Data assimilation with reduced basis and noisy measurement : Applications to nuclear reactor cores

Helin Gong

Électricité de France
EDF R&D

Sorbonne Université
Paris 6

Laboratoire Jacques-Louis Lions
UMR 7598

Thèse de Doctorat

de Sorbonne Université

pour l'obtention du grade de

Docteur en Mathématiques Appliquées

par

Helin Gong

Couplage de réduction de modèles et de mesures bruitées : Applications à l'assimilation de données pour les cœurs de centrales nucléaires

Data assimilation with reduced basis and noisy measurement: Applications to nuclear reactor cores

sous la direction de

Yvon Maday,

Olga Mula,

Jean-Philippe Argaud,

et Bertrand Bouriquet

Rapportée par

M.	Ludovic Chamoin	LMT Cachan
M.	Gianluigi Rozza	SISSA

Soutenue le 13 Juillet 2018 devant le jury composé de

M.	Ludovic Chamoin	Rapporteur
M.	Gianluigi Rozza	Rapporteur
M.	Ansar Calloo	Examinateur
M.	Bruno Després	Examinateur
M.	Jean-Philippe Argaud	Encadrant industriel
M.	Bertrand Bouriquet	Encadrant industriel
M.	Yvon Maday	Directeur de Thèse
Mme	Olga Mula	Co-Directeur de Thèse

École Doctorale de Sciences Mathématiques
de Paris Centre ED 386

Faculté de Mathématiques
UFR 929

Helin GONG :

Sorbonne Universités, UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France.
Électricité de France, R&D, 7 Boulevard Gaspard Monge, 91120 Palaiseau, France.

Adresse électronique: gong@ann.jussieu.fr, helin.gong@edf.fr

Remerciements
Acknowledgments

To Ruoqing Fan and Shuyan Gong,
my lovely wife and daughter.

Along with the writing of the acknowledgments of my PhD thesis, I recognize that one of the most important period of my life is coming to an end and a new career is about to start. At this end and start point, I am extremely grateful for many people that have made, in various ways, this PhD to be such an extraordinary journey of my life.

This thesis represents not only my work in the neutronic simulation domain and the applied mathematic domain, it is a milestone of my career. My passion is the field of large scale science and engineering computation. Thanks to the Laboratoire Jacques Louis Lions (LJLL, Sorbonne Universités, UPMC Univ Paris 06), which is one of the laboratories that stand at the very top world-wide in mathematics and the Électricité de France (EDF R&D), which is a global electricity company in the spotlight, I have been given opportunities to learn parts of the state-of-the-art of the applied mathematics in industry applications.

First and foremost I wish to thank my supervisor, Professor Yvon Maday, at LJLL. He has been supportive since the days I began working at LJLL as an assistant engineer in computing science since the September of 2014. Without his guidance, patience, enthusiasm, generosity and, above all, incredible scientific knowledge and inexhaustible mathematical imagination, none of this would have been possible. I remember he used to say something like "now you have to learn how to write a good article, you will finish your thesis in one to two years, then you have to write articles alone, further more, you might be a teacher and teach your students $\cdots$" to encourage me to improve. Yvon has supported me not only by providing a research assistantship over almost four years, but also academically and emotionally through the rough road to finish this thesis. Despite an agenda more complicated than the one of any minister, Yvon always made time for a meeting and managed to check my work. Yvon, I have been incredibly fortunate to have you as a supervisor and I am deeply grateful for opening the door to academic access, and all the support you have given me.

A huge thank you goes also to my co-supervisor, Professor Olga Mula, at Paris Dauphine Univer-

sity CEREMADE. With Olga, I have not only worked with an outstanding researcher in mathematics but also with an extraordinary person. She was always on-line to kindly answer my never-ending lists of questions about mathematics. She always kindly improve my writings, from the very rough, imprecise to the ultimate well-written one, for which I am most grateful. Olga, I have also been extremely fortunate to have had the chance to have such a great improvement in mathematic also thanks to you.

The most sincerest and deepest thanks I also give to my two industrial supervisors: Dr. Jean-Philippe Argaud and Dr. Bertrand Bouriquet, at EDF R&D. I was very lucky to have two very competent industrial supervisors in neutronic and computing domains. Thank you both for trusting me in this work despite all my defects that you knew. Jean-Philippe and Bertrand, you have both always been my references for any technical and administrative problem. Your guidance and your advice have helped me enormously in the organization of my work and in the progress of this thesis. I would mention that with your help, my life in France is going very well though the administrative process in France is complicated for an non-French speaker foreigner. I remember the first time I asked for my residence permit, when I arrived to the prefecture at 06:30 am, both of you already arrived, in order to help me to make the administrative process easy. Your availability, your patience and your time reserved for my work, all are immeasurable.

I would like to thank all the members of the jury: Professor Ludovic Chamoin and Professor Gianluigi Rozza for having accepted to be the reviewers of this thesis, for having conscientiously read my work and having spent your time to write the constructive reports. My sincere thanks also goes to Professor Bruno Després, researcher Ansar Calloo for your examination as well as your interest in my work. And thanks also goes to Professor Jean-Frédéric Gerbeau for the mid-term examination of my thesis.

A very big thanks is expressed to Professor Frédéric Hecht who has helped me a lot with the using of Freefem++, and he also gave me a lot of advice in mathematics. It would have been hard to accomplish this work without your help. I would also like to thank Professor Albert Cohen, for your interest in my work, for your enthusiasm and patience in enlightening me with knowledge about least-squares method, data assimilation knowledge, etc. The very big thanks is also expressed to Professor Kan Wang from Tsinghua University for your letter of recommendation, and the more than ten years advices and help in any circumstances.

This research was done in EDF R&D PERICLES (SINETICS) Department and the LJLL Laboratory of Sorbonne Université. I would like to thank these two establishments for hosting me during the entire research period. In particular, I would like to express my gratitude to the group leader of I23 at EDF, Emile Razafindrakoto, it because of you, the three years work at I23 goes rather well. Thank François De Caso and Francesco Silva, we have had a very pleasant discussing discussion during your internship, and large amount of industry related work of the thesis is also based on your excellent work. Thank Rebecca Jeffers for manly help in writings. I would also like to thank all the colleagues (Olivier Boiteau, Frank Hulsemann, Bruno Lathuilière, Romain Jonchière, Angelique Ponçot, Gérald Nicolas, Laurent Plagne, Soizic Peron, François Fèvotte, Lucie Berthon, Alexandre

Martin Sanchez, Romain Picot, Sibo Cheng, Dimitri Mottet et al.) at EDF that have been on my side during these three years of work. Further more, thank you to Professor Benoît Perthame, Professor Edwige Godlewski, Merbouha Lounic, Catherine Drouet, Malika Larcher, Khashayar Dadras, Suyan Zhou, Jingxuan Ma, Alain Yuan, Wenhui Du et al. for your support in science and administrative aspects at EDF and UPMC.

Special thanks to Pierre-Henri Tournier for the helps in Freefem++ and many software issues, to Marc Hoffmann for your discussions in statistics, to Anouk Nicolopoulos for lots of discussions on my PhD thesis and other helps for the easy life in the office at LJLL, to Romain Picot for the helps for the easy life in the office at EDF, to Tommaso Taddei for discussions and large amount of help in writing the thesis. Thank you to Chaoyu Quan, Yashan Xu, Haisen Zhang, Cailong Liu, Salli Moustafa, Matthew Paddick, Hongjun Ji, Martin Strugarek, Valentin Pages, Lilian Glaudin, Virgile Dubos, Anne-Françoise de Guerny, Jules Pertinand, Amaury Hayat, Shengquan Xiang, Shijie Dong, Yangyang Cao, Chengyou Jiang et al. for all your support, advices and help in any circumstances.

Very special thanks to all my ever advisors, leaders and colleagues from the Nuclear Power Institute of China (Qi Luo, Xiaogang Yuan, Qing Li, Fenggang Zang, Hongxing Yu, Bin Zhang, Yingrui Yu, Zhang Chen, Bangyang Xia, Yan Wen, Xiao Zheng, Zumao Yang, Hongbin Liu, Yandan Zhao, Ying Hu, Liuhui Huang et al.) for all your support, advices and help in any circumstances.

Very special thanks to all my lifelong friends in China (Wei Gong, Jing Gong, Qiming Zhai, Qiubo Cheng, Minggang Li, Zhengxin Li, Jingang Liang, Yan Sun, Junji Chen, Xiaoping Lu et al.) and also to all my friends from NPIC (Xiangyang Li, Qiwei Liu, Wenbo Zhao, Tianya Li, Di Lu, Hongzhi Xiang, Yun Cai, Xiao Tang, Xi Wu, Xingjie Peng et al.), and friends in Paris (Ying Zhang, Xin Sun, Yilin Chen, Yihong Ge, Jiahui Yuan, Qiang Xu, Zhanyang He), your support has been very important for me.

Last, but certainly not least, I would like to thank my parents and the rest of my family for the continuous encouragement, unlimited patience and daily support.

This book is for Ruoqing Fan and Shuyan Gong, my lovely wife and daughter.

Contents

Part I

Mathematic background and development

Chapter 1

Introduction

Contents

1.1 Objective of this work: state estimation

A nuclear power plant is a complex physical system to produce electricity. To both improve security and optimize the exploitation, it is necessary to know the neutronic state of the nuclear reactor core as well as possible. It is in particular mandatory to know the behavior of the core in current operation as in accidental case. This knowledge can be achieved through measurements, and also through mathematical/numerical models. Various numerical models have been developed since years, and give a very good estimation of the fine neutronic behavior of the core [81, 153]. Moreover, measurements, obtained during start-up tests as well as in operational context, allow to improve the quality of the knowledge of the whole core state and possibly correct bias from the model that can not take into account every single phenomenon occurring in the real process. This ensemble *computing + measurements* must then be considered as a whole in order to get the best estimation of this complex system state.

Integration of mathematical/numerical models with experimental observations is a general way to perform accurate predictions. Mathematical/numerical models of physical systems are often

deficient due to i) the uncertainty in the value of the parameters representing material properties and input forces, and ii) might also neglect some aspects of the system's behavior (model bias quoted above). On the other hand, experimental measurements are often scarce, corrupted by noise, and they might also provide only indirect measurements of the quantities we wish to predict.

Data assimilation is a set of mathematical and numerical methods that allow to obtain a better estimation of the true state of a physical system. In order to achieve this goal, these techniques are using both the information provided by experimental devices and the information provided by an a priori knowledge of the system (usually given by a numerical model with its associated parameters). Data assimilation is in particular the major source of improvement of the weather forecasting since the last 40 years, and it has been used successfully in several studies on the nuclear core analysis.

Data assimilation tasks present several challenges for applied mathematicians and engineers. Mathematical models involved in data assimilation often consist of (systems of) parameter-dependent Partial Differential Equations (PDEs) that are usually expensive to solve: since state-of-the-art data assimilation procedures are cast as optimization problems allowing to fit the best, which hence involve many model evaluations, the computational burden might be challenging for real applications, even unsustainable for real-time and in situ applications. This challenge becomes even more severe when the available mathematical model is affected by substantial data uncertainty: in this case, current research focuses on the development of numerical techniques that i) are directly informed by the specific data assimilation task at hand, and ii) meet the computational (time and memory) constraints, iii) are designed to appropriately take into account the data uncertainty in the model.

Recent advances in reduced order models (ROM) obtained from reduced basis (RB) methods for parameterized systems offer new opportunities for the integration of models and experimental data:

i) ROM allows extracting the principal component information of the system and also provides a guideline to set the amount and measurement locations needed, and also a way to learn the physical system,

ii) ROM techniques speed up computations, allowing better explorations of the parameter space at an acceptable computational cost.

iii) ROM provides actionable tools to compress our prior knowledge about the system coming from the parameterized mathematical model into low-dimensional and more manageable forms.

The goal of the thesis is to improve the physical and numerical interpretation of the information involved in data assimilation with efficient model reduction strategies for systems modeled by PDEs. More specifically, we shall focus on the data assimilation task: state estimation for stationary problems, especially neutronic state estimation in nuclear reactor applications.

1.2 State estimation with data assimilation

1.2.1 General overview

Let V be a Hilbert space over a domain $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ endowed with inner product $(\cdot,\cdot)$ and induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$. We next introduce the standard m-dimensional Euclidean space $\mathbb{R}^m$, $m \in \mathbb{N}$ with ℓ^p-norm $\|\boldsymbol{y}\|_{\ell^p(\mathbb{R}^m)} := \left(\sum_{i=1}^m |y_i|^p\right)^{1/p}$, $\forall \boldsymbol{y} \in \mathbb{R}^m, 1 \leq p < \infty$, for simplicity we further denote $\|\boldsymbol{y}\|_{\ell^p(\mathbb{R}^m)}$ by $\|\boldsymbol{y}\|_p$. In the framework of state estimation, the goal is to understand the inaccessible true value of a physical phenomenon/state $u^t \in V$, where the t stands for *true state* of interest. We can access this goal either through an *implicit* way based on the knowledge of a physical model or through an *explicit* way based on the knowledge from observations collected from the physical system itself.

i) Implicit knowledge acquisition from a physical model: In general, the physical model can be described in an implicit form through an operator equation e.g. a PED of the form

$$\mathcal{F}(u, \mu) = 0, \tag{1.2.1}$$

where $u \in V$ is a representation of the state, which depends on $p \geq 1$ essential parameters gathered in a vector $\mu \in \mathbb{R}^p$. To account for different possible working conditions for the particular physical phenomenon, we assume that these p parameters vary in a compact set $\mathcal{D} \subset \mathbb{R}^p$, by solving this implicit form numerically for each μ, we get a set of all possible states given by

$$\mathcal{M}_u \coloneqq \{u(\cdot, \mu) \ : \mu \in \mathcal{D}\}, \tag{1.2.2}$$

which is called the manifold of states. The knowledge of u thus directly depends on the knowledge of μ.

ii) Explicit knowledge acquisition from observations: The physical state of interest can also be represented explicitly as $u^t(x, \mu)$ at point $x \in \Omega$ for a particular parameter state μ by pointwise sensors that are installed in Ω, or less explicitly represented on average, e.g., a local average over u^t centered at x. More generally, one can estimate the state by an m-vector observation $\boldsymbol{y}^o \in \mathbb{R}^m$, where the o stands for *observation*. Let us assume that the observation $\boldsymbol{y}^o$ comes from an m-dimensional observation operator $\boldsymbol{\ell}(u^t) : V \to \mathbb{R}^m$ which maps the state space into the observation space, thus

$$\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o, \tag{1.2.3}$$

where $\boldsymbol{e}^o$ is an m-dimensional measurement error vector, considered stochastic, whose mean is $\bar{\boldsymbol{e}}^o = E[\boldsymbol{e}^o]$ that describes instrument errors, where $E[\cdot]$ is the expectation operator *.

Finally, there is a possibility to represent the physical state of interest based on the knowledge from a physical model combined with measurements. Thus, in this framework of state estimation, the problem can be described as:

Problem 1.2.1 (State estimation with data assimilation) *Find $u^* \in V$ as an approximation of the true state u^t from observation $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o$ combined with the information from a physical model (e.g., given by Equation* (1.2.1)*).*

Let us first assume that no model information is available, then the state estimation Problem 1.2.1 above can be e.g., defined as a minimization problem: find an approximation state $u^* \in V$ of u^t from observations $\boldsymbol{y}^o \in \mathbb{R}^m$ by solving a classical least-squares problem

$$u^* = \underset{u \in V}{\arg\min}\, J(u) = \underset{u \in V}{\arg\min}\, \frac{1}{2} \|\boldsymbol{y}^o - \boldsymbol{\ell}(u)\|_2^2 \ . \tag{1.2.6}$$

*We consider a random variable $\boldsymbol{z} \in \mathbb{R}^m$, defined as a function on an underlying probability space, and taking values in $\mathbb{R}^m$. Associated with this random variable is an induced probability measure ν on $\mathbb{R}^m$. Furthermore, we work with a sufficiently rich collection of subsets of $\mathbb{R}^m$ to each of which we are able to assign the probability that $\boldsymbol{z}$ is contained in it; this collection of subsets is termed a σ-algebra. Throughout these notes we work with the Borel σ-algebra generated by the open sets, $\mathcal{B}_\nu(\mathbb{R}^m)$, and abbreviate this σ-algebra by $\mathcal{B}_\nu$. We have defined a probability triple $(\mathbb{R}^m, \mathcal{B}_\nu, \nu)$. We further assume that $\boldsymbol{z}$ has a strictly positive probability density function (pdf) $\rho : \mathbb{R}^m \to \mathbb{R}^+$, which satisfies

$$\int_{\mathbb{R}^m} \rho(\boldsymbol{z}) d\boldsymbol{z} = 1. \tag{1.2.4}$$

In this setting, we denote by $E(\boldsymbol{z})$ the expected value of the random variable $\boldsymbol{z}$ on $\mathbb{R}^m$; this expectation is given by

$$E[\boldsymbol{z}] = \int_{\mathbb{R}^m} \rho(\boldsymbol{z}) \boldsymbol{z} d\boldsymbol{z}. \tag{1.2.5}$$

We can further denote the $m \times m$ covariance matrix $\mathbf{R} \equiv E\left[(\boldsymbol{z} - E[\boldsymbol{z}])(\boldsymbol{z} - E[\boldsymbol{z}])^T\right]$, where T stands for the transpose.

which is never mathematically well posed when V is infinite dimensional. To ensure the well posedness of problem (1.2.6) for all possible configurations, we need to introduce some information from an a priori knowledge on the state of the system, u^b, where b stands for *background* which can be the *initial state*, *first guess/forecast* from the model, e.g., PDEs. The background u^b is also an approximation to the true state u^t which satisfies

$$u^b = u^t + e^b \ , \tag{1.2.7}$$

where e^b is the background error, e.g., the model error. The optimal estimator u^* is therefore the result of the minimization of a quadratic function $J(u)$, composed of two terms, one associated with the background error, the other with the discrepancies between the observations and the image of the state in the observation space:

$$J(u) = \frac{1}{2}\lambda\|u - u^b\|^2 + \frac{1}{2}\|\boldsymbol{y}^o - \boldsymbol{\ell}(u)\|_2^2, \tag{1.2.8}$$

here $\lambda > 0$ is a penalty factor that needs to be carefully selected. This approach consists in penalizing the gap between the assimilation model and the observations by a back-pulling term proportional to the norm of the gap between the assimilation model and the background. To this end, the objective of *data assimilation* is to weigh each term in the cost function in Equation (1.2.8) by its precision, or equivalently by the inverse of its uncertainty (modeled as an error covariance matrix).

In practice, we always consider the discretization of the state u, for example, the solution $u^{\mathcal{N}}$ in a finite element subspace $V^{\mathcal{N}} \subset V$ of typically large dimension $\mathcal{N}$. Let us further assume $\boldsymbol{x}^t \in \mathbb{R}^{\mathcal{N}}$ is the coordinate vector of $u^{\mathcal{N}} \in V^{\mathcal{N}}$ for a given orthonormal basis. Thus, our interest is the discrete state $\boldsymbol{x}^t \in \mathbb{R}^{\mathcal{N}}$ instead of the continuous one $u^t \in V$. We follow the standard notation for data assimilation problem as formalized by Ide et al [113], denote by $\mathbf{H}$ the discretization of observation functional $\boldsymbol{\ell}$, we further have

$$\boldsymbol{y}^o = \mathbf{H}(\boldsymbol{x}^t) + \boldsymbol{e}^o, \tag{1.2.9}$$

$$\boldsymbol{x}^b = \boldsymbol{x}^t + \boldsymbol{e}^b, \tag{1.2.10}$$

where the observation error $\boldsymbol{e}^o \in \mathbb{R}^M$ consists of instrumentation error, and the background error $\boldsymbol{e}^b \in \mathbb{R}^{\mathcal{N}}$ consists of model errors and representation errors due to the discretization. In this setting, Problem 1.2.1 becomes:

Problem 1.2.2 (State estimation with data assimilation in discrete sense) *Find $\boldsymbol{x}^a \in \mathbb{R}^{\mathcal{N}}$ as an approximation of the true state $\boldsymbol{x}^t \in \mathbb{R}^{\mathcal{N}}$ from observation $\boldsymbol{y}^o = \boldsymbol{H}(\boldsymbol{x}^t) + \boldsymbol{e}^o$ combined with the information from a physical model.*

In discrete sense, Equation (1.2.8) becomes

$$\tilde{J}(\boldsymbol{x}) = \frac{1}{2}\lambda\|\boldsymbol{x} - \boldsymbol{x}^b\|_2^2 + \frac{1}{2}\|\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x})\|_2^2 \ . \tag{1.2.11}$$

In the above equation, we use the Euclidean norms in $\mathbb{R}^{\mathcal{N}}$ and $\mathbb{R}^m$ space respectively as measures. More generally, one can define two symmetric positive definite matrices $\mathbf{B} \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ and $\mathbf{R} \in \mathbb{R}^{m\times m}$ as weight matrices, to replace the penalty factor λ in (1.2.11):

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^b)^T\mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^b) + \frac{1}{2}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x}))^T\mathbf{R}^{-1}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x})), \tag{1.2.12}$$

where T stands for the transpose. One can easily find that (1.2.11) is just a special case of (1.2.12) when $\mathbf{B} = \lambda^{-1}\mathbb{I}^{\mathcal{N}\times\mathcal{N}}$ and $\mathbf{R} = \mathbb{I}^{m\times m}$, here $\mathbb{I}^{\mathcal{N}\times\mathcal{N}}$ and $\mathbb{I}^{m\times m}$ are identity matrices of dimension $\mathcal{N}\times\mathcal{N}$

and $m \times m$ respectively. To measure the weighted square difference between the observations and the equivalent of the state in the observation space on the one hand, the background and the state on the other hand statistically, the 3D variational (3D-Var) approach [204] takes the covariance matrices of observation and background errors as $\mathbf{B}$ and $\mathbf{R}$, which are defined as

$$\mathbf{R} := E\left[(\boldsymbol{e}^o - E[\boldsymbol{e}^o])(\boldsymbol{e}^o - E[\boldsymbol{e}^o])^T\right], \tag{1.2.13}$$

and

$$\mathbf{B} := E\left[(\boldsymbol{e}^b - E[\boldsymbol{e}^b])(\boldsymbol{e}^b - E[\boldsymbol{e}^b])^T\right], \tag{1.2.14}$$

where $E[\cdot]$ is the expectation operator. These matrices represent the prior assumptions about the errors associated with the background and observations and determine the relative weight given to $\boldsymbol{x}^b$ and $\boldsymbol{y}^o$ in the analysis.

