in term of complexity and computation costs. Here, we discuss how to extend such low-rank approximation methods to compute an approximation of the *solution map*  $\xi \mapsto u(t, \xi)$  for all  $t \in [0, T]$ .

##### Low-rank projection based approaches

In the context of RB methods (see, e.g., [99, 138]), the reduced approximation is classically obtained as the (Petrov-)Galerkin projection of the solution  $u$  of (2.1.1) onto a *time-independent* sub-space  $V_r$  of  $V$ . It writes under following form

$$u(t, \xi) \approx \sum_{i=1}^r v_i \alpha_i(t, \xi), \quad (2.1.2)$$

where  $\{v_1, \dots, v_r\}$  is a time-independent basis of  $V_r$  and  $\{\alpha_1(t), \dots, \alpha_r(t)\}$  are elements of some vector space  $S$  of functions defined on  $\Xi$ . Such an approximation, can be seen as a rank- $r$  approximation in the tensor space  $V \otimes S$  at each time  $t$ . If the flux  $f$  admits a low-rank representation, the approximation (2.1.2) can be computed efficiently by solving a reduced dynamical system of  $r$  equations satisfied by the coefficients  $\{\alpha_i(\xi)\}_{i=1}^r$ , using e.g. a time-stepping scheme. Sample-based methods have been proposed for the construction of  $V_r$  (see, e.g., [16,32,108,129]). In [94],  $V_r$  is obtained as the span of snapshots  $u(t^k, \xi^i)$  (in both time and parameter) of the solution of the full-order model. However, for a high-dimensional space  $V$ , it is not feasible (and far from optimal) to retain a large number of snapshots. Then, one can rely on a POD of the snapshots matrix in order to extract subspaces which are optimal in a mean-square sense. In particular, adaptive construction of  $V_r$  can be performed through a *POD-greedy* algorithm using a posteriori error estimates [66,70,95,97,98,165].

As pointed in [138], efficiency of such methods is limited by the effectivity of error estimators that may show bad long-time effectivity. Moreover, for parameter and time dependent problems with complex dynamics (e.g. advection dominated or hyperbolic parameter-dependent equations), the optimal subspace on which to approximate the solution at each time instant can considerably change over time. In that context, classical RB methods may require very high-dimensional (time-independent) reduced spaces for computing an approximation given by (2.1.2) with prescribed accuracy.

PGD method has also been considered in [131,132] for the low-rank approximation of the solution of stochastic evolution equations. In this context, the approximation is sought under the form

$$u(t, \xi) \approx \sum_{i=1}^r v_i(t) \alpha_i(\xi), \quad (2.1.3)$$

where  $\{v_1(t), \dots, v_r(t)\}$  is a time-dependent basis of  $V_r(t)$  and  $\{\alpha_1, \dots, \alpha_r\}$  are elements of  $S$ . Approximation (2.1.3) is a rank- $r$  element in the tensor space  $V^{[0,T]} \otimes S$ . The resulting approximation can be seen as a projection of  $u(t, \xi) \in V$  onto a *time-dependent* sub-space  $V_r(t) = \text{span}\{v_1(t), \dots, v_r(t)\} \subset V$  which allows to well capture transient phenomena. Contrary to RB basis method previously discussed, the approximation in (2.1.3) is obtained with a global in time variational principle that is not optimal at each instant  $t$ . Let us mention, that reduced basis method based on Petrov-Galerkin space-time (PG-ST) formulations has also been introduced, see for example [152,158]. Such an approach provides a low-rank approximation of the form (2.1.3). At the discrete level, previous approaches differ from RB approaches since they do not rely on time-stepping scheme except in very particular cases [84].

## Dynamical low-rank approximation methods

Dynamical low-rank approximation (DLRA) methods (see survey [161, Section 9.5]) aim at computing an approximation of  $u$  under the form

$$u(t, \xi) \approx \sum_{i=1}^r v_i(t) \alpha_i(t, \xi). \quad (2.1.4)$$

The approximation (2.1.4) can be interpreted as a projection onto a *time-dependent* reduced space  $V_r(t)$  obtained through principles which are local in time (e.g. Dirac Frenkel principle). DLRA methods are particular low-rank tensor methods since they rely on computing a rank- $r$  approximation in the tensor space  $V \otimes S$  at each time  $t$ . They first appeared in [114,130] in the case of low-rank matrices and then extended to low-rank approximation in tensor format (see,

e.g., [121, 159]) for the solution of high dimensional time dependent problems. Let us mention that dynamically (bi-)orthogonal methods that provide approximations under the form (2.1.4), have been also proposed [44, 45, 80, 149] and analysed [128] for the numerical solution of time-dependent stochastic equations.

### 2.1.2 Contributions

The contributions [MBF11, MBF13, MBF14, MBF16] presented in the next sections are the result of collaborations with A. Falcó (Univ. CEU Cardenal Herrera) and A. Nouy (ECN, LMJL) on dynamical low-rank approximation methods for the solution of parameter dependent dynamical systems and T. Heuzé (ECN, GeM) for possible extension to hyperbolic parameter dependent conservation laws.

- We have proposed a dynamical RB method for the solution of parameter-dependent dynamical systems of the form (2.1.1). It relies on the greedy construction of time-dependent sub-spaces  $V_r(t) \subset V$  generated from samples of the solution of the full-order model at some selected parameter values. The resulting method can be interpreted as a DLRA method with a subspace point of view and a uniform control of the error over the parameter set. Some prospects for extension of such dynamical RB to parameter dependent hyperbolic conservation laws, within finite volume framework, are also discussed.
- Regarding to matrix ODE (e.g., arising from (2.1.1) after discretization of the parameter domain), the DLRA relies on approximation in the set of fixed rank matrices. However, the representation of any element of this set is not unique. To dodge this undesirable property, we have proposed a geometric description of the set of fixed rank matrices endowed with the structure of analytic principal bundle, relying on an explicit description of local charts. From this description, we have derived a new splitting integration scheme suitable for DLRA.

*The work concerning dynamical RB methods has been partially supported by the GdR MOMAS through the project REDYN (2015). The work concerning application to hyperbolic conservation laws has been founded through the project PEPS : DROME by the Cellule Energie du CNRS (2019), co-driven with T. Heuzé.*

## 2.2 Projection based dynamical low-rank approximation

*Notations.* We consider in this section the resolution of a dynamical system under the form (2.1.1) in algebraic setting i.e.  $V = \mathbb{R}^N$ . In what follows,  $V$  is equipped with canonical inner product  $\langle u, v \rangle := u^T v$  and associated norm  $\|u\|$ .

The dynamical RB approach developed in [MBF11] performs in two steps. During an *offline* step, the time-dependent subspaces are constructed in greedy fashion from the trajectories  $t \mapsto u(t, \xi)$  of the exact solution for the selected values of the parameter  $\xi$ . Then the low-rank approximation is defined as the Galerkin projection of the solution  $u(t, \xi)$  in the time dependent subspace, and computed for any new instance of the parameter  $\xi \in \Xi$  during the *online* stage. For the sake of brevity, we only detail the projection step and greedy construction of the time-dependent reduced spaces.

## 2.2.1 Construction of the approximation

We consider that for any  $t > 0$ , the *reduced space*  $V_r(t) \subset V$  is given and satisfies

$$\dim V_r(t) = r \quad \text{for all } t > 0. \quad (2.2.1)$$

We denote by  $\{v_1(t), \dots, v_r(t)\}$  the an orthonormal basis of  $V_r(t)$  and  $\mathbf{V}(t) = [v_1(t), \dots, v_r(t)] \in \mathbb{R}^{N \times r}$  the associated orthogonal matrix. Finally  $\mathbf{\Pi}_{V_r}(t) = \mathbf{V}(t)\mathbf{V}^T(t) \in \mathbb{R}^{d \times d}$  is the orthogonal projector onto  $V_r(t)$ . We seek an approximation  $u_r(t, \xi) \in V_r(t)$  of  $u(t, \xi)$  under the form

$$u_r(t, \xi) := \sum_{i=1}^r \alpha_i(t, \xi) v_i(t) \quad \text{for all } t > 0. \quad (2.2.2)$$

where  $\boldsymbol{\alpha}_r = (\alpha_1, \dots, \alpha_r)^T$  is solution of the small  $r$ -order system obtained through the Galerkin projection of (2.1.1) onto  $V_r(t)$

$$\begin{cases} \boldsymbol{\alpha}'_r(t, \xi) = f_r(\boldsymbol{\alpha}_r(t, \xi), t, \xi), \\ \boldsymbol{\alpha}_r(0, \xi) = \boldsymbol{\alpha}_r^0(\xi), \end{cases} \quad (2.2.3)$$

with  $\boldsymbol{\alpha}_r^0(\xi) = \mathbf{V}^T(0)u^0(\xi)$  and a reduced flux  $f_r$  defined by

$$f_r(\boldsymbol{\alpha}_r, t, \xi) = \mathbf{V}(t)^T f(\mathbf{V}(t)\boldsymbol{\alpha}_r, t, \xi) - \mathbf{V}(t)^T \mathbf{V}'(t)\boldsymbol{\alpha}_r. \quad (2.2.4)$$

We have the following error estimate [MBF11, Proposition 2.4].

**Proposition 2.2.1.** *Let assume that  $f : V \rightarrow V$  is Lipschitz continuous function. The error norm satisfies*

$$\|u(t, \xi) - u_r(t, \xi)\| \leq \Delta_r(t, \xi) \quad (2.2.5)$$

for all  $t \geq 0$ , where  $\Delta_r(t, \xi)$  is the solution of the ordinary differential equation

$$\begin{cases} \Delta'_r(t, \xi) = L[f](u_r(t, \xi))\Delta_r(t, \xi) + \|\mathbf{r}(t, \xi)\|, \\ \Delta_r(0, \xi) = \|u^0(\xi) - \mathbf{\Pi}_{V_r}(0)u^0(\xi)\|, \end{cases} \quad (2.2.6)$$

with  $\mathbf{r}(t, \xi) = \mathbf{\Pi}_{V_r^\perp}(t)(\mathbf{V}'(t)\mathbf{V}(t)^T u_r(t, \xi) - f(u_r(t, \xi), t, \xi))$ , the log-lipschitz constant defined as

$$L[f](v) = \sup_{u \in V, u \neq v} \frac{\langle u - v, f(u) - f(v) \rangle}{\|u - v\|^2}.$$

**Remark 2.2.2.** *Note that the last term in (2.2.3) takes into account the time dependency of the subspace  $V_r$ , otherwise  $\mathbf{V}'(t) = 0$  and we recover the usual projected dynamical system within the framework of classical RB method (see, e.g., [97]).*

**Remark 2.2.3.** *In practice, the coefficients  $\boldsymbol{\alpha}_r$  as well as the error bound  $\Delta_r$  are computed using an appropriate integration scheme. The resulting numerical method can be implemented in an efficient time-stepping offline/online fashion provided suitable assumptions on the flux (see [MBF11, Section 4] for details).*

## 2.2.2 Selection of sub-spaces

We present greedy algorithms for the construction of an increasing sequence of reduced spaces  $\{V_r\}_{r>0}$ . These spaces are generated from successive samples of the solution  $u$  of the full-order model at parameters values  $\{\xi^r\}_{r>0}$  which are selected adaptively using the an a posteriori estimate  $\Delta_r$  among a finite training set  $\tilde{\Xi} \subset \Xi$ .

**T-Greedy** We have proposed in [MBF11], an adaptive greedy strategy called *T-greedy* algorithm. A step  $r$  of this algorithm is as follows:

$$\xi^{r+1} \in \arg \max_{\xi \in \Xi} \Delta_r^{(0,T)}(\xi), \quad (2.2.7)$$

$$V_{r+1}(t) = V_r(t) + \text{span}\{u(t, \xi^{r+1})\}. \quad (2.2.8)$$

with the following global error bound  $\Delta_r^{(0,T)}(\xi) := \|\Delta_r(\cdot, \xi)\|_{(0,T),p}$ , where  $\|\cdot\|_{(0,T),q}$  denotes the natural norm in  $L^q(0, T)$  with  $q = 2, \infty$ .

**POD-Greedy** For the adaptive construction of time-independent reduced spaces  $V_r$ , we recall the POD-greedy algorithm [95]. It selects a new parameters value

$$\xi^{r+1} \in \arg \max_{\xi \in \Xi} \Delta_r^{(0,T)}(\xi).$$

then the trajectory  $u(\cdot, \xi^{r+1})$  of the full-order model is computed and the space  $V_r$  is enriched by the  $\ell$ -dimensional subspace generated by the first  $\ell$  POD modes of  $u(\cdot, \xi^{r+1}) - \Pi_{V_r} u(\cdot, \xi^{r+1})$ .

**Remark 2.2.4.** *If the error estimate is such that*

$$c_r \Delta_r(t, \cdot) \leq \|u(t, \cdot) - u_r(t, \cdot)\| \leq C_r \Delta_r(t, \cdot), \quad (2.2.9)$$

with  $0 < c_r \leq C_r < \infty$  some positive constants, the POD-greedy algorithm has been proved to converge with quasi-optimal rates [98].

**Advantages of the proposed approach**

- For time-dependent subspace  $V_r(t)$ , the Galerkin projection  $u_r$  of the reduced dynamical system interpolates the solution  $u$  for the instances of the parameters  $\xi^1, \dots, \xi^r$  selected with the T-greedy algorithm. Whereas, for time-independent subspace  $V_r$  obtained through the POD-greedy algorithm, the approximation  $u_r$  does not in general interpolate the function  $\xi \mapsto u(\cdot, \xi)$  for  $\xi^1, \dots, \xi^r$ .
- POD-greedy algorithm may lead to high-dimensional reduced spaces for reaching a desired accuracy. A typical example is the advection problem for which a very high-dimensional time-independent reduced space may be required to approximate the solution even for one instance of the parameter. We will see on numerical examples (see Section 2.2.3), that using time-dependent reduced spaces allows to overcome such a limitation.

### 2.2.3 Numerical applications

The proposed approach is applied for the numerical approximation of the solution of the following nonlinear partial differential equation defined on a open interval  $D \subset \mathbb{R}$  and a time interval  $I = (0, T)$

$$\frac{\partial}{\partial t} u(\xi) + \frac{\partial}{\partial x} (c(u(\xi), \xi) u(\xi)) + a(\xi) \frac{\partial}{\partial x} u(\xi) - \mu(\xi) \frac{\partial^2}{\partial x^2} u(\xi) = g(\xi), \quad \text{on } D \times I, \quad (2.2.10)$$

with appropriate boundary conditions and a parameter-independent initial condition  $u^0(x)$ . Here  $\xi = (\xi^1, \dots, \xi^d)$  denotes a random vector with values in  $\Xi$  and with independent components. The functions  $c : X \times \Xi \rightarrow \mathbb{R}$  and  $a : D \times \Xi \rightarrow \mathbb{R}$  will be specified in each test case. Finally,  $\mu : \Xi \rightarrow \mathbb{R}$  is a parameter-dependent coefficient and  $g : D \times I \times \Xi \rightarrow \mathbb{R}$  is a given source term. We consider an approximation of the solution of (2.2.10), still noted  $u$ , obtained with an appropriate scheme yielding a system of  $N$  ordinary differential equations of the form (2.1.1).

In what follows, we denote by MTI (resp. MTD) the method using time-independent (resp. time-dependent) reduced spaces.

## Advection equation

Let  $D = (0, 1)$  and  $I = (0, 0.2)$ . The first application is an advection equation  $\mu(\xi) = 0$ ,  $g(x, t, \xi) = 0$ ,  $c(u, \xi) = 0$  and  $a(x, \xi) = a(\xi) = a_0 + a_1\xi^1$ , with  $\xi^1 \sim U(-1, 1)$ ,  $a_0 = 1$  and  $a_1 = 0.5$ . We impose periodic boundary conditions and consider a finite difference upwind scheme with  $N = 2000$ , together with an explicit Euler time integration scheme. We consider two functions as initial conditions. One is a smooth function given by  $u_{cont}^0(x) = \frac{1}{\sqrt{2\pi}} \exp(-(\frac{x-0.6}{0.05})^2)$  whereas the second is a discontinuous function given by  $u_{disc}^0(x) = \mathbf{1}_{[0.1, 0.9]}(x)(\lfloor 3x \rfloor + \sin(10x))$ .

**Deterministic case** We first consider a deterministic problem with  $\xi = 0.65$  and compare the approximations obtained by projections on reduced spaces constructed in two different ways. In the first method (MTI), the reduced space  $V_r$  is time-independent and generated by the first  $r$  modes of the POD of the trajectory  $t \mapsto u(t, 0.65)$ , with  $1 \leq r \leq 200$ . In the second method (MTD), we consider the one-dimensional time-dependent space  $V_r(t) = \text{span}\{u(t, 0.65)\}$ ,  $r = 1$ . For obtaining a very accurate precision, we have shown that the MTI requires a reduced space with rather high dimension (i.e. 50) when  $u^0 = u_{cont}^0$ . This is worst when  $u^0 = u_{disc}^0$  where a reduced space with higher dimension (i.e. 200) is required to well approximate the discontinuous solution. Concerning the MTD, we always obtain relative errors up to the machine precision, in both cases, with only one-dimensional time dependent reduced space.

**General case** We now consider the parameter-dependent problem and compare the approximations obtained with MTI and MTD for a subspace of dimension  $r = 20$  generated from a training set of size 30. We estimate the expectation  $\mathbb{E}(E_q(\xi))$  and maximum  $\max_{\xi \in \Xi}(E_q(\xi))$  of the relative errors  $E_q = \|u_r - u\|_{L,q} / \|u\|_{L,q}$ , with  $q = 2, \infty$ , by the empirical mean and by the maximum of the values of  $E_q$  taken at 50 randomly chosen values of the parameters. These quantities are depicted on Figure 2.2.1 for different values of  $r$  and for both continuous and discontinuous initial conditions. For the same dimension of the reduced spaces, MTD clearly provides a more accurate approximation than MTI in particular when considering discontinuous initial condition.

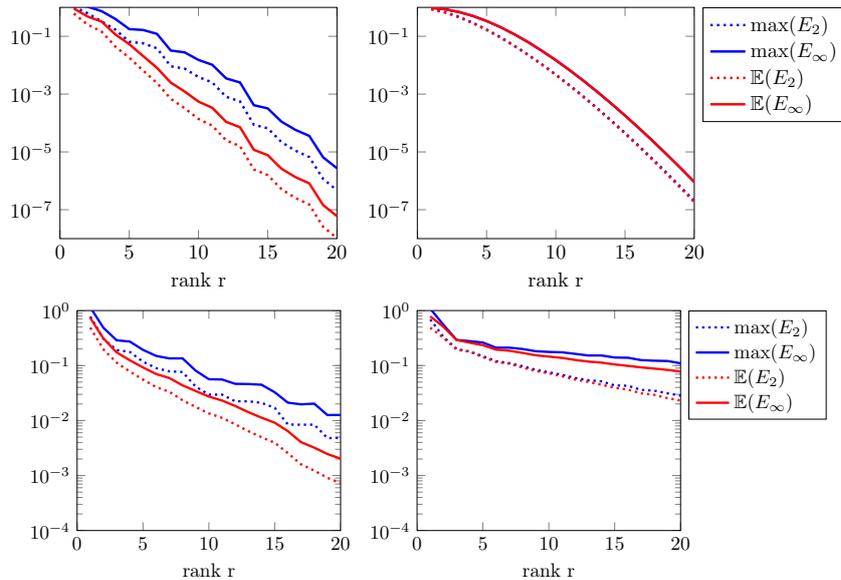


Figure 2.2.1: Advection equation: Statistical estimations of the expectation and maximum of the relative errors  $E_2$  and  $E_\infty$  with respect to the reduced dimension  $r$ , for MTD (left) and MTI (right) for both continuous (top) and discontinuous (bottom) initial conditions.

## Viscous Burger's equation

We study a nonlinear viscous Burger's equation with uncertain parameters [165]. We consider a spatial domain  $D = (0, 1)$ , a time interval  $I = (0, 1)$ , homogeneous Dirichlet boundary conditions and an initial condition  $u^0 = 0$ . We consider  $c(u, \xi) = \frac{1}{2}u$ ,  $a(x, \xi) = 0$  and a diffusion coefficient  $\mu(\xi) = \xi$ , with  $\xi \sim U(0.01, 0.06)$  a uniform random variable. The source term is defined by  $g(x, t, \xi) = g_1(x, t) + g_2(x, t)$ , with  $g_1(x, t) = 4e^{-\left(\frac{x-0.2}{0.03}\right)^2} \mathbf{1}_{[0.1, 0.3]}(x) \sin(4\pi t)$  and  $g_2(x, t) = 4 \cdot \mathbf{1}_{[0.6, 0.7]}(x) \mathbf{1}_{[0.2, 0.4]}(t)$ . We use a finite difference scheme in space ( $N = 298$ ) and a semi-implicit Euler scheme in time (with  $K = 200$  time steps). Reduced spaces are constructed with POD-greedy and T-greedy algorithms with a training set of size 60. The superiority of MTD over MTI is confirmed by Figure 2.2.2 where we have plotted statistical estimations (with a random sample of size 50) of the expectation and maximum of the relative errors. As we can see, MTD provides with  $r = 15$  an approximation with a relative error of  $10^{-10}$ , whereas MTI only provides an approximation with a relative error of  $10^{-2}$  for the same dimension of the reduced space. For  $r = 50$ , MTI provides an approximation with relative error  $10^{-6}$ , which is still higher than MTD with  $r = 15$ .

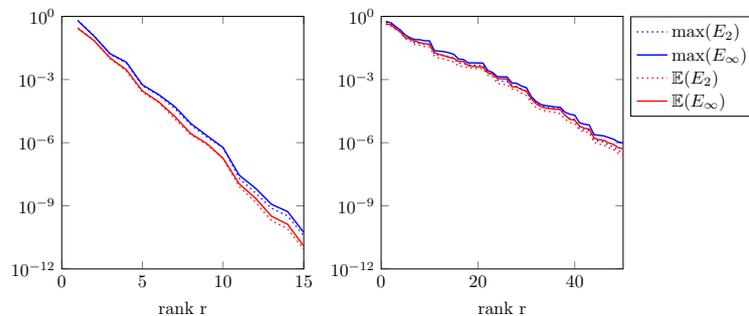


Figure 2.2.2: Burger's equation: Statistical estimations of the expectations and maxima of the relative errors  $E_2$  and  $E_\infty$  with respect to the reduced dimension  $r$ , for MTD (left) and MTI (right).

## 2.2.4 Prospects for parameter-dependent conservation laws

In this section, we give some elements of discussion concerning the extension of the proposed dynamical RB for parameter-dependent hyperbolic conservation laws. Let us consider the following one dimensional parameter-dependent scalar conservation equation. The space domain is the bounded open interval  $D \subset \mathbb{R}$ , and  $I = [0, T]$  is the time domain. The function  $u(\cdot, t; \xi)$ , belonging to the space  $V$ , is solution of the equation

$$\partial_t u(x, t, \xi) + \partial_x f(u(x, t, \xi), \xi) = 0, \quad (x, t) \in D \times I, \quad (2.2.11)$$

with initial condition  $u^0 : \bar{D} \rightarrow \mathbb{R}$  and suitable boundary conditions. The considered initial boundary value problem (2.2.11) depends on parameters  $\xi \in \Xi \subset \mathbb{R}^p$  through the conservative flux  $f(\cdot, \xi)$ .

### Context and challenges

It is now well understood that classical low-rank approximation methods are linear approximation methods which are not suitable for solving (parameter-dependent) hyperbolic problems because the solution manifold can not be well approximated with a linear space of small dimension  $r$ . Indeed, for such problems, the Kolmogorov  $r$ -width decreases slowly with  $r$  (see, e.g., [92, 164]). To overcome this issue, new ROM approaches have emerged this past years. A seminal work for dealing with transport dominated problem has been considered in [137] using a freezing method. It opened the route to various MOR approaches most of them relying on constructing suitable approximations from transformed snapshots see, e.g., [25, 34, 141, 145, 146, 164].

Moreover, defining suitable projection based ROM methods providing approximations that preserve as much as possible the mathematical features of solutions (conservative, entropic, monotonic, total variation diminishing...) of hyperbolic PDEs remain an opened question. In this direction, it was proposed in [1, 2] an approximation problem based on the minimization of the residual of the discretized equations in  $L^1$ -norm for ROM of hyperbolic conservation laws. It has the advantage to provide non oscillatory approximation, especially in presence of shocks. Projection-based hyper-reduced models of nonlinear conservation laws globally conservative that inherits a semi-discrete entropy inequality has also been proposed in [42].

### Transformed dynamical reduced basis

As illustrated in Section 2.2.3, the dynamic RB method discussed previously has shown to be interesting candidate when dealing with parameter-dependent transport dominated problems. We have considered, within the project PEPS DROME, to combine this approach with adapted local basis, with respect to the parameter, for solving Equation (2.2.11). The idea is to design reduced spaces spanned by local basis functions  $\{\phi_i(x, t; \xi)\}_{i=1}^r$  deduced from  $\{u(\varphi_i(x, t, \xi), t), t; \xi_i)\}_{i=1}^r$ , which correspond to transformed snapshots of the solution for given parameter instances with  $\varphi(\cdot, t, \xi, \xi_i) : D \rightarrow D$  a parameter-dependent space transformation at the instant  $t$ . Then, the solution  $u(x, t, \xi)$  of Equation(2.2.11) is approximated by the following *rank-r* approximation

$$u(x, t, \xi) \approx \sum_{i=1}^r \alpha_i(t, \xi) \phi_i(x, t; \xi). \quad (2.2.12)$$

From a physical viewpoint, a relevant choice is to derive the parameter-dependent transformations  $\varphi(\xi, \xi_i)$  from the characteristic associated with the hyperbolic system (2.2.11), such that the approximation (2.2.12) captures well the features of the original solution. Such approximation is exact with  $r = 1$  for parameter-dependent linear transport equation when the space transformation  $\varphi_i(\xi, \xi_i)$  is a suitable chosen parameter and time dependent space shift (see, e.g., [146, Example 3.6]). Thus, efficiently computing the transformed snapshots of the trajectory of the true solution for given instances of the parameter is crucial in that context.

### Reconstruction of finite volume solution from snapshots

In the paper [MBF16], we have focused on the design of a robust approximation for discontinuous solutions generated by parameter-dependent linear hyperbolic systems. The proposed approach is derived in Finite Volume (FV) framework where the snapshots consist of trajectories of the numerical approximation provided by a known FV scheme. To this end, the Reconstruct-Translate-Average (RTA) method have been introduced, which is inspired from Godunov's method also interpreted as Reconstruct-Evolve-Average (REA) [120]. The proposed method is detailed and analyzed for one dimensional parameter-dependent linear scalar transport equation, and an application to system of linear hyperbolic conservation laws is shown. In that case, one snapshot is sufficient to approximate the FV solution for any parameter values. This leads to a simple and efficient method that does not require any projection step neither time stepping procedure. The proposed approach has been applied for the reconstruction of solutions for the parameter-dependent transport problem and the wave equation.

### 2.2.5 Summary and discussion for future research

We have discussed previously projection-based model order reduction approach for the solution of parameter-dependent dynamical systems. This generalizes classical reduced basis methods with time-independent reduced spaces and performs better in practice. In particular, the proposed T-Greedy provides smaller reduced spaces  $V_r(t)$  allowing reduced computational cost

during the online stage.

Here are some some perspectives and questions that follow from the presented contributions.

### Towards nonlinear approximation.

- Obviously the RTA approach proposed in [MBF16] is only a tool to compute efficiently a robust approximation of the finite volume solution of linear conservation laws, for any parameter values, from precomputed snapshots. It is a first step toward efficient nonlinear ROM strategies, in particular to design dynamical RB method with adapted local basis in finite volume framework e.g. from some precomputed dictionary. In the lines of [34, 146], using RTA approach together with efficient strategies to compute both transformations and reduced basis would be interesting for attacking general hyperbolic conservation laws.
- More generally, designing approximation methods for solving parameter-dependent evolution equations has become an active field of research in the MOR community leading to the emergence of *nonlinear* approximation methods. Beyond approximation relying on transformed snapshots, many effort has also been put on finding suitable approximation format for attacking the approximation of the solution itself but the question of constructing such an approximation, when only the equation is available, remains a challenging task. In this direction, recent approaches involving machine learning techniques such as e.g. Neural Galerkin schemes [33], that are dynamical nonlinear approximation methods, seem appealing.

**Randomized linear algebra for dynamical reduced basis.** Despite the proposed time-dependent RB approach yields a significant reduction of online computation costs, it requires during the offline step the storage of the whole trajectory of the snapshots as well as related reduced quantities (matrices, error estimates etc.) which can be limiting for very large complex numerical model. Recently, novel projection methods based on randomized linear algebra have been proposed in [10, 11]. The idea behind is to approximate the ROM from its random sketch, which is a set of low-dimensional random projections of the reduced approximation space and the spaces of associated residuals. An interesting fact is that such an approach can be used for reducing both complexity and memory requirements of classical ROM approaches. Moreover, it allows to provide sharp error estimates that can be computed efficiently [106, 150, 151] and improve projection stability [10]. It would be of interest of extending such approaches within the framework of dynamical RB proposed in this section for parameter and time dependent problem.

## 2.3 Geometry based dynamical low-rank approximation

We now focus on low-rank approximation of time-dependent matrices  $A(t) \in \mathbb{R}^{n \times m}$ . Introducing  $\dot{A}(t) = \frac{d}{dt}A(t)$  the time derivative, the matrix  $A(t)$  is defined as the solution of the following matrix ODE

$$\dot{A}(t) = F(A(t), t), \quad A(0) = A^0, \quad (2.3.1)$$

given  $A^0 \in \mathbb{R}^{n \times m}$  and  $F : \mathbb{R}^{n \times m} \times [0, T] \rightarrow \mathbb{R}^{n \times m}$ . In view of MOR, the goal of DLRA methods is to approximate the solution  $A(t)$  of (2.3.1) with a matrix  $Z(t)$ , cheaper to compute, which belongs to the nonlinear manifold of fixed rank matrices

$$\mathcal{M}_r(\mathbb{R}^{n \times m}) = \{Z \in \mathbb{R}^{n \times m} : \text{rank}(Z) = r\},$$

where  $r \ll \min(n, m)$  stands for the rank. When  $A(t)$  is known,  $Z(t)$  can be defined as the best rank- $r$  approximation solution of

$$Z(t) = \arg \min_{W \in \mathcal{M}_r(\mathbb{R}^{n \times m})} \|A(t) - W\|, \quad (2.3.2)$$

with  $\|\cdot\|$  the Frobenius norm. In that case,  $Z$  is obtained through a truncated SVD of  $A(t)$  for each instant  $t$ . Nevertheless, as  $A$  is implicitly given by (2.3.1), it is more relevant to introduce low-rank approximation using  $\dot{A}$ . To that goal, the approximation  $Z$  is classically obtained through its derivative  $\dot{Z}$  which satisfies the Dirac-Frenkel variational principle

$$\dot{Z}(t) = \arg \min_{\delta W \in T_{Z(t)}\mathcal{M}_r(\mathbb{R}^{n \times m})} \|\delta W - F(Z(t), t)\|, \quad (2.3.3)$$

given  $Z(0) = Z^0 \in \mathcal{M}_r(\mathbb{R}^{n \times m})$  the best rank- $r$  approximation of  $A(0)$  and  $T_{Z(t)}\mathcal{M}_r(\mathbb{R}^{n \times m})$  the tangent space to  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  at  $Z(t)$ . Equivalently,  $\dot{Z}(t)$  corresponds to the orthogonal projection of  $F(Z(t), t)$  on the solution dependent tangent space, i.e.

$$\dot{Z}(t) = P_{T_{Z(t)}}F(Z(t), t), \quad Z(0) = Z^0, \quad (2.3.4)$$

where  $P_{T_Z}$  denotes the projection onto  $T_{Z(t)}\mathcal{M}_r(\mathbb{R}^{n \times m})$ .

**Towards numerics.** One way to deal with the numerical integration of (2.3.4) is to use projection based methods (see, e.g., [113]), in the lines of Riemannian optimization. Such methods work as follows. Perform one step of the numerical scheme (e.g. Runge Kutta) leaving  $\mathcal{M}_r(\mathbb{R}^{n \times m})$ , and then project back to it by means of retraction. This latter step is usually performed using a  $r$ -terms truncated SVD. As these approaches work directly in the ambient space  $\mathbb{R}^{n \times m}$ , they do not exploit the geometry of the set of fixed rank matrices  $\mathcal{M}_r(\mathbb{R}^{n \times m})$ . Following [114], we adopt a second point of view and propose numerical integration of the differential equation (2.3.4) using suitable geometric description of  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  which requires to deal with the subsequent difficulties.

1. The first difficulty relies on the proper geometric description of the set of fixed rank matrices  $\mathcal{M}_r(\mathbb{R}^{n \times m})$ . In practice, a way to compute the rank- $r$  matrix  $Z(t)$  is done through its parametrization

$$Z(t) = U(t)G(t)V(t)^T, \quad (2.3.5)$$

with  $U(t) \in \mathbb{R}^{n \times r}$ ,  $V(t) \in \mathbb{R}^{m \times r}$  and  $G(t) \in \mathbb{R}^{r \times r}$ . Such a parametrization of the matrix  $Z(t)$  is not unique. A way to dodge this undesirable property is to properly define the tangent space  $T_{Z(t)}\mathcal{M}_r(\mathbb{R}^{n \times m})$  imposing the so-called *gauge conditions* on  $U, V$ . In that case, the matrix  $Z(t)$  admits a unique decomposition (see [114, Proposition 2.1]) of the form (2.3.5). In addition, the system (2.3.4) results in a system ODEs driving the evolution of the parameters  $U, G, V$ .

2. The second difficulty appears when numerical integration is performed for solving the resulting system of ODEs governing the evolution of parameters  $U, V$  and  $G$  [114, 130]. In presence of small singular values for  $Z(t)$ , the matrix  $G(t)$  may be ill-conditioned. As consequence, classical integration schemes may be unstable for large rank and time steps (see, e.g., [112, Section 2.1]). Moreover, in case of *overapproximation*, i.e. when the approximation  $Z(t)$  has a rank  $r$  greater than the rank of the exact solution  $A(t)$ , these methods fail since  $G(t)$  becomes singular.

Using a chart based geometric description of  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  relying on a natural parametrization of matrices, we have proposed numerical integrator in local coordinates for solving (2.3.4), suitable for MOR.

### 2.3.1 Geometry of the set of fixed-rank matrices

In what follows, we present briefly the geometric description of the fixed-rank matrix manifold and the associated tangent space proposed in [MBF13].

Any matrix  $Z \in \mathcal{M}_r(\mathbb{R}^{n \times m})$  can be represented (in non-unique way) by means of the factorization

$$Z = UGV^T, \quad (2.3.6)$$

with  $U \in \mathcal{M}_r(\mathbb{R}^{n \times r})$ ,  $V \in \mathcal{M}_r(\mathbb{R}^{m \times r})$  and  $G \in \text{GL}_r$  with  $\text{GL}_r$  the Lie group of  $r \times r$  invertible matrices. We consider  $U_\perp \in \mathcal{M}_{n-r}(\mathbb{R}^{n \times (n-r)})$ ,  $V_\perp \in \mathcal{M}_{m-r}(\mathbb{R}^{m \times (m-r)})$ , the matrices such that  $U_\perp^T U = 0$  and  $V_\perp^T V = 0$ . The *neighborhood*  $\mathcal{U}_Z$  of  $Z$  in  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  is defined as the set

$$\mathcal{U}_Z = \{(U + U_\perp X)H(V + V_\perp Y)^T : (X, Y, H) \in \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \text{GL}_r\}.$$

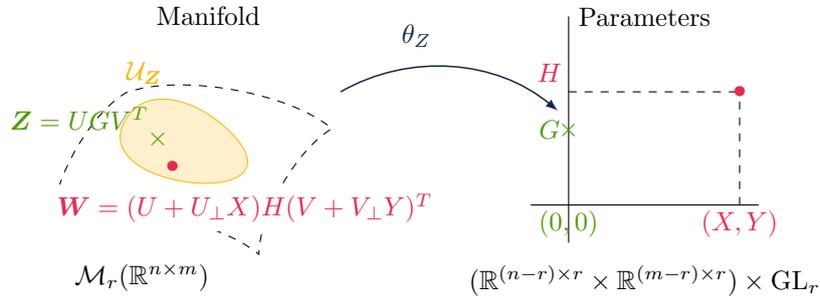


Figure 2.3.1: Representation of the local chart  $\theta_Z$  that associates to  $W = (U + U_\perp X)H(V + V_\perp Y)^T$  in  $\mathcal{U}_Z \subset \mathcal{M}_r(\mathbb{R}^{n \times m})$  the parameters  $(X, Y, H) \in \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \text{GL}_r$ .

We associate to the neighborhood  $\mathcal{U}_Z$  of  $Z$  the *local chart*  $\theta_Z : \mathcal{U}_Z \rightarrow \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \text{GL}_r$  (see Figure 2.3.1) which is given by

$$\theta_Z(W) = (U_\perp^+ W (V^+)^T (U^+ W (V^+)^T)^{-1}, V_\perp^+ W^T (U^+)^T (V^+ W^T (U^+)^T)^{-1}, U^+ W (V^+)^T)$$

for any  $W \in \mathcal{U}_Z$ . Here  $U^+$  and  $V^+$  stand for the Moore-Penrose pseudo-inverses<sup>1</sup> of  $U$  and  $V$  respectively. This means that any matrix  $W$  belonging to the neighborhood  $\mathcal{U}_Z$  admits a unique parametrization

$$W = (U + U_\perp X)H(V + V_\perp Y)^T,$$

with parameters  $(X, Y, H) \in \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \text{GL}_r$ . In this description, the parameters are not longer  $U, V, G$  but  $X, Y, H$ .

Such geometric description confers the set  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  the structure of an analytic  $r(n + m - r)$ -dimensional manifold. Moreover,  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  is an analytic principal bundle with typical fiber  $\text{GL}_r$  and base<sup>2</sup>  $\mathbb{G}_r(\mathbb{R}^n) \times \mathbb{G}_r(\mathbb{R}^m)$  (see [MBF13, Theorem 7]).

We now define the tangent map at  $Z \in \mathcal{M}_r(\mathbb{R}^{n \times m})$  noted  $T_Z i$  by

$$\begin{aligned} T_Z i : \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \mathbb{R}^{r \times r} &\rightarrow \mathbb{R}^{n \times m}, \\ (\delta X, \delta Y, \delta H) &\mapsto U_\perp \delta X G V^T + U G (V_\perp \delta Y)^T + U \delta H V^T. \end{aligned}$$

<sup>1</sup>For any  $A \in \mathbb{R}^{n \times m}$ , the Moore-Penrose pseudo inverse is given by  $A = (A^T A)^{-1} A^T$ .

<sup>2</sup>Here  $\mathbb{G}_r(\mathbb{R}^p) = \{V_r \subset \mathbb{R}^p : \dim(V_r) = r\}$  denotes the Grassmann manifold.

Then, the tangent space to  $\mathcal{M}_r(\mathbb{R}^{n \times m})$  at  $Z$  is defined as the image through  $T_Z i$  of the tangent space in the local coordinates in  $\mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \mathbb{R}^{r \times r}$

$$T_Z \mathcal{M}_r(\mathbb{R}^{n \times m}) = \{U_\perp \delta X G V^T + U G (V_\perp \delta Y)^T + U \delta G V^T : \delta X \in \mathbb{R}^{(n-r) \times r}, \delta Y \in \mathbb{R}^{(m-r) \times r}, \delta G \in \mathbb{R}^{r \times r}\}.$$

As stated in [MBF13, Proposition 6],  $T_Z i$  is an isomorphism between  $T_Z \mathcal{M}_r(\mathbb{R}^{n \times m})$  and  $\mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \mathbb{R}^{r \times r}$ . Thus, any tangent matrix  $\delta Z \in T_Z \mathcal{M}_r(\mathbb{R}^{n \times m})$  admits a unique parametrization of the form

$$\delta Z = T_Z i(\delta X, \delta Y, \delta H) = U_\perp \delta X G V^T + U G (V_\perp \delta Y)^T + U \delta H V^T, \quad (2.3.7)$$

where  $(\delta X, \delta Y, \delta H) \in \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \mathbb{R}^{r \times r}$  are uniquely given through

$$\begin{aligned} \delta X &= U_\perp^+ \delta Z (V^+)^T G^{-1}, \\ \delta Y &= V_\perp^+ \delta Z^T (U^+)^T G^{-T}, \\ \delta H &= U^+ \delta Z (V^+)^T. \end{aligned} \quad (2.3.8)$$

### 2.3.2 Dynamical low-rank approximation

In the context of dynamical low-rank approximation, we recall that  $Z$  is given through the projected Equation (2.3.4). By definition of the tangent map (2.3.7), the tangent matrix is given by  $\dot{Z} = T_Z i(\dot{X}, \dot{Y}, \dot{H})$  where the parameters satisfies

$$(\dot{X}, \dot{Y}, \dot{H}) = T_Z i^{-1}(P_{T_Z} F(Z)) \quad (2.3.9)$$

with

$$P_{T_Z} F(Z) = P_U^\perp F(Z) P_V^T + P_U^T F(Z) (P_V^\perp)^T + P_U F(Z) P_V^T.$$

Here  $P_U = U U^+$ ,  $P_V = V V^+$  denote the projections associated to  $U, V$  respectively, and their related orthogonal projections  $P_U^\perp = I - P_U$ ,  $P_V^\perp = I - P_V$ . Equation (2.3.9) yields equivalently to the following system of ODEs on the parameters:

$$\begin{aligned} \dot{X} &= U_\perp^+ F(Z) (V^+)^T G^{-1}, \\ \dot{Y} &= V_\perp^+ F(Z)^T (U^+)^T G^{-T}, \\ \dot{H} &= U^+ F(Z) (V^+)^T. \end{aligned} \quad (2.3.10)$$

Unlike [114, Proposition 2.1], the proposed geometrical description allows to avoid the non uniqueness of the parametrization (2.3.5) without imposing some gauge conditions a posteriori through the tangent space. More precisely, if we impose  $(\dot{U}, \dot{V}, \dot{G}) = (U_\perp \dot{X}, V_\perp \dot{Z}, \dot{H})$  and assuming that  $U, V$  orthogonal, our parametrization naturally fulfills the gauge conditions  $U^T \dot{U} = 0$  and  $V^T \dot{V} = 0$  and we recover the following system of ODEs [114] governing the evolution of the factors  $U, G, V$

$$\begin{aligned} \dot{U} &= (I - U U^T) F(Z) V G^{-1}, \\ \dot{V} &= (I - V V^T) F(Z)^T U G^{-T}, \\ \dot{G} &= U^T F(Z) V. \end{aligned} \quad (2.3.11)$$

**Remark 2.3.1.** *Let us remark that if  $F : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$  is a  $\mathbb{C}^p$  vector field then (2.3.1) admits a unique maximal solution. With the proposed chart-based based description, the projected dynamical system (2.3.9) inherits such properties. Indeed,  $T_Z i^{-1} P_{T_Z} F : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{(n-r) \times r} \times \mathbb{R}^{(m-r) \times r} \times \mathbb{R}^{r \times r}$  is also a  $\mathbb{C}^p$  vector field [78, Theorem 2.6, Theorem 3.5].*

### 2.3.3 Integration schemes

In seminal work [122], and more recently in [38, 39], projection-splitting integrators have been proposed to deal with numerical integration of (2.3.4). They have the advantage to be accurate, robust and stable independently of the rank  $r$  and do not need any regularization in case of over-approximation. The integrator initially proposed in [122], is based on an explicit scheme using a Lie-Trotter splitting of the projection operator  $P_{T_Z(t)}$ . In addition to its simplicity, it has the advantage to remain robust in case of small singular values and especially for overapproximation as it avoids the inversion of  $G(t)$ . Following the same idea, we have derived in [MBF14] a first order numerical integrator in local coordinates for solving (2.3.4) that can be interpreted as a splitting integrator.