Thus, the 3D-Var statement of Problem 1.2.2 is: find $\boldsymbol{x}^a$ such that

$$\boldsymbol{x}^a = \min_{\boldsymbol{x}\in\mathbb{R}^{\mathcal{N}}} J(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbb{R}^{\mathcal{N}}} \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^b)^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b) + \frac{1}{2}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x}))^T\mathbf{R}^{-1}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x})). \tag{1.2.15}$$

Note here that in the data assimilation literature, a in $\boldsymbol{x}^a$ stands for *analysis*. The optimal solution of 3D-Var, i.e. the analysis $\boldsymbol{x}^a$, is the closest to the true state $\boldsymbol{x}^t$ in an root mean square sense. Its determination requires that the gradient of the cost function with respect to its variable $\boldsymbol{x}$ at the analysis $\boldsymbol{x}^a$ be equal to zero

$$\nabla J(\boldsymbol{x}^a) = \frac{dJ}{d\boldsymbol{x}}|_{\boldsymbol{x}^a} = 0. \tag{1.2.16}$$

Under linear assumption of the observation operator $\mathbf{H}$, we have:

$$\mathbf{B}^{-1}(\boldsymbol{x}^a - \boldsymbol{x}^b) - \mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x}^a)) = 0, \tag{1.2.17}$$

we then obtain

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y}^o - \mathbf{H}(\boldsymbol{x}^b)) \ . \tag{1.2.18}$$

Thus we have the well-known 3D-Var solution to Problem 1.2.2 :

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \mathbf{K}(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x}^b)), \tag{1.2.19}$$

where the difference $\boldsymbol{y} - \mathbf{H}(\boldsymbol{x}^b)$ is called innovation and

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{1.2.20}$$

is a weight or gain computed on the estimated statistical error covariances of the background and the observations, here we already applied the Sherman-Woodbury-Morrinson equation [195] to $\left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ in (1.2.18). It is proven through

$$\begin{aligned}
&\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\
&\quad= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\
&\quad= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}(\mathbf{H}^T + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\
&\quad= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\
&\quad= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}.
\end{aligned} \tag{1.2.21}$$

This equality of $\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ and $(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ turns out to be useful, both from a theoretical and practical standpoint. For instance, the observation space is quite often much

smaller than the dimension of the state space, so that the inversion of the matrix $\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is much more costly than the inversion of the matrix $\mathbf{HBH}^T + \mathbf{R}$. Therefore, it is often useful to resort to (1.2.20).

We now make several remarks.

i) The estimation in Equation (1.2.19) is the same to the *Best Linear Unbiased Estimator, BLUE*[†] [18, 40]. In the framework of BLUE, one takes the linear hypothesis on the observation operator $\mathbf{H}$, and assume the background error $\boldsymbol{e}^b$ and observation error $\boldsymbol{e}^o$ are unbiased when deriving Equation (1.2.19).

ii) Equation (1.2.20) represents the Kalman filter [117] in the simplest static case. The Kalman filtering is an algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone, by estimating a joint probability distribution over the variables for each timeframe. More details on this could be found for example in reference [40, 32].

iii) If the background and observation error probability density functiond (pdfs) are Gaussian, then $\boldsymbol{x}^a$ is also the maximum likelihood estimator of $\boldsymbol{x}^t$ [40, 132].

iv) When the observations are distributed over time, the 3D-Var approach is generalized to the 4D-Var approach [194, 137, 133, 205, 71, 209]. The equations are similar, provided by the operators which are generalized including a forecasting model that allows comparison the state of the model with the observations at time t_k.

The matrices involved in computing Equation (1.2.19) and (1.2.20) cannot be stored because of their very large dimensions, which can be as much as $10^7 - 10^9$ state variables for classical meteorology representation. Thus the direct calculation of the gain matrix, $\mathbf{K}$, is unfeasible. The basic principle of the 3D-Var is to avoid explicitly calculating the gain matrix and make its inversion using a minimization procedure of the cost function $J(\boldsymbol{x})$. In this case the solution to (1.2.15) is obtained iteratively doing various evaluations of the equation and its gradient to get the minimum using a suited descent algorithm. The iterative 3D-Var Algorithm 1.2.1 is a classical case of an optimization algorithm [163]. For the gradient, there is a wide choice of algorithmic approached, and quasi-Newton methods [163, 178] are generally used and recommended.

Algorithm 1.2.1 Iterative 3D-Var algorithm

1: $k = 0, \boldsymbol{x} = \boldsymbol{x}_0, J(\boldsymbol{x}_0), \nabla J(\boldsymbol{x}_0)$
2: **while** $\|\nabla J(\boldsymbol{x}_k)\| > \epsilon$ or $k \leq k_{max}$ **do**
3: gradient descent and update of $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \lambda \nabla J(\boldsymbol{x}_k)$, where the step size λ is small enough
4: compute $J(\boldsymbol{x}_{k+1})$
5: compute $\nabla J(\boldsymbol{x}_{k+1})$
6: k=k+1

The 3D-Var/4D-Var approaches are just a small part among all the different optimization methods used in data assimilation applications. In this work, we only focus on the variational approaches, and take the terminology *data assimilation* as an analysis technique for *state estimation* in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of physical properties (e.g., [73, 25]). There are many different types of data assimilation algorithm, each varies in formulation, complexity, optimality and suitability for practical

[†]Definition: The Best Linear Unbiased Estimator (BLUE) of a parameter $\theta \in \mathbb{R}^p$ based on data $\mathbf{Y} \in \mathbb{R}^q$ is 1) a linear function of $\mathbf{Y}$, that is, the estimator can be written as $\mathbf{MY}$, where $\mathbf{M} \in \mathbb{R}^{p\times q}$ is an $p \times q$-dimensional operator, 2) unbiased ($E[\mathbf{MY}] = \theta$), and 3) has the smallest variance among all unbiased linear estimators.

application. In general, solving data assimilation problems in practice can be classically performed in two ways: as variational data assimilation and as smoothing/filtering data assimilation [40]. Newer approaches are also becoming available, e.g., nudging methods, reduced methods, ensemble methods and hybrid methods that combine variational and statistical approaches. We refer to the recent surveys [18, 130, 40, 113] for a thorough introduction to data assimilation and the way to solve the optimal problems.

1.2.2 Roles and modeling of covariance matrices in data assimilation

A correct specification of observation and background error covariances is crucial to the quality of the analysis, because they determine to what extent the background state will be corrected to match the observations. Indeed, the $\mathbf{B}$ matrix has multiple roles in the 3D-Var framework:

i) Information spreading: $\mathbf{B}$ matrix spreads information from an observed point to an unobserved one;

ii) Balance properties: on one hand, $\mathbf{B}$ matrix spreads information from an observed point to an unobserved one, on the other hand, it helps preserve physical properties;

iii) Information smoothing in space;

iv) Preconditioning of the assimilation: as the cost function is generally ill conditioned, a change of variable involving $\mathbf{B}$ usually helps reduce the condition number of the problem.

By definition, exact values of $\mathbf{R}$ and $\mathbf{B}$ would require the knowledge of the true state $\boldsymbol{x}^t$. This is not possible, and both matrices have to be estimated in practice. Often, the $\mathbf{R}$ matrix is assumed to be diagonal, i.e., have uncorrelated observation errors with empirically prescribed variances. Notice also that the dimension of $\mathbf{B}$ matrix is the square of the state dimension $\mathcal{N}$. For typical engineering problems, the size of $\mathbf{B}$ matrix has around $10^8 \times 10^8 = 10^{16}$ entries. It is too large to be calculated explicitly and to be stored in present-day computer memories. As a result, the $\mathbf{B}$ matrix needs to be modeled. Additionally, $\mathbf{B}$ is often required to have some physical properties which are required to be reflected in the analysis, e.g., i) it must be smooth in physical space, on sensible scales, ii) it should go to zero for very large separations if it is believed that observations should only have a local effect on the increments, iii) it should not exhibit physically unjustifiable variations according to direction or location, etc. A favoured method of producing background error statistics is the ensemble (Monte Carlo) method [111], in which an ensemble of N members of background are generated which are intended to have the same spread as the true state error pdf. Then $\mathbf{B}$ can be approximated by taking covariance statistics of the differences between each member and the mean:

$$\mathbf{B} \approx \langle (\boldsymbol{x}^b - \langle \boldsymbol{x}^b \rangle)(\boldsymbol{x}^b - \langle \boldsymbol{x}^b \rangle)^T \rangle, \tag{1.2.22}$$

where $\langle \cdot \rangle$ denotes an average operator over m, for instance an ensemble average $\langle \boldsymbol{x} \rangle = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_i$. The $\mathbf{B}$ matrix can also be estimated as the product of a correlation matrix $\mathbf{C}$ by a scaling factor, namely a suitable variance coefficients $\mathbf{B}_0$,

$$\mathbf{B} \approx \mathbf{B}_0 \mathbf{C}, \tag{1.2.23}$$

where $\mathbf{C}$ is a correlation matrix which can be e.g., the Second Order Auto-Regressive (SOAR) function [75, 173]

$$\mathbf{C}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (1 + \frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2}{L}) e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2}{L}}, \tag{1.2.24}$$

where L is the correlation length depends on the specified problem. In [91], the authors studied the construction of correlation functions, especially the correlation models that are legitimate correlation

functions on the sphere. For more information on modeling $\mathbf{B}$ we refer to the recent articles and surveys [40, 20, 75, 92], examples can also be found in [173, 165, 91].

1.2.3 Reduced methods in data assimilation

As data assimilation often deals with large-scale and non-linear problems, e.g., with order $10^7 - 10^9$ state variables for classical meteorology representation, the tools offered by reduction can be very useful. Here we mention some approaches in data assimilation with reduced modeling:

i) The first class of approaches reduces the model, but not the data assimilation scheme. In other words, all the machinery of model order reduction methods, algorithms, and libraries can be used. The general aim is to approximate a full-scale high-dimensional problem by a new, much smaller dimensional problem. Examples in this topic can be found in [53], in which POD was developed to build a reduced model of the Pacific tropical ocean, and the 4D-Var assimilation was performed with this reduced model and it adjoint. On the real-time side, the work in [17] combined POD and 4D-Var assimilation for fluid flow control applications, and the authors in [52] designed reduced direct and adjoint models to be used within the 4D-Var minimization algorithm, etc. The idea to reduce the model is widespread in applied mathematics, we refer Section 1.3.1 for an introduction of the methodology.

ii) The second class of approaches reduces the covariance matrices, but still do not change the data assimilation scheme. One of the major drawbacks of the data assimilation algorithms in large dimensions is the prohibitive cost of the covariance matrices: they are too large to be fully estimated or even stored in memory, and also too large to be used directly for numerical computations. Thus most of the solutions rely on a simple principle: find a reduced and cheaper way to compute and represent the covariance matrices. As a square root of $\mathbf{B}$ matrix is often required in variational assimilation, factorizations of $\mathbf{B}^{1/2}$ using sparse matrices have been proposed, $\mathbf{B}^{1/2} = \mathbf{L}\Sigma\mathbf{C}$, where $\mathbf{L}$ is a physical balance operator, Σ is a spectral transform and $\mathbf{C}$ is a diagonal or block diagonal matrix containing the correlations in spectral space, for further information in this topic, we refer to [20, 89, 44]. Besides, the empirical orthogonal functions (EOFs) have also been used to reduced the rank of $\mathbf{B}$ matrix and provide a low-cost approximation in a variational context [183], more recently, the ensemble method has been applied in oceanography to estimate the $\mathbf{B}$ matrix in a variational framework [72], etc.

iii) The third class of approaches reduces the data assimilation method, i.e., reduce the working dimension of the algorithm cores. In this aspect, many works have been done with the reduced methods within variational assimilation, mostly 4D-Var. In [83], the reduced 4D-Var aims at reducing the dimension in the core part of the algorithm, namely the minimization. The state $\boldsymbol{x} \in \mathbb{R}^{\mathcal{N}}$ is project onto a lower-dimensional space around the background

$$\boldsymbol{x} = \boldsymbol{x}^b + \sum_{i=1}^{n} c_i q_i, \tag{1.2.25}$$

where n is the reduced dimension, $\{c_i\}_i$ are the real coefficients, and $\{q_i\}_i$ are suitable reduced basis functions which can be EOFs [183], forward/backward singular vectors, Lyapunov vectors [136], etc. However, it was unable to reach a good enough minimum, contrary to the classical unreduced 4D-Var. This seems to be due mostly to the presence of a significant model error. We refer to [103, 43, 170, 206] for more applications in this aspect.

Another approach, benefits from RB methods, instead of introducing the covariance matrices, one searches the optimal solution from a problem dependent solution set (also-called *manifold*) combining with the measurement information, thus largely simplifies the assimilation

algorithm, one can refer to [29, 148, 144, 145, 31, 78]. In this thesis, we mainly focus on this approach.

1.2.4 State-of-the-art on the use of data assimilation for state estimation in nuclear engineering domain

Data assimilation has found an increasingly wide applications in nuclear engineering domain mainly in i) *uncertainty quantification/evaluation* [‡] of simulation codes [104, 4, 119, 109], ii) *parameters estimation* such as the basic nuclear data evaluation and adjustment [106, 105, 34, 115, 165, 174], and iii) *field reconstruction / state estimation* to improve the accuracy of simulation codes [173, 60, 59, 4, 182, 2], etc.

Particularly, in [15], the authors give a review of data assimilation applications in nuclear reactor physics, focused on *field reconstruction* and *parameters estimation*, based on the standard simulation of neutron fields which are already very accurate. In the framework of field reconstruction, the differential influence of instruments in nuclear core activity evaluation with a 3D-Var approach is studied in [36, 38] based on a EDF 900 MWe nuclear Pressurized Water Reactor (PWR900). In [35], the 3D-Var and BLUE methodologies are used in optimal design of measurement network for neutron flux/power field reconstruction, also in the standard configuration of a PWR900. The robustness of these data assimilation methodologies used in nuclear field reconstruction is reported in [37].

Other applications of field reconstruction with data assimilation related to nuclear domain are such as off site nuclear emergency management [184, 87, 22], the atmospheric dispersion modeling for nuclear accident assessments [218] and numerical reconstruction of high dose rate zones due to the Fukushima Dai-ichi Nuclear Power Plant accident [118], etc.

Furthermore, at EDF R&D, data assimilation methods are systematically implemented in the platform SALOME [182, 2] as a module called Assimilation de Données et Aide à l'Optimisation (ADAO), designed to offer different tools (e.g., BLUE, 3D-Var, 4D-Var, etc.) to solve data assimilation problems, in various fields such as mechanics, or nuclear physics. ADAO can be used for field reconstruction applications in neutronic and to elaborate advanced methodology, for instance to define an optimal set of measurement points to study the influence of a measurement network on the accuracy of the reconstructed neutron flux field [36, 37, 38], or to define an optimal set of measurement points recovering the best reconstruction of a specified field [35].

Through ADAO with current implemented data assimilation methods is able to solve large amounts of data assimilation problems at EDF, there are still many challenges when solving large scale problem, especially in field reconstruction applications where the dimension of the spatial discretization might amount to 10^6 or even more, here we list a few.

i) There is no general way to model the problem-dependent covariance matrix $\mathbf{B}$. Thus, for each specified engineering problem, one has to remodel the covariance matrix, which makes the essential part of data assimilation framework.

ii) The iteration procedure when solving the minimum cost function $J(x)$ is time-consuming, which makes the data assimilation method not feasible for on-line purpose.

iii) Even with a well-modeled covariance matrix or accelerated iteration procedure, solving the

[‡]Uncertainty quantification (UQ) is the science of quantitative characterization and reduction of uncertainties in both computational and real world applications. It tries to determine how likely certain outcomes are if some aspects of the system are not exactly known. Many problems in the natural sciences and engineering are also rife with sources of uncertainty. Computer experiments on computer simulations are the most common approach to study problems in uncertainty quantification [69, 187, 114, 215].

field reconstruction problem with data assimilation method is still costly in computing resource (such as time, memory, etc.).

The data assimilation with reduced basis approach is a solution to address parts of the above challenges. Much more ambitious is that this approach is based on a learning strategy concept for the specified problem, which is hopeful to improve the result of data assimilation.

1.3 Data assimilation with reduced basis

1.3.1 Fundamentals of reduced basis methods

In this scenario, we are interested in solving parametric problems providing a solution map

$$u : \mathcal{D} \mapsto V \tag{1.3.1}$$

from a compact parameter set $\mathcal{D} \subset \mathbb{R}^p$ onto V. Such problems arise in many situations, e.g., optimization, control or parameter identification problems, response surface or sensitivity analysis. The problem may be stationary or time dependent but in all these cases, a solution $u(\mu)$ has to be evaluated or computed for many instances of $\mu \in \mathcal{D}$. In all the problems we consider, $u(\mu)$ will be given as the solution of some parametric partial differential equations of the form

$$\mathcal{F}(u, \mu) = 0, \tag{1.3.2}$$

where $\mathcal{F}$ is a partial differential operator, and μ is a parameter that varies in the set $\mathcal{D}$.

Even well optimized, the favorite discretization method used to solve the underlying problem will lead to very heavy computations in order to approximate all these solutions and decision may not be taken appropriately due to too large computer time for reliable simulations. Instead of presenting an alternative to the favorite discretization, one can construct a surrogate method with the current discretization that simplifies the complexity of the equations resulting in very fast solution algorithms. The method is based on a learning strategy concept, and, for a specified problem, the preliminary *off-line* preparation is much time-consuming. It is only after this learning step that the full speed of the method can be appreciated *on-line*, paying off the cost of the off-line preparation step.

We look for a sequence of nested finite dimensional spaces $V_0 \subset V_1 \subset \cdots \subset V_n \subset \cdots$, $\dim(V_n) = n$, to approximate all the functions in the solution manifold

$$\mathcal{M} := \{u(\mu) : \mu \in \mathcal{D}\}, \tag{1.3.3}$$

such that each V_n approximates $\mathcal{M}$ to a known tolerance ε_n, where ε_n is defined as

$$\varepsilon_n \equiv \sup_{u \in \mathcal{M}} \inf_{v \in V_n} \|u - v\|. \tag{1.3.4}$$

Namely, for a given basis $\{q_i\}_{i=1}^n$ of V_n, our interest is to find $u_n \in V_n$ to approximate $u(\mu) \in \mathcal{M}$:

$$u(\mu) \simeq u_n(\mu) \equiv \sum_{i=1}^{n} \alpha_i(\mu) q_i \ , \tag{1.3.5}$$

which satisfies:

$$\|u(\mu) - u_n(\mu)\| \leq C_r \inf_{v_n \in V_n} \|u(\mu) - v_n\|, \tag{1.3.6}$$

where $C_r > 0$ is a scalar that reflects how good the selection method is. In order to apprehend in which sense the good behavior of $\mathcal{M}$ and the corresponding approximation should be understood, it is helpful to introduce the notion of Kolmogorov n-width $d_n(\mathcal{M}, V)$ of $\mathcal{M}$ in V be defined as

$$d_n(\mathcal{M}, V) := \inf_{\substack{X \subset V \\ \dim(X) \leq n}} \max_{u \in \mathcal{M}} \min_{v \in X} \|u - v\|_V. \tag{1.3.7}$$

We assume in the following that it decays fast with n. This quantity, first introduced in Kolmogorov [124] , describes the best achievable accuracy, in the norm of V , when approximating all possible elements of $\mathcal{M}$ by elements from a linear n-dimensional space V_n. If $d_n(\mathcal{M}, V)$ decays fast, then for a given target accuracy $\varepsilon > 0$, one can build a space V_n s.t. $\text{dist}(\mathcal{M}, V_n) := \max_{u \in \mathcal{M}} \min_{v \in V_n} \|u - v\|_V \leq \varepsilon$, where $\dim(V_n) = n \equiv n(\varepsilon)$ is moderate.