#### Symmetric splitting method

We first consider the setting of the classical description [114]. To perform time integration, a symmetric Lie-Trotter splitting method [122] is applied to Equation (2.3.3). This integration scheme relies on a decomposition of the projection  $P_{T_Z(t)}$  as follows

$$P_{T_Z(t)} = Q_1 - Q_2 + Q_3, \quad (2.3.12)$$

where  $Q_1$ ,  $Q_2$  and  $Q_3$  are three projections respectively defined by

$$Q_1 A = A P_V^T, \quad Q_2 A = P_U A P_V^T, \quad Q_3 A = P_U A, \quad (2.3.13)$$

for any  $A \in \mathbb{R}^{n \times m}$ . Using this splitting, one integration step from  $t_0$  to  $t_1$  starting from the factorized rank- $r$  matrix  $Z_0 = Z(t_0)$  under the form  $Z_0 = U_0 G_0 V_0^T$  reads as follows.

1. Integrate on  $[t_0, t_1]$  the  $n \times r$  matrix differential equation

$$\frac{d}{dt}(UG) = F((UG)V^T)V, \quad \dot{V} = 0,$$

with initial conditions  $(UG)(t_0) = U_0 G_0$ ,  $V(t_0) = V_0$ . Then set  $U_1$  and  $\hat{G}_1$  such that  $U_1 \hat{G}_1 = (UG)(t_1)$ .

2. Integrate on  $[t_0, t_1]$  the  $r \times r$  matrix differential equation

$$\dot{G} = -U^T F(UGV^T)V, \quad \dot{U} = 0, \quad \dot{V} = 0,$$

with initial conditions  $G(t_0) = \hat{G}_1$ ,  $U(t_0) = U_1$ ,  $V(t_0) = V_0$ . Set  $\tilde{G}_1 = G(t_1)$ .

3. Integrate on  $[t_0, t_1]$  the  $m \times r$  matrix differential equation

$$\frac{d}{dt}(VG^T) = F(U(VG^T)^T)^T U, \quad \dot{U} = 0,$$

with initial conditions  $(VG^T)(t_0) = V_0 \tilde{G}_1^T$ ,  $U(t_0) = U_1$ . Set  $V_1$  and  $G_1$  such that  $V_1 G_1^T = (VG^T)(t_1)$ .

After these steps, the obtained approximation is  $Z(t_1) = U_1 G_1 V_1^T$ . Each step corresponds to the integration of the right hand side of (2.3.4) associated to the projections  $Q_1$ ,  $Q_2$  and  $Q_3$  respectively (for details see [122, Lemma 3.1]).

## Chart based splitting method

In [MBF14], we have proposed a numerical integrator relying on the fibre bundle structure of the manifold of fixed rank matrices. Thus, the guiding idea is to perform some update of the parameters  $(X, Y, H)$ , instead of  $(U, G, V)$  directly, whose dynamic is governed by the system of ODEs (2.3.10). Working in a fixed neighborhood  $\mathcal{U}_Z$  of  $Z$ , the matrices  $U, V, G$  are fixed and Equation (2.3.10) writes equivalently

$$\begin{aligned}\dot{H} &= U^+ F(Z)(V^+)^T, \\ \dot{X}G &= U_{\perp}^+ F(Z)(V^+)^T, \\ \dot{Y}G^T &= V_{\perp}^+ F(Z)^T(U^+)^T.\end{aligned}\tag{2.3.14}$$

Then, we integrate the system (2.3.14) from  $t_0$  to  $t_1$  in three steps. Letting  $U(t) = U(t_0) + U(t_0)_{\perp}X(t)$ ,  $V(t) = V(t_0) + V(t_0)_{\perp}Y(t)$  and  $Z(t) = U(t)H(t)V(t)^T$ , we start from  $(X(t_0), Y(t_0), H(t_0)) = (0, 0, G(t_0))$  and we proceed as follows.

1. Integrate on  $[t_0, t_1]$  the  $r \times r$  matrix differential equation

$$\dot{H} = U^+ F(Z)(V^+)^T, \quad \dot{X} = 0, \quad \dot{Y} = 0,$$

with initial conditions  $X(t_0) = 0, H(t_0) = G(t_0)$  and  $Y(t_0) = 0$ . Set  $H_1 = H(t_1)$ .

2. Integrate on  $[t_0, t_1]$  the  $n \times r$  matrix differential equation

$$\dot{X}H = U_{\perp}^+ F(Z)(V^+)^T, \quad \dot{H} = 0, \quad \dot{Y} = 0,$$

with initial conditions  $X(t_0) = 0, H(t_0) = H_1$  and  $Y(t_0) = 0$ . Then set  $X_1 = X(t_1)$ .

3. Integrate on  $[t_0, t_1]$  the  $m \times r$  matrix differential equation

$$\dot{Y}H^T = V_{\perp}^+ F(Z)^T(U^+)^T, \quad \dot{H} = 0, \quad \dot{X} = 0.$$

with initial conditions  $X(t_0) = X_1, H(t_0) = H_1$  and  $Y(t_0) = 0$ . Then set  $Y_1 = Y(t_1)$ .

After these three steps, we obtain an approximation  $Z(t_1) := U_1 H_1 V_1^T$  with  $U_1 = U(t_0) + U(t_0)_{\perp}X_1$  and  $V_1 = V(t_0) + V(t_0)_{\perp}Y_1$ .

The chart based method can be interpreted as a Lie-Trotter splitting that relies on the following decomposition of the projection

$$P_{T_Z(t)} = P_1 + P_2 + P_3,\tag{2.3.15}$$

where  $P_1, P_2$  and  $P_3$  are three projections respectively defined by

$$P_1 A = P_U A P_V^T, \quad P_2 A = P_U^{\perp} A P_V^T, \quad P_3 A = P_U A (P_V^{\perp})^T,\tag{2.3.16}$$

for any  $A \in \mathbb{R}^{n \times m}$ . Here each term  $P_i$  of the projection is associated to the ODE solved at Step  $i$  (see [MBF14, Appendix 1]).

## Comparison of the two approaches

- The inversion of  $G$  in the symmetric splitting method (and  $H$  in the chart based splitting method) is avoided. This convenient choice allows to deal with the case of over-approximation.
- The symmetric splitting and the chart based methods differ by the update order. Indeed, the chart based method first updates  $H$  and then  $X, Y$  (or equivalently  $G$  and then  $U, V$ ) whereas the symmetric splitting updates  $UG$ , then  $G$  and finally  $VG$ . Moreover, Step 2 of the symmetric splitting method described in Section 2.3.3 can be interpreted as a backward evolution problem that can be ill-conditioned, as pointed out in [9, Section 5]. In the chart based method, the update of  $G$  at Step 1 is still a forward evolution problem due to our splitting choice.
- In practice, we provide practical formulation of those methods amenable for numerical use. Explicit approximation of the flux is performed leading first order accurate schemes. The two methods are proven to coincide and are exact for matrix approximation [MBF14, Lemma 3.11].

## 2.3.4 Numerical applications

We consider the approximation of the parameter-dependent Burger's viscous equation in one dimension. To that goal, let  $\Omega \times I = (0, 1) \times [0, 1]$  be a space time domain. We seek  $u(\xi)$  the solution of

$$\partial_t u(\xi) - \mu(\xi) \partial_{xx}^2 u(\xi) + u(\xi) \partial_x u(\xi) = f(\xi), \quad \text{on } \Omega \times I, \quad (2.3.17)$$

with the initial data  $u^0(\xi) : \Omega \rightarrow \mathbb{R}$  and supplemented with homogeneous Dirichlet boundary conditions. The solution  $u(\cdot; \xi)$  depends on the parameter  $\xi \in \mathbb{R}^3$  through the viscosity  $\mu(\xi) = \xi_1$ , the initial condition and the source term defined by means of the function  $f(\xi) : \Omega \times I \rightarrow \mathbb{R}$  given by

$$f(x, t; \xi) = \xi_2 \exp(-(x - 0.2)^2 / 0.03^2) \sin(\xi_2 \pi t) \mathbf{1}_{[0.1, 0.3]}(x).$$

The problem (2.3.17) is semi-discretized in space by means of finite difference (FD) schemes with  $n$  nodes and  $m$  instances of the parameter  $\xi$  such that we get the following dynamical system

$$\dot{X}(t) = LX(t) + h(X, t), \quad X(0) = 0, \quad (2.3.18)$$

with solution  $X(t) \in \mathbb{R}^{n \times m}$ . We denote  $L$  the operator obtained by tensorization of discrete Laplacian and diagonal matrix obtained from  $m$  instances of  $\xi_1$ . Moreover, we define the matrix valued function  $h$  obtained by the discrete version of the first derivative.

In what follows we denote KSL the symmetric approach and chart based approach our approach.

### Single parameter Burger's problem

In this section, the parameter  $\xi_1$  takes its values in  $[0.01, 0.06]$  while the others are fixed to  $\xi_2 = 4$  and  $\xi_3 = 0$ . We chose the following initial condition

$$u^0(x; \xi) = \sin(x \xi_1) e^{-100(x - 10 \xi_1)^2}.$$

and set  $n = 100$  and  $m = 60$ . For solving the matrix ODE given by Equation (2.3.18), we first confront the chart based and KSL algorithms for various ranks, fixing  $\Delta t = 10^{-4}$ . The approximations obtained are compared to a reference solution noted  $\{X_{ref}^k\}_{k=0}^K$  computed with

an explicit Euler scheme with  $\Delta t = 5 \cdot 10^{-6}$ .

On Figure 2.3.2, the evolution of  $e^k = \|X_{ref}^k - Z^k\|$  for both algorithms is studied for different ranks. As we can observe, both methods seem to provide an approximation with similar accuracy, except for ranks 15 and 20 where the chart based algorithm provides a slightly more accurate approximation than the KSL algorithm.

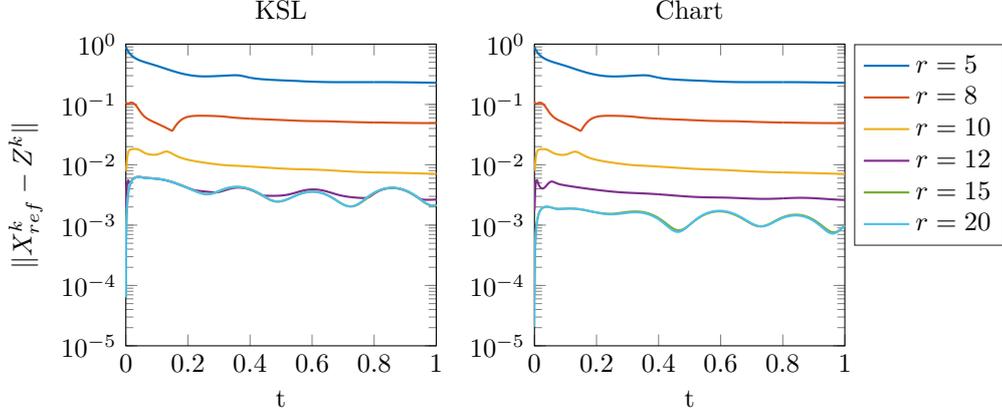


Figure 2.3.2: Burgers's equation: evolution of  $e^k$  for both algorithms for  $r \in \{5, 8, 10, 12, 15, 20\}$ .

This observation is confirmed by Figure 2.3.3 where we analyze the convergence with respect to rank and time step. As we can see, the error  $e^K = \|X_{ref}^K - Z^K\|$  decreases with respect to the rank and time step, and is the smallest (by up to one order of magnitude) for the chart based algorithm for larger rank.

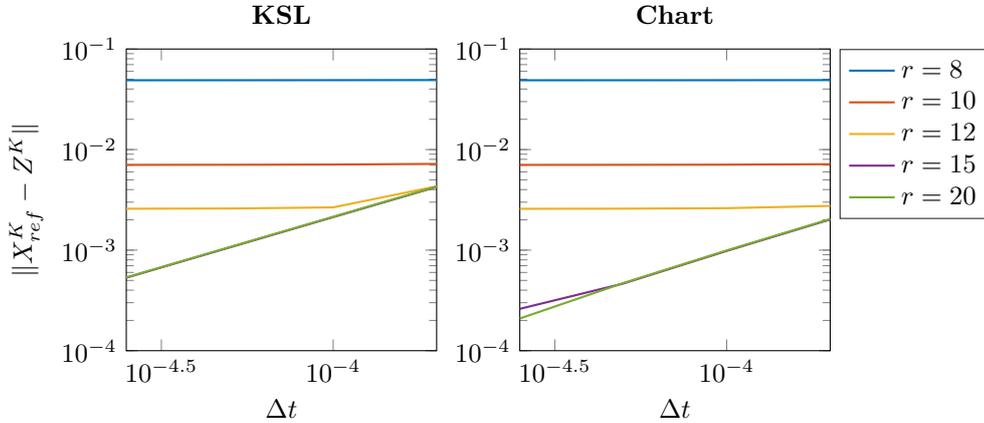


Figure 2.3.3: Burgers's equation: error  $e^K$  at final time for different  $(r, \Delta t)$ .

### Multiple varying parameters Burger's problem

To conclude this section, we illustrate the behavior of the two methods for the case where  $\xi$  is a vector of independent random parameters uniformly distributed on  $[0.01, 0.06] \times [2, 4] \times [0.01, 0.1]$  with  $u^0(x; \xi) = \xi_3 e^{-100(x-10\xi_2)^2}$ . The numerical simulations are performed for  $m = 60$  and  $n = 100$ . Here, the two KSL and Chart methods are run with  $\Delta t = 10^{-4}$  and compared to true numerical reference solution  $\{X_{ref}^k\}_{k=0}^K$  obtained with the explicit Euler scheme with the same time step. We represent on Figure 2.3.4 the approximation error at final time for the three approaches with respect to  $r \in \{5, 10, 12, 15, 20, 25, 30\}$ . We clearly observe that the chart method provides a better approximation than KSL for which the error seems to stagnate after  $r \geq 12$ .

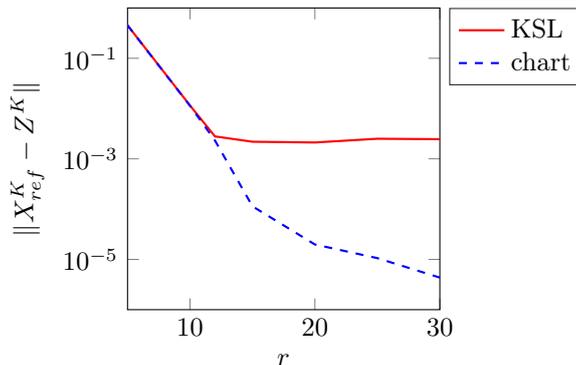


Figure 2.3.4: Burgers's equation (multiple parameter case): error  $e^K$  at final time for different  $r$ .

### 2.3.5 Summary and discussion for future research

We have introduced and compared some geometry based algorithms for dynamical low-rank approximation. Using a different geometry description of the set of fixed rank matrices relying on charts, we generalized the description of [114]. Then, from this description we derive a new splitting algorithm motivated by fibre bundle structure of the manifold of fixed rank matrices.

Let us now outline some possible perspectives and applications.

**Analysis of convergence order and stability of the splitting scheme.** The presented results demonstrate the feasibility of an alternative splitting naturally provided by the proposed geometric description. It would clearly be interesting to conduct further numerical experiments and theoretical analysis, to better understand the differences with KSL splitting algorithm, especially in term of accuracy with respect to time step and rank following [112] as well as a study of the influence of backward/forward integration for  $G$  on the scheme stability in the lines of [9, 110].

**Alternating scheme relying on the fibre bundle structure of the set of fixed rank matrices.**

The proposed splitting scheme is a first step towards designing new algorithms integrating the geometric structure of the manifold of fixed rank matrices, by working in neighborhoods. Especially, deriving a numerical scheme working alternatively with the fiber and base, (with possible parallelization as in [39] to enhance computation efficiency) should be interesting.

**Toward adaptive DLRA.** Presented methods compute low-rank approximations with a fixed rank  $r$ . This choice, done a priori, may not be optimal at each time  $t$ . Especially if the actual rank of  $X(t)$  varies with time (e.g. for applications arising from the discretization of a particular PDE). This is a first reason why allowing the rank of the approximation to evolve with time is interesting. Moreover, adapting the rank through the iterative process enables some computational time saving. DLRA with rank adaptivity has been the object of recent prospects. In [40] the authors propose a variant of the splitting integrator introduced in [39] with an additional truncation step, to allow-rank adaptivity.

- To our knowledge, rank adaptivity is introduced at numerical level through efficient algorithms in view of preserving some particular features as : exactness, robustness or symmetry properties. An interesting question concerns the possibility of formulating an adaptive DLRA method based on a global optimization in a "dynamical" low-rank manifold with suitable geometry description (e.g. integrating the rank).
- Finally, when considering the general case of tensors, as in [41, 60, 67, 147], one could

imagine the development of adaptive algorithms for the construction of DLRA that adapts dynamically not only the rank but also the tensor format, e.g., with dynamical choice of tree-based formats.

**Revisiting classical Riemannian optimization algorithm.** To finish, let us mention that the geometric description of the low-rank matrix manifold, proposed in [MBF13], was motivated within the context of DLRA for parameter-dependent dynamical systems. However, it can be carried over to optimization in the manifold of fixed rank matrices, in particular to revisit classical Riemannian optimization algorithm. This has been object of investigation during the internship of M. M. Abedin Nejad [3].

# Chapter 3

## APPROXIMATION FOR PDES WITH PROBABILISTIC INTERPRETATION

*In the previous chapters, we focused on low-rank approximation methods for the MOR of complex numerical models typically arising from the discretization of parameter-dependent PDEs. Here, we consider another direction by addressing directly the problem of constructing an approximation of the solution of the considered PDEs (possibly parameter-dependent) using approaches that rely on pointwise evaluations (possibly noisy) of the function to approximate.*

*This chapter is organized as follows. After giving a brief discussion on probabilistic approximation methods in Section 3.1, we give required notations and results concerning Monte Carlo method for estimating the solution of elliptic PDEs in Section 3.2. In section 3.3, we introduce a probabilistic sparse polynomial interpolation methods to deal with high dimensional equations. Then, in section 3.4, we discuss some probabilistic RB method for parameter dependent equations.*

### 3.1 General context

In this chapter, we focus on computing an approximation of the solution  $u : \bar{D} \rightarrow \mathbb{R}$  of the following continuous boundary problem (possibly parameter dependent)

$$\begin{aligned} -\mathcal{A}u &= g && \text{in } D, \\ u &= f && \text{on } \partial D, \end{aligned} \tag{3.1.1}$$

where  $D$  is an open bounded domain in  $\mathbb{R}^d$  with boundary  $\partial D$ . Here,  $\mathcal{A}$  is some linear elliptic partial differential operator and the functions  $f : \partial D \rightarrow \mathbb{R}$  and  $g : D \rightarrow \mathbb{R}$  are the source and boundary terms respectively. Here, we assume that  $u$  lies in  $V$ , some high dimensional vector space.

#### 3.1.1 Probabilistic approaches for approximation

When solving a problem of the form (3.1.1), standard numerical approaches (e.g. finite difference, finite element ...) are usually considered to provide a global approximation  $u_n$  of  $u$ , for a given mesh of  $D$ , in a finite dimensional space  $V_n \subset V$ . Such methods meet their limit when addressing parameter dependent problems, when many evaluations of the complex numerical solution  $u_n$  is required, or are no longer suitable in high-dimension. As discussed in previous chapters, approximations in low-rank tensor formats are good candidates to tackle those limitations. Classically, the presented methods are *deterministic* and rely on projection principles or optimization problems, that provide an approximation  $u_r$  of the complex numerical solution  $u_n$  in linear subspace  $V_r \subset V_n$  of small dimension. Here, we follow another path, by considering *probabilistic* approaches, in the sense that they integrate some randomness, which only use pointwise evaluations of  $u$ .

## Monte Carlo methods for approximation

Monte Carlo (MC) methods are well known to be efficient probabilistic approaches when computing numerically an integral or an expectation (see, e.g., monographs [88, 89]), especially in high dimension.

When turning to the computation of the solution of a PDE under the form (3.1.1), MC method allows to provide an estimate of the solution  $u(x)$  at any point  $x \in D$  without suffering from the curse of dimensionality. Such an estimate relies on the so-called *Feynman-Kac* probabilistic representation formula. In brief, for any  $x \in D$ ,  $u(x)$  can be interpreted as the expectation of some given stochastic functional depending on the operator  $\mathcal{A}$  and the data  $f, g$ . The main limitations of such approaches are that they only provide pointwise estimates of  $u$  and may suffer from slow convergence of MC estimate. To tackle the first limitation, it is possible to compute approximations in a given approximation format through classical interpolation or regression (see, e.g., [18, 19, 163]) within deep learning context. Let us mention, that it is also possible to consider some general approximation setting relying on the minimization of an empirical risk on a selected model class [68, 69]. The resulting method called *variational Monte Carlo method* is shown to be applicable to a broad range of problems including e.g. approximation using neural networks, or approximation in low-rank tensor format. Concerning the second difficulty, variance reduction techniques (see Section 3.2) can be considered. In particular let us mention that in [85, 86], the authors have proposed to combine some polynomial interpolation for reconstructing a global approximation  $u_n$  of  $u$  in some polynomial space  $V_n \subset V$  from MC estimates, such that the approximation error  $e_n = u - u_n$  can be used as control variate during some iterative procedure.