Several reasons can account for a rapid decrease of $d_n(\mathcal{M}, V)$ (see [145]): if $\mathcal{M}$ is a set of functions defined over a domain, we can refer to regularity, or even to analyticity, of these functions with respect to the domain variable (as analyzed in the example in [124]). Thus the approximation of any element $u \in \mathcal{M}$ by finite expansions is a classical problem addressed by, among others, reduced basis methods as the results of [150] and [64, 63] show.

Several methods exist to find spaces V_n that yield an error $e_n := \max_{u \in \mathcal{M}} \|u - P_{V_n} u\|$ comparable to d_n. The first choice is polynomials, with high order polynomial approximation like $u_\Lambda(y, \mu) = \sum_{\nu \in \Lambda} a_\nu y^\nu$ [65], one is able to get a good approximation with a moderate n for some specified problem. Anyway, in polynomial approximation, the polynomials are unable to adapt the specific problem. The RB methods [150, 143] with well selected basis, on the contrary, are able to catch more information for the underlying problem. It generates V_n from particular snapshots $u_i = u(\mu_i), i = 1, \cdots, n$, of the solution manifold. These spaces are not optimal, however it has been shown in [30] and [76] that, whenever $d_n(\mathcal{M}, V)$ is $\mathcal{O}(n^{-r})$ or $\mathcal{O}(e^{-c_0 n^\alpha})$, then a certain greedy recursive selection of the μ_i in the reduced basis method gives similar convergence rates for the spaces V_n generated by the optimal algorithm. The performance of other model reduction methods such as the generalized empirical interpolation method (GEIM [144]), the generalized reduced basis methods [131] give also a decay close to d_n of the Kolmogorov width of the solution manifold.

Over the last decade, the reduced order models obtained from RB methods have become increasingly popular for simulations requiring:

i) many-queries (optimal control, inverse problems, uncertainty quantification etc.),

ii) real-time computations (monitoring, parameter estimation, etc.).

We emphasize again that, with ROM, a computationally intensive *off-line* phase typically involve approximation spaces of only a few hundreds or even less dimensions is necessary, that leads to vast savings in computation time when these models are solved during the *on-line* phase. We refer to the recent surveys [13, 23, 56] for a thorough introduction to ROM. We further refer to [24, 164, 107, 176, 185] for a complete introduction to the RB methods. The recent advances in ROM for parameterized systems offer new opportunities for the integration of *models* and *experimental data*. In this framework, ROM i) provides a way to learn the physical system, which allows extracting the principal component information of the system and also providing a guideline to set the amount and locations of the measurement needed; ii) speeds up computations allowing better explorations of the parameter space at an acceptable computational cost; iii) provides actionable tools to compress our prior knowledge about the system coming from the parameterized mathematical model into low-dimensional and more manageable forms, which makes the combination with experimental data more efficiently.

1.3.2 State estimation with reduced basis and measurements

Recall that the objective of the state estimation as described in Problem 1.2.1 is to approximate the true state u^t from m observations $\boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$, where the $\{\ell_i\}_i$ are, e.g., linear functionals.

To achieve this approximation, the first information that we have in hand is the observations. Let us denote by w_i the Riesz representation of ℓ_i which satisfies:

$$\forall u \in V,\ (w_i, u) = \ell_i(u), \tag{1.3.8}$$

where $(\cdot,\cdot)$ is the inner product defined in V. Let us then introduce the space

$$W_m = \text{span}\{w_1, \cdots, w_m\}. \tag{1.3.9}$$

In particular, we assume that the $\{w_j\}_j$ are linearly independent i.e., dim $W_m = m$. We then define the Gramian matrix $\mathbb{W}$ with entries

$$\mathbb{W}_{i,j} = \ell_i(w_j) = (w_i, w_j),\ i,j = 1, \cdots m. \tag{1.3.10}$$

Note that $\mathbb{W}$ is a symmetric matrix. Any element $w \in W_m$ can be represented as

$$w = \sum_{i=1}^{m} \beta_i w_i. \tag{1.3.11}$$

For notational simplicity, we denote by $w = \mathcal{W}_m\boldsymbol{\beta}$ the linear combination of $w = \sum_{i=1}^m \beta_i w_i$, where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)^T \in \mathbb{R}^m$ is an m-dimensional column vector and $\mathcal{W}_m := (w_1, \cdots, w_m)$ is a permutation of elements $\{w_1, \cdots, w_m\}$ in space W_m. For any $k > 0$ elements $\{w'_1 = \mathcal{W}_m\boldsymbol{\beta}'_1, \cdots, w'_k = \mathcal{W}_m\boldsymbol{\beta}'_k\}$, the permutation $(w'_1, \cdots, w'_k)$ can be represented as $(w'_1, \cdots, w'_k) = \mathcal{W}_m\mathbf{B}'$, where $\mathbf{B}'$ is an $m \times k$-dimensional matrix with entries $(\mathbf{B}')_{i,j} = (\boldsymbol{\beta}'_i)_j,\ \ 1 \le i \le m,\ 1 \le j \le k$. Let us then denote by $v^T u$ the inner product $(v, u), \forall v, u \in V$. In this spirit, we can denote by $\mathcal{W}_m^T u \in \mathbb{R}^m$ the m-dimensional column vector with elements $(\mathcal{W}_m^T u)_j := w_j^T u = (w_j, u),\ 1 \le j \le m, u \in V$, thus we have $\boldsymbol{\ell}(u) = \mathcal{W}_m^T u$. We can further define an $m \times m$-dimensional matrix $\mathcal{W}_m^T\mathcal{W}_m$ with entries $(\mathcal{W}_m^T\mathcal{W}_m)_{i,j} := w_i^T w_j = (w_i, w_j),\ 1 \le i, j \le m$ such that $\mathbb{W} = \mathcal{W}_m^T\mathcal{W}_m$. At last, we denote by $w^o = P_{W_m}u^t$ the orthogonal projection of $u^t \in V$ onto the space W_m, where throughout this thesis P_X denotes the orthogonal projection onto X for any closed subspace $X \subset V$. Note that $w^o \in W_m$ is another representation of the m observations of u^t which is different from $\boldsymbol{\ell}(u^t)$. In this setting, we first have the following proposition.

Proposition 1.3.1 *The orthogonal projection of any function $u \in V$ onto the subspace W_m can be represented as*

$$P_{W_m}u = \mathcal{W}_m\mathbb{W}^{-1}\mathcal{W}_m^T u = \mathcal{W}_m\mathbb{W}^{-1}\boldsymbol{\ell}(u). \tag{1.3.12}$$

Proof: By recognizing the singular value decomposition of matrix $\mathcal{W}_m^T\mathcal{W}_m = \mathbb{W}$ is $\mathbb{W} = USU^T$, we have the permutation $\hat{\mathcal{W}}_m := (\hat{w}_1, \cdots, \hat{w}_m) = \mathcal{W}_m US^{-1/2}$ where $\{\hat{w}_1, \cdots, \hat{w}_m\}$ are orthonormal basis of W_m. Thanks to the notations quoted above, we can represent $P_{W_m}u$ by $P_{W_m}u = \hat{\mathcal{W}}_m\hat{\mathcal{W}}_m^T u = \mathcal{W}_m US^{-1/2}(\mathcal{W}_m US^{-1/2})^T u = \mathcal{W}_m US^{-1/2}S^{-1/2}U^T\mathcal{W}_m^T u = \mathcal{W}_m\mathbb{W}^{-1}\mathcal{W}_m^T u = \mathcal{W}_m\mathbb{W}^{-1}\boldsymbol{\ell}(u)$.

□

To this end, we get the first approximation $P_{W_m}u^t$ of u^t which satisfies

$$\|P_{W_m}u^t - u^t\| \le \inf_{w \in W_m} \|w - u^t\|. \tag{1.3.13}$$

Note that, from this proposition, the information from observation $\boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$ is equivalent to knowing $w^o = P_{W_m} u^t$. Note also that w^o is totally computable thanks to the Riesz representation theory and the observation information. But the estimation (1.3.13) may be a very rough estimation since there is no reason that the observations generate a space W_m with good approximation properties. Additional information from e.g., model is necessary for further improvement.

As pointed out in Section 1.2, the estimation u^b of the true state u^t can be acquired through a physical model $\mathcal{F}(u, \mu) = 0, \mu \in \mathcal{D}$. But generally speaking, there is no reason that the model is perfect to describe the true state, thus the following relation always holds:

$$0 < \eta < \|u^b - u^t\| \ , \tag{1.3.14}$$

where u^b is the solution from the physical model with the correct parameter $\mu \in \mathcal{D}$ corresponding to the true state u^t, the left part $\eta < \|u^b - u^t\|$ is necessary which reflects the *model bias*, because of the imperfect model. Recall that with RB method, with a carefully selected n-dimensional space $V_n \subset V$ where n is relatively small, one is able to get a very good estimation to u^b (so that u^t) with $u_n \in V_n$ which satisfies

$$0 < \eta < \|u_n - u^t\| \leq C_r \inf_{v \in V_n} \|v - u^t\| = C_r \ \|P_{V_n} u^t - u^t\|, \tag{1.3.15}$$

where C_r is some scalar. The Galerkin method is a standard way to calculate u_n once the finite dimensional subspace V_n on V is given, but one should be able to identify the correct parameter $\mu \in \mathcal{D}$ corresponding to the true state u^t.

It is natural to combine the two estimations mentioned above, namely, the estimation $P_{W_m} u^t$ from observation information and the estimation u_n from the reduced model, that provides an algorithm

$$A : W_m \to V, \tag{1.3.16}$$

which assigns to any $w^o \in W_m$ the estimation $u^* = A(w^o) = A(P_{W_m} u^t)$ and, hopefully is such that

$$\|u^* - u^t\| \leq C_A(w^o) \ \|P_{W_m + V_n} u^t - u^t\| \tag{1.3.17}$$

for some scalar $C_A(w^o)$. The best algorithm A, for a given fixed value of w^o, would give the smallest constant $C_A(w^o)$ and the algorithm which gives this smallest constant is said to be *instance optimal* with constant $C_A(w^o)$. There are two different approaches in this data assimilation scenario.

The first approach is the so-called *parametrized-background data-weak (PBDW) state estimation* [148]. In particular, in order to find the optimal estimate of u^t for a given measurement $\boldsymbol{\ell}(u^t)$, we first write the state estimate u as the sum of two contributions:

$$u = v + \eta. \tag{1.3.18}$$

The first contribution v, is called *deduced background estimate* which represents anticipated model part; and V_n is interpreted as a background or prior space which approximates the best-knowledge manifold on which we hope the true state resides. As already discussed, non-zero model error is a virtual certainty, and thus we cannot realistically assume that u^t lies exactly on the best-knowledge manifold, which thus motivates the second contribution to u. This second contribution η is called *update estimate* which accommodates unanticipated update or model bias. We then search for η of minimum norm, that is to say, we look for the smallest correction to the best-knowledge parametric manifold, subject to the observation constraints $\boldsymbol{\ell}(\eta + v) = \boldsymbol{\ell}(u^t)$. Thus the PBDW statement is

that: find $u^* = v^* + \eta^*$ such that

$$\begin{aligned} (v^*, \eta^*) &= \arg\inf_{\substack{v\in V_n \\ \eta\in V}} \|\eta\|^2 \\ \text{subject to:} \quad & \boldsymbol{\ell}(\eta^* + v^*) = \boldsymbol{\ell}(u^t). \end{aligned} \tag{1.3.19}$$

Note that the constraint $\eta \in V$ in (1.3.19) can be weakened, by recognizing that the component $P_{W_m^\perp}\eta^*$ doesn't contribute to $(1.3.19)_2$ but increases the infimum of $(1.3.19)_1$, one can immediately have $\eta* \in W_m$ [148]. The simplified Euler-Lagrange equation associated with the PBDW minimization statement (1.3.19) is: find $u^* = v^* + \eta^*$ such that

$$\begin{aligned} &(\eta^*, w) + (v^*, w) = (u^t, w) \quad \forall w \in W_m \ , \\ &(\eta^*, v) = 0 \quad \forall v \in V_n \end{aligned} \tag{1.3.20}$$

This estimation u^* gives the error bound

$$\|u^* - u^t\| \leq \frac{1}{\beta(V_n, W_m)} \|P_{V_n+W_m} u^t - u^t\|, \tag{1.3.21}$$

where the inf-sup constant $\beta(V_n, W_m)$ is defined as

$$\beta(V_n, W_m) := \inf_{v\in V_n} \sup_{w\in W_m} \frac{(v, w)}{\|v\|\|w\|} \in [0, 1], \tag{1.3.22}$$

and plays the role of a stability constant [29, 148, 31].

The second approach is the so-called *optimal recovery* as proposed in [29]. This approach relies on the facts that i) the observations $\boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$ provide information of the true state u^t and ii) the true state u^t is in the solution manifold $\mathcal{M}$ of a parametric PDE.

i) We first recognize that, from Proposition 1.3.1, knowing the observations $\boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$ is equivalent to knowing $w^o = P_{W_m} u^t$. In going further, we think of measurements as simply providing the knowledge of this projection, on the contrary, all elements of the orthogonal complement $W_m^\perp \subset V$ of W_m have zero measurements. Indeed, if $u \in V$ satisfies the measurements then u could be any of the functions $u + \eta$, with $\eta \in W_m^\perp$, and each of these functions would be assigned the same approximation if all the information we have about u^t is that $w^o = P_{W_m} u^t$. Thus we define

$$H_{w^o} = \{u \in V \mid P_{W_m} u = w^o\} = w^o + W_m^\perp. \tag{1.3.23}$$

ii) As explained above, we need additional information about u^t to have a much more meaningful problem. Our interest here is the solution manifold $\mathcal{M}$ of a parametric PDE. If the Kolmogorov n-width $d_n(\mathcal{M}, V)$ of the manifold $\mathcal{M}$ decays fast with n, the typical way of resolving $\mathcal{M}$ is through a finite sequence of spaces $\{V_1, \cdots, V_n\}$ with V_i of dimension i where the spaces are known — by a learning strategy, e.g., the off-line phase of the RB methods — to approximate $\mathcal{M}$ to some known accuracy ε_i, where $\varepsilon_i \equiv \sup_{u\in\mathcal{M}} \|P_{V_i} u - u\|$, $1 \leq i \leq n$. Accordingly, we define

$$\mathcal{K}_i := \{u \in V \mid \|P_{V_i} u - u\| \leq \varepsilon_i\}, \quad 1 \leq i \leq n. \tag{1.3.24}$$

The optimal recovery thus becomes: find an optimal solution u^* which satisfies $u^* \in H_{w^o} \cap \mathcal{K}_n$ (the so-called *one-space problem* [29]) or more precisely $u^* \in \cap_{i=1}^n \mathcal{K}_i \cap H_{w^o}$ (the so-called *the multi-space problem* [29]). For the one-space problem, the optimal solution u^* is found by searching

$u^* \in H_{w^o}$ such that u^* is closest to V_n, i.e.,

$$u^* = \arg\min_{u \in H_{w^o}} \|u - P_{V_n} u\|. \tag{1.3.25}$$

This estimation u^* gives the error bound

$$\|u^* - u^t\| \leq \nu(V_n, W_m) \|P_{V_n + W_m} u^t - u^t\| \ , \tag{1.3.26}$$

where $\nu(V_n, W_m)$ is defined as

$$\nu(V_n, W_m) := \sup_{\eta \in W_m^\perp} \frac{\|\eta\|}{\|\eta - P_{V_n}\eta\|} = \sup_{\eta \in W_m^\perp} \frac{\|\eta\|}{\|P_{V_n^\perp}\eta\|} \ . \tag{1.3.27}$$

Note that, in this scenario, the authors proved that the error bound of PBDW estimation in (1.3.21) can be improved and has the same form as in (1.3.26), and further pointed out that

$$\nu(V_n, W_m) = \beta(V_n, W_m)^{-1}, \tag{1.3.28}$$

which it is the best constant in estimates of this form. The lower bound applies to both linear and nonlinear algorithms, that is, (1.3.26) cannot be improved also using nonlinear mappings, thus,

$$\nu(V_n, W_m) = \min_{\substack{A \\ w^o \in W_m}} |C_A(w^o)| \ . \tag{1.3.29}$$

We denote the solution to (1.3.19) by u_1^* and the solution to (1.3.25) by u_2^*. The following proposition shows the equivalence of the above two approaches.

Proposition 1.3.2 *The solution u_1^* to* (1.3.19) *is equivalent to the solution u_2^* to* (1.3.25).

Proof: Let $u_1^* = \eta_1^* + v_1^*$ be the solution to (1.3.19). We first prove that $\eta_1^* \in V_n^\perp \cap W_m$ from the original Equation (1.3.19). We split $\eta_1^* = P_{V_n}\eta_1^* + P_{V_n^\perp}\eta_1^*$, then $(v_1^* + P_{V_n}\eta_1^*, P_{V_n^\perp}\eta_1^*)$ also satisfies $(1.3.19)_2$, and we have $\|P_{V_n^\perp}\eta_1^*\| \leq \|\eta_1^*\|$, thus $P_{V_n}\eta_1^* = 0$ and $\eta_1^* \in V_n^\perp \cap W_m$. Second, from $(1.3.19)_2$, we have $\boldsymbol{\ell}(\eta_1^* + v_1^*) = \boldsymbol{\ell}(u^t)$, from Proposition 1.3.1, it is equivalent to $P_{W_m}(\eta_1^* + v_1^*) = P_{W_m} u^t$. Recall that $\eta_1^* \in V_n^\perp \cap W_m$, then we have

$$\begin{aligned} P_{W_m} u^t &= P_{W_m}(\eta_1^* + v_1^*) \\ &= P_{W_m}\eta_1^* + P_{W_m} v_1^* \\ &= \eta_1^* + P_{W_m} v_1^* \end{aligned} \tag{1.3.30}$$

and we get

$$\eta_1^* = P_{W_m} u^t - P_{W_m} v_1^* = w^o - P_{W_m} v_1^* \ . \tag{1.3.31}$$

In other words, the component v_1^* is chosen such that its projection onto the observable space explains the observed data for a minimal η_1^*:

$$v_1^* = \arg\min_{v \in V_n} \|w^o - P_{W_m} v\| \tag{1.3.32}$$

From $\eta_1^* \in V_n^\perp \cap W_m$ and (1.3.31) we further get

$$0 = P_{V_n}\eta_1^* = P_{V_n}(w^o - P_{W_m} v_1^*) = P_{V_n} w^o - P_{V_n} P_{W_m} v_1^* = P_{V_n} w^o - P_{W_m} v_1^*, \tag{1.3.33}$$

thus we have the following relation

$$P_{V_n} w^o = P_{W_m} v_1^*. \tag{1.3.34}$$

Now we focus on (1.3.25). Let $w^o + v' = u \in H_{w^o}$, where $v' \in W_m^\perp$, from (1.3.25) we have

$$\begin{aligned} u_2^* &= \underset{u \in H_{w^o}}{\arg\min} \|u - P_{V_n} u\| \\ &= \underset{u \in H_{w^o}, v \in V_n}{\arg\min} \|u - v\|^2 \\ &= \underset{v' \in W_m^\perp, v \in V_n}{\arg\min} \|w^o + v' - v\|^2 \\ &= \underset{v' \in W_m^\perp, v \in V_n}{\arg\min} \|w^o + v' - (P_{W_m} v + P_{W_m^\perp} v)\|^2 \\ &= \underset{v' \in W_m^\perp, v \in V_n}{\arg\min} \|w^o - P_{W_m} v + v' - P_{W_m^\perp} v\|^2 \\ &= \underset{v' \in W_m^\perp, v \in V_n}{\arg\min} \|w^o - P_{W_m} v\|^2 + \|v' - P_{W_m^\perp} v\|^2, \end{aligned} \tag{1.3.35}$$

the last equation holds because $w^o - P_{W_m} v \in W_m$ and $v' - P_{W_m^\perp} v \in W_m^\perp$. Then the solution $(u_2^*, v_2^*, v_2'^*)$ to (1.3.35) satisfies

$$v_2^* = \underset{v \in V_n}{\arg\min} \|w^o - P_{W_m} v\| \ , \tag{1.3.36}$$

and

$$v_2'^* = P_{W_m^\perp} v_2^* \ , \tag{1.3.37}$$

and

$$u_2^* = w^o + v_2'^* \ . \tag{1.3.38}$$

From (1.3.32) and (1.3.36) we find that $v_1^* = v_2^*$. Then we denote $u_2^* - P_{V_n} u_2^*$ by η_2^*. To prove the equivalence of (1.3.19) and (1.3.25), we only need to point out that $P_{V_n} u_2^* = v_2^*$ and $\eta_2^* = \eta_1^*$. We first have

$$\begin{aligned} \eta_2^* &= u_2^* - P_{V_n} u_2^* \\ &= w^o + P_{W_m^\perp} v_2^* - P_{V_n}(w^o + v_2'^*) \\ &= w^o + P_{W_m^\perp} v_2^* - P_{V_n}(w^o + P_{W_m^\perp} v_2^*) \\ &= w^o + P_{W_m^\perp} v_2^* - P_{V_n} w^o - P_{V_n} P_{W_m^\perp} v_2^* \\ &= w^o + P_{W_m^\perp} v_2^* - P_{V_n} w^o - P_{W_m^\perp} v_2^* \\ &= w^o - P_{V_n} w^o. \end{aligned} \tag{1.3.39}$$

Then we have $\eta_2^* = w^o - P_{W_m} v_1^* = \eta_1^*$ from (1.3.31) and (1.3.34). At last we prove $v_2^* = P_{V_n} u_2^*$.

$$\begin{aligned} P_{V_n} u_2^* &= P_{V_n}(w^o + v_2'^*) \\ &= P_{V_n}(w^o + P_{W_m^\perp} v_2^*) \\ &= P_{V_n} w^o + P_{V_n} P_{W_m^\perp} v_2^* \\ &= P_{W_m} v_1^* + P_{V_n} P_{W_m^\perp} v_2^* \\ &= P_{W_m} v_2^* + P_{W_m^\perp} v_2^* \\ &= v_2^* \end{aligned} \tag{1.3.40}$$

The last third relation follows from $P_{V_n} w^o = P_{W_m} v_1^*$, the last second relation follows from $v_1^* = v_1^*$. Thus we have proved that $u_1^* = \eta_1^* + v_1^*$ and $u_2^* = \eta_2^* + v_2^*$ subject to $v_1^* = v_2^*$ and $\eta_1^* = \eta_2^*$.