## Probabilistic approaches within MOR

In MOR, various probabilistic approaches have recently emerged in view of improving the efficiency of the underlying methods either in term of their approximation properties or their numerical complexity and stability.

Within the framework of RB method for parameter-dependent problems, the authors in [56] have proposed a probabilistic greedy algorithm that uses, at each step, different training sets with moderate cardinality, randomly chosen. In practice, the resulting method allows to deal efficiently with high dimensional parametric problems. In [35], the authors derive a similar probabilistic EIM using sequential sampling in the parameter set, which provides an interpolation with a prescribed precision with high probability. Probabilistic approaches can be also used to provide numerically stable and efficient error estimates for low-rank approximations, see, e.g., [150, 151] for contributions using randomized (a posteriori) error estimates. In [10, 11], this idea has been generalized within the framework of randomized linear algebra to enhance the efficiency and stability of projection-based model order reduction methods for solving parameter-dependent equations. In particular, it allows drastic computational savings. Finally, let us also mention that a control variate using a reduced basis paradigm, has been proposed in [26, 27] for MC estimate of the expectation of some parameter dependent random variable.

### 3.1.2 Contributions

Within this chapter, our main concern is the development of probabilistic approximation methods, as previously discussed, that use on pointwise estimates, over  $\bar{D}$ , of the solution  $u$  of (3.1.1). To that goal, probabilistic interpretation of PDEs using the *Feynman-Kac formula* was the com-

mon thread in contributions [MBF12, MBF15, MBF17] within the PhD of A. Macherey [123] supervised together with A. Nouy (LMJL, ECN) and C. Prieur (LJK, UGA). Especially, we have followed two subsequent directions.

- From pointwise estimates of the solution  $u$  over  $\bar{D}$ , we propose to compute a global approximation through polynomial interpolation overcoming possible limitations due to high dimension and slow convergence rates of Monte Carlo estimates.
- The idea is to propose a RB method in full sampling setting, i.e. that uses pointwise estimates of the solution  $u$  over  $D$  for given instances of the parameter. In particular, we propose a greedy algorithm using some probabilistic MC error estimate. In that case, the difficulty is to perform such algorithm keeping a reasonable budget (for MC sampling) in particular during the selection of the parameter and ensuring that the resulting procedure is a weak greedy algorithm with high probability.

## 3.2 Monte Carlo methods for PDEs

In this section, we introduce essential ingredients and notations concerning numerical probabilistic methods for PDEs that will be considered in the next sections. A deeper discussion on that topic can be found e.g. in [88].

*Notations.* In what follows,  $(\Omega, \mathcal{F}, \mathbb{P})$  denotes a probability space, with  $\Omega$  is the sample space,  $\mathcal{F}$  a  $\sigma$ -algebra and  $\mathbb{P}$  a probability measure.

### 3.2.1 Monte Carlo methods

Let  $Z$  be a random variable with finite expectation  $\mathbb{E}(Z)$  and variance  $\mathbb{V}(Z)$ . MC methods are simple methods that compute an estimate of  $\mathbb{E}(Z)$  through the empirical mean

$$\bar{Z}_M := \frac{1}{M} \sum_{m=1}^M Z_m, \quad (3.2.1)$$

where  $\{Z_1, \dots, Z_M\}$  are  $M$  independent random variables having the same distribution as  $Z$ . It is well known, that  $\bar{Z}_M$  is an unbiased estimate that converges almost surely to  $\mathbb{E}(Z)$ , by the strong law of large numbers. At this point, a natural question is to understand how good is the estimate  $\bar{Z}_M$  for the expectation  $\mathbb{E}(Z)$ , in which sense (e.g. error bounds via confidence interval, root mean square...) and as  $M \rightarrow \infty$ .

#### Error bounds via confidence intervals

It is possible to derive some asymptotic error bounds, through confidence intervals, of the error  $|\mathbb{E}(Z) - \bar{Z}_M|$ , as  $M \rightarrow \infty$ . Using Central Limit theorem (other asymptotic controls are also possible see [88, Chapter 2.3]), we get the following result. Let  $x \in (0, 1)$  and  $c(M, x) = z_x \sqrt{\frac{\bar{V}_M}{M}}$  with, we have

$$\mathbb{P}(|\bar{Z}_M - \mathbb{E}[Z]| \leq c(M, x)) \rightarrow 1 - x,$$

as  $M \rightarrow \infty$ . Here  $z_x \in \mathbb{R}$  satisfies  $F(z_x) = 1 - x/2$  with  $F$  the cumulative distribution function of the centered reduced normal distribution and  $\bar{V}_M = \frac{1}{M} \sum_{m=1}^M (Z_m - \bar{Z}_M)^2$  is the empirical variance. This result means that for  $M$  large enough, the error is bounded by  $c(M, x)$  with a

probability  $1 - x$ . In practice, the question of how to choose  $M$ , to use these error bounds, remains a delicate task. Alternatively, non-asymptotic error bounds, such that

$$\mathbb{P}(|\bar{Z}_M - \mathbb{E}[Z]| \leq c(M, x)) \geq 1 - x, \quad (3.2.2)$$

is valid for any value of  $M$ , are preferable. Some of this non-asymptotic error bounds are related to *concentration inequalities* (e.g. Hoeffding, Bernstein). We refer the interested reader to [28] for a detailed overview.

### Root mean square error and variance reduction techniques

Another relevant measure of the quality of the estimate (3.2.1) is given by

$$\mathbb{E}(|\mathbb{E}(Z) - \bar{Z}_M|^2) = \mathbb{V}(\bar{Z}_M) = \frac{\mathbb{V}(Z)}{M}. \quad (3.2.3)$$

This result says that  $\bar{Z}_M$  is an estimate of the expectation  $\mathbb{E}(Z)$  for which the root mean square error behaves like  $O(M^{-1/2})$ . This shows the limitations of the MC method that requires a large number of samples of  $Z$  to converge. In practice, this may induce high computation cost when the samples of  $Z$  are costly to compute (e.g. when  $Z$  is related to the evaluation of the solution of a PDEs).

The so-called *variance reduction* methods [88, Chapter 3], are aimed at improving the variance in (3.2.3). Among these approaches, let mention the *antithetic sampling*, *conditioning* or *stratification* and *importance sampling* methods. Some variance reduction methods use *control variate*. They consist in using an additional centered square integrable random variable  $Y$  that is simulated jointly with  $Z$  such that

$$\bar{Z}_M^c := \frac{1}{M} \sum_{m=1}^M (Z_m - Y_m), \quad (3.2.4)$$

with  $\{Y_1, \dots, Y_M\}$  is a  $M$ -sample of  $Y$ , is a new unbiased estimate of  $\mathbb{E}(Z)$ , that converges to  $\mathbb{E}(Z)$ . In practice,  $Y$  is chosen such that the new estimate variance

$$\mathbb{V}(\bar{Z}_M^c) = \mathbb{E}(|\mathbb{E}(Z) - \bar{Z}_M^c|^2) = \frac{\mathbb{V}(Z - Y)}{M},$$

is smaller than  $\mathbb{V}(Z)$ . A possible optimal choice, is to seek  $Y$  as a linear combination of finite number of random variables, that minimize  $\mathbb{E}((Z - Y)^2)$ . This approach has been extended to multilevel approaches that construct sequences of control variates via Multilevel Monte-Carlo method (MLMC) (see survey [82]). Let us mention that other reduction variance approaches, using control variates, have been also developed in the context of the numerical simulation of PDEs, possibly parameter-dependent (see, e.g., [26, 85, 86]).

### 3.2.2 Probabilistic representation of solution of PDEs

*Notations.* Let  $W = (W_t)_{t \geq 0}$  be a  $d$ -dimensional brownian motion, i.e.  $W = (W_1, \dots, W_d)^T$  where each  $W_i, i = 1, \dots, d$  are independent standard brownian motions defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  endowed with its natural filtration  $(\mathcal{F}_t)_{t \geq 0}$ .

Given  $D$  be an open bounded domain in  $\mathbb{R}^d$ , we consider the following boundary problem

$$\begin{aligned} -\mathcal{A}u &= g \quad \text{in } D, \\ u &= f \quad \text{on } \partial D, \end{aligned} \quad (3.2.5)$$

with solution  $u : \bar{D} \rightarrow \mathbb{R}$  and  $f : \partial D \rightarrow \mathbb{R}$ ,  $g : \bar{D} \rightarrow \mathbb{R}$  the boundary condition and source terms, respectively. Here, we assume that  $\mathcal{A}$  is linear and elliptic partial differential operator given by

$$\mathcal{A} = \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^T)_{ij} \partial_{x_i x_j}^2 + \sum_{i=1}^d b_i \partial_{x_i}. \quad (3.2.6)$$

It corresponds to the *infinitesimal generator* associated to the  $d$ -dimensional diffusion process  $X^x = (X_t^x)_{t \geq 0}$  which is the solution of the *stochastic differential equation*

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dW_t, \quad X_0^x = x \in \bar{D}, \quad (3.2.7)$$

where  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  are the drift and diffusion coefficients, respectively.

Before recalling Feynman-Kac formula, we first need to introduce additional assumptions and notations. Denoting by  $\|\cdot\|$  both euclidean norm on  $\mathbb{R}^d$  and Frobenius norm on  $\mathbb{R}^{d \times d}$ , we introduce the following assumption on  $b$  and  $\sigma$ .

**Assumption 1 (A1).** *There exists a constant  $0 < M < +\infty$  such that for all  $x, y \in \bar{D}$  we have*

$$\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \leq M\|x - y\|, \quad (3.2.8)$$

Under Assumption 1, by [57, Theorem 5.3], there exists an unique *strong* solution to the Equation (3.2.7).

Denoting  $a = \sigma \sigma^T$ , we introduce the following uniform ellipticity assumption.

**Assumption 2 (A2).** *There exists  $c > 0$  such that*

$$y^T a(x)y \geq c\|y\|^2, \quad y \in \mathbb{R}^d, \quad x \in \bar{D}.$$

As problem (3.2.5) is defined on a bounded domain, we have to define the *first exit time* of  $D$  for the process  $X^x$  as

$$\tau^x = \inf \{s > 0 : X_s^x \notin D\}. \quad (3.2.9)$$

Also, we assume some regularity property on the spatial domain  $\bar{D}$  and data.

**Assumption 3 (A3).** *The domain  $D$  is an open connected bounded domain of  $\mathbb{R}^d$ , regular according to the two following assumptions:*

i)

$$\mathbb{P}(\tau^x = 0) = 1, \quad x \in \partial D,$$

ii) *each point of  $\partial D$  satisfies the exterior cone condition which means that, for all  $x \in \partial D$ , there exists a finite right circular cone  $K$ , with vertex  $x$ , such that  $\bar{K} \cap \bar{D} = \{x\}$ .*

**Assumption 4 (A4).** *We assume that  $f$  is continuous on  $\partial D$ ,  $g$  is Hölder-continuous on  $\bar{D}$ .*

The following probabilistic representation theorem [57, Theorem 2.4] holds.

**Theorem 3.2.1** (Feynman-Kac formula). *Under assumptions (A1)-(A4) there exists a unique solution of (3.2.5) in  $\mathcal{C}(\bar{D}) \cap \mathcal{C}^2(D)$ , which satisfies for all  $x \in \bar{D}$*

$$u(x) = \mathbb{E} \left( f(X_{\tau^x}^x) + \int_0^{\tau^x} g(X_t^x) dt \right), \quad (3.2.10)$$

with  $X^x$  the stopped diffusion process solution of (3.2.7).

### 3.2.3 Numerical estimation of the solution of a PDEs

We now discuss the numerical computation of an estimate of  $u(x)$ . We perform a suitable integration scheme, to get an approximation of the stopped diffusion process  $X^x$  with a MC method to evaluate the expectation in formula (3.2.10).

#### Integration of diffusion process

An approximation of the diffusion process is obtained using an Euler-Maruyama scheme. More precisely, setting  $t_n = n\Delta t$ ,  $n \in \mathbb{N}$ ,  $X^x$  is approximated by a piecewise constant process  $X^{x,\Delta t}$ , where  $X_t^{x,\Delta t} = X_n^{x,\Delta t}$  for  $t \in [t_n, t_{n+1}[$  and

$$\begin{aligned} X_{n+1}^{x,\Delta t} &= X_n^{x,\Delta t} + \Delta t b(X_n^{x,\Delta t}) + \sigma(X_n^{x,\Delta t}) \Delta W_n, \\ X_0^{x,\Delta t} &= x, \end{aligned} \quad (3.2.11)$$

where  $\Delta W_n = W_{n+1} - W_n$  is an increment of the standard Brownian motion.

Numerical computation of  $u(x)$ , for all  $x \in \bar{D}$ , also requires the computation of a stopped process  $X^{x,\Delta t}$  at  $\tau^{x,\Delta t}$  an estimation of the first exit time of  $D$ . Here, we consider the simplest way to define this discrete exit time

$$\tau^{x,\Delta t} = \min \left\{ t_n > 0 : X_{t_n}^{x,\Delta t} \notin D \right\}. \quad (3.2.12)$$

Such a discretization choice may leads to over-estimation of the exit time with an error in  $O(\Delta t^{1/2})$ . More sophisticated approaches [88, Chapter 6] exist to improve the order of converge, as Brownian bridge, boundary shifting or Walk On Sphere (WOS) methods.

#### Monte Carlo estimate

Letting  $\{X^{x,\Delta t}(\omega_m)\}_{m=1}^M$  be independent samples of  $X^{x,\Delta t}$ , we obtain an the MC estimate noted  $u_{\Delta t, M}(x)$  for  $u(x)$  defined as

$$u_{\Delta t, M}(x) = \frac{1}{M} \sum_{m=1}^M \left[ f(X_{\tau^{x,\Delta t}}^{x,\Delta t}(\omega_m)) + \int_0^{\tau^{x,\Delta t}} g(X_t^{x,\Delta t}(\omega_m)) dt \right]. \quad (3.2.13)$$

#### Approximation error

Introducing the following pointwise estimate,

$$u_{\Delta t}(x) = \mathbb{E} \left( f(X_{\tau^{x,\Delta t}}^{x,\Delta t}) + \int_0^{\tau^{x,\Delta t}} g(X_t^{x,\Delta t}) dt \right),$$

the error is the sum of two contributions,

$$u(x) - u_{\Delta t, M}(x) = \underbrace{u(x) - u_{\Delta t}(x)}_{\varepsilon_{\Delta t}} + \underbrace{u_{\Delta t}(x) - u_{\Delta t, M}(x)}_{\varepsilon_{MC}}, \quad (3.2.14)$$

where  $\varepsilon_{\Delta t}$  is the time integration error and  $\varepsilon_{MC}$  is the Monte-Carlo estimation error.

Under assumptions (A1)-(A4), in particular assumption (A3-ii) which ensures that  $D$  does not have singular points, it can be proven [86, §4.1] that the time integration error  $\varepsilon_{\Delta t}$  converges to zero. It can be improved to  $O(\Delta t^{1/2})$  by adding differentiability assumptions on the boundary

[87]. The estimation error  $\varepsilon_{MC}$  is a random variable with zero mean and standard deviation converging as  $O(M^{-1/2})$ . The computational complexity for computing a pointwise evaluation of  $u_{\Delta t, M}(x)$  is in  $O(M\Delta t^{-1})$  in expectation for  $\Delta t$  sufficiently small, so that the computational complexity for achieving a precision  $\epsilon$  (root mean squared error) behaves as  $O(\epsilon^{-4})$ . This does not allow to obtain a very high accuracy in a reasonable computational time. The convergence with  $\Delta t$  can be improved to  $O(\Delta t)$  by suitable boundary corrections [87], therefore yielding a convergence in  $O(\epsilon^{-3})$ . Variance reduction methods, discussed in Section 3.2, can be used to further improve the convergence.

### 3.3 High dimensional PDEs

Keeping the same notations as in Section 3.2.2, we assume that  $u : \bar{D} \rightarrow \mathbb{R}$  is solution of (3.2.5). We recall that under the assumptions (A1)-(A4),  $u$  admits, by Theorem 3.2.1, the following probabilistic representation

$$u(x) = \mathbb{E}[F(u, X^x)] := \mathbb{E} \left( u(X_{\tau^x}^x) - \int_0^{\tau^x} \mathcal{A}u(X_t^x) dt \right), \quad x \in \bar{D}$$

where  $F$  depends on the data of the problem and  $X^x$  is the stopped diffusion process solution of Equation (3.2.7). An MC estimate of  $u(x)$  is

$$u_{\Delta t, M}(x) := \frac{1}{M} \sum_{m=1}^M F(u, X^{x, \Delta t}(\omega_m)).$$

with  $X^{x, \Delta t}$  the approximated stopped diffusion process obtained through Euler-Maruyama scheme.

Within this section, we consider the problem of reconstructing a global approximation of the solution  $u$  from pointwise estimates  $u_{\Delta t, M}(x)$  for any  $x \in \bar{D}$ . It is structured as follows. In the Section 3.3.1, we discuss possible strategies for global reconstruction of a function from pointwise evaluations. Then, we present the control variate method for high dimensional PDEs proposed in [MBF12]. Some numerical experiments illustrating the behavior of the proposed approach are given in Section 3.3.3.

#### 3.3.1 Approximation from pointwise evaluations

We consider the problem of constructing a global approximation  $u_n$  of  $u$ , in some finite dimensional linear space  $V_n \subset V$ , from pointwise evaluations of  $u$  at given points  $\Gamma_m = \{x_1, \dots, x_m\} \subset D$  with  $m \geq n$ . Two projection based strategies are classically used to construct  $u_n$  namely: the *interpolation* and *least-squares* methods.

##### Interpolation method

Well known interpolation methods include polynomial and spline interpolation as well as Kriging. Let  $u$  be continuous and  $\Gamma_n$  be an unisolvent grid of interpolation points (here  $m = n$ ) for  $V_n$ , Lagrange interpolation methods aim at constructing the approximation  $u_n$  of  $u$  as the unique function in  $V_n$  such that

$$u_n(x_i) = u(x_i), \quad x_i \in \Gamma_n.$$

Such an approximation satisfies the following error bound in  $L^\infty$ -norm

$$\|u - u_n\|_\infty \leq (1 + \mathcal{L}_n) \inf_{v \in V_n} \|u - v\|_\infty,$$

where  $\mathcal{L}_n$  stands for the Lebesgue constant. The choice of the points  $\Lambda_n$  is crucial to control the growth of the Lebesgue constant with respect to  $n$ . In dimension one, interpolation is quite well understood for polynomial approximation (see, e.g., [107, 153]). In that context, it is well known that taking uniformly distributed interpolation points leads to a Lebesgue constant  $\mathcal{L}_n$  that grows exponentially with  $n$ . A suitable choice of interpolation points, is the sequence of Chebyshev points for which  $\mathcal{L}_n$  grows only logarithmically with  $n$ . When turning to general multivariate domains with arbitrary linear approximation space (i.e.  $V_n$  is not necessary polynomial space), a general procedure consists in selecting the interpolation points by maximizing over  $D^n$  of the determinant of the collocation matrix. This provides the sequence of Fekete points, for which the Lebesgue constant grows linearly with  $n$ . However, it remains computationally untractable in high dimension as it requires to solve a multivariate non convex optimization problem. Another procedure, first proposed in the context of reduced basis [14], is to select interpolation points in greedy fashion providing sequence of nested interpolations points  $\Gamma_n \subset \Gamma_{n+1}$ . This method has a reasonable computational cost for moderate  $d$  and is independent on  $n$ . The resulting interpolation points, called *magic points* [124] coincide, in the univariate case for polynomial approximation, with *Leja* point. In that case, an upper bound of  $\mathcal{L}_n$  behaves quadratically with  $n$ . When turning to high dimensional problem, sparse polynomial interpolation methods [49, 50] are suitable approaches that allow to construct adaptively the sequence of interpolation point  $\{\Gamma_n\}_{n \geq 0}$  for which the Lebesgue constant behaves well and do not suffer from the curse of dimensionality.

### Least-square method

Let  $u$  be a fonction in  $L^2_\mu(D)$  with some probability measure  $\mu$  on  $D$ . Least square fitting methods construct the approximation  $u_n$  of  $u$  in  $V_n$  as solution of the following least square regression problem

$$\min_{v \in V_n} \sum_{i=1}^m |u(x_i) - v(x_i)|^2 \quad (3.3.1)$$

where the  $\{x_1, \dots, x_m\}$  is  $m$ -sample for the probability measure  $\mu$ . The least squares method is stable and accurate for large  $m$  large enough [53]. Weighted least squares have been proposed to overcome this limitation. In that case, the approximation  $u_n$  is sought as the solution of

$$\min_{v \in V_n} \sum_{i=1}^m w(x_i) |u(x_i) - v(x_i)|^2 \quad (3.3.2)$$

where the  $w(x_i)$  are some weights related to  $\{x_1, \dots, x_m\}$ . Here the  $x_i$  are independent and identically distributed copies of a new measure  $\rho$ . It is shown in [54], that for an optimal choice of weight  $w$  and sampling measure  $\rho$ , stability and optimal accuracy are achieved if  $m$  scales linearly with  $n$  up to an additional logarithmic factor. To that goal the weight  $w$  and sampling measure  $\rho$  are given through the inverse Christoffel function that depends on the space  $V_n$  and the measure  $\mu$ . It can be *boosted* when the measure  $\rho$  is chosen so as to ensure the stability of the weighted least-squares projection [100]. In that case, it has the advantage to requires a sample of size close to the ideal  $m = n$ , as for interpolation.

### 3.3.2 Control variate method for high dimensional PDEs

We seek a global approximation of  $u$  from pointwise estimates  $u_{\Delta t, M}(x)$  for any  $x \in \bar{D}$ . In that context, we have proposed in [MBF12] algorithms that combine the variance reduction method proposed in [85, 86] with adaptive sparse polynomial interpolation [49, 50] for approximation of  $u$  for high-dimensional problem.

*Notations.* Let  $\Lambda \subset \mathbb{N}^d$  be a finite set of multi-indices and  $V_\Lambda \subset \mathcal{C}^2(\bar{D})$  an approximation space of finite dimension  $\#\Lambda$  and let  $\mathcal{J}_\Lambda : \mathbb{R}^D \rightarrow V_\Lambda$  be the interpolation operator associated with an unisolvent grid  $\Gamma_\Lambda = \{x_\nu : \nu \in \Lambda\}$ . We let  $\{\varphi_\nu\}_{\nu \in \Lambda}$  denote the (unique) basis of  $V_\Lambda$  that satisfies the interpolation property  $\varphi_\nu(x_\mu) = \delta_{\nu\mu}$  for all  $\nu, \mu \in \Lambda$ . The interpolation  $\mathcal{J}_\Lambda(w) = \sum_{\nu \in \Lambda} w(x_\nu) \varphi_\nu(x)$  of function  $w$  is then the unique function in  $V_\Lambda$  such that

$$\mathcal{J}_\Lambda(w)(x_\nu) = w(x_\nu), \quad \nu \in \Lambda.$$

### A sequential control variate method

The *sequential control variate* procedure of [85] provides a sequence of approximations  $\{\tilde{u}^k\}_{k \geq 1}$  of  $u$  in  $V_\Lambda$ , which are defined for all  $k \geq 1$  by  $\tilde{u}^k = \tilde{u}^{k-1} + \tilde{e}^k$ , where  $\tilde{e}^k$  is an approximation of  $e^k$ , solution of

$$\begin{aligned} -\mathcal{A}(e^k)(x) &= g(x) + \mathcal{A}(u^{k-1})(x), & x \in D, \\ e^k(x) &= f(x) - u^{k-1}(x), & x \in \partial D. \end{aligned}$$

Note that  $e^k$  admits a Feynman-Kac representation  $e^k(x) = \mathbb{E}(F(e^k, X^x))$ , where  $F(e^k, X^x)$  depends on the residuals  $g + \mathcal{A}(u^{k-1})$  on  $D$  and  $f - u^{k-1}$  on  $\partial D$ . Here, the approximation

$$\tilde{e}^k = \mathcal{J}_\Lambda(e_{\Delta t, M}^k)$$

is defined as the interpolation of the Monte-Carlo estimate  $e_{\Delta t, M}^k(x)$  of  $e_{\Delta t}^k(x) = \mathbb{E}(F(e^k, X^{x, \Delta t}))$  (using  $M$  samples of  $X^{x, \Delta t}$ ).

Now, let us introduce the time integration error at point  $x$  for a function  $h$

$$e^{\Delta t}(h, x) = \mathbb{E}[F(h, X^{\Delta t, x})] - \mathbb{E}[F(h, X^x)].$$

Then the following theorem [MBF12, Theorem 2] gives a control of the error  $\tilde{u}^{k+1}(x_\nu) - u(x_\nu)$  in expectation at step  $k$  of the control variate algorithm.

**Theorem 3.3.1.** *Assuming (A2)-(A4), it holds*

$$\sup_{\nu \in \Lambda} |\mathbb{E} [\tilde{u}^{k+1}(x_\nu) - u(x_\nu)]| \leq C(\Delta t, \Lambda) \sup_{\nu \in \Lambda} |\mathbb{E} [\tilde{u}^k(x_\nu) - u(x_\nu)]| + C_1(\Delta t, \Lambda)$$

with  $C(\Delta t, \Lambda) = \sup_{\nu \in \Lambda} \sum_{\mu \in \Lambda} |e^{\Delta t}(l_\mu, x_\nu)|$  and  $C_1(\Delta t, \Lambda) = \sup_{\nu \in \Lambda} |e^{\Delta t}(u - \mathcal{J}_\Lambda(u), x_\nu)|$ .

Moreover if  $C(\Delta t, \Lambda) < 1$ , it holds

$$\limsup_{n \rightarrow \infty} \sup_{\nu \in \Lambda} |\mathbb{E} [\tilde{u}^k(x_\nu) - u(x_\nu)]| \leq \frac{C_1(\Delta t, \Lambda)}{1 - C(\Delta t, \Lambda)}. \quad (3.3.3)$$

For  $\Delta t$  sufficiently small (i.e. under the condition  $C(\Delta t, \Lambda) < 1$ ), the error at interpolation points uniformly converges geometrically up to a threshold term depending on time integration errors for interpolation functions  $l_\nu$  and the interpolation error  $u - \mathcal{J}_\Lambda(u)$ .

Theorem 3.3.1 can be extended [MBF12, Corollary 1] to provide a convergence result in  $L^\infty(D)$  using Lebesgue constants in  $L^\infty$ -norm associated to  $\mathcal{J}_\Lambda$ .

**Corollary 3.3.2** (Convergence in  $L^\infty$ ). *Assuming (A2)-(A4), it holds*

$$\limsup_{n \rightarrow \infty} \|\mathbb{E} [\tilde{u}^k - u]\|_\infty \leq \frac{C_1(\Delta t, \Lambda)}{1 - C(\Delta t, \Lambda)} \mathcal{L}_\Lambda + \|u - \mathcal{J}_\Lambda(u)\|_\infty, \quad (3.3.4)$$

with the Lebesgue constant  $\mathcal{L}_\Lambda = \sup_{x \in \bar{D}} \sum_{\nu \in \Lambda} |\varphi_\nu(x)|$ .

## Adaptive sparse (polynomial) interpolation

When  $D$  admits some cartesian product structure, multivariate interpolation can be performed on tensor product of univariate bases. For high dimensional problems, there can be some advantage in considering approximation spaces that are sparse and anisotropic with respect to the different variables. To that goal, adaptive selection of the interpolation points can be envisaged. In such a procedure the set of interpolation points is successively enriched, as the polynomial dimension increases, at different stages of the computation. In what follows, we recall in brief the main elements of *sparse polynomial interpolation* following [49, 50].

**Notations** Here, we assume that the function to interpolate  $u$  is defined on  $D = D_1 \times \dots \times D_d$ , where the  $D_k$  are open bounded intervals of  $\mathbb{R}$ . For  $1 \leq i \leq d$ , we consider  $\{\varphi_k^{(i)}\}_{k \in \mathbb{N}}$  where the  $\varphi_k^{(i)}$  are univariate polynomial of degree  $k$  defined on  $D_i$ . For a multi-index  $\nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}^d$ , we introduce the multivariate function

$$\varphi_\nu(x) = \prod_{i=1}^d \varphi_{\nu_i}^{(i)}(x_i).$$

Let  $\Lambda \subset \mathbb{N}^d$  *downward closed* i.e. for all  $\nu \in \Lambda$ , such that  $\mu \leq \nu$  then  $\mu \in \Lambda$ . To define a set of points  $\Gamma_\Lambda$  unisolvent for  $\mathcal{P}_\Lambda$ , we can proceed as follows. For each dimension  $1 \leq i \leq d$ , we introduce a sequence of points  $\{z_k^{(i)}\}_{k \in \mathbb{N}}$  in  $\overline{D}_i$  such that for any  $p \geq 0$ ,  $\Gamma_p^{(i)} = \{z_k^{(i)}\}_{k=0}^p$  is unisolvent for  $\mathcal{P}_p = \text{span}\{\varphi_k^{(i)} : 0 \leq k \leq p\}$  (therefore defining an interpolation operator  $\mathcal{J}_p^{(i)}$ ). Then we let

$$\Gamma_\Lambda = \{z_\nu = (z_{\nu_1}^{(1)}, \dots, z_{\nu_d}^{(d)}) : \nu \in \Lambda\} \subset \overline{D}.$$

**Main idea** The goal of sparse interpolation is to provide a sequence of approximations  $\{\mathcal{J}_{\Lambda_n}(u)\}_{n \geq 1}$ , in increasing polynomial spaces  $\{\mathcal{P}_{\Lambda_n}\}_{n \geq 1}$ , associated with a nested sequence of downward closed subsets  $\{\Lambda_n\}_{n \geq 1}$ . The sequence of downward closed subsets  $(\Lambda_n)_{n \geq 1}$  is selected such that the best approximation error and the Lebesgue constant satisfy

$$\mathcal{L}_{\Lambda_n} \inf_{w \in \mathcal{P}_{\Lambda_n}} \|u - w\|_\infty \rightarrow 0, \text{ as } n \rightarrow \infty.$$

In practice, sparse interpolation algorithm with bulk chasing procedure introduced in [48] provides a practical algorithm that constructs a good sequence of sets  $\Lambda_n$ .

## Proposed algorithms

We have proposed (see [MBF12] for details) two algorithms combining sparse adaptive interpolation and sequential control variate procedure discussed previously. Here we give the general guidelines of the proposed algorithms.

1. The first algorithm called **Perturbed Adaptive Sparse Interpolation** algorithm. It is a perturbed version of the sparse adaptive algorithm where at each step  $n$  the computation of the exact interpolant  $\mathcal{J}_{\Lambda_n}(u)$  from  $u(x_\nu)$  is replaced by an approximation denoted  $u_n$  in  $V_n = \mathcal{P}_{\Lambda_n}$  based on MC estimate  $u_{\Delta t, M}(x_\nu)$ . In particular, this approximation approximation is computed with the sequential control variate procedure stopped for a given number of iterations or according to some stopping criterion.
2. As a second algorithm, we consider an **Adaptive Sequential Control Variates** algorithm where an approximation  $e_{n_k}^k$  of  $e^k$  is obtained by applying the adaptive interpolation algorithm to the function  $e_{\Delta t, M}^k$ , which uses Monte-Carlo estimations  $e_{\Delta t, M}^k(x_\nu)$  of  $e^k(x_\nu)$  at interpolation points. At each iteration,  $e_{n_k}^k$  therefore belongs to a different approximation space  $V_{n_k} = \mathcal{P}_{\Lambda_{n_k}}$ .

### 3.3.3 Numerical applications

In this section, we give a short illustration of the behavior of the perturbed adaptive sparse algorithm for polynomial interpolation in high dimension. A detailed study of the different algorithms discussed previously is available in [MBF12, Section 4.3].

Let us consider the following simple diffusion equation

$$\begin{aligned} -\Delta u(x) &= g(x), & x \in D, \\ u(x) &= f(x), & x \in \partial D, \end{aligned} \quad (3.3.5)$$

where  $D = ]-1, 1[^d$ . The source terms and boundary conditions will be specified later for each test case. The stochastic differential equation associated to (3.3.5) is the following

$$dX_t^x = \sqrt{2}dW_t, \quad X_0^x = x, \quad (3.3.6)$$

where  $(W_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion. We use tensorized grids of magic points for the selection of interpolation points evolved in adaptive algorithms.

We consider two test cases noted respectively (TC1) and (TC2).

(TC1) As first test case in dimension  $d = 10$ , we define (3.3.5) such that its solution is the Henon-Heiles potential

$$u(x) = \frac{1}{2} \sum_{i=1}^d x_i^2 + 0.2 \sum_{i=1}^{d-1} (x_i x_{i+1}^2 - x_i^3) + 2.5 \cdot 10^{-3} \sum_{i=1}^{d-1} (x_i^2 + x_{i+1}^2)^2, \quad x \in \bar{D}.$$

(TC2) We also consider the problem (3.3.5) whose exact solution is a sum of non-polynomial functions in dimension  $d = 20$ , given by

$$u(x) = x_1^2 + \sin(x_{12}) + \exp(x_5) + \sin(x_{15})(x_8 + 1).$$

MC estimate are computed for  $(\Delta t, M) = (10^{-4}, 1000)$  and use set to 30 the maximal number of iterates in the control variate procedure. Since for both test cases the exact solution is known, we propose to compare the behavior exact and perturbed sparse adaptive algorithms. In Table 3.1 and Table 3.2, we summarize the results associated to both algorithms for (TC1) and (TC2) respectively. We observe that the perturbed sparse adaptive algorithm performs well in comparison to sparse adaptive algorithm, for (TC1). Indeed, we get an approximation with a precision below the prescribed value  $\varepsilon$  for both algorithms.

$\#\Lambda_n$	$\ u - \mathcal{J}_{\Lambda_n}(u)\ _\infty$	$\ u - \mathcal{J}_{\Lambda_n}(u)\ _2$	$\#\Lambda_n$	$\ u - u_n\ _\infty$	$\ u - u_n\ _2$
1	3.0151e+00	1.2094e+00	1	8.3958e-01	6.9168e-01
17	1.8876e+00	5.9579e-01	17	5.2498e-01	3.4420e-01
36	7.0219e-01	2.0016e-01	36	1.9209e-01	1.2594e-01
46	1.6715e-01	4.9736e-02	46	4.6904e-02	2.8524e-02
53	2.9343e-02	4.8820e-03	53	7.8754e-03	2.8960e-03
60	1.5475e-02	4.1979e-03	61	3.0365e-03	1.7610e-03
71	8.4575e-03	2.1450e-03	71	2.3486e-03	1.2395e-03
77	3.9968e-15	1.5784e-15	77	6.2172e-15	1.2874e-15

Table 3.1: (TC1) Comparison of sparse adaptive algorithm (first four columns) and perturbed sparse adaptive algorithm (last four columns).

Similar observation holds for (TC2) in Table 3.2 and this despite the fact that the test case involves higher dimensional problem.

$\#\Lambda_n$	$\ u - \mathcal{J}_{\Lambda_n}(u)\ _\infty$	$\ u - \mathcal{J}_{\Lambda_n}(u)\ _2$	$\#\Lambda_n$	$\ u - u_n\ _\infty$	$\ u - u_n\ _2$
1	3.9361e+00	1.2194e+00	1	7.2832e-01	7.0771e-01
6	2.2705e+00	5.4886e-01	6	2.7579e-01	5.1539e-01
11	2.8669e-01	1.0829e-01	11	4.4614e-02	4.1973e-02
15	1.6425e-01	4.7394e-02	15	1.5567e-02	2.5650e-02
20	2.7715e-03	7.2230e-04	20	5.6927e-04	5.3597e-04
24	1.5564e-04	7.5314e-05	24	2.5952e-05	3.0835e-05
30	8.0838e-06	2.1924e-06	30	2.8808e-06	1.9451e-06
35	8.9815e-08	2.4651e-08	35	2.1927e-08	1.5127e-08
40	2.0152e-09	6.9097e-10	40	2.8455e-10	2.6952e-10
45	2.4783e-10	4.1600e-11	45	3.3188e-11	1.7911e-11
49	4.6274e-13	8.5980e-14	49	8.6362e-14	5.0992e-14
54	2.2681e-13	3.1952e-14	54	4.8142e-14	2.6617e-14
56	8.4376e-15	3.0438e-15	56	8.4376e-15	6.3039e-15

Table 3.2: (TC2) Comparison of sparse adaptive algorithm (first four columns) and perturbed sparse adaptive algorithm (last four columns).

### 3.4 Parameter-dependent PDEs

Now, we consider a parameter-dependent variant of the model problem (3.2.5). Let  $\Xi \subset \mathbb{R}^p$  be a set of parameter. Keeping the same notations as in Section 3.2, the problem of interest is

$$\begin{aligned} -\mathcal{A}(\xi)u(\xi) &= g(\xi), \quad \text{in } D, \\ u(\xi) &= f(\xi), \quad \text{on } \partial D, \end{aligned} \tag{3.4.1}$$

where the boundary and source terms  $f$  and  $g$  as well as the linear elliptic operator  $\mathcal{A}$  (through the coefficients  $b, \sigma$  in (3.2.6)), depend on  $\xi$ . Also, we assume that, for any  $\xi \in \Xi$ , there is a unique solution  $u(\xi)$  to (3.4.1). Moreover, if for all parameter  $\xi \in \Xi$ , the assumptions (A1)-(A4) are fulfilled, by Theorem 3.2.1, the solution  $u(\xi)$  admits a probabilistic representation under the form

$$u(x, \xi) = \mathbb{E} \left( F(x, X^{x, \xi}, \xi) \right) := \mathbb{E} \left( f(X_{\tau^{x, \xi}}^{x, \xi}, \xi) + \int_0^{\tau^{x, \xi}} g(X_t^{x, \xi}, \xi) dt \right), \quad x \in \bar{D}. \tag{3.4.2}$$

Here,  $X^{x, \xi}$  is a parameter-dependent diffusion process, defined for all  $t \geq 0$  by

$$X_t^{x, \xi} = X_0^{x, \xi} + \int_0^t b(X_s^{x, \xi}, \xi) ds + \int_0^t \sigma(X_s^{x, \xi}, \xi) dW_s, \quad X_0^{x, \xi} = x \in \bar{D}, \tag{3.4.3}$$

where  $W$  stands for the standard brownian motion. A MC estimate of  $u(x, \xi)$  is then

$$u_{\Delta t, M}(x, \xi) := \frac{1}{M} \sum_{m=1}^M F(x, X^{x, \Delta, \xi}(\omega_m), \xi), \tag{3.4.4}$$

for  $m$ -independent realisations of  $X^{x, \Delta, \xi}$  the approximated diffusion process obtained through Euler-Maruyama scheme.

This section concerns the approximation of the solution manifold  $\mathcal{M} = \{u(\xi) : \xi \in \Xi\}$ . First, we motivate a probabilistic RB method in full sampling setting in Section 3.4.1. It leads us to consider some probabilistic greedy procedure with a bandit algorithm for the selection of a new snapshot detailed in Section 3.4.2.

*Here, we present some working directions and preliminary results obtained in [123]. Practical implementation and numerical illustrations are under consideration for future publication [MBF17].*

### 3.4.1 Sample based probabilistic reduced basis method

*Notations.* Throughout this section, we assume that  $u(\xi)$  belongs to  $V \subset L^2(D)$  a Hilbert space equipped with its inner product  $\langle \cdot, \cdot \rangle_V$  and the related norm  $\|\cdot\|_V$ .

#### Classical framework

We recall that RB method (see Section 1.3) is aimed at providing a low-rank approximation  $u_r$  of  $u$  in a low-dimensional subspace  $V_r \subset V$ , of small dimension  $r$ , following two steps: the offline and online stages.

During the offline step, the reduced basis space  $V_r$  is deduced from snapshots of  $u$  computed for  $r$  instances of the parameter  $\xi$  selected in a training set  $\tilde{\Xi} \subset \Xi$  using a greedy procedure relying on some computable error estimate  $\Delta(u_r(\xi), \xi)$ . It is summarized in Algorithm 3.4.1.

**Algorithm 3.4.1** (Deterministic greedy algorithm). *Let  $\tilde{\Xi} \subset \Xi$  be a finite training set and  $V_0 = \{0\}$ . For  $r \geq 1$  proceed as follows.*

(Step 1.) *Select*

$$\xi_r \in \arg \max_{\xi \in \tilde{\Xi}} \Delta(u_{r-1}(\xi), \xi).$$

(Step 2.) *Compute  $u(\xi_r)$  and update  $V_r = \text{span}\{u(\xi_1), \dots, u(\xi_r)\}$ .*

Usually, the Algorithm 3.4.1 is stopped when  $\Delta(u_{r-1}(\xi), \xi)$  is below some target precision  $\varepsilon > 0$  or for a given space dimension  $r$ . Then, during the offline stage,  $u_r$  is computed as an approximation of  $u$  through suitable projection in  $V_r$  at low complexity depending only on  $r$ .

We recall that, if the projection  $u_r \in V_r$  satisfies the following quasi-optimality condition

$$\|u(\xi) - u_r(\xi)\|_V \leq C \|u(\xi) - P_{V_r} u(\xi)\|_V, \quad \xi \in \Xi \quad (3.4.5)$$

for some constant  $C$  independent from  $V_r$  and  $\xi$ , and if the error estimate  $\Delta(u_r(\xi), \xi)$  is chosen such that

$$\|u(\xi_r) - P_{V_r} u(\xi_r)\|_V \geq \gamma \sup_{\xi \in \tilde{\Xi}} \|u(\xi) - P_{V_r} u(\xi)\|_V, \quad \gamma \in (0, 1], \quad (3.4.6)$$

thus the Algorithm 3.4.1 is a *weak greedy* algorithm with respect to the discrete solution manifold  $\tilde{\mathcal{K}} = \{u(\xi) : \xi \in \tilde{\Xi}\}$ .

#### Probabilistic and sample based framework

As the exact solution  $u(\xi)$  of (3.4.1) is not computable in general, a numerical approximation  $u_n(\xi)$  in a finite dimensional subspace  $V_n \subset V$  is usually considered. It is usually related to some discretization a priori for a given mesh of  $D$  of the boundary problem (3.4.1). Instead, we assume

that we can only compute pointwise estimates  $u_{\Delta t, M}(x, \xi)$  of  $u(x, \xi)$  for any  $(x, \xi) \in \Xi \times D$ . In view of practical applications, it rises several questions listed below.

- 1) How to compute the snapshots required for generating the reduced spaces  $V_r$  and the approximation  $u_r$  ?
- 2) How to choose a computable error estimator  $\Delta(u_r(\xi), \xi)$  in Algorithm 3.4.1 ?