□

1.3.3 Algebraic form of PBDW and further development

Now we consider the algebraic form of the PBDW state estimation of problem (1.3.20). The practical computation of u^* in algebraic form with matrices can be found in [148]. Let $\{q_1, \cdots, q_n\}$ be the basis of V_n, any $v \in V_n$ can be expressed as $v = \sum_{i=1}^n \alpha_i q_i$. For notational convenience, we follow Section 1.3.2 and denote $v = \mathcal{B}_n \boldsymbol{\alpha}$, where $\mathcal{B}_n := (q_1, \cdots, q_n)$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)^T \in \mathbb{R}^n$. We then introduce the $m \times n$-dimensional matrices $\mathbb{M} = \mathcal{W}_m^T \mathcal{B}_n \in \mathbb{R}^{m \times n}$ with entries $\mathbb{M}_{i,j} = w_i^T q_j = (w_i, q_j)$, which is the inner product of w_i and q_j. We further introduce the $n \times n$-dimensional matrix $\mathbb{Q} = \mathcal{B}_n^T \mathcal{B}_n \in \mathbb{R}^{n \times n}$ with entries $\mathbb{Q}_{i,j} = q_i^T q_j = (q_i, q_j)$.

Considering $u^* \in V_n \oplus (W_m \cap V_n^\perp)$, any $u \in V_n \oplus (W_m \cap V_n^\perp)$ has the form $u = v + \eta$, where $v = \sum_{j=1}^n \alpha_j q_j = \mathcal{B}_n \boldsymbol{\alpha} \in V_n$ and $\eta = \sum_{i=1}^m \beta_i w_i = \mathcal{W}_m \boldsymbol{\beta} \in W_m \cap V_n^\perp$. With these notations, we have

$$u^* = \sum_{i=1}^m \beta_i^* w_i + \sum_{j=1}^n \alpha_j^* q_j = \mathcal{W}_m \boldsymbol{\beta}^* + \mathcal{B}_n \boldsymbol{\alpha}^*. \tag{1.3.41}$$

Substituting (1.3.41) into the Euler-Lagrange equation (1.3.20) of the PBDW minimization statement, one can easily get the values of $\boldsymbol{\alpha}^* = (\alpha_1^*, \cdots, \alpha_n^*)^T$ and $\boldsymbol{\beta}^* = (\beta_1^*, \cdots, \beta_n^*)^T$ by solving the following problem [148]:

$$\begin{bmatrix} \mathbb{W} & \mathbb{M} \\ \mathbb{M}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^* \\ \boldsymbol{\alpha}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}^t \\ 0 \end{bmatrix}. \tag{1.3.42}$$

Note that the necessary and sufficient conditions for solvability of (1.3.42) are: i) $\mathbb{M}$ is full column rank matrix and ii) $\mathbb{W}$ is invertible, these conditions can be satisfied by careful selections of V_n and W_m. Note also that, in particular, the data assimilation process proposed by (1.3.19) or (1.3.25) proposes an approximation of the solution u^* written as the sum of two contributions: the *deduced background estimate* $v^* \in V_n$ and the *update estimate* $\eta^* \in W_m \cap V_n^\perp$. Both [29] and [148] point out that the deduced background estimate v^* is chosen such that its projection onto the observable space W_m is the closest one to the observed data $w^o = P_{W_m} u^t$. We present it here as a proposition and show the proof for the sake of completeness.

Proposition 1.3.3 *The* deduced background estimate v^* *of the solution* u^* *of PBDW which lies on the reduced space* V_n *is the solution to the problem: find* $v^\dagger \in V_n$ *such that*

$$v^\dagger = \underset{v \in V_n}{\arg\min} \|w^o - P_{W_m} v\| \ . \tag{1.3.43}$$

Proof: One can find the proof in the first part of the proof procedure of Proposition 1.3.2. Here we provide another proof by using the algebraic form. From (1.3.42), we have

$$\mathbb{W}\boldsymbol{\beta}^* + \mathbb{M}\boldsymbol{\alpha}^* = \boldsymbol{y}^t, \tag{1.3.44}$$

and

$$\mathbb{M}^T \boldsymbol{\beta}^* = 0. \tag{1.3.45}$$

Since $\mathbb{W}$ is invertible, then from (1.3.44) we have

$$\boldsymbol{\beta}^* + \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}^* = \mathbb{W}^{-1}\boldsymbol{y}^t. \tag{1.3.46}$$

Substitute (1.3.46) into (1.3.45) :

$$\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}^* = \mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t. \tag{1.3.47}$$

Knowing that $\mathbb{M}$ is a full column rank matrix, we further have

$$\boldsymbol{\alpha}^* = (\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M})^{-1}\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t, \tag{1.3.48}$$

and

$$v^* = \mathcal{B}_n\boldsymbol{\alpha}^* = \mathcal{B}_n(\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M})^{-1}\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t. \tag{1.3.49}$$

We now derive the algebraic form of $v^\dagger$ of (1.3.43). From Proposition 1.3.1, we have $w^o = P_{W_m}u^t = \mathcal{W}_m\mathbb{W}\boldsymbol{\ell}(u^t) = \mathcal{W}_m\mathbb{W}^{-1}\boldsymbol{y}^t$, and $P_{W_m}v = \mathcal{W}_m\mathbb{W}^{-1}\mathcal{W}_m^T\mathcal{B}_n\boldsymbol{\alpha} = \mathcal{W}_m\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}$. Thus we have:

$$\begin{aligned}
\|w^o - P_{W_m}v\|^2 &= \|\mathcal{W}_m\mathbb{W}^{-1}\boldsymbol{y}^t - \mathcal{W}_m\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}\|^2 \\
&= \|\mathcal{W}_m(\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha})\|^2 \\
&= (\mathcal{W}_m(\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}))^T(\mathcal{W}_m(\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha})) \\
&= (\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha})^T\mathcal{W}_m^T\mathcal{W}_m(\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}) \\
&= (\boldsymbol{y}^{tT}\mathbb{W}^{-1} - \boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{W}^{-1})\mathbb{W}(\mathbb{W}^{-1}\boldsymbol{y}^t - \mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}) \\
&= \boldsymbol{y}^{tT}\mathbb{W}^{-1}\boldsymbol{y}^t - 2\boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t + \boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}
\end{aligned} \tag{1.3.50}$$

Then problem (1.3.43) is equivalent to: find $\boldsymbol{\alpha}^\dagger$ such that

$$\boldsymbol{\alpha}^\dagger = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min}\, J(\boldsymbol{\alpha}) = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min}\, \boldsymbol{y}^{tT}\mathbb{W}^{-1}\boldsymbol{y}^t - 2\boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t + \boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha} \tag{1.3.51}$$

The determination of the optimal solution $\boldsymbol{\alpha}^\dagger$ requires that the gradient of the cost function $J(\boldsymbol{\alpha})$ with respect to its variable $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}^\dagger$ be equal to zero:

$$\nabla J(\boldsymbol{\alpha}^\dagger) = \frac{dJ}{d\boldsymbol{\alpha}}|_{\boldsymbol{\alpha}^\dagger} = 2\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M}\boldsymbol{\alpha}^\dagger - 2\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t = 0. \tag{1.3.52}$$

Thus we have

$$\boldsymbol{\alpha}^\dagger = (\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M})^{-1}\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t, \tag{1.3.53}$$

and

$$v^\dagger = \mathcal{B}_n\boldsymbol{\alpha}^\dagger = \mathcal{B}_n(\mathbb{M}^T\mathbb{W}^{-1}\mathbb{M})^{-1}\mathbb{M}^T\mathbb{W}^{-1}\boldsymbol{y}^t. \tag{1.3.54}$$

From (1.3.49) and (1.3.54), one can easily find that $v^\dagger = v^*$, this completes the proof.

□

It thus appears that this part v^* of the approximation provided by PBDW or the optimal recovery approach is one of the possible solutions of Problem 1.3.1 written as follows:

Problem 1.3.1 (Recovery with reduced basis, RRB) *We assume that we are given a series of linearly independent vectors $(q_i)_{i\in\mathbb{N}} \in V$ which span a sequence of reduced space $V_n = \text{span}\{q_1, \cdots, q_n\} \in V$ of dimension n, and linear measurements $\boldsymbol{y}^t = \boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$ of $u^t \in \mathcal{M}$. The recovery problem is to find $v \in V_n$ from observations $\boldsymbol{y}^t$ to approximate the true state u^t.*

For the sake of completeness, we also note that, another solution to Problem 1.3.1 is from the well-known least-squares approximation which is stated as: find v_{LS}^* such that

$$v_{LS}^* = \underset{v\in V_n}{\arg\min}\, \|\boldsymbol{\ell}(v) - \boldsymbol{y}^t\|_2\ . \tag{1.3.55}$$

We next derive the algebraic form of v_{LS}^* of (1.3.55). By denoting $v = \mathcal{B}_n\boldsymbol{\alpha} \in V_n$, we first have

$$\begin{aligned}\|\boldsymbol{\ell}(v) - \boldsymbol{y}^t\|_2^2 &= \|\mathcal{W}_m^T\mathcal{B}_n\boldsymbol{\alpha} - \boldsymbol{y}^t\|_2^2 \\ &= \|\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^t\|_2^2 \\ &= \boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{M}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T\mathbb{M}^T\boldsymbol{y}^t + \boldsymbol{y}^{tT}\boldsymbol{y}^t\end{aligned} \tag{1.3.56}$$

Problem (1.3.55) is equivalent to: find $\boldsymbol{\alpha}_{LS}^*$ such that

$$\boldsymbol{\alpha}_{LS}^* = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min}\, J_{LS}(\boldsymbol{\alpha}) = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min}\, \boldsymbol{y}^{tT}\boldsymbol{y}^t - 2\boldsymbol{\alpha}^T\mathbb{M}^T\boldsymbol{y}^t + \boldsymbol{\alpha}^T\mathbb{M}^T\mathbb{M}\boldsymbol{\alpha} \tag{1.3.57}$$

The determination of the optimal solution $\boldsymbol{\alpha}_{LS}^*$ requires that the gradient of the cost function $J_{LS}(\boldsymbol{\alpha})$ with respect to its variable $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}_{LS}^*$ be equal zero:

$$\nabla J_{LS}(\boldsymbol{\alpha}^*) = \frac{dJ_{LS}}{d\boldsymbol{\alpha}}|_{\boldsymbol{\alpha}_{LS}^*} = 2\mathbb{M}^T\mathbb{M}\boldsymbol{\alpha}_{LS}^* - 2\mathbb{M}^T\boldsymbol{y}^t = 0. \tag{1.3.58}$$

Thus we have

$$\boldsymbol{\alpha}_{LS}^* = (\mathbb{M}^T\mathbb{M})^{-1}\mathbb{M}^T\boldsymbol{y}^t, \tag{1.3.59}$$

and

$$v_{LS}^* = \mathcal{B}_n\boldsymbol{\alpha}_{LS}^* = \mathcal{B}_n(\mathbb{M}^T\mathbb{M})^{-1}\mathbb{M}^T\boldsymbol{y}^t. \tag{1.3.60}$$

If we further define a weighted norm $\|\cdot\|_{\mathbb{W}}$ in $\mathbb{R}^m$, such that for all $\boldsymbol{y} \in \mathbb{R}^m$ we have $\|\boldsymbol{y}\|_{\mathbb{W}} := \sqrt{\boldsymbol{y}^T\mathbb{W}^{-1}\boldsymbol{y}}$ [41]. From (1.3.53) and (1.3.59) we have the following remark:

Remark 1.3.1 *The problem in* (1.3.43) *is equivalent to: find $v^* \in V_n$ that satisfies the weighted least-squares problem [41]:*

$$v^* = \underset{v\in V_n}{\arg\min}\, \|\boldsymbol{\ell}(v) - \boldsymbol{y}^t\|_{\mathbb{W}}. \tag{1.3.61}$$

Furthermore, if $w_1, \cdots, w_m$ are orthonormal basis, then $\mathbb{W} = \mathbb{I}$, which is the identity matrix, thus the problem in (1.3.43) *is equivalent to the least-squares problem* (1.3.55).

1.3.4 Stability properties of PBDW

We first state a proposition on the inf-sup constant $\beta(V_n, W_m)$ of PBDW which gives a practical computation framework for any given basis $\{q_1, \cdots, q_n\}$ of V_n and $\{w_1, \cdots, w_m\}$ of W_m. Let $\mathbb{W} = U_{\mathcal{W}}S_{\mathcal{W}}U_{\mathcal{W}}^T$, $\mathbb{Q} = U_{\mathcal{Q}}S_{\mathcal{Q}}U_{\mathcal{Q}}^T$ be the singular value decomposition of $\mathbb{W} = \mathcal{W}_m^T\mathcal{W}_n$ and $\mathbb{Q} = \mathcal{Q}_n^T\mathcal{Q}_n$, where $\mathcal{W}_m = (w_1, \cdots, w_m)$ and $\mathcal{Q}_n = (q_1, \cdots, q_n)$ respectively. Then we have

$$\hat{\mathcal{B}}_n = \mathcal{B}_n U_{\mathcal{Q}} S_{\mathcal{Q}}^{-\frac{1}{2}}, \tag{1.3.62}$$

and

$$\hat{\mathcal{W}}_m = \mathcal{W}_m U_{\mathcal{W}} S_{\mathcal{W}}^{-\frac{1}{2}}, \tag{1.3.63}$$

where $\hat{\mathcal{B}}_n = (\hat{q}_1, \cdots, \hat{q}_n)$ and $\hat{\mathcal{W}}_n = (\hat{w}_1, \cdots, \hat{w}_m)$ are orthonormalization representation of V_n and W_m respectively. We further define $\hat{\mathbb{M}} := \hat{\mathcal{W}}_m^T\hat{\mathcal{B}}_n$ and have the relation

$$\hat{\mathbb{M}} = S_{\mathcal{W}}^{-\frac{1}{2}}U_{\mathcal{W}}^T\mathcal{W}_m^T\mathcal{B}_n U_{\mathcal{Q}} S_{\mathcal{Q}}^{-\frac{1}{2}} = S_{\mathcal{W}}^{-\frac{1}{2}}U_{\mathcal{W}}^T\mathbb{M}U_{\mathcal{Q}}S_{\mathcal{Q}}^{-\frac{1}{2}}. \tag{1.3.64}$$

Now we introduce the following proposition on the inf-sup constant $\beta(V_n, W_m)$.

Proposition 1.3.4 *The inf-sup constant $\beta(V_n, W_m)$ equals to the minimum singular value of $\hat{\mathbb{M}}$.*

Proof: By convention, we organize the singular values of $\hat{\mathbb{M}}$ in decreasing order $0 < s_n \leq \cdots, \leq s_1 \leq 1$. Let $v = \hat{\mathcal{B}}\hat{\alpha} \in V_n$ and $w = \hat{\mathcal{W}}\hat{\beta} \in W_m$, and $P_{W_m} = \hat{\mathcal{W}}\hat{\mathcal{W}}^T$ is the orthogonal projection operator onto W_m, then we have

$$\begin{aligned}
\beta(V_n, W_m) &\equiv \inf_{v\in V_n} \sup_{w\in W_m} \frac{(v,w)}{\|v\|\|w\|} \\
&= \inf_{v\in V_n, \|v\|=1} \sup_{w\in W_m, \|w\|=1} (v,w) \\
&= \inf_{v\in V_n, \|v\|=1} \|P_{W_m} v\| \\
&= \inf_{\hat{\boldsymbol{\alpha}}\in\mathbb{R}^n, \|\hat{\boldsymbol{\alpha}}\|_2=1} \|\hat{\mathcal{W}}\hat{\mathcal{W}}^T\hat{\mathcal{B}}\hat{\boldsymbol{\alpha}}\| \\
&= \inf_{\hat{\boldsymbol{\alpha}}\in\mathbb{R}^n, \|\hat{\boldsymbol{\alpha}}\|_2=1} \|\hat{\mathcal{W}}^T\hat{\mathcal{B}}\hat{\boldsymbol{\alpha}}\|_2 \\
&= \inf_{\hat{\boldsymbol{\alpha}}\in\mathbb{R}^n, \|\hat{\boldsymbol{\alpha}}\|_2=1} \|\hat{\mathbb{M}}\hat{\boldsymbol{\alpha}}\|_2 \\
&= s_n[\hat{\mathbb{M}}]
\end{aligned} \tag{1.3.65}$$

this completes the proof.

□

Note that in *Remark 2.13* of [29], the authors also pointed out the relation $\beta(V_n, W_m) = s_n[\hat{\mathbb{M}}]$ by using the so-called 'favorable bases' $\{\hat{q}_1, \cdots, \hat{q}_n\}$ of V_n and $\{\hat{w}_1, \cdots, \hat{w}_m\}$ of W_m. Here we provide a computable scheme for any given basis $\{q_1, \cdots, q_n\}$ of V_n and $\{w_1, \cdots, w_m\}$ of W_m.

As stated in [29, 148, 31], the inf-sup constant $\beta(V_n, W_m)$ plays the key role in the stability performance of the recovery algorithm, especially with respect the model bias quantified by $\|P_{V_n+W_m} u^t - u^t\|$, see (1.3.21), (1.3.26) and (1.3.28).

In practical data assimilation procedure, the observations are always polluted by measurement noise and have the form $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o$, where $\boldsymbol{e}^o \in \mathbb{R}^m$ is the m-dimensional noise vector. Indeed, the PBDW state estimation (1.3.19) or (1.3.25) is obtained in the absence of the measurement noise $\boldsymbol{e}^o$, i.e., the observation we used is $\boldsymbol{y}^t = \boldsymbol{\ell}(u^t)$ or $w^o = P_{W_m} u^t$, thus, we denote the corresponding solution u^* by u^{nf}, the superscript "nf" denotes "noise-free". In [149], the authors attribute the PBDW estimation error to two distinct contributions:

- The *deterministic error*: this part arises from the fact that modeling error is inevitable (i.e. $u^t \notin V_n$) and that the update space W_m and V_n are finite dimensional; this error is present even if each observation is noise-free – that is, even if one is able to probe the true deterministic state. The error estimation on this part is represented in (1.3.21) or (1.3.26).

- The *stochastic error*: this part arises from the fact that each observation is noisy, corrupted by the random observation noise. The error estimation on this part will be shown in (1.3.67) and (1.3.68) later.

In case the observations suffer from noise, we apply the same PBDW statement but subject to the observation constraints $\boldsymbol{\ell}(\eta + v) = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o$. Thus the PBDW statement in the noisy case is:

find $u^* = v^* + \eta^*$ such that

$$\begin{aligned} (v^*, \eta^*) &= \arg\inf_{\substack{v \in V_n \\ \eta \in W_m}} \|\eta\|^2 \\ \text{subject to:} \quad & \boldsymbol{\ell}(\eta^* + v^*) = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o. \end{aligned} \tag{1.3.66}$$

In this case, the stochastic component of the error is given in [149, Proposition 3], here we list out for convenience,

$$E[u^* - u^{nf}] = 0 \ , \tag{1.3.67}$$

$$E[\|u^* - u^{nf}\|^2] \leq \left(1 + \frac{2}{\beta(V_n, W_m)}\right)^2 \sigma^2 \text{trace}(\mathbb{W}^{-1}) \ , \tag{1.3.68}$$

where $E[\cdot]$ is the expectation operator, σ^2 is the variance of the measurement noise.

To this end, the left important issue is that: how to find *good* V_n and W_m. There are at least two options:

1) V_n and W_m are sought independently.
   - The reduced space V_n can be obtained through several different model reduction processes, such as Proper orthogonal decomposition (POD) [128], Proper Generalized Decomposition (PGD) [129], different kinds of Weak Greedy RB methods [152, 131, 185, 56, 107], etc.
   - For a given V_n, there are several processes looking for W_m from a dictionary Σ of observation functions, such as i) greedy methods like GEIM [144] ($m = n$) and greedy stability maximization methods [148, 31] ($m \geq n$), ii) quasi-uniform or random methods, etc.
2) One can obtain V_n and W_m simultaneously. In the special case is $m = n$, one can obtain V_n and W_m with the greedy algorithm from GEIM [144]. We briefly summarise the procedure in Algorithm 1.3.2.

Algorithm 1.3.2 Greedy Algorithm of GEIM

1: $\mu_1 = \arg\sup_{\mu \in \mathcal{D}} \|u(\cdot, \mu)\|$
2: $\ell_1 = \arg\sup_{\ell \in \Sigma} |\ell(u(\cdot, \mu_1))|$
3: $q_1 = u(\cdot, \mu_1)/\ell_1(u(\cdot, \mu_1))$
4: **for** $n \geq 2$ **do** ▷ $\mathcal{J}_n := \sum_{i=1}^n \alpha_i q_i \in V_n$ s.t. $\ell_i(\mathcal{J}_n[u]) = \ell_i(u)$
5: $\quad \mu_n = \arg\sup_{\mu \in \mathcal{D}} \|(u - \mathcal{J}_{n-1}[u])(\mu)\|$
6: $\quad \ell_n = \arg\sup_{\ell \in \Sigma} |\ell\left((u - \mathcal{J}_{n-1}[u])(\mu_n)\right)|$
7: $\quad q_n = (u - \mathcal{J}_{n-1}[u])(\mu_n)/\ell_n\left((u - \mathcal{J}_{n-1}[u])(\mu_n)\right)$

1.3.5 State-of-the-art on the use of reduced order modeling for nuclear reactor simulations

We give here a brief state of the art on previous works on nuclear reactor simulations that make use of reduced order modeling techniques. The method which seems to have been more extensively employed is the POD [166, 177] method. Examples of applications to neutronics are

[216] and [47, 46, 100]. In addition to these contributions, one can find in [140] an evaluation of the temperature reactivity feedbacks in a lead-cooled fast reactor with a POD basis.

Less numerous are the contributions using the reduced basis method. The works [190, 189, 192] apply the method to nuclear reactor core spatial kinetics and dynamics. In particular, [51] models the motion of control rods by applying reduced bases to parametrized multi-group neutron diffusion equations both in the time-dependent and stationary formulations. In addition to this, a new numerical algorithm based on reduced basis techniques is proposed in [93] to improve the computational performances for the homogenization of a coupled elliptic-parabolic system describing flows in heterogeneous porous media in nuclear waste storage. As already brought up, our main interest in this thesis is to couple the reduced basis techniques with measurement data for state estimation/field reconstruction purposes.

1.4 Contributions of this work

There are mainly three contributions in this thesis.

1) **Stability analysis and noise control** (Chapter 2).

We first recognize that there are two different recovery algorithms to solve Problem 1.3.1.

- One is to find an element $v_{LC}^* \in V_n$ as described in (1.3.43) such that its projection onto the observable space W_m is the closest one to the observed data w^o, where the subscript 'LC' stands for 'Lebesgue constant' [144, 148, 146] which controls the error amplification caused by model bias. For an operator $A : V \to V$, the Lebesgue constant is defined as $\Lambda \coloneqq \sup_{u\in V} \frac{\|A(u)\|}{\|u\|}$. Later we will show that this algorithm aims at optimizing the Lebesgue constant of the recovery algorithm.
- The other is to find an element $v_{LS}^* \in V_n$ as described in (1.3.55) such that its observations $\boldsymbol{\ell}(v_{LS}^*)$ is the closest one to the observed data $\boldsymbol{y}^o$ in the m-dimension Euclidean space $\mathbb{R}^m$. For an operator $Q : \mathbb{R}^m \to V$ which maps the noise $\boldsymbol{e}^o \in \mathbb{R}^m$ onto the space V, we define the operator norm $\|Q\| := \sup_{\boldsymbol{y}\in\mathbb{R}^m} \frac{\|Q(\boldsymbol{y})\|}{\|\boldsymbol{y}\|_2}$ and call it the noise amplifying factor or stability factor. Later we will show that this algorithm aims at optimizing the noise amplification factor of the recovery algorithm.

We find that the two algorithms have different stability properties with respect to the mode bias and the measurement noise. By doing the stability analysis, we are able to optimize the observation space W_m and amount of sensors for a given accuracy. Furthermore, we find that the improvement of this optimization is rather limited, this inspired us to find other solutions to stabilize the recovery algorithms.

In detail, if we are given noisy observations, $\boldsymbol{y}^o = \boldsymbol{\ell}_i(u^t) + \boldsymbol{e}^o$, where $\boldsymbol{e}^o = (e_1, \cdots, e_m)^T \in \mathbb{R}^m$ is the m-dimensional noise vector, and solve u^* with (1.3.66), then we have the following error estimation

$$\|u^* - u^t\| \leq \frac{1}{\beta(V_n, W_m)} \|P_{V_n+W_m} u^t - u^t\| + \|Q\| \|\boldsymbol{e}^o\|_2, \tag{1.4.1}$$

where the operator $Q : \mathbb{R}^m \to V$ is determined by (1.3.66) which maps the noise $\boldsymbol{e}^o \in \mathbb{R}^m$ onto the space V. Numerical results show that $\|Q\|$ largely depends on the V_n, W_m and the specified observation functions $w_i,\ i = 1, \cdots, m$. We also find that $\|Q\|$ diverges when $m \to \infty$, thus $u^* \nrightarrow u^t$ in the noise case. Our contributions to improve this behavior are:

i) The search for a proper way to select ℓ_i or adapt the algorithm in (1.3.66) in order to have a relatively good $\|Q\|$.

ii) To see if including additional information on where $\mathcal{M}$ projects onto V_n improves the behavior.

Note that, in the noisy case, the adaptive parametrized-background data-weak (APBDW) approach [202] adapts the algorithm in (1.3.66) as follows:

$$u^* = \underset{u\in V, v\in V_n}{\arg\min}\ \xi_m\|u-v\|^2 + \|\boldsymbol{\ell}(u)-\boldsymbol{y}^o\|_2^2 \tag{1.4.2}$$

where $\xi_m > 0$ is a regularization coefficient that balances the relative importance of the mismatch from the reduced space V_n. The numerical results show that, in some cases, especially when the dimension n of the reduced space is relatively large, the performance of the above algorithm in controlling noise amplification is limited.

In some case, we observe that for specified basis $\{q_i\}_{i=1}^n$ of V_n, in absence of noise, the part of recovery u^* which lies on V_n, i.e. v^* is constrained by a hypercube:

$$\mathcal{K}_n(\tau,\alpha^0) := \{v\in V_n\ ;\ v=\sum_{i=1}^n \alpha_i q_i,\quad |\alpha_i|\le \alpha_i^0\}, \tag{1.4.3}$$

where $\alpha_i^0 = \sup_{u\in\mathcal{M}}\{|\alpha_i^*|\ |\ \sum_{i=1}^n \alpha_i^* q_i = v^*\}$ is the upper bound of α_i.

If the sensors suffer from noise, our approach for the underlying recovery problem is: find u^* such that

$$u^* = \underset{u\in V, v\in\mathcal{K}_n(\tau,\alpha^0)}{\arg\min}\ \xi_m\|u-v\|^2 + \|\boldsymbol{\ell}(u)-\boldsymbol{y}^o\|_2^2\ , \tag{1.4.4}$$

and call (1.4.4) the constrained stabilized PBDW (CS-PBDW). A first application of (1.4.4) is we might neglect the correction term η and thus simply optimize with respect to V_n:

$$u^* = \underset{v\in\mathcal{K}_n(\tau,\alpha^0)}{\arg\min}\ \|\boldsymbol{\ell}(u)-\boldsymbol{y}^o\|_2^2 \tag{1.4.5}$$

We call (1.4.5) the constrained stabilized recovery with reduced basis (CS-RRB). Observe that for $m = n$, (1.4.5) corresponds to the GEIM ([144, 146]) in absence of noise. The case $m \ge n$ is also analyzed in noise case in [16] which shows the possibility to improve the quality of the state estimation by adding more measurements for a fixed amount of reduced basis from GEIM greedy algorithm.

2) **Sensor failure and constraint on the sensor locations** (Chapter 5).

In some applications, the locations of the sensors might be restricted, also during real-time reconstruction, some sensors might fail for some reasons. These cases need a stable interpolation/extrapolation scheme, or relatively robust sensor deployment and a better reduced basis choosing scheme.

Our study is based on GEIM. In case of k sensor failures, we propose an algorithm to reselect the reduced space $\hat{V}_{n-k}$ spanned by $\{\hat{q}_1\cdots\hat{q}_{n-k}\}$, and make use of all the observations from the sensors without failures. Numerical results confirm the very stable and robust performance of the proposed method.

3) **Parameter sampling for multi-dimensional parameter problems** (Chapter 6).

The success of the 'classical' reduced basis methods rely on the assumption that the high fidelity model in the off-line phase is sufficiently accurate for all parameters. In practice, the construction of V_n is done with greedy algorithm [143, 48] when applying the reduced basis method. Such greedy approaches have proven themselves to be particularly valuable for the approximation of high-dimensional problems where simple approaches are excluded due to the curse of dimension. For a detailed recent overview of such ideas in a general context, we refer to [207]. Here we take

GEIM greedy algorithm for example, see Table 1.3.2. In the generic greedy approach, $\mathcal{D}, \Sigma$ can be infinite so we replace them with discrete training sets $\mathcal{D}^{(training)} \in \mathcal{D}$ and $\Sigma^{(training)} \in \Sigma$ to make this algorithm feasible.

Algorithm 1.4.3 Weak Greedy Algorithm of GEIM

1: $\mu_1 = \arg\max_{\mu \in \mathcal{D}^{(\text{training})}} \|u(\cdot, \mu)\|$
2: $\ell_1 = \arg\max_{\ell \in \Sigma^{(\text{training})}} |\ell(u(\cdot, \mu_1))|$
3: $q_1 = u(\cdot, \mu_1)/\ell_1(u(\cdot, \mu_1))$
4: **for** $n \geq 2$ **do** ▷ $\mathcal{J}_n := \sum_{i=1}^{n} \alpha_i q_i \in V_n$ s.t. $\ell_i(\mathcal{J}_n[u]) = \ell_i(u)$
5: $\quad \mu_n = \arg\max_{\mu \in \mathcal{D}^{(\text{training})}} \|(u - \mathcal{J}_{n-1}[u])(\mu)\|$
6: $\quad \ell_n = \arg\max_{\ell \in \Sigma^{(\text{training})}} |\ell\left((u - \mathcal{J}_{n-1}[u])(\mu_n)\right)|$
7: $\quad q_n = (u - \mathcal{J}_{n-1}[u])(\mu_n)/\ell_n\left((u - \mathcal{J}_{n-1}[u])(\mu_n)\right)$

The question is: how to choose $\mathcal{D}^{(training)}$ without degrading the algorithm? Or in detail, this can be addressed in the subsequent exposition:

i) Overfitting: In case of too small training set, the error on $\mathcal{D}^{(training)}$ may be nicely decreasing or can ideally be reduced to 0 with growing basis dimension, but the error for other test parameters $\mu \in \mathcal{D}$ remains large.

ii) Training time: In case of too large training set, the training time can be exorbitant.

The selection of $D^{(training)}$ is especially important in high-dimensional problems, the size of this training set quickly becomes considerable, rendering the computational cost substantial and perhaps even prohibitive. As a consequence, a fine enough train set is not realistic in practice, one is faced to the problem of ensuring the quality of the basis set under a non-rich enough train set. It is worth noting that when dealing with certain high-dimensional problems, one may encounter the situation that the optimal basis set itself is also of large size. This situation is caused by the general complexity of the real engineering problems, there are many strategies and various flavors of adaptivity, e.g., sampling set randomization, adaptive refinement of training sets, hp-RBM, time-partitioning etc., see, e.g., [84, 85, 5, 49, 54, 80, 108, 151, 10].

We propose a method to enrich the training set $D^{(training)}$ in a greedy way. In short, once we have a training set with $k > 0$ sampling points $\{\mu_1, \cdots, \mu_k\}$, the next sampling point μ_{k+1} is determined by $\mu_{k+1} = (\mu_i + \mu_j)/2$ s.t. $(i, j) = \arg\max_{1 \leq i', j' \leq k} \frac{\text{dist}(u(\mu_{i'}), u(\mu_{j'}))}{\text{dist}(\mu_{i'}, \mu_{j'})}$. This greedy algorithm is able to take the typical snapshots $u(\mu)$ into account in the every beginning. Furthermore, in case the computational cost of each snapshot is remarkable which is the general case in engineering domain, this algorithm is more efficient than uniform or random sampling methods.

1.5 Summary of the results by chapters

Chapter 1: We overview in the first chapter some relevant mathematical background that will be used frequently throughout the thesis. We analyzed and adapted the generalized empirical interpolation method (GEIM) and the parametrized-background data-weak (PBDW) approach to the problem of state estimation with data assimilation. These approaches aim at reducing the problem's complexity for state estimation based on the fact that the Kolmogorov n-width of the manifold of all solutions decays fast with n. The GEIM [144] that allows, from values of the measurements, to build a linear combination of some precomputed solutions associated to some well chosen parameters. The linear combination is determined in such a way that it takes the same values on the sensors. This concept generalizes the classical — e.g. polynomial or radial basis

— interpolation procedure, the main difference is that the interpolating functions are not a priori known but depend on the quantity we want to represent. The PBDW approach was originally proposed in [148] for perfect observations then the noisy observations case was analyzed in [149]. It is characterized by the following characteristics.

i) *Variational formulation*: the most important feature of PBDW is that, it is a special case of 3D-Var variational data assimilation for a parametrized background and a particular choice of (penalized-update) background covariance. Benefiting from the background space V_n, PBDW avoids modeling the covariance matrices in 3D-Var variational data assimilation formulation.

ii) *Background and update spaces*: PBDW provides the state estimate u as the sum of two contributions: $u = v + \eta$. The first contribution v, is called *deduced background estimate* which represents anticipated model part; and V_n is interpreted as a background or prior space which approximates the best-knowledge manifold on which the true state resides. This second contribution to η, is called *update estimate* which accommodates unanticipated update that the model couldn't represent.

Chapter 2: In this chapter, we present a *stability analysis* for the state estimation / field reconstruction with reduced basis. Then we propose the so-called constrained stabilized recovery methods in order to control the noise.

Stability analysis (Section 2.2): We formulate the posteriori error analysis respect to measurement noise for GEIM, RRB and PBDW. The estimation error is attributed to two distinct contributions: i) The *deterministic error*, which arises from the fact that modeling error is inevitable (i.e. $u^t \notin V_n$) and that the update space W_m and the background space V_n are finite dimensional; this error is present even if each observation is noise-free, though one is able to probe the true deterministic state. We use the so-called Lebesgue constant Λ to evaluate the deterministic error; ii) The *stochastic error*, which arises from the fact that each observation is noisy, corrupted by the random observation noise. We use the stability factor $\|Q\|$ to evaluate the stochastic error. We will also propose algorithms to optimize the update space W_m with respect to different measures, i.e., diminishing the deterministic error or diminishing the stochastic error.

Constrained Stabilized Reconstruction (Section 2.3 - 2.5): We propose the so-called constrained stabilized generalized empirical interpolation method (CS-GEIM) and the constrained stabilized parametrized-background data-weak (CS-PBDW) approaches to further improve the stability and robust performance for state estimation with data assimilation with noisy measurements. The constraint relies on the coefficients of the reduced basis functions, are not the same order of quantity, on the contrary, they decay in certain rate with the reduced dimension n increases. We demonstrate that with the coefficients constraint, the recovery with reduced basis dramatically decreases the noise amplifying, with good stability performance with respect to the reduced basis dimension n and the number of measurement m. To go further, we propose the regularized versions (R-GEIM, R-PBDW) to diminish the computational cost in the on-line phase, while retaining the same accuracy as the constrained stabilized formulations.

Chapter 3: In the second part of the present thesis, we applied the methodologies for sensor placement and field reconstruction applications in nuclear domain. The framework of GEIM/PBDW makes it possible for us to focus on the need for effective, many-query design evaluation in nuclear reactor domain: i) either real-time scenarios (control, parameter estimation, monitoring), where the solution of the problem needs to be known very quickly under limited resources for a previously unknown parameter, ii) or multi-query scenarios (design and optimization, multi-model/scale simulation), where the problem has to be solved repeatedly for many different parameters. In the nuclear reactor applications, we set the goal to reconstruct the thermal/fast flux and power distributions for a wide range of parameters variation domain by using the measurements from sensors with combination of reduced basis methods, instead of solving the original neutronic governing equations with

high resolution. The purpose of this chapter is to present the physical context and adapt the GEIM methodology to neutronic field.

Chapter 4: We provide three contributions in this chapter.

Sensor placement in a nuclear reactor core (Section 4.1): We first adapt the GEIM greedy algorithm, to be able to reconstruct the flux and power fields only from the thermal flux measurements. Then we show its application in optimizing the sensor placement in nuclear reactor cores. Numerical results on 1D/2D benchmarks confirm its feasibility. Furthermore, GEIM, can be applied efficiently to the analysis of real nuclear cores. With the GEIM approach and the sensor information, we are able to provide an optimized instrumental network to measure the neutronic flux inside the core. The locations of the instruments show a global coherence with respect to what is known from the physical behavior and process. This determination of the optimal localizations is done over all the geometries of reactors operated by EDF: PWR 900 MWe, PWR 1300 MWe and PWR 1450 MWe. Several strategies are developed in order to emphasize the advantages and limitations of the determination of the instrumental setup. Numerical results show that the method gives some instrumental setup that are consistent with what is expected from a physical point of view. Moreover, we demonstrate that, with respect to the random case, the results of optimal networks are clearly better. Finally, we conclude that the strategy which is the closest to the actual instrumentation setup gives the best result.

Neutronic field reconstruction (Section 4.2, 4.3): We first apply the GEIM to control rods movements problem to improve the flux and the power distribution knowledge in the whole domain without knowing the control rods position (step) parameters thanks to the measurements. To reach this goal, we first construct a set of snapshots by solving large amounts of PDEs for the underlying physical problem. Then we select the sensor places (interpolation points) and the basis functions in a greedy way with GEIM greedy algorithm. With the selected measurements, we are able to reconstruct the flux and power distribution in high accuracy. Numerical results confirm that the proposed GEIM methodology is hopefully acceptable in engineering usage, especially the online-offline computational framework is good for on-line monitoring purpose.

Then we apply the GEIM method to reconstruct the Pin-by-Pin flux/power distribution during the life cycle of a nuclear reactor core. All the simulations are based on the PWR 1450 MWe nuclear reactor operated by EDF. Numerical result shows that, the manifold of the underlying problem (even in Pin-by-Pin wise), which reflects the burnup and power evolution of the reactor core of the whole life, is rather regular. This makes it possible to reconstruct the fields with only a few basis functions and measurements. Further more, by considering the measurement system which is close to the real engineering case, the proposed method still provides an acceptable accuracy.

Neutronic field reconstruction with noisy measurements (Section 4.4): The extension of the method is to take into account the effect of the noise to qualify its impact when few instruments remains and also to control it through various technique, either data pre-processing or mathematical correction of the induced error. In this aspect, we propose the methodologies so called CS-GEIM and CS-PBDW to control the noise and also the model bias. Based on the stability analysis, we are able to optimize the sensor size and location. With the stabilized algorithms, we are able to control the measurement noise amplification. Further more, by adding more measurements, the proposed methods are able to provide even better accuracy than the noise level. The algorithms we proposed are robust with respect to the measurement noise and model bias. Finally, we emphasize that the regularized versions (R-GEIM, R-PBDW) with low computational cost are more suitable for on-line calculation in case the shape of the noise is known, i.e., Gaussion noise.

In the third part of the thesis, we present the supplementary developments for industrial applications of the methodologies.

Chapter 5: Measurement failures for data assimilation with reduced basis, particularly, EIM, is a practical issue. We propose a so called measurement failure tolerant EIM (mftEIM) approach to

address this issue. Compared with other possible approaches, numerical results confirmed that, the mftEIM approach provides the best interpolation in case of measurement failures. Need to point out that, the basis functions for mftEIM are not fixed, and depend on the detailed failure scheme. Thus, one can either calculate the specified new basis functions when the sensor failures happen, or pre-calculate the basis functions for all possible measurement failures cases, in the latter case, a relatively large storage is necessary.