The first point could be addressed by computing a snapshot as a global approximation  $u_n(\xi)$  over  $D$  obtained through suitable projection in some finite dimensional vector space  $V_n \subset V$ , from pointwise estimate  $u_{\Delta t, M}(x, \xi)$ . To that goal we can, for example, consider a least-square or interpolation procedure in polynomial space as discussed in Section 3.3. Following the same path, it is possible to compute  $u_r(\xi)$  the approximation of  $u(\xi)$  in  $V_r$ . At this step, the goal is to compute such a projection using a small number of samples in comparison to seeking directly an approximation  $u_n(\xi)$  in some approximation space  $V_n$ .

Concerning the second point, the error estimate  $\Delta(u_r(\xi), \xi)$  required in Algorithm 3.4.1 is usually related to some norm of the equation residual in classical RB methods (see Section 1.3.1). Here, we follows another path by considering the  $L^2$ -norm of the current approximation error  $e_r(\xi) = u(\xi) - u_r(\xi)$  i.e.

$$\Delta(u_r(\xi), \xi) = \|e_r(\xi)\|_{L^2}^2.$$

By construction,  $u_r(\xi)$  is a linear approximation in  $V_r$ , regular enough (inherited from the snapshots), so that  $e_r(\xi)$  can be written as the solution, for all  $\xi$  in  $\Xi$ , of

$$\begin{aligned} -\mathcal{A}(\xi)e_r(\xi) &= g_r(\xi) \quad \text{on } D, \\ e_r(\xi) &= f_r(\xi) \quad \text{on } \partial D, \end{aligned} \tag{3.4.7}$$

where  $f_r(\xi) := f(\xi) - u_r(\xi)$  and  $g_r(\xi) = g(\xi) + \mathcal{A}(\xi)u_r(\xi)$ . By Theorem 3.2.1 for all  $\xi$  in  $\Xi$ ,  $e_r(\xi)$  is the unique solution of (3.4.7) in  $\mathcal{C}(\bar{D}) \cap \mathcal{C}^2(D)$  and satisfies for all  $x \in \bar{D}$

$$e_r(x, \xi) = \mathbb{E} \left( F_r(x, X^{x, \xi}, \xi) \right) = \mathbb{E} \left( f_r(X_{\tau^{x, \xi}}^{x, \xi}, \xi) + \int_0^{\tau^{x, \xi}} g_r(X_t^{x, \xi}, \xi) dt \right). \tag{3.4.8}$$

Then, we have [123, Theorem 5.7] the following probabilistic reinterpretation of  $\|e_r(\xi)\|_{L^2}^2$ .

**Theorem 3.4.2.** *Taking  $Y \sim \mathcal{U}(D)$  uniformly distributed on  $D$ , we have for any  $\xi$  in  $\Xi$*

$$\|e_r(\xi)\|_{L^2}^2 = |D| \mathbb{E} (Z_r(\xi)). \tag{3.4.9}$$

Here  $Z_r(\xi) = F_r(Y, X^{Y, \xi}, \xi) F_r(Y, \tilde{X}^{Y, \xi}, \xi)$  where  $X^{x, \xi}$  and  $\tilde{X}^{x, \xi}$  are two i.i.d random stopped diffusion processes solution of (3.4.3), for any  $x \in \bar{D}$ , and  $|D|$  is the Lebesgue measure of  $D$ .

### 3.4.2 Probabilistic greedy algorithm

In this section, we present and analyze a probabilistic variant of Algorithm 3.4.1, in general setting, motivated by the applications discussed in the previous section. As a starting point, we assume that the error estimator required at the  $r$ -th iteration is under the form

$$\Delta(u_r(\xi), \xi) := \|u(\xi) - u_r(\xi)\|_V^2 = \mathbb{E}(Z_r(\xi)), \tag{3.4.10}$$

where  $Z_r(\xi)$  is some parameter-dependent real valued random variable, defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Solving the following optimization problem,

$$\xi_r \in \arg \max_{\xi \in \Xi} \mathbb{E}(Z_{r-1}(\xi)), \tag{3.4.11}$$

at step  $r$  of Algorithm 3.4.1, is in general out of reach, since  $\mathbb{E}(Z_{r-1}(\xi))$  is unknown a priori or too costly to compute.

In view of numerical applications, a first practical and naive approach is to seek  $\xi_r$  as the maximum of a crude MC estimate (empirical mean) of  $\mathbb{E}(Z_{r-1}(\xi))$ , computed using i.i.d. samples of  $Z_{r-1}(\xi)$ ,  $\xi \in \tilde{\Xi}$ . Despite its simplicity, we have seen that such an estimate for the expectation suffers from low convergence with respect to the number of samples leading to possible high computational costs especially here where  $Z_{r-1}(\xi)$  can be expensive to evaluate. Moreover, nothing ensures that returned (random) parameter  $\xi_r$  is a (quasi-)optimum for (3.4.11), almost surely or at least with high probability.

### Probably Approximately Correct maximum

Instead, the so-called *bandit algorithms* (see, e.g., monograph [119]) are good candidates to address (3.4.11). In particular *probably approximately correct (PAC)* bandit algorithms are designed not to return an optimum almost surely but a quasi-optimum, either in absolute or relative precision, with high probability. Here, we particularly focus on PAC bandit algorithms that return  $\xi_r$  (random) as a *PAC maximum in relative precision* for  $\mathbb{E}(Z_r(\xi))$  over  $\tilde{\Xi}$ , i.e.,

$$\mathbb{P}(\mathbb{E}(Z_{r-1}(\xi_r^*)) - \mathbb{E}(Z_{r-1}(\xi_r)) \leq \varepsilon \mathbb{E}(Z_{r-1}(\xi_r^*))) \geq 1 - \lambda, \quad (3.4.12)$$

for any prescribed threshold  $\varepsilon$  in  $(0, 1)$ , probability of failure  $\lambda$  in  $(0, 1)$  and

$$\xi_r^* \in \arg \max_{\xi \in \tilde{\Xi}} \mathbb{E}[Z_{r-1}(\xi)].$$

In what follows, we denote  $\xi_r := \text{PAC}_{\lambda, \varepsilon}(Z_{r-1}, \tilde{\Xi})$  a PAC maximum in relative precision for (3.4.11), if it satisfies (3.4.12).

**Practical algorithms** Classical PAC bandit algorithms are aimed at returning a PAC maximum in *absolute precision* [6]. Here, we are rather interested in having a PAC maximum in *relative precision* as proposed in [MBF15]. In this paper, we studied two variants : a naive and an adaptive one. The interest of the adaptive algorithm is that it outperforms the mean complexity of the naive algorithm in terms of number of required samples. Such an algorithm preferentially samples the random variable  $Z_r(\xi)$  for the parameter values for which it is more likely to find a maximum. Thus it is particularly well suited for applications with high sampling cost. Both algorithms have been proven [MBF15, Propositions 3.1 & 3.2] to provide PAC solution in relative precision for a family of random variables  $\{Z_r(\xi), \xi \in \tilde{\Xi}\}$ , indexed by  $\xi$ , satisfying suitable concentration inequalities under the form (3.2.2).

### A weak greedy algorithm with high probability

Now, we present a probabilistic weak greedy algorithm where the parameter  $\xi_r$  is a PAC maximum in relative precision at each step  $r$ .

**Algorithm 3.4.3** (Probabilistic greedy algorithm). *Let  $\tilde{\Xi} \subset \Xi$  be a discrete training set and  $(\lambda_r)_{r \geq 1} \subset (0, 1)$ . Starting from  $V_0 = \{0\}$ , proceed, for  $r \geq 1$ , as follows.*

(Step 1.) *Select*

$$\xi_r := \text{PAC}_{\lambda_r, \varepsilon}(Z_{r-1}, \tilde{\Xi})$$

(Step 2.) *Compute  $u(\xi_r)$  and update  $V_r = \text{span}\{u(\xi_1), \dots, u(\xi_r)\}$ .*

We have proven in [123, Chapter 5] that Algorithm 3.4.3 is a weak greedy algorithm with high probability.

**Theorem 3.4.4.** *Take  $(\lambda_r)_{r \geq 1} \in (0, 1)$  such that  $\sum_{r \geq 1} \lambda_r = \lambda < 1$ ,  $\varepsilon \in (0, 1)$  and  $\tilde{\Xi} \subset \Xi$  a discrete training set. Moreover, suppose that for  $r \geq 1$ , the approximation  $u_r$  of  $u$  in  $V_r$  satisfies the following quasi-optimality condition (3.4.5). Then, Algorithm 3.4.3 is a weak-greedy algorithm of parameter  $\frac{\sqrt{1-\varepsilon}}{C}$ , with probability at least  $1 - \lambda$  i.e.*

$$\mathbb{P} \left( \|u(\xi_r) - P_{V_{r-1}}u(\xi_r)\|_V \geq \frac{\sqrt{1-\varepsilon}}{C} \max_{\xi \in \tilde{\Xi}} \|u(\xi) - P_{V_{r-1}}u(\xi)\|_V, \forall r \geq 1 \right) \geq 1 - \lambda. \quad (3.4.13)$$

### General comments

- At a step  $r$  of the Algorithm 3.4.3, the reduced space  $V_r = \text{span}\{u(\xi_1), \dots, u(\xi_r)\}$ , as well as the approximation  $u_r(\xi)$  are no longer deterministic. Indeed, they are related to the selected parameters  $\xi_1, \dots, \xi_r$  depending themselves on the errors at the previous steps through i.i.d. samples of the random variables  $Z_i(\xi)$  for all  $\xi \in \tilde{\Xi}$  and  $i < r$  (required during PAC selection of  $\xi_r$ ). Thus, the result stated in Theorem 3.4.4 holds with high probability instead of with certainty, which is the price to pay when introducing some randomness in the greedy procedure. Especially, it demonstrates that Algorithm 3.4.3 is a weak greedy algorithm, with high probability, with respect to the discrete solution manifold  $\mathcal{K} = \{u(\xi) : \xi \in \tilde{\Xi}\}$ .
- In the lines of [56], it is also possible to consider a fully probabilistic variant of Algorithm 3.4.3, in which a training set  $\Xi_r$  randomly chosen is used at each step  $r$  of Algorithm 3.4.3 instead of  $\tilde{\Xi}$ . For a particular class of functions, of  $\xi$ , that can be approximated by polynomials with a certain algebraic rate, it can be proven that, for suitable chosen size of random training set  $\Xi_r$ , the resulting algorithm is a weak greedy algorithm with high probability with respect to the continuous solution manifold  $\mathcal{K} = \{u(\xi) : \xi \in \Xi\}$  (some elements of proof are given in [123, Chapter 5]).

## 3.5 Summary and discussion for future research

In this chapter, we have addressed the problem of approximating the solution of PDEs, whose solution admits a probabilistic representation. Within this framework, we have proposed a probabilistic sparse interpolation method for high dimensional problems.

Ongoing work concerns some probabilistic reduced basis method using samples of the function to approximate. Beyond the validation of the proposed approach, one pending and challenging question that originally motivated this work concerns the feasibility of such an approach for parameter-dependent problems in high dimension.

In what follows we give additional elements concerning possible extensions to these contributions.

**Learning approaches with tensor using F-K representation for PDEs.** Several recent works on developing algorithms that are learning approaches in tensor format i.e. that use random and unstructured evaluation of the function to approximate have been proposed see e.g. [93, 101, 136] or [68]. Within this framework, an interesting idea would be to combine it with Feynman-Kac pointwise estimations of the solution to perform suitable approximation of high dimensional PDEs.

**Applications of the probabilistic reduced basis framework.** The approach, discussed in Section 3.4.1, has been motivated by the solution of PDEs with probabilistic representation. It is none the less usable for the general problem of approximating any family of functions (possibly random)  $\{u(\xi) : \xi \in \Xi\}$ , indexed by the parameter  $\xi$ . In particular, if  $u(\xi)$  is known a priori but costly to evaluate, the proposed probabilistic greedy algorithm can be used to generate a sequence of spaces  $V_r$  and corresponding interpolant  $u_r$  of  $u$  onto  $V_r$ . Then, the proposed method leads to a probabilistic EIM as in [35]. In particular, in a fully discrete setting,  $u$  can be identified with a matrix and the resulting algorithm is a probabilistic version of adaptive cross approximation for low-rank matrix approximation [15, 157], where the proposed PAC bandit algorithm provides a particular column-selection strategy that sparsely explores the matrix entries.

**Time dependent problems.** The proposed approaches were mainly applied for the approximation of linear (parameter-dependent) elliptic problems. However, Feynman-Kac representation formula still holds for parabolic problems (see e.g. [57, 88]). Thus, applying for example the proposed interpolation method, or a learning approach with low-rank tensor format as previously discussed, for dealing with time dependent high-dimensional and/or parameter-dependent problem approaches is straight-forward. However, extending the presented probabilistic and sampling based RB paradigm for time-dependent and parameter-dependent problems (as discussed in Chapter 2) is a challenging and open question.



## PAST RESEARCH CONTRIBUTIONS

*In this chapter we give a concise summary of the contributions [MBF2, MBF4, MBF3, MBF5, MBF6, MBF7] concerning the development of numerical methods for the simulation of multi-component or multi-phase flows.*

### Compressible-incompressible interface flows

In the literature, several approaches are available to treat diphasic flows constituted of immiscible gaz and liquid. Some of them consider both fluids as e.g. incompressible [183] or compressible [186]. However, in one approach the one does not ensure the volume variation of the gaz whereas the second does not respect the volume conservation of the liquid. As [171], we consider an approach that consider the gaz as compressible fluid and the liquid as an incompressible fluid. To deal numerically with such interface flows, one faces at least to the following difficulties: (1) transition from a gas density linked to the local temperature and pressure by an equation of state to a liquid density mainly constant in space, (2) proper approximation of the divergence constraint in incompressible regions and (3) wave transmission at the interface.

We proposed in [MBF2, MBF3] a **global** numerical approach (i.e. the same for each phase) that allow to easily simulate compressible-incompressible on unstructured meshes (allowing complex geometry) which possible high order precision. This work was related to my PhD thesis [MBF1] made at CEA-CESTA (Commissariat à l'énergie atomique et aux énergies alternatives, Centre d'études scientifiques et techniques d'Aquitaine), under the supervision of Jean-Paul Caltagirone (Trèfle, I2M, Université de Bordeaux), Gérard Gallice (CEA-CESTA, retired), Boniface Nkonga (Laboratoire J.A. Dieudonné, Université de Nice Sophia Antipolis).

### Global model

The proposed numerical approach relies on the formulation of a **global model** using the same set of primitive unknowns and equations used everywhere in the flow, but with a dynamic parameterization that changes from compressible to incompressible regions.

Flows under consideration are viscous, Newtonian and consequently described by the Navier-Stokes equations. On one hand, the compressible Navier-Stokes equations are considered under weakly compressibility assumption so that a non-conservative formulation can be used. On the other hand, the incompressible non-isothermal model is retained. In addition, the level set transport equation is used to capture the interface position needed to identify the local characteristics of the fluid and to recover the adequate local modelling.

The first step of unification consists in choosing a same unknown suitable for both compressible and incompressible models. This is possible for both primitive and entropic unknowns. For flows without shocks, we consider quasi-linear **unified** equations [179, 180], for primitives unknowns  $\mathbf{Y} = (p, \mathbf{u}^t, \mathbf{T})^t$  (with  $p$  the pressure,  $\mathbf{u}$  the velocity and  $T$  the temperature)

$$L(\mathbf{Y}, \boldsymbol{\chi}, \rho, )\mathbf{Y} = \mathbf{S}(\mathbf{Y}, \rho), \quad (1)$$

with

$$L(\mathbf{Y}, \boldsymbol{\chi}, \rho, \mu, \kappa) = A_0(\mathbf{Y}, \boldsymbol{\chi}, \rho) \partial_t + \sum_{i=1}^d A_i(\mathbf{Y}, \boldsymbol{\chi}, \rho) \partial_{x_i} - \sum_{i,j=1}^d \partial_{x_i} (K_{ij}(\mathbf{Y}, \mu, \kappa) \partial_{x_j}).$$

Here,  $A_0, A_i$  correspond the jacobian matrices derived from Navier-Stokes equations under their conservative form. The  $K_{ij}$  are the diffusion matrices (see [MBF3, Section 2.2]). Here, the jacobian matrices  $A_0, A_i$  depend on three quantities namely the density  $\rho$ , and two thermodynamical coefficients

$$\boldsymbol{\chi} = (\alpha_p, \beta_T) = \left( -\frac{1}{\rho} \left( \frac{\partial \rho}{\partial T} \right)_p, \frac{1}{\rho} \left( \frac{\partial \rho}{\partial p} \right)_T \right)$$

that represent respectively the volumic dilation effects ( $\alpha_p$ ) and isothermal compressibility ( $\beta_T$ ). In the particular case of perfect gaz (indicated with subscript  $_g$ ) we have

$$\boldsymbol{\chi}_g = \left( \frac{1}{T}, \frac{1}{p} \right) \text{ and } \rho_g = \frac{p}{RT}, \quad (2)$$

whereas for an incompressible (indicated with subscript  $_l$ ) fluid it holds

$$\boldsymbol{\chi}_l = (0, 0) \text{ et } \rho_l = \text{constant}. \quad (3)$$

The diffusion matrices also depend on the viscosity coefficient  $\mu$  and,  $\kappa$  the thermal conductivity which are different for each fluid. For the liquid (gas), they are noted  $\mu_l, \kappa_l$  (respectively  $\mu_g, \kappa_g$ ).

The interface is captured through Level Set method [185]. It is given as the zero isovalue of a continuous function  $\phi$ , called Level Set function that satisfies  $\phi < 0$  in the liquid and  $\phi > 0$  in the gaz. Here, we consider diphasic flow without any state change, so that the Level Set  $\phi$  satisfies the following transport equation

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0, \quad (4)$$

where  $\mathbf{u}$  is the fluid flow velocity.

Combining (1)-(4), we derive the following global system of equations

$$\mathcal{L}(\mathbf{Y}, \phi) \mathbf{Y} = \mathbf{S}(\mathbf{Y}, \phi), \quad (5)$$

$$\partial_t \phi + \mathbf{u} \cdot \nabla \phi = 0, \quad (6)$$

where  $\mathcal{L}$  is defined by

$$\mathcal{L}(\mathbf{Y}, \phi) = \mathcal{A}_0(\mathbf{Y}, \phi) \partial_t + \sum_{i=1}^d \mathcal{A}_i(\mathbf{Y}, \phi) \partial_{x_i} - \sum_{i,j=1}^d \partial_{x_i} [\mathcal{K}_{ij}(\mathbf{Y}, \phi) \partial_{x_j}],$$

with the global matrices

$$\left\{ \begin{array}{l} \mathcal{A}_0(\mathbf{Y}, \phi) = A_0(\mathbf{Y}, \boldsymbol{\chi}(\phi), \rho(\phi)), \\ \mathcal{A}_i(\mathbf{Y}, \phi) = A_i(\mathbf{Y}, \boldsymbol{\chi}(\phi), \rho(\phi)), \\ \mathcal{K}_{ij}(\mathbf{Y}, \phi) = K_{ij}(\mathbf{Y}, \mu(\phi), \kappa(\phi)), \\ \mathcal{S}(\mathbf{Y}, \phi) = \mathbf{S}(\mathbf{Y}, \rho(\phi)), \end{array} \right. \text{ and } \left\{ \begin{array}{l} \boldsymbol{\chi}(\phi) = (1 - H(\phi)) \boldsymbol{\chi} + H(\phi) \boldsymbol{\chi}_g, \\ \rho(\phi) = (1 - H(\phi)) \rho + H(\phi) \rho_g, \\ \mu(\phi) = (1 - H(\phi)) \mu + H(\phi) \mu_g, \\ \kappa(\phi) = (1 - H(\phi)) \kappa + H(\phi) \kappa_g, \end{array} \right.$$

In the previous definition  $H(\phi) = \mathbf{1}_{\phi \geq 0}$  is the characteristic function equals to 1 in the gas.

## Global finite element approximation

The resolution of (5)-(6) relies on a time splitting. At each time step, we first solve the global equation (5) then the interface position is updated through (6). The two related subproblems are solved independently with different numerical schemes but with the same mesh in time and space (thus the time step fulfills a global CFL condition). As they are adapted to treat complex geometry and provide high order approximation, Stabilized Finite Element (SFE) have been used for the space discretization of (5) together with explicit Euler scheme in time. A Runge Kutta / Discontinuous Galerkin (RKDG) [174,184] has been considered for solving the transport equation on the level set (6). The combination, a posteriori, of these two solvers has been done following the procedure proposed in [183] for incompressible and diphasic flows (see details in [MBF3]). Here, we only present the "coupled" and "global" numerical method used for the numerical solution of (5).

For single fluid flow Lagrange FE method are not suitable for numerical solution of (1) as they provide unstable approximations in case of large advection, and violate the inf-sup condition (for the pressure). To overcome these limitations, it is possible to add to the Galerkin formulation some additional term called *stabilization*. To that goal, a Galerkin Least Square (GLS) [179,180] method is retained since it allows to deal with those problems for both incompressible and compressible equations. GLS method is a Petrov-Galerkin method, where the test function is  $\mathbf{W}_h$  plus an additional term  $\mathcal{L}\mathbf{W}_h$  multiplied by a matrix containing coefficients  $\tau_e$ . These coefficients defined locally on each element  $e$  of the mesh act as stabilization term. This matrix is defined differently depending on the compressible [178] or incompressible case [175]. Thus, the resulting numerical scheme is said to be **unified**, as it allows to solve numerically both compressible and incompressible Navier-Stokes equations.