Chapter 6: In the framework of reduced basis method, the crucial ingredient is the choice of reduced basis $\{q_n\}$ from the discrete manifold $\mathcal{M}^{(training)} := \{u(\mu)\ ;\ \mu \in \mathcal{D}^{(training)}\}$ to span the reduced basis space V_n. We propose an adaptive sampling algorithm to sampling the parameter in $\mathcal{D}$ to construct the discrete manifold. Compared with other adaptive methods e.g. [5, 49, 54, 80, 108, 151], the the method we proposed saves computing cost, it reduces the number of function evaluation which might be very costly in the real engineering problem. This gives the adaptive sampling method more potential for engineering problem with high-dimensional parameter space. Nevertheless, we need to point out that, the adaptive sampling method may fail for some special case, e.g., the snapshots are from the the function in [151] with periodicity or symmetrical parameter space. The algorithm with pseudo-grads as an indicator may fail because of periodicity or symmetry. In this case, one can set the initial points with LHS method or one can refer to [151] with locally adaptive approach for anisotropic parameter space.

Chapter 7: Finally, we conclude in chapter 7 with summary of the thesis and some suggestions for future work at EDF or even for the whole nuclear sector.

Chapter 2

State estimation with reduced basis and noisy measurements

Contents

In this chapter, we introduce the formulations of data assimilation with reduced basis and noisy measurements. First in Section 2.1 we derive the problem statement, and we highlight the key issue that will be addressed. In Section 2.2, we give a stability analysis. Then in Section 2.3 we derive the formulation of the constrained stabilized reconstruction based on reduced basis i.e. CS-GEIM. In Section 2.4 we apply the constraint to the PBDW method, and provide several formulations with respect to different kinds of noise distribution. We introduce in Section 2.5 the regularized stabilized GEIM and regularized stabilized PBDW, which are able to provide a very high efficiency computing framework. In the end, we draw a short conclusion for this chapter in Section 2.6.

We indicate here that the formulation of CS-GEIM has already been presented in a published paper with J.-P. Argaud, B. Bouriquet, Y. Maday and O. Mula. Its reference in the manuscript is [16].

2.1 Introduction and problem setting

We first recall the notations introduced in the previous chapter: V is a Hilbert space over a domain $\Omega \subset \mathbb{R}^d$ ($d \geq 1$) endowed with inner product $(\cdot,\cdot)$ and induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$, $V_n \in V$ stands for the reduced space of dimension n, and $W_m \in V$ stands for the observation space spanned by the Riesz representations of the m functionals $\{\ell_i\}_{i=1}^m$.

Recall that the objective of the state estimation as described in Problem 1.2.1 is to approximate the true state u^t from m observations $\boldsymbol{\ell}(u^t) = (\ell_1(u^t), \cdots, \ell_m(u^t))^T$. As already stated in the previous chapter, in the framework of data assimilation with reduced basis, PBDW provides the optimal algorithm in the sense that its outcome u^* satisfies

$$\|u^* - u^t\| \leq \nu(V_n, W_m)\|P_{V_n+W_m}u^t - u^t\|\,, \tag{2.1.1}$$

where $\nu(V_n, W_m)$ is the best constant among all possible reconstruction schemes (see (1.3.29)). We recall also that $\nu(V_n, W_m) = \beta(V_n, W_m)^{-1}$, where $\beta(V_n, W_m)$ can be evaluated by the computation of an inf-sup constant (1.3.22).

The estimate in (2.1.1) doesn't take into account the noise from measurements, in [149], the authors provide the error estimation to PBDW with respect to noisy observations. In [202, 203], the author proposed an adaptive parametrized-background data-weak approach to variational data assimilation, in which the noisy measurements are considered, and in [26] the authors analyzed the stability performance of the operators with respect to noisy measurements from an oblique projection point of view. In this chapter, we will analyze stability performance for the recovery problems with reduced basis, then provide some recovery methods which show robust stability performance with respect to noise.

Now, let us focus on the case that the observations suffer from noise. We are given noisy observations $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o = (\ell_1(u^t) + e_1, \cdots, \ell_m(u^t) + e_m)^T$, where $\boldsymbol{e}^o = (e_1, \cdots, e_m)^T \in \mathbb{R}^m$ is the noise vector. We assume that $\{e_i\}_{i=1}^m$ are independent, have density p_i on $\mathbb{R}$. More precisely we make the following three assumptions on the noise term:

i) zero mean: $E[e_i] = 0,\ i = 1, \cdots, m$;

ii) covariance: $E[e_i^2] = \sigma_i^2,\ i = 1, \cdots, m$;

iii) uncorrelated: $E[e_i e_j] = 0,\ i \neq j$.

We denote the covariance matrix by $\mathbb{D} := E[\boldsymbol{e}^T\boldsymbol{e}] \in \mathbb{R}^{m\times m}$, which is thus diagonal. We do not assume that the observation error follows any particular distribution, but only assume that the mean and the covariance of the distribution exist. Note that in practice the mean and covariance of the data acquired is more readily quantifiable than the distribution. We first study the following problem:

Problem 2.1.1 (Recovery with reduced basis and noisy measurements, RRB) *Let $\mathcal{M} := \{u^t(\mu) : \mu \in \mathcal{D}\}$ be the solution manifold of a physical problem, where $\mathcal{D}$ is the parameter set. Assume that we are given a series of linearly independent vectors $(q_i)_{i\in\mathbb{N}} \in V$ which span a reduced space $V_n = \operatorname{span}\{q_1, \cdots, q_n\} \in V$ of dimension n, and linear measurements $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o = (\ell_1(u^t) + e_1, \cdots, \ell_m(u^t) + e_m)^T$ of $u^t \in \mathcal{M}$. The recovery problem is: find $v \in V_n$ from observations $\boldsymbol{y}^o$ to approximate the true state u^t.*

In the following parts of this chapter, we will first analyze the stability performance for two typical methods for the above recovery problem. Based on the stability analysis, we provide two methods to optimize the observation space W_m with respect to different stability objectives. We will

show that the improvement is rather limited by optimizing the observation space. Thus we propose a constrained stabilized GEIM (CS-GEIM) approach for the recovery problem 2.1.1, which shows robust stability performance with respect to noisy measurements. The natural extension of CS-GEIM is that, by introducing the update term $\eta \in W_m$, we obtain the constrained stabilized PBDW (CS-PBDW), which again stabilizes the PBDW approach with respect to noisy measurements.

2.2 Stability analysis for recovery with reduced basis

2.2.1 Stability factors with respect to model bias and noisy measurements

The recovery problem 2.1.1 needs to be considered from two different aspects point of view:

- With respect to observations $\boldsymbol{y}^o$ (noisy or not), we denote by Q the operator: $\mathbb{R}^m \to V_n$, that to any $\boldsymbol{y}^o \in \mathbb{R}^m$ associates $v^*_{\boldsymbol{y}^o} = Q(\boldsymbol{y}^o)$ a solution to Problem 2.1.1.
- If the data are not noisy $\boldsymbol{y} = \boldsymbol{\ell}(u^t)$, we denote the associated operator $A = Q \circ \boldsymbol{\ell} \ : \ V \to V_n$, $\forall u^t \in V$, $v^*_{\boldsymbol{\ell}(u^t)} = A(u^t) = Q(\boldsymbol{\ell}(u^t))$.

We assume Q (and thus A) is a linear operator, i.e., we only focus on the recovery algorithms which are linear mappings. In this setting, we use two quantities to measure the quality of this recovery operator with respect to observation noise $\boldsymbol{e}^o$ and model bias/mismatch $\delta u^t = u^t - P_{V_n} u^t$.

- As a measure of stability with respect to measurement noise we use the operator norm

$$\|Q\| \coloneqq \sup_{\boldsymbol{y}^o \in \mathbb{R}^m} \frac{\|Q(\boldsymbol{y}^o)\|}{\|\boldsymbol{y}^o\|_2}, \tag{2.2.1}$$

  and call it stability factor. (Remind that $\|\cdot\|$ is the V-norm.)

- As a measure of stability with respect to model mismatch, we follow [144, 148, 146] and define the Lebesgue constant Λ of A as

$$\Lambda \coloneqq \sup_{u \in V, u \neq 0} \frac{\|A(u)\|}{\|u\|}. \tag{2.2.2}$$

In fact, the Λ defined in (2.2.2) is just the norm of the reconstruction operator $A = Q \circ \boldsymbol{\ell}$. In the context of interpolation, it is usually called Lebesgue constant. When not causing confusion and for the sake of convenience, we call this norm Lebesgue constant in this thesis, particularly in the context of reconstruction with reduced basis. Note also that we may follow [29, 7, 26] and use the so-called quasi-optimality constant $\nu(A)$ $(0 < \nu(A) < \infty)$ which is the smallest value for a given A that satisfies

$$\forall u^t \in V, \quad \|A(u^t) - u^t\| \leq \nu(A) \|P_{V_n} u^t - u^t\| \ , \tag{2.2.3}$$

here we say that A is quasi-optimal if $\nu(A)$ is close to 1. We remind that $P_{V_n} u^t$ denotes the orthogonal projection of u^t onto the space V_n. Thus, for recovery problem 2.1.1 in the linear case, from (2.2.1) and (2.2.3) we derive an error estimate of the form:

$$\forall u^t \in V, \forall \boldsymbol{e}^o \in \mathbb{R}^m, \quad \|u^t - Q(\boldsymbol{\ell}(u^t) + \boldsymbol{e}^o)\| \leq \nu(A) \|P_{V_n} u^t - u^t\| + \|Q\| \|\boldsymbol{e}^o\|_2 \ . \tag{2.2.4}$$

Furthermore, (2.2.2) implies that such a recovery operator Q (and A) is exact on the space V_n, i.e., $u^t = Q(\boldsymbol{\ell}(u^t))$ for all $u^t \in V_n$, this is straightforward, since otherwise $\nu(A) = \infty$. Following

[26], we call recovery operators fulfilling this property *perfect recovery operators*. From (2.2.2) and (2.2.3) we further have

$$\nu(A) \leq (1 + \Lambda). \tag{2.2.5}$$

The relation (2.2.5) can be proved through

$$\begin{aligned} \forall u^t \in V, \quad \|A(u^t) - u^t\| &= \|A(u^t) - P_{V_n} u^t + P_{V_n} u^t - u^t\| \\ &\leq \|A(u^t) - P_{V_n} u^t\| + \|P_{V_n} u^t - u^t\| \\ &= \|A(u^t) - A(P_{V_n} u^t)\| + \|P_{V_n} u^t - u^t\| \\ &= \|A(u^t - P_{V_n} u^t)\| + \|P_{V_n} u^t - u^t\| \\ &\leq \Lambda \|u^t - P_{V_n} u^t\| + \|P_{V_n} u^t - u^t\| \\ &= (1 + \Lambda) \|P_{V_n} u^t - u^t\|. \end{aligned} \tag{2.2.6}$$

The third relation follows from the fact that A is exact on the space V_n, the fourth relation follows from the linearity of A, and the last second relation follows from the definition of Λ. Then from the definition of $\nu(A)$ we have (2.2.5). Later we will show the equivalence between the Lebesgue constant Λ and the quasi-optimality constant $\nu(A)$ for some particular operators.

We observe that, the error bound in (2.2.4) raises several questions: i) Does the operator have the smallest possible operator norm $\|Q\|$? ii) Does the operator have the smallest possible quasi-optimality constant $\nu(A)$? To answer these questions, we will analyze two typical approaches to the recovery problem 2.1.1, and show some numerical results of these stability factors, one can refer to [26] for further analysis. Note also that the PBDW approach and GEIM approach have the smallest possible quasi-optimality constant $\nu(A)$ [29, 145].

2.2.2 Recovery aiming to optimize the Lebesgue constant Λ

We begin by recalling some useful notations introduced in Chapter 1. Let $\mathcal{B}_n = (q_1, \cdots, q_n)$ and $\mathcal{W}_m = (w_1, \cdots, w_m)$ be the representation of V_n and W_m respectively, where $\{w_i\}_{i=1}^m$ are the Riesz representations of the observation functionals $\{\ell_i\}_{i=1}^m$. Furthermore, let $\hat{\mathcal{B}}_n = (\hat{q}_1, \cdots, \hat{q}_n)$ and $\hat{\mathcal{W}}_m = (\hat{w}_1, \cdots, \hat{w}_m)$ be the basis set derived from $\mathcal{B}_n$ and $\mathcal{W}_m$ respectively by an orthonormalization process. For notational simplicity, we suppress the subscript n and m of $\mathcal{B}_n, \mathcal{W}_m, \hat{\mathcal{B}}_n, \hat{\mathcal{W}}_m$. We further denote $\mathbb{W} = \mathcal{W}^T\mathcal{W}, \mathbb{Q} = \mathcal{B}^T\mathcal{B}$, $\mathbb{M} = \mathcal{W}^T\mathcal{B}$, $\mathbb{M}_r = \mathcal{W}^T\hat{\mathcal{B}}$, $\hat{\mathbb{W}} = \hat{\mathcal{W}}^T\hat{\mathcal{W}}$, $\hat{\mathbb{M}} = \hat{\mathcal{W}}^T\hat{\mathcal{B}}$.

Stability factors of Lebesgue Constant approach

We first analyze the algorithm as stated in (1.3.43) or (1.3.61) to the recovery problem 2.1.1 that we recall: find $v^* \in V_n$ s.t.

$$v^* = \arg\min_{v \in V_n} \|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_{\mathbb{W}} \,, \tag{2.2.7}$$

here we recall that the weighted norm $\|\cdot\|_{\mathbb{W}}$ in $\mathbb{R}^m$ is defined such that for all $\boldsymbol{y} \in \mathbb{R}^m$ we have $\|\boldsymbol{y}\|_{\mathbb{W}} := \sqrt{\boldsymbol{y}^T\mathbb{W}^{-1}\boldsymbol{y}}$. As pointed out in [29], the above algorithm provides an operator Q, such that the associated operator A has the smallest quasi-optimality constant $\nu(A)$ (also the smallest Lebesgue constant Λ) in the sense that no other linear Q operator leads to a better $\nu(Q \circ \boldsymbol{\ell})$, similar proof can also be found in [26]. We thus denote this approach as stated in (2.2.7) by *LC approach*, where 'LC' stands for the optimal Lebesgue constant. We refer to (1.3.49), but replace $\mathcal{B}$ with $\hat{\mathcal{B}}$ which is the orthonormal basis set, then the solution to (2.2.7) is given by:

$$v^* = \hat{\mathcal{B}}\boldsymbol{\alpha}^* = \hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\boldsymbol{y}^o. \tag{2.2.8}$$

Now we state a proposition on the recovery (2.2.7).

Proposition 2.2.1 *The recovery operator $Q_{LC} : \mathbb{R}^m \to V_n$ from (2.2.7) can be represented as*

$$Q_{LC} = \hat{\mathcal{B}}(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1}, \tag{2.2.9}$$

with norm

$$\|Q_{LC}\| = s_1[(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1}]. \tag{2.2.10}$$

And the associated linear operator $A_{LC} : V \to V_n$ of Q_{LC} can be represented as

$$A_{LC} = \hat{\mathcal{B}}(\hat{\mathbb{M}}^T \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T \hat{\mathcal{W}}^T, \tag{2.2.11}$$

with the Lebesgue constant

$$\Lambda_{LC} = (s_n[\hat{\mathbb{M}}])^{-1}. \tag{2.2.12}$$

The notation $s_k[\mathbb{B}]$ represents the $k-$th singular value of matrix $\mathbb{B}$ in descending order, the subscript 'LC' of Q, A and Λ reflect that they come from the recovery (2.2.7).

Proof: The relation (2.2.9) is straightforward from (2.2.8). Now we prove (2.2.10).

$$\begin{aligned} \forall \boldsymbol{e}^o \in \mathbb{R}^m,\ \|Q_{LC}\boldsymbol{e}^o\| &= \|\hat{\mathcal{B}}(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \boldsymbol{e}^o\| \\ &= \|(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \boldsymbol{e}^o\|_2. \end{aligned} \tag{2.2.13}$$

The second relation follows from the fact that $\hat{\mathcal{B}}$ is an orthonormal representation of V_n, thus we have

$$\begin{aligned} \|Q_{LC}\| &:= \sup_{\boldsymbol{e}^o \in \mathbb{R}^m} \frac{\|Q_{LC}(\boldsymbol{e}^o)\|}{\|\boldsymbol{e}^o\|_2} \\ &= \sup_{\boldsymbol{e}^o \in \mathbb{R}^m} \frac{\|(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \boldsymbol{e}^o\|_2}{\|\boldsymbol{e}^o\|_2} \\ &= \sup_{\boldsymbol{e}^o \in \mathbb{R}^m, \|\boldsymbol{e}^o\|_2 = 1} \|(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \boldsymbol{e}^o\|_2 \ . \end{aligned} \tag{2.2.14}$$

The right side of the last relation is the definition of matrix norm of $\mathbb{B} := (\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1}$ induced by the Euclidean vector norm [158], thus we have

$$\sup_{\boldsymbol{e}^o \in \mathbb{R}^m, \|\boldsymbol{e}^o\|_2 = 1} \|\mathbb{B}\boldsymbol{e}^o\|_2 = s_1[\mathbb{B}]. \tag{2.2.15}$$

We now prove (2.2.11). $\forall u \in V,\ \boldsymbol{y} = \boldsymbol{\ell}(u) = \mathcal{W}^T u$, we have

$$A_{LC}(u) = Q_{LC}(\boldsymbol{\ell}(u)) = \hat{\mathcal{B}}(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \mathcal{W}^T u. \tag{2.2.16}$$

We recall that $\hat{\mathcal{W}} = \mathcal{W} U_{\mathcal{W}} S_{\mathcal{W}}^{-\frac{1}{2}}$, $\hat{\mathcal{B}} = \mathcal{B} U_Q S_Q^{\frac{1}{2}}$, where $\mathbb{W} = U_{\mathcal{W}} S_{\mathcal{W}} U_{\mathcal{W}}^T$ and $\mathbb{Q} = U_{\mathcal{Q}} S_{\mathcal{Q}} U_{\mathcal{Q}}^T$ are the singular value decompositions of $\mathbb{W}$ and $\mathbb{Q}$ respectively. We further have $\mathbb{W}^{-1} = U_{\mathcal{W}} S_{\mathcal{W}}^{-1} U_{\mathcal{W}}^T$, then

$$\begin{aligned} A_{LC}(u) &= \hat{\mathcal{B}}(\mathbb{M}_r^T \mathbb{W}^{-1} \mathbb{M}_r)^{-1} \mathbb{M}_r^T \mathbb{W}^{-1} \mathcal{W}^T u \\ &= \hat{\mathcal{B}}(\hat{\mathcal{B}}^T \mathcal{W} U_{\mathcal{W}} S_{\mathcal{W}}^{-1} U_{\mathcal{W}}^T \mathcal{W}^T \mathcal{B})^{-1} \hat{\mathcal{B}}^T \mathcal{W} U_{\mathcal{W}} S_{\mathcal{W}}^{-1} U_{\mathcal{W}}^T \mathcal{W}^T u \\ &= \hat{\mathcal{B}}(\hat{\mathbb{M}}^T \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T \hat{\mathcal{W}}^T u \ . \end{aligned} \tag{2.2.17}$$

Let $\hat{\mathbb{M}} = \tilde{U} \tilde{S} \tilde{V}^T$ be the singular value decomposition of $\hat{\mathbb{M}}$ where $\tilde{S}$ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal in descending order, i.e., $s_1 > \cdots > s_n > 0$. Then $\tilde{\mathcal{B}} = \hat{\mathcal{B}}\tilde{V} = (\tilde{q}_1, \cdots, \tilde{q}_n), \tilde{\mathcal{W}} = \hat{\mathcal{W}}\tilde{U} = (\tilde{w}_1, \cdots, \tilde{w}_m)$ are another orthonormal representations

of V_n, W_m respectively. Furthermore we have the notation $\tilde{\mathbb{M}} := \tilde{\mathcal{W}}^T\tilde{\mathcal{B}} = \tilde{S}$, and $(\tilde{\mathbb{M}}^T\tilde{\mathbb{M}})^{-1}\tilde{\mathbb{M}}^T$ is an $n \times m$ rectangular diagonal matrix with entries $({s_1}^{-1}, \cdots, {s_n}^{-1})$ on the diagonal. In these settings, we further have

$$\begin{aligned} A_{LC}(u) &= \tilde{\mathcal{B}}(\tilde{\mathbb{M}}^T\tilde{\mathbb{M}})^{-1}\tilde{\mathbb{M}}^T\tilde{\mathcal{W}}^T u \\ &= (\tilde{q}_1, \cdots, \tilde{q}_n)(\tilde{\mathbb{M}}^T\tilde{\mathbb{M}})^{-1}\tilde{\mathbb{M}}^T \left((\tilde{w}_1, u), \cdots, (\tilde{w}_m, u)\right)^T \\ &= \sum_{i=1}^{n} \frac{1}{s_i}(\tilde{w}_i, u)\tilde{q}_i \ . \end{aligned} \tag{2.2.18}$$

Thus we have

$$\forall u \in V,\ \|A_{LC}(u)\|^2 = \|\sum_{i=1}^{n} \frac{1}{s_i}(\tilde{w}_i, u)\tilde{q}_i\|^2 = \sum_{i=1}^{n} \frac{1}{s_i^2}|(\tilde{w}_i, u)|^2\|\tilde{q}_i\|^2 = \sum_{i=1}^{n} \frac{1}{s_i^2}|(\tilde{w}_i, u)|^2 \ . \tag{2.2.19}$$

Then

$$\Lambda_{LC} := \sup_{u\in V} \frac{\|A_{LC}(u)\|}{\|u\|} = \sup_{u\in V} \frac{\left(\sum_{i=1}^{n} \frac{1}{s_i^2}|(\tilde{w}_i, u)|^2\right)^{1/2}}{\|u\|} = s_n^{-1}. \tag{2.2.20}$$

For the last relation we invoke Cauchy-Schwarz inequality and obtain the desired bound.