This unified FE scheme is a good candidate for numerical solution of (5), however it requires some particular care for the treatment of the elements crossed by the interface (thus containing incompressible and compressible fluid). Indeed, it requires the computation of "discontinuous" integrals (i.e. where the quantity to integrate is discontinuous due to density, viscosity and conductivity) that allows correct wave transmission at the interface and the derivation of a suitable stabilization term adapted to each phase. To this end, we use averaged quantities in each element  $e$  relying on the volume fraction of the compressible fluid in each element noted  $\varepsilon_e^n$  (obtained using  $\phi$ ). We take

- arithmetical average for the unknown  $\bar{\mathbf{Y}}_e^n$ , the matrix of stabilization  $\bar{\tau}_e^n$ , and for the thermodynamical coefficients  $\bar{\chi}_e^n$ ,
- while harmonic average is retained for the density  $\bar{\rho}_e^n$ , viscosity  $\bar{\mu}_e^n$  and conductivity  $\bar{\kappa}_e^n$ .

We finally get a general numerical method relying on the following global discrete variational formulation. Let be given  $\mathbf{Y}_h^n$  at  $t^n$  and the interface position through  $\phi^n$ , Equation (5) is solved from  $t^n$  to  $t^{n+1}$  for  $1 \leq n \leq N - 1$  by seeking  $\mathbf{Y}_h^{n+1} \in \mathbf{S}_h^p \times \mathbf{S}_h^u \times \mathbf{S}_h^T$  such that for all  $\mathbf{W}_h \in \mathbf{T}_h^p \times \mathbf{T}_h^u \times \mathbf{T}_h^T$  we have

$$\left\{ \begin{array}{l} \underbrace{\sum_{e \in \mathcal{T}_h} \int_e \mathbf{W}_h \cdot \left( \frac{\mathcal{A}_0^n \mathbf{Y}_h^{n+1} - \mathbf{Y}_h^n}{\delta t} + \sum_{i=1}^d \mathcal{A}_{i_e}^n \partial_{x_i} \mathbf{Y}_h^{n+1} - \bar{\mathcal{S}}_e^n \right) + \sum_{i,j=1}^d \partial_{x_i} \mathbf{W}_h \cdot \bar{\mathcal{K}}_{ij_e}^n \partial_{x_j} \mathbf{Y}_h^{n+1}}_{\text{Galerkin part}} d\Omega \\ \quad + \underbrace{\sum_{e \in \mathcal{T}_h} \int_e \bar{\mathcal{L}}_e^{nt} \mathbf{W}_h \cdot \bar{\tau}_e^n \left( \bar{\mathcal{L}}_e^n \mathbf{Y}_h^{n+1} - \bar{\mathcal{S}}_e^n \right)}_{\text{GLS stabilization}} d\Omega = 0, \\ \text{+ boundary conditions .} \end{array} \right. \quad (7)$$

Here  $\mathbf{Y}_h^n$  is the FE approximation at  $t^n$  and  $\mathbf{W}_h$  the test function. Finally the average matrices  $\overline{\mathcal{A}}_{0e}^n, \overline{\mathcal{A}}_{ie}^n, \overline{\mathcal{K}}_{ije}^n, \overline{\mathcal{S}}_e^n, \overline{\boldsymbol{\tau}}_e^n$  are given through

$$\begin{aligned}\overline{\mathcal{A}}_{0e}^n &= A_0(\overline{\mathbf{Y}}_e^n, \overline{\boldsymbol{\chi}}_e^n, \overline{\rho}_e^n), \\ \overline{\mathcal{A}}_{ie}^n &= A_i(\overline{\mathbf{Y}}_e^n, \overline{\boldsymbol{\chi}}_e^n, \overline{\rho}_e^n), \\ \overline{\mathcal{K}}_{ije}^n &= K_{ij}(\overline{\mathbf{Y}}_e^n, \overline{\boldsymbol{\mu}}_e^n, \overline{\kappa}_e^n), \\ \overline{\boldsymbol{\tau}}_e^n &= \varepsilon_e^n \boldsymbol{\tau}_{e,g}(\overline{\mathbf{Y}}_e^n, \overline{\rho}_e^n, \overline{\boldsymbol{\mu}}_e^n, \overline{\kappa}_e^n) + (1 - \varepsilon_e^n) \boldsymbol{\tau}_{e,l}(\overline{\mathbf{Y}}_e^n, \overline{\rho}_e^n, \overline{\boldsymbol{\mu}}_e^n, \overline{\kappa}_e^n), \\ \overline{\mathcal{S}}_e^n &= \mathbf{S}(\overline{\mathbf{Y}}_e^n, \overline{\rho}_e^n),\end{aligned}$$

with  $\boldsymbol{\tau}_{e,g}, \boldsymbol{\tau}_{e,l}$  the stabilization matrices for a compressible and incompressible fluid respectively.

## Numerical results

First numerical results have concerned numerical simulation of smooth flow for which the global behavior is driven by the gas compressibility. For the considered cases, comparisons with analytical computation of relevant quantity related to the flow permitted the validation of the proposed approach as well as convergence study, in both one and two dimensions. Then more general test cases have been considered as rising bubble or oscillation water column (see [MBF3] for details).

## Multi-material compressible flows

My post-doctorate was dedicated to the simulation of multi-material compressible flows arising in astrophysics or inertial confinement fusion (ICF). Several approaches are available for numerical simulation of such flows, using either interface capturing or tracking methods. Here we have adopted subsequently each point of view leading to several publications [MBF4, MBF5, MBF6, MBF7]. These contributions are related to several collaborations with the CEA with Edouard Audit (Maison de la simulation, CEA), Jérôme Breil (CEA-CESTA, Celia), Guy Schurtz (CEA-CESTA, Celia, retired) and Samuel Kokh (CEA, Saclay) but also during research stay during CEMRACS 2010 with the team SIMCAPIAD.

**Eulerian methods:** We consider problems restricted to multi-material flows with  $m$  component on cartesian mesh. In this framework, some methods [169, 170] consider a global model to represent the flow with particular Equation of States (EOS). Choosing suitable EOS allows to treat the diffusion region (it is only numerical i.e. not related to any physics) appearing near the interface when solving numerically those equations (e.g. using finite volume (FV)). In particular, it allows the numerical solution to converge to the correct solution. Some approaches have been developed to limitate the interface diffusion [172, 173] using **anti-diffusive** numerical scheme. In [169, 182], a 5 equation model (quasi-conservative) is proposed to modelize the evolution of  $m = 2$  compressible and immiscible fluids separated by an interface. It corresponds to Euler equations combined to partial mass equations, and the transport of a color function. In [182], this system is solved using **anti-diffusive** FV scheme, in order to capture accurately the interface (less numerical diffusion). In that goal, a splitting is proposed 1) the acoustic part of the equations is solved using a classical FV scheme (e.g. relaxation, acoustic scheme), 2) the transport is solved using the anti-diffusive scheme [172, 173]. This approach is only proposed to **two materials** compressible flows and is only **first order** accurate.

1. In [MBF4] we have proposed to improve the accuracy of this method to second order using MUSCL reconstruction scheme in space (for acoustic fluxes) together with Heun scheme

in time. The approach has been validated in 1D, and 2D dimensional case have been treated using directional splitting.

2. Then, to deal with flows with a arbitrary number of component, we have proposed an extension of the five equation model [169], leading to the following ***m*-component model**

$$\left\{ \begin{array}{l} \partial_t(\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p = 0, \text{ moment conservation} \\ \partial_t(\rho E) + \nabla \cdot ((\rho E + p)\mathbf{u}) = 0, \text{ energie conservation} \\ \partial_t(\rho \mathcal{Y}_k) + \nabla \cdot (\rho \mathbf{u} \mathcal{Y}_k) = 0, \text{ partial mass conservation} \\ \partial_t \mathcal{Z}_k + \mathbf{u} \cdot \nabla \mathcal{Z}_k = 0, \text{ color function transport,} \end{array} \right. \quad (8)$$

where  $\rho$  is the density,  $\mathbf{u}$  the velocity,  $p$  the pressure,  $E$  total energy and the  $\mathcal{Y}_k$  are mass fractions for each fluid  $k$ ,  $1 \leq k \leq m$ . This system, together with properly chosen equation of states is hyperbolic. Following [182], we proposed to apply a lagrange-projection scheme with an anti-diffusive scheme for the projection step when solving (8). In that case, the main difficulty was to generalize the proposed scheme to  $m > 2$  color function  $\mathcal{Z}_k$  preserving the maximum principle and such that the sum is equal to 1. To that goal, a recursive construction of the fluxes for the color function has been proposed in the line of [181]. The resulting method has been tested and validated on several test cases in 1D and 2D for flows with 3 and 5 materials [MBF5,MBF7] for different equations of state.

**Contribution on ALE method:** Here we focus on the simulation of multi-material compressible flows on unstructured meshes. In that context, we consider an interface capturing approach that consists in explicit reconstruction of interfaces from color functions see e.g. VOF (volume of fluid) method ( $m = 2$ ) or MOF (moment of fluid) method ( $m \geq 2$ ) which are very efficient when combined with ALE (Arbitrary Lagrangian Eulerian) method to track the interface. The latter allows to combine the advantages of an eulerian approach adapted for complex flows with distortion and a lagrangian approach that allows to exactly capture material fronts as interfaces. Here, we retained a CCALE-MOF (Cell-Centered Arbitrary Lagrangian Eulerian MOF), where the color functions are the fluid moment (e.g. mass fraction) [176, 177]. This algorithm works according to the following steps.

1. **Lagrangian step:** First, the mesh as well as conservative unknowns ( $\rho, \rho \mathbf{u}, \rho E$ ) are updated during a Lagrangian step using a centered VF: Explicit Unstructured Cell-Centered Lagrangian HYDrodynamics (EUC-CLHYD). The color functions (e.g. mass fraction) remain unchanged during this step.
2. **Reconstruction step:** Interfaces are reconstructed from fluid moment allowing perfect recovering of interfaces.
3. **Rezoning step:** During this step the mesh is "reshaped", i.e, if the mesh is very distorted during lagrangian step it is rectified.
4. **Projection step:** Finally, a projection step is done. It consists in interpolating in conservative way the physical quantities on the new mesh.

The CCALE-MOF approach, works well for cartesian coordinates and can be extended quite easily for cylindrical coordinates. The objective of the work [MBF6] was to develop some "hybrid" approach valid for both coordinates with axisymmetric geometries. The method was tested with success for the numerical simulation of complex multimaterial flows with axisymmetric geometries for cartesian, polar or unstructured meshes.



## BIBLIOGRAPHY

- [1] R. Abgrall, D. Amsallem and R. Crisovan *Robust model reduction by  $L^1$ -norm minimization and approximation via dictionaries: application to nonlinear hyperbolic problems*. Advanced Modeling and Simulation in Engineering Sciences, 3(1):1–16, 2016.
- [2] R. Abgrall and R. Crisovan *Model reduction using  $L^1$ -norm minimization as an application to nonlinear hyperbolic problems* International Journal for Numerical Methods in Fluids, 87(12):628–651, 2018.
- [3] M. M. Abedin Nejad, *Optimization Algorithms on Tensor Manifolds using Charts* Internship report, 2017.
- [4] A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings, *A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modelling of complex fluids*. Journal of Non-Newtonian Fluid Mechanics, 139(3):153–176, 2006.
- [5] A. Ammar, F. Chinesta, and A. Falco, *On the convergence of a greedy rank-one update algorithm for a class of linear systems*. Archives of Computational Methods in Engineering, 17(4):473–486, 2010.
- [6] J.-Y. Audibert, and S. Bubeck, R.M., *Best arm identification in multi-armed bandits*. In: Annual Conference on Learning Theory (COLT), 2010.
- [7] M. Bachmayr, A. Cohen, and G. Migliorati, *Sparse polynomial approximation of parametric elliptic pdes. Part i: affine coefficients*. ESAIM Math. Model. Numer. Anal., 51(1):321–339, 2017.
- [8] M. Bachmayr, A. Cohen, R. DeVore, and G. Migliorati, *Sparse polynomial approximation of parametric elliptic pdes. Part ii: lognormal coefficients*. ESAIM Math. Model. Numer. Anal., 51(1):341–363, 2017.
- [9] M. Bachmayr, H. Eisenmann H., E. Kieri, A. Uschmajew, *Existence of dynamical low-rank approximations to parabolic problems*. Mathematics of Computation, AMS Early View articles, preprint (2021)
- [10] O. Balabanov and A. Nouy. *Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation*. Advances in Computational Mathematics. 45(5-6):2969–3019, 2019.
- [11] O. Balabanov and A. Nouy. *Randomized linear algebra for model reduction. Part II: minimal residual methods and dictionary-based approximation*. Advances in Computational Mathematics. 47, 26, 2021.
- [12] O. Balabanov and A. Nouy. *Preconditioners for model order reduction by interpolation and random sketching of operators*. arXiv, arXiv:2104.12177, 2021.
- [13] J. Ballani and L. Grasedyck, *A projection method to solve linear systems in tensor format*. Numerical Linear Algebra with Applications, 20(1):27-43, 2013.

- [14] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations*. *Comptes Rendus Mathématique*, 339(9):667–672, 2002.
- [15] M., Bebendorf, *Approximation of boundary element matrices*. *Numerische Mathematik*, 86(4), 565–589, 2000.
- [16] U. Baur, P. Benner, B. Haasdonk, C. Himpe, I. Maiern, and M. Ohlberger, *Comparison of Methods for Parametric Model Order Reduction of Instationary Problems*. In preprint, 2015.
- [17] M. Bebendorf, Y. Maday, and B. Stamm, *Comparison of some reduced representation approximations*. In Alfio Quarteroni and Gianluigi Rozza, editors, *Reduced Order Methods for Modeling and Computational Reduction*, volume 9 of *MS and A Modeling, Simulation and Applications*, pages 67–100. Springer International Publishing, 2014.
- [18] C. Beck, E. Weinan and A. Jentzen. *Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations*. *Journal of Nonlinear Science*, 29:1563–1619, 2019.
- [19] C. Beck, S. Becker, P. Grohs, N. Jaafari and A. Jentzen. *Solving the Kolmogorov PDE by means of deep learning*. arXiv preprint arXiv:1806.00421v2, 2021.
- [20] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.
- [21] P. Benner, V. Mehrmann and D.C. Sorensen, *Dimension reduction of large-scale systems*. Springer, 45, 2005.
- [22] M. Billaud-Friess, *A tensor-based algorithm for the model reduction of high dimensional problems*, Enumath (talk), 2013 .
- [23] P. Benner, S. Gugercin and K. Willcox, *Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems*, *SIAM Review*, Vol. 57, 2015.
- [24] G. Beylkin and M. J. Mohlenkamp, *Algorithms for numerical analysis in high dimensions*. *SIAM J. Sci. Comput.*, 26(6):2133–2159, 2005.
- [25] F. Black, P. Schulze, and B. Unger, *Projection-based model reduction with dynamically transformed modes*. *ESAIM: Mathematical Modelling and Numerical Analysis*, 54(6), 2011–2043, 2020.
- [26] M.-R. Blel, V. Ehrlacher and T. Lelièvre, *Influence of sampling on the convergence rates of greedy algorithms for parameter-dependent random variables*, arXiv:2105.14091, 2021.
- [27] S. Boyaval and T. Lelièvre *A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm*. *Communications in Mathematical Sciences*, 8(3):735–762, 2010.
- [28] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Non asymptotic Theory of Independence*, Oxford University Press, 2013.
- [29] L. Boucinha, A. Ammar, A. Gravouil, and A. Nouy, *Ideal minimal residual-based proper generalized decomposition for non-symmetric multi-field models – Application to transient elastodynamics in space-time domain*. *Comput. Methods Appl. Mech. Eng.*, 273, 56–76, 2014.

- [30] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova and P. Wojtaszczyk, *Convergence Rates for Greedy Algorithms in Reduced Basis Methods*. SIAM Journal on Mathematical Analysis, 43(3):1457–1472, 2011.
- [31] A. Buffa, Y. Maday, A.T. Patera, C. Prud’homme and G. Turinici, *A priori convergence of the greedy algorithm for the parametrized reduced basis*, ESAIM-Math. Model. Numer. Anal.,46(3):595–603,2012.
- [32] T. T. Bui, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*. SIAM J. Sci. Comput., 6 , 3270–3288, 2008.
- [33] J. Bruna, B. Peherstorfer and E. Vanden-Eijnden *Neural Galerkin Scheme with Active Learning for High-Dimensional Evolution Equations*. ArXiv e-prints, 2203.01360, 2022.
- [34] N. Cagniard, R. Crisovan, Y Maday, and R. Abgrall, *Model Order Reduction for Hyperbolic Problems: a new framework*. hal-01583224, 2017.
- [35] D. Cai, C. Yao, and Q. Liao , *A Stochastic Discrete Empirical Interpolation Approach for Parameterized Systems*. Symmetry, 14(3), 556, 2022.
- [36] E. Cances, V. Ehrlacher, and T. Lelievre, *Convergence of a greedy algorithm for high-dimensional convex nonlinear problems*. Mathematical Models and Methods In Applied Sciences, 21(12):2433–2467, 2011.
- [37] E. Cances, V. Ehrlacher, and T. Lelievre, *Greedy algorithms for high-dimensional non-symmetric linear problems*. ESAIM Proc., 41, 95–131, 2013.
- [38] G. Ceruti and C.Lubich, *Time integration of symmetric and anti-symmetric low-rank matrices and Tucker tensors*. Bit Numer. Math., 60(3), 5–614, 2020.
- [39] G. Ceruti and C.Lubich, *An unconventional robust integrator for dynamical low-rank approximation*. Bit Numer. Math., 1–22, 2021.
- [40] G. Ceruti , J. Kusch and C.Lubich, *A rank-adaptive robust integrator for dynamical low-rank approximation*. BIT Numerical Mathematics, 1–26, 2022.
- [41] G. Ceruti, C.Lubich, and D. Sulz *Rank-adaptive time integration of tree tensor networks*. ArXiv e-prints, 2201.10291, 2022.
- [42] J. Chan, *Entropy stable reduced order modeling of nonlinear conservation laws*. Journal of Computational Physics, 423:109789, 2020.
- [43] S. Chaturantabut and D.C. Sorensen, *Nonlinear Model Reduction via Discrete Empirical Interpolation*, SIAM Journal on Scientific Computing, 42(5):2737–2764, 2010.
- [44] M. Cheng, T.Y. Hou and Z. Zhang and D.C. Sorensen, *A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations I: Derivation and algorithms*, Journal of Computational Physics, 242(0):843-868, 2013.
- [45] M. Cheng, T.Y. Hou and Z. Zhang and D.C. Sorensen, *A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations II: Adaptivity and generalizations* , Journal of Computational Physics, 242(0):753 - 776, 2013.
- [46] Y. Chen, J. S. Hesthaven, Y. Maday and J. Rodríguez, *Certified Reduced Basis Methods and Output Bounds for the Harmonic Maxwell’s Equations*. SIAM J. Sci. Comput., 32(2):970–996, 2010.

- [47] F. Chinesta, P. Ladeveze, and E. Cueto, *A short review on model order reduction based on proper generalized decomposition*. Archives of Computational Methods in Engineering, 18(4):395–404, 2011.
- [48] A. Chkifa, A. Cohen and R. DeVore. *Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs*. ESAIM: Mathematical Modelling and Numerical Analysis, 47(1):253–280,2013.
- [49] A. Chkifa, A. Cohen, A., and C. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*. J. Math. Pures Appl., 103(2), 400–428, 2014.
- [50] A. Chkifa, A. Cohen, A., and C. Schwab, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs*. Found. Comput. Math., 14, 601–633, 5, 2014.
- [51] A. Cohen, W. Dahmen, and G. Welper, *Adaptivity and variational stabilization for convection-diffusion equations*. ESAIM: Mathematical Modelling and Numerical Analysis, 46:1247–1273, 8 2012.
- [52] A. Cohen and R. DeVore *Approximation of high-dimensional parametric PDEs* Acta Numer., 24, 1–159, 2015.
- [53] A. Cohen, M.A. Davenport and D. Leviatan, *On the Stability and Accuracy of Least Squares Approximations*. Found. Comput. Math., 13(5):819–834, 2013.
- [54] A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*. SMAI journal of computational mathematics, 3:181–203, 2017.
- [55] A. Cohen and G. Migliorati, *Multivariate approximation in downward closed polynomial spaces*. Contemporary Computational Mathematics-A Celebration of the 80th Birthday of Ian Sloan. Springer, Cham, 233–282, 2018.
- [56] A. Cohen, W. Dahmen, R. DeVore, and J. Nichols, J. *Reduced Basis Greedy Selection Using Random Training Sets*. ESAIM: Mathematical Modelling and Numerical Analysis, 54(5):1509–1524, 2020.
- [57] F. Comets and T. Meyre. *Calcul stochastique et modèles de diffusions*. 2ème éd. Dunod 2015.
- [58] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive petrov–galerkin methods for first order transport equations*. SIAM Journal on Numerical Analysis, 50(5):2420–2445, 2012.
- [59] W. Dahmen, C. Plesken and G. Welper, *Double greedy algorithms: Reduced basis methods for transport dominated problems*. ESAIM Math. Model. Numer. Anal., 48(03):623–663, May 2013.
- [60] A. Dektor, A. Rodgers and D. Venturi, *Rank-adaptive tensor methods for high-dimensional nonlinear PDEs* J. Sci. Comput., 2021, 88 (36),1–27, 2021.
- [61] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*. SIAM J. Matrix Anal. Appl., 21(4):1253–1278, 2000.
- [62] R. DeVore, G. Petrova and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constructive Approximation, 37(3):455–466, 2013.

- [63] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way* Acta Numer., Cambridge University Press, 22:133–288, 2013.
- [64] C. Desceliers, R. Ghanem and C. Soize, *Polynomial chaos representation of a stochastic preconditioner*. Int. J. Numer. Methods Eng., 64(5), 618–634, 2005.
- [65] A. Doostan and G. Iaccarino, *A least-squares approximation of partial differential equations with high-dimensional random inputs*. Journal of Computational Physics, 228(12):4332–4345, 2009.
- [66] M. Drohmann, B. Haasdonk and M. Ohlberger, *Reduced Basis Approximation for Non-linear Parametrized Evolution Equations based on Empirical Operator Interpolation*. SIAM Journal on Scientific Computing, 34(2):A937–A969, 2012.
- [67] V. Ehrlacher, L. Grigori, D. Lombardi and H. Song, *Adaptive Hierarchical Subtensor Partitioning for Tensor Compression*. SIAM Journal on Scientific Computing, 2021.
- [68] M. Eigel, R. Schneider, P. Trunschke and S. Wolf, *Variational Monte Carlo—bridging concepts of machine learning and high-dimensional partial differential equations*. WIAS preprint 2714, 2019.
- [69] M. Eigel, R. Schneider, and P. Trunschke, *Convergence bounds for empirical nonlinear least-squares*. Advances in Computational Mathematics, 45(5), 2503–2532, 2020.
- [70] J. L. Eftang, D. J. Knezevic, and A. T. Patera, *An hp certified reduced basis method for parametrized parabolic partial differential equations*. Math. Comput. Model. Dyn. Syst., 17, 395–422, 2011.
- [71] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, volume 159 of applied mathematical sciences, 2004.
- [72] M. Espig and W. Hackbusch, *A regularized newton method for the efficient approximation of tensors represented in the canonical tensor format*. Numerische Mathematik, 122:489–525, 2012.
- [73] A. Falcó and A. Nouy, *Proper generalized decomposition for nonlinear convex problems in tensor banach spaces*. Numerische Mathematik, 121:503–530, 2012.
- [74] A. Falcó and A. Nouy, *A Proper Generalized Decomposition for the solution of elliptic problems in abstract form by using a functional Eckart-Young approach*. Journal of Mathematical Analysis and Applications, 376(2):469–480, 2011.
- [75] A. Falcó and W. Hackbusch, *On minimal subspaces in tensor representations*. Foundations of Computational Mathematics, 12:765–803, 2012.
- [76] A. Falcó, W. Hackbusch, and A. Nouy, *Geometric Structures in Tensor Representations (Final Release)*. arXiv:1505.03027, 2015.
- [77] A. Falcó, W. Hackbusch, and A. Nouy, *Tree-based tensor formats*. SeMA Journal, 1-15, 2018.
- [78] A. Falcó, and F. Sánchez, *Model order reduction for dynamical systems: a geometric approach*. Comptes Rendus Mécanique, 346(7), 515–523, 2018.
- [79] 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.

- [80] F. Feppon, and P.F.J. Lermusiaux, *Dynamically Orthogonal Numerical Schemes for Efficient Stochastic Advection and Lagrangian Transport*. SIAM Review, 60(3), 595-625, 2018.
- [81] L. Figueroa and E. Suli, *Greedy approximation of high-dimensional Ornstein-Uhlenbeck operators*. Foundations of Computational Mathematics, 12:573–623, 2012.
- [82] M. B. Giles. Multilevel Monte Carlo methods *Acta Numer.*, Cambridge University Press, 259–328, 2015.
- [83] L. Giraldi, A. Nouy, G. Legrain, and P. Cartraud, *Tensor-based methods for numerical homogenization from high-resolution images*. Computer Methods in Applied Mechanics and Engineering, 254(0):154 – 169, 2013.
- [84] S. Glas, A. Mayerhofer, K. Urban, *Two Ways to Treat Time in Reduced Basis Methods*. In Model Reduction of Parametrized Systems, 2017.
- [85] E. Gobet and S. Maire. A spectral Monte Carlo method for the Poisson equation. *Monte Carlo Methods and Applications mcma*, 10:275-285, 2004.
- [86] E. Gobet and S. Maire. Sequential control variates for functionals of Markov processes. *SIAM Journal on Numerical Analysis*, 43:1256-1275, 2005.
- [87] E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Processes and Their Applications*, 120(2):130–162, 2010.
- [88] E. Gobet, *Monte-Carlo methods and stochastic processes: from linear to non-linear*. Chapman and Hall/CRC, 2016.
- [89] C. Graham and D. Talay, *Stochastic Simulation and Monte Carlo Methods*. Springer, 2013.
- [90] L. Grasedyck, *Hierarchical singular value decomposition of tensors*. SIAM J. Matrix Anal. Appl., 31:2029–2054, 2010.
- [91] L. Grasedyck, D. Kressner and C. Tobler, *A literature survey of low-rank tensor approximation techniques*. GAMM–Mitteilungen., 36(1), 53–78, 2013.
- [92] C. Greif and K. Urban, *Decay of the Kolmogorov  $N$ -width for wave problems*. Applied Mathematics Letters, 96, 216–222, 2019.
- [93] E. Grelier and A. Nouy and M. Chevreuril. *Learning with tree-based tensor formats*. arXiv preprint arXiv:1811.04455, 2018.
- [94] M. A. Grepl and A. T. Patera, *A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations*. ESAIM Math. Model. Numer. Anal., 39(1):157–181, 2005.
- [95] B. Haasdonk and M. Ohlberger, *Reduced basis method for finite volume approximations of parametrized linear evolution equations*. ESAIM Math. Model. Numer. Anal., 42:277–302, 2008.
- [96] B. Haasdonk and M. Ohlberger, *Reduced basis method for explicit finite volume approximations of nonlinear conservation laws*, Proc. 12th International Conference on Hyperbolic Problems: Theory, Numerics, Application, 2008.

- [97] B. Haasdonk and M. Ohlberger, *Efficient reduced models and a posteriori error estimation for parametrized dynamical systems by offline/online decomposition*. Math. Comput. Model. Dyn. Syst., 17:145–161, 2011.
- [98] B. Haasdonk, *Convergence rates of the POD-greedy method*. M2AN Math. Model. Numer. Anal., 47 859–873, 2015.
- [99] B. Haasdonk, *Chapter II : Reduced basis methods for parametrized PDEs - A tutorial introduction for stationary and instationary problems.*, In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (eds.), Model Reduction and Approximation: Theory and Algorithms. SIAM, Philadelphia, PA, 2017.
- [100] C. Haberstich, A. Nouy, and G. Perrin, *Boosted optimal weighted least-squares*, *arXiv preprint arXiv:1912.07075v2*, 2019.
- [101] C. Haberstich, A. Nouy, and G. Perrin, *Active learning of tree tensor networks using optimal least-squares*, *arXiv preprint arXiv:2104.13436v1*, 2021.
- [102] W. Hackbusch and S. Kuhn, *A New Scheme for the Tensor Representation*. Journal of Fourier analysis and applications, 15(5):706–722, 2009.
- [103] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, Springer Science and Business Media, 2012.
- [104] S. Holtz, T. Rohwedder, and R. Schneider, *On manifolds of tensors with fixed TT rank*. Numer. Math., 120(4):701–731, 2012.
- [105] S. Holtz, T. Rohwedder, and R. Schneider, *The Alternating Linear Scheme for Tensor Optimisation in the TT format*. SIAM Journal on Scientific Computing, 34(2):683–713, 2012.
- [106] C. Homescu, L. R. Petzold and R. Serban, *Error Estimation for Reduced-Order Models of Dynamical Systems*. SIAM Journal on Numerical Analysis, 43(4):1693–1714, 2006.
- [107] B.A. Ibrahimoglu, *Lebesgue functions and Lebesgue constants in polynomial interpolation*, Journal of Inequalities and Applications: 2016:93, 2016.
- [108] A. Janon, M. Nodet and C. Prieur, *Certified reduced-basis solutions of viscous Burgers equation parametrized by initial and boundary values*. ESAIM: M2AN, 47(2):317–348, 2013.
- [109] M. Kahlbacher and S. Volkwein, *Galerkin proper orthogonal decomposition methods for parameter dependent elliptic systems*. Discussiones Mathematicae: Differential Inclusions, Control and Optimization, (27):95–117, 2007.
- [110] Y. Kazashi, F. Nobile, F. and E. Vidličková, *Stability properties of a projector-splitting scheme for dynamical low rank approximation of random parabolic equations*. Numerische Mathematik, 149(4), 973–1024, 2021.
- [111] B. N. Khoromskij and C. Schwab, *Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs*. SIAM Journal on Scientific Computing, 33(1):364–385, 2011.
- [112] E. Kieri, C. Lubich, H. Walach, *Discretized dynamical low-rank approximation in the presence of small singular values*. SIAM Journal on Numerical Analysis, 54(2), 1020-1038, 2016.

- [113] E. Kieri, B. Vandereycken, B., *Projection methods for dynamical low-rank approximation of high-dimensional problems*. Computational Methods in Applied Mathematics, 19(1), 73-92, 2019.
- [114] O. Koch and C. Lubich, *Dynamical low-rank approximation*. SIAM Journal on Matrix Analysis and Applications, 29(2), 434-454, 2007.
- [115] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer Science and Business Media, 2013.
- [116] D. Kressner and C. Tobler, *Low-rank tensor krylov subspace methods for parametrized linear systems*. SIAM Journal on Matrix Analysis and Applications, 32(4):1288–1316, 2011.
- [117] P. Ladevèze, *Nonlinear Computational Structural Mechanics - New Approaches and Non-Incremental Methods of Calculation*. Springer Verlag, 1999.
- [118] P. Ladevèze, J.C. Passieux, and D. Néron, *The LATIN multiscale computational method and the Proper Generalized Decomposition*. Computer Methods in Applied Mechanics and Engineering, 199(21-22):1287–1296, 2010.
- [119] T. Lattimore, and C. Szepesvári, *Bandit Algorithms*. Free online edition, 2022.
- [120] R.J. Leveque, *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.
- [121] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken, *Dynamical approximation by hierarchical tucker and tensor-train tensors*. SIAM Journal on Matrix Analysis and Applications, 34(2):470–494, 2013.
- [122] C. Lubich C. and I.V. Oseledets. *A projection-splitting integrator for dynamical low-rank approximation*. BIT Numerical Mathematics, 54(1), 171-188, 2014.
- [123] A. Macherey, *Approximation and model reduction for partial differential equations with probabilistic interpretation*. PhD Thesis, Ecole Centrale de Nantes, [tel-03351933](https://tel.archives-ouvertes.fr/tel-03351933), 2021.
- [124] Y. Maday, N. C. Nguyen A.T. Patera and S. H. Pau, *A general multipurpose interpolation procedure: the magic points*, Communications on Pure and Applied Analysis, 8(1):383–404, 2009.
- [125] H. G. Matthies and E. Zander, *Solving stochastic systems with low-rank tensor compression*. Linear Algebra and its Applications, 436(10), 2012.
- [126] H. G. Matthies and E. Zander, *Solving stochastic systems with low-rank tensor compression*. Linear Algebra and its Applications, 436(10), 2012.
- [127] O. Mula, Y. Maday, and G. Turicini, *Convergence analysis of the generalized empirical interpolation method*, SIAM J. Numer. Anal., 54(3):1713–1731, 2016.
- [128] E. Musharbash E., F. Nobile F. and T. Zhou, *On the dynamically orthogonal approximation of time-dependent random PDEs*, SIAM Journal on Scientific Computing, 2015, 37(2):A776-A810, 2015.
- [129] N.C. Nguyen, G. Rozza and A.T. Patera, *Reduced basis approximation and a posteriori error estimation for the time-dependent viscous Burgers' equation*. Calcolo, 46(3):157–185, 2009.

- [130] A. Nonnenmacher, C. Lubich, Dynamical low-rank approximation: applications and numerical experiments. *Mathematics and Computers in Simulation*, 79(4):1346–1357, 2008.
- [131] A. Nouy, *A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations*, *Computer Methods in Applied Mechanics and Engineering*, 196(45-48) : 4521–4537, 2007.
- [132] A. Nouy, *Proper Generalized Decompositions and Separated Representations for the Numerical Solution of High Dimensional Stochastic Problems*, *Archives of Computational Methods in Engineering*, 17(4):403–434, 2010.
- [133] A. Nouy, *A priori model reduction through proper generalized decomposition for solving time-dependent partial differential equations*. *Computer Methods in Applied Mechanics and Engineering*, 199(23-24):1603–1626, 2010.
- [134] A. Nouy, *Low-rank tensor methods for model order reduction*. In R. Ghanem, D. Higdon, H. Owhadi (Eds), *Handbook of Uncertainty Quantification*. Springer International Publishing, Cham, 2017.
- [135] A. Nouy, *Chapter IV : Low-rank methods for high-dimensional approximation and model order reduction*. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (eds.), *Model Reduction and Approximation: Theory and Algorithms*. SIAM, Philadelphia, PA, 2017.
- [136] A. Nouy. *Higher-order principal component analysis for the approximation of tensors in tree-based low rank formats*. *Numerische Mathematik*, 141:743–789, 2019.
- [137] M. Ohlberger, M. and S. Rave, *Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing*. *Comptes Rendus Mathématique*, 351(23-24):901–906, 2013.
- [138] M. Ohlberger, M. and S. Rave, *Reduced Basis Methods: Success, Limitations and Future Challenges*. arXiv:1511.02021, 2015.
- [139] I. V. Oseledets and E. E. Tyrtysnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*. *SIAM J. Sci. Comput.*, 31(5):3744–3759, 2009.
- [140] I. V. Oseledets, *Tensor-train decomposition*. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.
- [141] B. Peherstorfer, *Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling*, arXiv:1812.02094, 2018.
- [142] N. A. Pierce and M. B. Giles, *Adjoint Recovery of Superconvergent Functionals from PDE Approximations*. *SIAM Rev.*, 42(2):247–264, 2000.
- [143] C. Prud’homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, a. T. Patera and G. Turinici, *Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods*. *J. Fluids Eng.*, 124(1):70, 2002.
- [144] A. Quarteroni, A. Manzoni and F. Negri, *Reduced Basis Methods for Partial Differential Equations*. Springer International Publishing., 2016.
- [145] J. Reiss, P. Schulze, J. Sesterhenn and V. Mehrman, *The Shifted Proper Orthogonal Decomposition: A Mode Decomposition for Multiple Transport Phenomena*. *SIAM Journal on Scientific Computing*, 40(3):A1322–A1344, 2018.

- [146] D. Rim, B. Peherstorfer, and K. Mandli Manifold Approximations via Transported Subspaces: Model reduction for transport-dominated problems. arXiv:1912.13024, 2019.
- [147] A. Rodgers, A. Dektor and D. Venturi, *Adaptive integration of nonlinear evolution equations on tensor manifolds*, arXiv, arXiv:2008.00155, 2021.
- [148] T. Rohwedder and A. Uschmajew, *On local convergence of alternating schemes for optimization of convex problems in the tensor train format*. *SIAM J. Numer. Anal.*, 51(2):1134–1162, 2013.
- [149] T.P. Sapsis and P.F.J. Lermusiaux, *Dynamically orthogonal field equations for continuous stochastic dynamical systems*. *Physica D: Nonlinear Phenomena*, 238(23-24):2347–2360, 2009.
- [150] K. Smetana, O. Zahm and A.T. Patera, *Randomized Residual-Based Error Estimators for Parametrized Equations*. *SIAM J. Sci. Comput.*, 41(2), A900-A926, 2019.
- [151] K. Smetana and O. Zahm, *Randomized residual-based error estimators for the proper generalized decomposition approximation of parametrized problems*. *International Journal for Numerical Methods in Engineering*, 121(23), 5153–5177, 2020.
- [152] K. Steih and K. Urban, *Space-time Reduced Basis methods for time-periodic parametric partial differential equations*. Ulm University, 2011.
- [153] S.J. Smith, *Lebesgue constants in polynomial interpolation* *Annales Mathematicae et Informaticae*, 33:109–123, 2006.
- [154] , T. Taddei, S. Perotto, A. Quarteroni, *Reduced basis techniques for nonlinear conservation laws*, *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(3):787–814, 2015.
- [155] V. Temlyakov, *Greedy approximation*. *Acta Numerica*, 17:235–409, 2008.
- [156] V. Temlyakov, *Greedy Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2011.
- [157] E. Tyrtyshnikov, *Incomplete Cross Approximation in the Mosaic-Skeleton Method*. *Computing*, 64(4), 367–380, 2000.
- [158] K. Urban and A.T. Patera, *A new error bound for reduced basis approximation of parabolic partial differential equations*. *Comptes Rendus Mathématique*, 350(3): 203–207, 2012.
- [159] A. Uschmajew and B. Vandereycken, *The geometry of algorithms using hierarchical tensors*. Technical report, ANCHP-MATHICSE, Mathematics Section, EPFL, 2012.
- [160] A. Uschmajew, *Local Convergence of the Alternating Least Squares Algorithm for Canonical Tensor Approximation*. *SIAM J. Matrix Anal. Appl.*, 33(2), 639–652, 2012.
- [161] A. Uschmajew and B. Vandereycken, *Geometric Methods on Low-Rank Matrix and Tensor Manifolds*. *Handbook of Variational Methods for Nonlinear Geometric Data*. Springer, 2020.
- [162] S. Volkwein, *Model reduction using proper orthogonal decomposition*, Lecture notes, 2008.
- [163] E. Weinan, H. Jiequn and A. Jentzen. *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.
- [164] G. Welper, *Interpolation of Functions with Parameter Dependent Jumps by Transformed Snapshots*. *SIAM Journal on Scientific Computing*, 39(4): A1225–A1250, 2017.

- [165] D. Wirtz, D.C. Sorensen and B. Haasdonk, *A-posteriori error estimation for DEIM reduced nonlinear dynamical systems*, SIAM J. Sci. Comput., 36(2):A311–A338, 2014.
- [166] J. Xu, and L. Zikatanov, *Some observations on Babuska and Brezzi theories*. Numerische Mathematik, 94(1):195–202, 2003.
- [167] O. Zahm, *Reduced Order Modeling for parameter dependent equations - Application to uncertainty quantification*, PhD Thesis, Ecole Centrale de Nantes, [tel-01256411](tel:01256411), 2015.
- [168] O. Zahm and A. Nouy, *Interpolation of Inverse Operators for Preconditioning Parameter-Dependent Equations*. SIAM J. Sci. Comput., 38(2), A1044–A1074, 2016.

## References related to PhD and post-doctorate

- [169] G. Allaire, S. Clerc, and S. Kokh, S. *A Five-Equation Model for the Simulation of Interfaces between Compressible Fluids*. Journal of Computational Physics, 181(2):577–616, 2002.
- [170] G. Allaire, S. Clerc, and S. Kokh, S. *A five-equation model for the numerical simulation of interfaces in two-phase flows*. Comptes Rendus de l’Academie des Sciences Series I Mathematics, 12 Part 1(331):1017–1022, 2000.
- [171] R. Caiden R. P. Fedkiw, and C. Anderson, *A numerical method for two phase flow consisting of separate compressible and incompressible regions*, Journal of Computational Physics, 166(1):1–27, 2001.
- [172] B. Després and F. Lagoutière, *Un schéma non linéaire anti-dissipatif pour l’équation d’advection linéaire*, C. R. Acad. Sci. Paris Sér. I Math. 328(10):939–944, 1999.
- [173] B. Després, F. Lagoutière, *Contact discontinuity capturing schemes for linear advection and compressible gas dynamics* J. Sci. Comput. 16(4):479–524, 2002.
- [174] D.A. Di Pietro, D. Lo Forte, N. Parolini, *Mass preserving finite element implementations of the level set method*, Applied Numerical Mathematics, 56(9):1179–1195, 2006.
- [175] L. P. Franca, and S. L. Frey, *Stabilized finite element methods : II. The incompressible Navier-Stokes equations*, Computer Methods in Applied Mechanics and Engineering, 99(2): 209–233, 1992.
- [176] S. Galera, P.H. Maire and J. Breil, *A two-dimensional unstructured cell-centered multi-material ALE scheme using VOF interface reconstruction*, Journal of Computational Physics, 229(16):5755–5787, 2010.
- [177] S. Galera, J. Breil and P.H. Maire, *A 2D unstructured multi-material Cell-Centered Arbitrary Lagrangian-Eulerian (CCALE) scheme using MOF interface reconstruction* Computers and Fluids, 1(46): 237–244, 2011.
- [178] G. Hauke, *Simple stabilizing matrices for the computation of compressible flows in primitive variables*. Computer Methods in Applied Mechanics and Engineering, 190:6881–6893, 2001.
- [179] G. Hauke and T. J. R. Hughes, *A unified approach to compressible and in-compressible flows*, Computer Methods in Applied Mechanics and Engineering, 113:389–395, 1994.

- [180] G. Hauke and T. J. R. Hughes, *A comparative study of different sets of variables for solving compressible and incompressible flows*, Computer Methods in Applied Mechanics and Engineering, 153(1–2) :1–44, 1998.
- [181] S. Jaouen and F. Lagoutière, *Numerical transport of an arbitrary number of components*. Computer Methods in Applied Mechanics and Engineering, 196(33–34): 3127–3140, 2007.
- [182] S. Kokh, S. and F. Lagoutière, *An anti-diffusive numerical scheme for the simulation of interfaces between compressible fluids by means of a five-equation model*. Journal of Computational Physics, 229(8):2773–2809, 2010.
- [183] E. Marchandise and J.-F. Remacle, *A stabilized finite element method using a discontinuous level set approach for solving two phase incompressible flows*. Journal of Computational Physics, 219(2):780–800, 2006.
- [184] E. Marchandise, J.-F. Remacle and N. Chevaugeon, *A quadrature-free discontinuous Galerkin method for the level set equation*, Journal of Computational Physics, 212(1):338–357, 2006.
- [185] S. Osher and J.A. Sethian, *Fronts propagating with curvature dependant speed: algorithms based on based on Hamilton-Jacobi formulations*. Journal of Computational Physics, 79(1):12–49, 1988.
- [186] R. Saurel, and R. Abgrall, *A multiphase Godunov method for compressible multifluid and multiphase flows*. Journal of Computational Physics, 150(2):425–467, 1999.