□

Remark 2.2.1 *From* (2.2.12) *and Proposition 1.3.4, we observe that the Lebesgue constant* Λ *of the recovery* (2.2.7) *is equivalent to the inverse of inf-sup constant* $\beta(V_n, W_m)$*, i.e.,* $\Lambda = \beta(V_n, W_m)^{-1}$*. A special case is that when* $m = n$*,* Λ *is the Lebesgue constant of GEIM algorithm [145, Theorem 1.2]. Another case is that when* $V_n \subset W_m$*, from the definition of* $\beta(V_n, W_m)$ *we have* $\beta(V_n, W_m) = 1$*, thus* $\Lambda_{LC} = 1$*. This is the ideal case such that any element in* V_n *can be represented by an element in* W_m*.*

Stability factors of PBDW

As already stated, PBDW looks for the solution in $V_n + W_m$ which has the same measurements as $\boldsymbol{y}^o$ and with a minimal updated term $\eta^* \in W_m \cap V_n^{\perp}$ in the sense of norm. We first have the following statement.

Statement 2.2.1 *The algebraic representation of the solution* u^* *to the PBDW approach in noisy case as stated in* (1.3.66) *can be represented as:*

$$u^* = A_{PBDW}u^t + Q_{PBDW}\boldsymbol{e}^o \ . \tag{2.2.21}$$

The linear operator $Q_{PBDW} : \mathbb{R}^m \to V_n + W_m$ *of PBDW can be represented as:*

$$Q_{PBDW} := \hat{\mathcal{B}}\mathbb{E} + \mathcal{W}\mathbb{H}, \tag{2.2.22}$$

and the associated linear operator $A_{PBDW} : V \to V_n + W_m$ *of* Q_{PBDW} *can be represented as:*

$$A_{PBDW} := \hat{\mathcal{B}}\mathbb{E}\mathcal{W}^T + \mathcal{W}\mathbb{H}\mathcal{W}^T, \tag{2.2.23}$$

where $\mathbb{E} := (\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}$*, and* $\mathbb{H} := \mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}$*.*

Proof: For a given space V_n and W_m, and the corresponding basis $\{q_1, \cdots, q_n\}$ and $\{w_1, \cdots, w_m\}$, we follow all the notations introduced previously. We denote the solution by $u^* = v^* + \eta^* =$

$\hat{\mathcal{B}}\boldsymbol{\alpha}^* + \mathcal{W}\boldsymbol{\beta}^*$. We refer to Proposition 1.3.3 and Remark 1.3.1 for the analysis with respect to noise-free measurements, it is easy to find that v^* is the solution to Problem 2.1.1: find v^* s.t.

$$v^* = \arg\min_{v \in V_n} \|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_{\mathbb{W}}. \tag{2.2.24}$$

Hence we follow Proposition 2.2.1 and obtain

$$v^* = \hat{\mathcal{B}}\boldsymbol{\alpha}^*, \quad \text{with } \boldsymbol{\alpha}^* = (\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\boldsymbol{y}^o. \tag{2.2.25}$$

From $(1.3.66)_2$, we have $\mathcal{W}^T\mathcal{W}\boldsymbol{\beta}^* + \mathcal{W}^T\hat{\mathcal{B}}\boldsymbol{\alpha}^* = \boldsymbol{y}^o$, i.e.,

$$\mathbb{W}\boldsymbol{\beta}^* + \mathbb{M}_r\boldsymbol{\alpha}^* = \boldsymbol{y}^o, \tag{2.2.26}$$

Substituting (2.2.25) into (2.2.26), we obtain

$$\boldsymbol{\beta}^* = \mathbb{W}^{-1}\boldsymbol{y}^o - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\boldsymbol{y}^o. \tag{2.2.27}$$

Recall the denotations of $\mathbb{E}$ and $\mathbb{H}$, from (2.2.25) and (2.2.27) we have

$$u^* = \hat{\mathcal{B}}\mathbb{E}\boldsymbol{y}^o + \mathcal{W}\mathbb{H}\boldsymbol{y}^o, \tag{2.2.28}$$

Substituting $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o = \mathcal{W}^T u^t + \boldsymbol{e}^o$ into (2.2.28), we further obtain

$$u^* = \hat{\mathcal{B}}\mathbb{E}\mathcal{W}^T u^t + \mathcal{W}\mathbb{H}\mathcal{W}^T u^t + \hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o + \mathcal{W}\mathbb{H}\boldsymbol{e}^o = A_{PBDW}u^t + Q_{PBDW}\boldsymbol{e}^o\ . \tag{2.2.29}$$

This completes the proof. □

Based on the linear representation (2.2.29) of u^*, we then have the following proposition.

Proposition 2.2.2 *The stability factor $\|Q_{PBDW}\|$ is*

$$\|Q_{PBDW}\| = (s_1[\mathbb{S}])^{1/2}, \tag{2.2.30}$$

and the Lebesgue constant is

$$\Lambda_{PBDW} = \left(s_n[\hat{\mathbb{M}}]\right)^{-1}, \tag{2.2.31}$$

where

$$\mathbb{S} = \mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} + \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-2}\mathbb{M}_r^T\mathbb{W}^{-1}\ , \tag{2.2.32}$$

and $s_k[\mathbb{B}]$ represents the $k-$th singular value of matrix $\mathbb{B}$ in descending order.

Proof: First we recognize that $\hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o \perp \mathcal{W}\mathbb{H}\boldsymbol{e}^o$, i.e., $(\mathcal{W}\mathbb{H}\boldsymbol{e}^o, \hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o) = 0$, this follows from the fact that $\forall \boldsymbol{e}^o \in \mathbb{R}^m$, if $u^*(\boldsymbol{e}^o) = v^*(\boldsymbol{e}^o) + \eta^*(\boldsymbol{e}^o)$ is the PBDW solution, we have $\eta^*(\boldsymbol{e}^o) \perp V_n$ and we know that $\eta^*(\boldsymbol{e}^o) \in V_n^\perp \cap W_m$. Here we present the proof for the sake of convenience.

$$\begin{aligned}
(\mathcal{W}\mathbb{H}\boldsymbol{e}^o, \hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o) &= \boldsymbol{e}^{oT}\mathbb{H}^T\mathcal{W}^T\hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o \\
&= \boldsymbol{e}^{oT}\mathbb{H}^T\mathbb{M}_r\mathbb{E}\boldsymbol{e}^o \\
&= \boldsymbol{e}^{oT}\left[\mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\right]\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\boldsymbol{e}^o \\
&= \boldsymbol{e}^{oT}\left[\mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\right]\boldsymbol{e}^o \\
&= 0
\end{aligned} \tag{2.2.33}$$

Then we prove $\mathbb{H}^T\mathbb{W}\mathbb{H} = \mathbb{H}$, this follows:

$$\begin{aligned}\mathbb{H}^T\mathbb{W}\mathbb{H} &= \left[\mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\right]\mathbb{W}\left[\mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\right] \\ &= \mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} - \\ &\qquad \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} \\ &= \mathbb{W}^{-1} - \mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1} \\ &= \mathbb{H}\end{aligned} \tag{2.2.34}$$

By definition, we have

$$\begin{aligned}\forall \boldsymbol{e}^o \in \mathbb{R}^m,\ \|Q_{PBDW}\boldsymbol{e}^o\|_2^2 &= (\hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o + \mathcal{W}\mathbb{H}\boldsymbol{e}^o, \hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o + \mathcal{W}\mathbb{H}\boldsymbol{e}^o) \\ &= \boldsymbol{e}^{oT}(\mathbb{E}^T\hat{\mathcal{B}}^T + \mathbb{H}^T\mathcal{W}^T)(\hat{\mathcal{B}}\mathbb{E} + \mathcal{W}\mathbb{H})\boldsymbol{e}^o \\ &= \boldsymbol{e}^{oT}(\mathbb{E}^T\hat{\mathcal{B}}^T\hat{\mathcal{B}}\mathbb{E} + \mathbb{H}^T\mathcal{W}^T\mathcal{W}\mathbb{H})\boldsymbol{e}^o \\ &= \boldsymbol{e}^{oT}(\mathbb{E}^T\mathbb{E} + \mathbb{H}^T\mathbb{W}\mathbb{H})\boldsymbol{e}^o \\ &= \boldsymbol{e}^{oT}(\mathbb{E}^T\mathbb{E} + \mathbb{H})\boldsymbol{e}^o.\end{aligned} \tag{2.2.35}$$

The third relation follows from $\hat{\mathcal{B}}\mathbb{E}\boldsymbol{e}^o \perp \mathcal{W}\mathbb{H}\boldsymbol{e}^o$, and the fourth relation follows from $\mathbb{H}^T\mathbb{W}\mathbb{H} = \mathbb{H}$. We further obtain

$$\mathbb{E}^T\mathbb{E}+\mathbb{H} = \mathbb{W}^{-1}-\mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathbb{W}^{-1}+\mathbb{W}^{-1}\mathbb{M}_r(\mathbb{M}_r^T\mathbb{W}^{-1}\mathbb{M}_r)^{-2}\mathbb{M}_r^T\mathbb{W}^{-1} = \mathbb{S}. \tag{2.2.36}$$

Note that from $\mathbb{W}^{-1/2}\mathbb{M}_r = \hat{\mathcal{W}}^T\hat{\mathcal{B}} = \hat{\mathbb{M}}$, we may also present $\mathbb{S}$ by

$$\mathbb{S} = \mathbb{W}^{-1} - \mathbb{W}^{-1/2}\hat{\mathbb{M}}(\hat{\mathbb{M}}^T\hat{\mathbb{M}})^{-1}\hat{\mathbb{M}}^T\mathbb{W}^{-1/2} + \mathbb{W}^{-1/2}\hat{\mathbb{M}}(\hat{\mathbb{M}}^T\hat{\mathbb{M}})^{-2}\hat{\mathbb{M}}^T\mathbb{W}^{-1/2}. \tag{2.2.37}$$

We finally invoke Rayleigh quotient theory [110] to the positive definite hermitian matrix $\mathbb{S}$ and obtain the desired bound

$$\|Q_{PBDW}\|^2 = \sup_{\boldsymbol{e}^o\in\mathbb{R}^m, \boldsymbol{e}^o\neq\mathbf{0}} \frac{\boldsymbol{e}^{oT}\,\mathbb{S}\,\boldsymbol{e}^o}{\boldsymbol{e}^{oT}\boldsymbol{e}^o} = s_1[\mathbb{S}]. \tag{2.2.38}$$

We now prove (2.2.31). Recall that $\forall u \in V,\ \boldsymbol{y} = \boldsymbol{\ell}(u) = \mathcal{W}^T u$, we have

$$A_{PBDW}(u) = \hat{\mathcal{B}}\mathbb{E}\mathcal{W}^T u + \mathcal{W}\mathbb{H}\mathcal{W}^T u. \tag{2.2.39}$$

We recall that $\hat{\mathcal{W}} = \mathcal{W}_m U_{\mathcal{W}} S_{\mathcal{W}}^{-\frac{1}{2}}$, $\hat{\mathcal{B}} = \mathcal{B}U_Q S_Q^{\frac{1}{2}}$, where $\mathbb{W} = U_{\mathcal{W}}S_{\mathcal{W}}U_{\mathcal{W}}^T$ and $\mathbb{Q} = U_{\mathcal{Q}}S_{\mathcal{Q}}U_{\mathcal{Q}}^T$ are the singular value decompositions of $\mathbb{W}$ and $\mathbb{Q}$ respectively. We further have $\mathbb{W}^{-1} = U_{\mathcal{W}}S_{\mathcal{W}}^{-1}U_{\mathcal{W}}^T$, then

$$A_{PBDW}(u) = \hat{\mathcal{B}}(\hat{\mathbb{M}}^T\hat{\mathbb{M}})^{-1}\hat{\mathbb{M}}^T\hat{\mathcal{W}}^T u + \hat{\mathcal{W}}\hat{\mathcal{W}}^T u - \hat{\mathcal{W}}\hat{\mathbb{M}}(\hat{\mathbb{M}}^T\hat{\mathbb{M}})^{-1}\hat{\mathbb{M}}^T\hat{\mathcal{W}}^T u\ . \tag{2.2.40}$$

Let $\hat{\mathbb{M}} = \tilde{U}\tilde{S}\tilde{V}^T$ be the singular value decomposition of $\hat{\mathbb{M}}$ where $\tilde{S}$ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal in descending order, i.e., $s_1 > \cdots > s_n > 0$. Then $\tilde{\mathcal{B}} = \hat{\mathcal{B}}\tilde{V} = (\tilde{q}_1, \cdots, \tilde{q}_n), \tilde{\mathcal{W}} = \hat{\mathcal{W}}\tilde{U} = (\tilde{w}_1, \cdots, \tilde{w}_m)$ are another orthonormal representations of V_n, W_m respectively, furthermore $\tilde{\mathbb{M}} = \tilde{\mathcal{W}}^T\tilde{\mathcal{B}} = \tilde{S}$. In this setting, we further have

$$\begin{aligned}A_{PBDW}(u) &= \tilde{\mathcal{B}}(\tilde{\mathbb{M}}^T\tilde{\mathbb{M}})^{-1}\tilde{\mathbb{M}}^T\tilde{\mathcal{W}}^T u + \tilde{\mathcal{W}}\tilde{\mathcal{W}}^T u - \tilde{\mathcal{W}}\tilde{\mathbb{M}}(\tilde{\mathbb{M}}^T\tilde{\mathbb{M}})^{-1}\tilde{\mathbb{M}}^T\tilde{\mathcal{W}}^T u \\ &= \sum_{i=1}^{n}\frac{1}{s_i}(\tilde{w}_i, u)\tilde{q}_i + \sum_{i=n+1}^{m}(\tilde{w}_i, u)\tilde{w}_i\end{aligned} \tag{2.2.41}$$

Thus we have

$$\begin{aligned}\forall u \in V,\ \|A_{PBDW}(u)\|^2 &= \|\sum_{i=1}^{n}\frac{1}{s_i}(\tilde{w}_i,u)\tilde{q}_i + \sum_{i=n+1}^{m}(\tilde{w}_i,u)\tilde{w}_i\|^2 \\ &= \sum_{i=1}^{n}\frac{1}{s_i^2}|(\tilde{w}_i,u)|^2\|\tilde{q}_i\|^2 + \sum_{i=n+1}^{m}|(\tilde{w}_i,u)|^2\|\tilde{w}_i\|^2 \\ &= \sum_{i=1}^{n}\frac{1}{s_i^2}|(\tilde{w}_i,u)|^2 + \sum_{i=n+1}^{m}|(\tilde{w}_i,u)|^2\ .\end{aligned} \tag{2.2.42}$$

Then

$$\Lambda_{PBDW} := \sup_{u\in V}\frac{\|A_{PBDW}(u)\|}{\|u\|} = \sup_{u\in V}\frac{\left(\sum_{i=1}^{n}\frac{1}{s_i^2}|(\tilde{w}_i,u)|^2 + \sum_{i=n+1}^{m}|(\tilde{w}_i,u)|^2\right)^{1/2}}{\|u\|} = s_n^{-1}. \tag{2.2.43}$$

For the last relation we invoke Cauchy-Schwarz inequality and obtain the desired bound. □

Remark 2.2.2 *From* (2.2.12) *and* (2.2.31)*, we observe that the Lebesgue constant* Λ_{PBDW} *of PBDW is the same to the Lebesgue constant* Λ_{LC} *of the LC approach* (2.2.7)*. Furthermore, we have* $\Lambda_{PBDW} = \Lambda_{LC} = \beta(V_n, W_m)^{-1}$*. It is easy to observe that the error bound of LC approach is*

$$\forall u^t \in V,\quad \|A_{LC}(u^t) - u^t\| \leq \beta(V_n, W_m)^{-1}\|P_{V_n}u^t - u^t\|,$$

and from (1.3.26) *we observe that PBDW narrows the above bound to*

$$\|A_{PBDW}(u^t) - u^t\| \leq \beta(V_n, W_m)^{-1}\|P_{V_n+W_m}u^t - u^t\|.$$

2.2.3 Recovery aiming to optimize the stability factor $\|Q\|$

Now let us turn to the analysis of the second approach to the recovery problem 2.1.1 that we recall: find $v_{LS}^* \in V_n$ s.t.

$$v_{LS}^* = \arg\min_{v\in V_n}\|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_2, \tag{2.2.44}$$

which is a classical least-squares (LS) approximation, thus we call it *LS approach*. We refer to (1.3.60), but replace $\mathcal{B}$ with the orthonormal one $\hat{\mathcal{B}}$, then the solution to (2.2.44) is given by:

$$v_{LS}^* = \hat{\mathcal{B}}\boldsymbol{\alpha}_{LS}^* = \hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\boldsymbol{y}^o\ . \tag{2.2.45}$$

We first state a proposition on the estimations of the Lebesgue constant and stability factor of (2.2.44) as follows.

Proposition 2.2.3 *The recovery operator* $Q_{LS} : \mathbb{R}^m \to V_n$ *from* (2.2.44) *can be represented as*

$$Q_{LS} = \hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T, \tag{2.2.46}$$

with norm

$$\|Q_{LS}\| = s_1[(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T] = (s_n[\mathbb{M}_r])^{-1}. \tag{2.2.47}$$

And the associated linear operator $A_{LS} : V \to V_n$ *of* Q_{LS} *can be represented as*

$$A_{LS} = \hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathcal{W}^T\ , \tag{2.2.48}$$

and the Lebesgue constant of A_{LS} satisfies

$$\Lambda_{LS} \leq s_1[(\hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T S_{\mathcal{W}}], \tag{2.2.49}$$

where $s_k[\mathbb{B}]$ represents the $k-$th singular value of the matrix $\mathbb{B}$ in descending order, the subscript 'LS' of Q, A and Λ means they come from LS approach (2.2.44).

Proof: The relation (2.2.46) is straightforward from (2.2.45). Now we prove (2.2.47).

$$\forall \boldsymbol{e}^o \in \mathbb{R}^m,\ \|Q_{LS}\boldsymbol{e}^o\| = \|\hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\boldsymbol{e}^o\| = \|(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\boldsymbol{e}^o\|_2\ . \tag{2.2.50}$$

The second relation follows from the fact that $\hat{\mathcal{B}}$ is an orthonormal representation of V_n, thus we have

$$\begin{aligned}\|Q_{LS}\| &:= \sup_{\boldsymbol{e}^o\in\mathbb{R}^m} \frac{\|Q_{LS}(\boldsymbol{e}^o)\|}{\|\boldsymbol{e}^o\|_2} \\ &= \sup_{\boldsymbol{e}^o\in\mathbb{R}^m} \frac{\|(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\boldsymbol{e}^o\|_2}{\|\boldsymbol{e}^o\|_2} \\ &= \sup_{\boldsymbol{e}^o\in\mathbb{R}^m,\|\boldsymbol{e}^o\|_2=1} \|(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\boldsymbol{e}^o\|_2\ .\end{aligned} \tag{2.2.51}$$

The right side of the last relation is the definition of matrix norm of $\mathbb{B} := (\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T$ induced by the Euclidean vector norm [158], thus we have

$$\sup_{\boldsymbol{e}^o\in\mathbb{R}^m,\|\boldsymbol{e}^o\|_2=1} \|\mathbb{B}\boldsymbol{e}^o\|_2 = s_1[\mathbb{B}]. \tag{2.2.52}$$

We now prove (2.2.48). $\forall u \in V,\ \boldsymbol{y} = \boldsymbol{\ell}(u) = \mathcal{W}^T u$, we have

$$A_{LS}(u) = Q_{LS}(\boldsymbol{\ell}(u)) = \hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathcal{W}^T u. \tag{2.2.53}$$

Then we prove (2.2.49).

$$\begin{aligned}\forall u \in V,\ \frac{\|A_{LS}(u)\|}{\|u\|} &= \frac{\|\hat{\mathcal{B}}(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathcal{W}^T u\|}{\|u\|} \\ &= \frac{\|(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T\mathcal{W}^T u\|_2}{\|u\|} \\ &= \frac{\|(\mathbb{M}_r^T\mathbb{M}_r)^{-1}\mathbb{M}_r^T U_{\mathcal{W}} S_{\mathcal{W}}^{\frac{1}{2}}\hat{\mathcal{W}}^T u\|_2}{\|u\|} \\ &= \frac{\|(\hat{\mathbb{M}}^T S_{\mathcal{W}}\hat{\mathbb{M}})^{-1}\hat{\mathbb{M}}^T S_{\mathcal{W}}\hat{\mathcal{W}}^T u\|_2}{\|u\|}\end{aligned} \tag{2.2.54}$$

The second relation follows from the orthonormal of $\hat{\mathcal{B}}$, the third relation follows from $\hat{\mathcal{W}}_n = \mathcal{W}_m U_{\mathcal{W}} S_{\mathcal{W}}^{-\frac{1}{2}}$, and from the definitions of $\mathbb{M}_r$ and $\hat{\mathbb{M}}$, we have the last relation.

Again we recall the definition of matrix norm of $\mathbb{B}' := (\hat{\mathbb{M}}^T S_{\mathcal{W}}\hat{\mathbb{M}})^{-1}\hat{\mathbb{M}}^T S_{\mathcal{W}}$ induced by the Euclidean vector norm [158], and we have

$$\sup_{\boldsymbol{y}\in\mathbb{R}^m,\|\boldsymbol{y}\|_2=1} \|\mathbb{B}'\,\boldsymbol{y}\|_2 = s_1[\mathbb{B}']. \tag{2.2.55}$$

Thus we have

$$\begin{aligned}\forall u \in V,\ \frac{\|A_{LS}(u)\|}{\|u\|} &= \frac{\|(\hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathcal{W}}^T u\|_2}{\|u\|} \\ &\leq s_1[(\hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T S_{\mathcal{W}}] \frac{\|\hat{\mathcal{W}}^T u\|_2}{\|u\|} \\ &= s_1[(\hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T S_{\mathcal{W}}] \frac{\|P_{W_m} u\|}{\|u\|} \\ &\leq s_1[(\hat{\mathbb{M}}^T S_{\mathcal{W}} \hat{\mathbb{M}})^{-1} \hat{\mathbb{M}}^T S_{\mathcal{W}}]\end{aligned} \tag{2.2.56}$$

The third relation follows from the fact that $\hat{\mathcal{W}}$ is an orthonormal representation of W_m, thus we obtain the desired bound. □

Note that we are unable to present the exact formula of Λ_{LS}, but rather an upper bound as in (2.2.49), so that the estimate of Λ_{LS} we present in the next context are all from (2.2.49).

Remark 2.2.3 *As pointed out in [26] that Q_{LS} provides the smallest possible operator norm among all perfect recovery operators which satisfy* (2.2.4).

2.2.4 Numerical results and analysis

Test case 2.2.1 *Consider the family of functions from [145],*

$$g(x, \mu) \equiv \frac{1}{\sqrt{1 + (25 + \mu_1 cos(\mu_2 x))x^2}}, \tag{2.2.57}$$

where $\mu = (\mu_1, \mu_2) \in \mathcal{D} = [0.01, 24.9] \times [0.01, 15], x \in \Omega = [0, 1]$.

We work with $L^2(\Omega) = \{u \mid \|u\|_{L^2(\Omega)} < \infty\}$ where the norm $\|\cdot\|_{L^2(\Omega)}$ is induced by the inner product $(w, v)_{L^2(\Omega)} = \int_\Omega w(x)v(x)dx$. Also, $H^1(\Omega) = \{u \mid \|u\|_{H^1(\Omega)} < \infty\}$ where the norm $\|\cdot\|_{H^1(\Omega)}$ is induced by the inner product $(w, v)_{H^1(\Omega)} = \int_\Omega w(x)v(x)dx + \int_\Omega \nabla w(x) \nabla v(x)dx$. Here we set Gaussian convolution as the measurement functionals, namely,

$$\ell_i(v) = \text{Gauss}(v, x_i^c, r_i) \equiv \int_\Omega \Big\{(2\pi r_i^2)^{-d/2} \exp\Big(- \frac{\|x - x_i^c\|^2_{\ell^2(\mathbb{R}^d)}}{2r_i^2}\Big)\Big\} v(x)dx, \tag{2.2.58}$$

where $x_i^c \in \mathbb{R}^d$ is the center of the sensor, and $r_i \in \mathbb{R}_{>0}$ is the width of the sensor; localized observation is of particular interest in this work.

We work with spaces V_n and W_m chosen as follows:

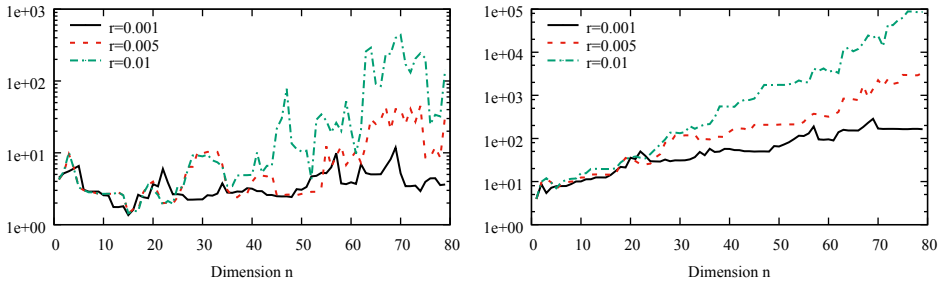
- V_n and the first n linear functionals are built with GEIM.
- The remaining $n+1$ to m functionals are selected with location which maximizes the shortest distance to the set of points at the current step.

Once the linear functionals $\{\ell_i\}_i$ have been determined, the corresponding Riesz representation $\{w_i\}_i$ are used to span the measurement space $W_m = \text{span}\{w_1, \cdots, w_m\}$. Our calculations were done with the finite element code $Freefem++$ [102].

Stability with respect to sensor width

Here we analyze the dependence of the stability factors with respect to the sensor width. We list the variations of the Lebesgue constant Λ and the stability factor $\|Q\|$ for sensor width $r = 0.001, 0.005$ and 0.01 in Figure 2.2.1 for LC approach. The reduced basis $\{q_i\}_i$ and the measurements are selected with GEIM greedy algorithm based on a fine enough training set $\mathcal{M}^{(training)} := \{g(x,\mu) \mid \mu \in \mathcal{D}^{(training)}\}$. In all cases, the number of sensors equals to the reduced dimension i.e. $m = n$.

From Figure 2.2.1 we observe that, the smaller the sensor width is, the more stable either for Λ or for $\|Q\|$. Furthermore, the two factors are of different orders of magnitude, e.g., for the most stable case ($r = 0.001$) in the figure, the Lebesgue constant Λ is of $\mathcal{O}(10)$, but $\|Q\|$ is of $\mathcal{O}(100)$, i.e., the stability factor with respect to noise is much more sensitive than the stability factor with respect to model mismatch. In figure 2.2.1b, the explosion of $\|Q\|$ comes from the almost linear dependence of the observation functionals $\{\ell_i\}_i$ when r is too large.



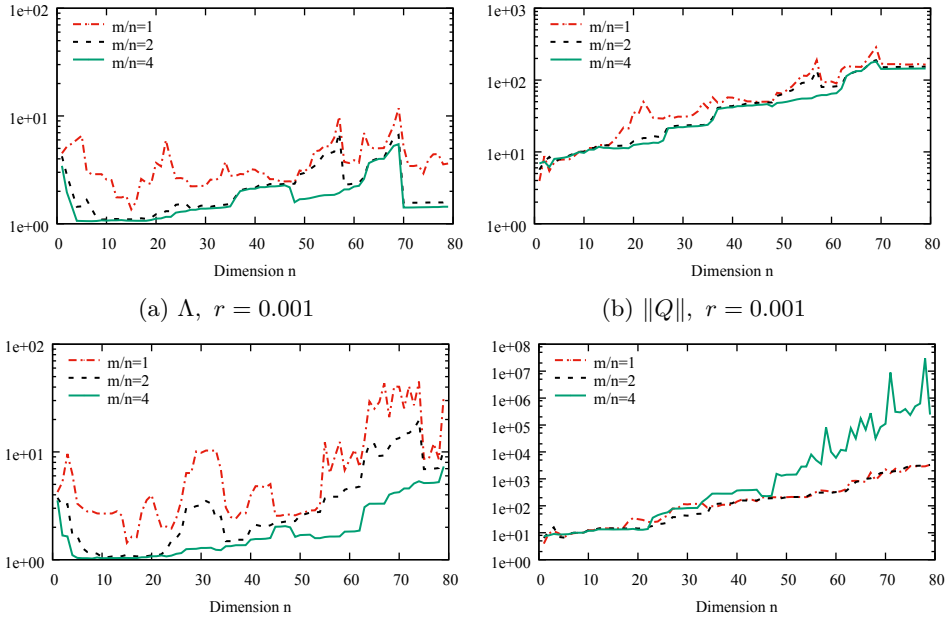
(a) Λ, $m = n$ (b) $\|Q\|$, $m = n$

Figure 2.2.1: Variations of the Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the reduced dimension n for LC approach, for three different values of the sensor width r. In all cases, we keep the number of sensors m equal to the number of reduced basis n.

Stability with respect to the number of observations

Now we study the effect of different number of observations. We determine the m measurements by i) selecting the first n measurements with GEIM greedy algorithm, and ii) selecting the left $m - n$ measurements with the quasi-uniform method [148], i.e., at step i, we insert a new point at the location x_i^c which maximizes the minimum distance to the set of points $\{x_1^c, \cdots, x_{i-1}^c, \}$ at step $i - 1$. Figure 2.2.2 shows the variations of the Lebesgue constant Λ and stability factor $\|Q\|$ for $m/n = 1, 2, 4$ for LC approach with sensor width $r = 0.001, 0.005$.

We observe that, by adding more measurements, on one hand, the Lebesgue constant is well improved, especially for a 'low resolution' measurement system e.g. $r = 0.005$ in our case; on the other hand, the stability factor $\|Q\|$ explodes, that is to say, by adding more measurements doesn't improve $\|Q\|$, see Figure 2.2.2d. The explosion comes from the almost linear dependence of the $\{\ell_i\}_i$ when m is too large for a relatively large r. In any case, the recovery approach (2.2.7) provides a better Lebesgue constant Λ rather than $\|Q\|$ with respect to measurement noise. This leads us considering other approaches for the recovery problem 2.1.1.



(a) $\Lambda,\ r = 0.001$ (b) $\|Q\|,\ r = 0.001$

(c) $\Lambda,\ r = 0.005$ (d) $\|Q\|,\ r = 0.005$

Figure 2.2.2: Variations of the Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the reduced dimension n for LC approach, for three different values of m/n and two values of sensor width r.

Comparison between LC approach and LS approach

From Proposition 2.2.3, we are able to present a numerical example to illustrate how Λ and $\|Q\|$ behave for LS approach as stated in (2.2.46). We show the comparison between LS approach and LC approach for three different values of m/n with the sensor width $r = 0.001,\ 0.005$ in Figure 2.2.3. All the calculations are based on Test case 2.2.1. We can draw several conclusions from the numerical results as follows.

i) The LC approach always provides the smallest Lebesgue constant Λ among all simulated cases.

ii) The LS approach always provides the smallest stability factor $\|Q\|$ among all simulated cases. That is to say, LS approach stabilizes the noise with respect to LC approach, especially when the number of measurements m is large or the sensor width r is large, i.e., the cases which lead the observation functionals almost linearly dependent.

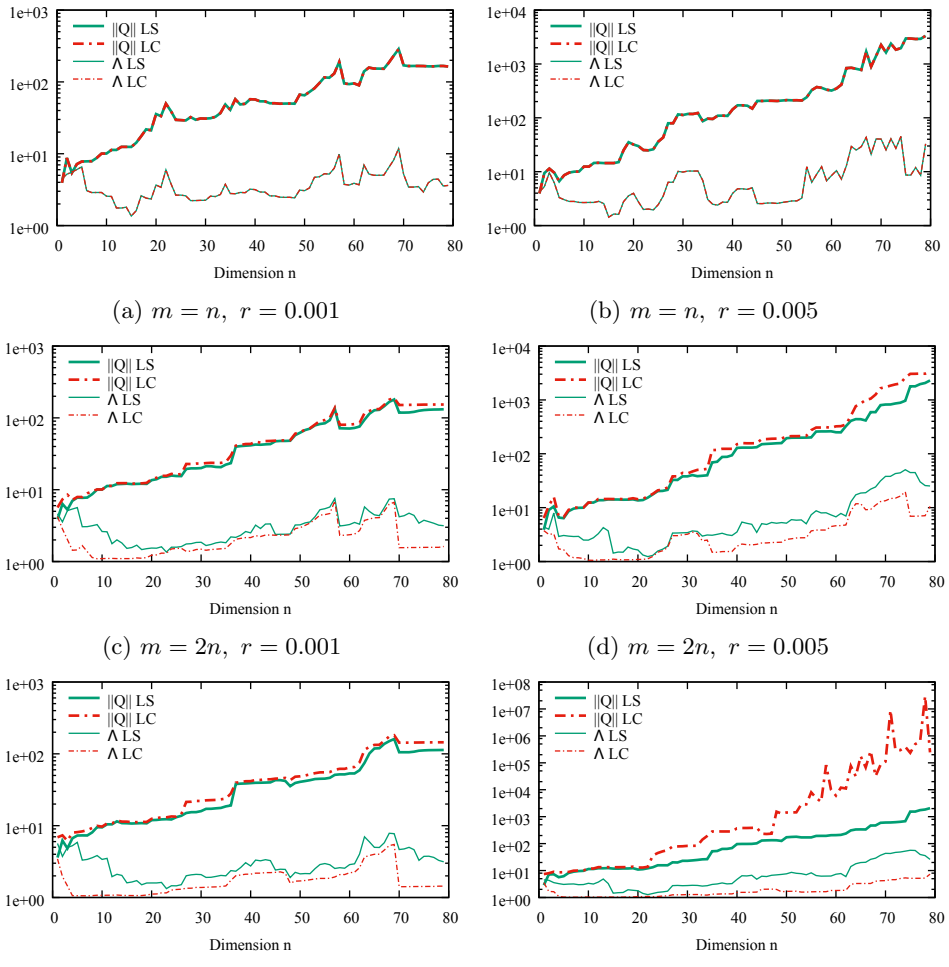
iii) Figure 2.2.3a and Figure 2.2.3b confirm that there is no difference for both approaches when $m = n$, this is the classical GEIM approach.

Furthermore, we built a set of snapshot $\mathcal{M}^{(test)} := \{g(x,\mu) \mid \mu \in \mathcal{D}^{(test)}\}$ which is different from the training set $\mathcal{M}^{(training)}$ with which we obtain the reduced space V_n. Then we test the two approaches in i) noise-free case and ii) the case with a uniform distributed noise with noise level $\sigma = 0.01$ on the test set. In order to verify the effects of Λ and $\|Q\|$ with respect the number of observations and the sensor width, we list the maximum relative recovery errors for $m = n,\ 4n$ and $r = 0.001,\ 0.005$ in Figure 2.2.4 from which we again confirm that, with respect to the recovery error:

i) the LC approach is better than LS approach when the measurements are noise-free. Furthermore, by adding more measurements, LC approach is able to improve the recovery accuracy, on the contrary, LS approach might enlarge the error, see Figure 2.2.4a and 2.2.4c;

ii) the LS approach is better than the LC approach when the measurements are polluted by noise. Particularly when the observation functions are almost linear dependent, the error in the LC approach explodes in presence of the noise (e.g. see 'LC m=4n' in Figure 2.2.4d), on the contrary, LS approach stabilizes the noise (e.g. see 'LS m=4n' in Figure 2.2.4d).

In short, LC approach is a more stable recovery method with respect to model mismatch, on the contrary, LS approach is a stable recovery method with respect to measurement noise.

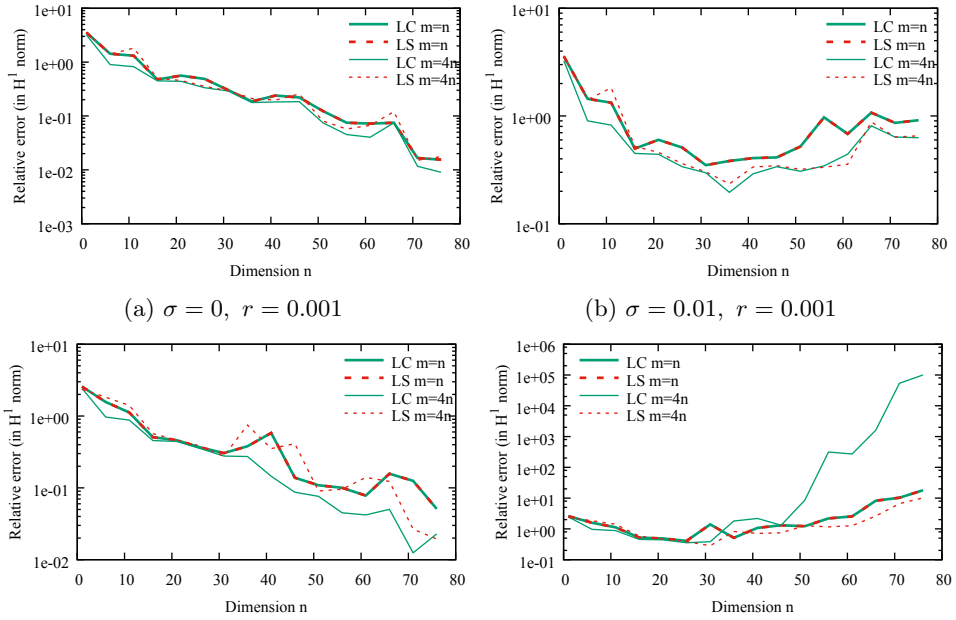


(a) $m = n,\ r = 0.001$ (b) $m = n,\ r = 0.005$

(c) $m = 2n,\ r = 0.001$ (d) $m = 2n,\ r = 0.005$

(e) $m = 4n,\ r = 0.001$ (f) $m = 4n,\ r = 0.005$

Figure 2.2.3: Variations of the Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the reduced dimension n for LC approach and LS approach, for three different values of m/n and two values of sensor width r.



(a) $\sigma = 0,\ r = 0.001$ (b) $\sigma = 0.01,\ r = 0.001$

(c) $\sigma = 0,\ r = 0.005$ (d) $\sigma = 0.01,\ r = 0.005$

Figure 2.2.4: Variations of the recovery errors (in H^1 norm) with respect to the reduced dimension n for LC approach and LS approach, for two different values of noise level σ and two values of sensor width r.

2.2.5 Optimizing the observation space W_m

Methodologies

From Proposition 2.2.1 and 2.2.3 and the corresponding numerical results, we observe that, for a given space V_n, the selection of an optimal observation space W_m largely depends on the following two criteria:

- The minimization of the Lebesgue constant $\Lambda = (s_n[\hat{\mathbb{M}}])^{-1}$ for LC approach. To improve the stability with respect to model mismatch, we wish to choose W_m such that any element in V_n is well approximated by an element in W_m (see Remark 2.2.1).

- The minimization of the stability factor $\|Q\| = (s_n[\mathbb{M}_r])^{-1}$ for LS approach. To improve the stability with respect to measurement noise, we wish to choose $\{\ell_1, \cdots, \ell_m\}$ (and thus W_m) such that any noise in $\mathbb{R}^m$ is well controlled.

Even though the underlying framework may accommodate any observation functional that is consistent with the data-acquisition procedure, in this work we focus on localized observations. As noted before, for localized observations using a given observation functional $\ell_i(\cdot) \equiv Gauss(\cdot, x_i^c, r_i)$, the locations of the centers $\{x_i^c\}_{i=1}^m$ and the feature $\{r_i\}_{i=1}^m$ largely determine the space W_m. We may select the observation functionals (and more specifically the observation centers) using several different processes: $Process_m^V(V_n) \to W_m$. Here we list a few:

- *Quasi-uniform or random method.* The algorithm aims at minimizing the approximation error by providing a uniform coverage of the domain, see [148]. The Quasi-uniform is a deterministic

sequential procedure: at step i, we insert a new point at the location which maximizes the minimum distance to the set of points at step $i-1$.

- *Generalized empirical interpolation method* [144]. The algorithm works for $m = n$.
- *Op* Λ *method*. The procedure is described in Algorithm 2.2.4. In short, the algorithm minimizes the Lebesgue constant Λ in a greedy manner. Note that, this algorithm is a little different from the greedy stability maximization procedure in [148], in which that algorithm maximizes the inf-sup constant $\beta(V_n, W_m)$ in a greedy manner. Furthermore, there is another way to add next observation functionals: in [31], a greedy algorithm to ensure that $\beta(V_n, W_m) \geq \beta_{min} > 0$ with a small number m of observations is given. [31] also gives convergence rates of the behavior of $\beta(V_n, W_m)$ with respect to m.
- *Op* $\|Q\|$ *method*. The procedure is described in Algorithm 2.2.5. In short, the algorithm minimizes the operator norm $\|Q\|$ in a greedy manner.

Algorithm 2.2.4 Op Λ greedy algorithm

input: the reduced space V_n, the observation functional set Ξ_M
output: the optimal observation space W_m
Set $W_n = \text{span}\{w_1, \cdots, w_n\}$ with GEIM greedy algorithm from Ξ_M
for $i = n+1, \cdots, m$ **do**
    Set $W_i = \text{span}\{W_{i-1}, w_j\}$, where $w_j \in \Xi_M$
    With V_n, W_i, compute the Lebesgue constant with (2.2.12), i.e., $\Lambda(V_n, W_i) = (s_n[\hat{\mathbb{M}}])^{-1}$
    $w_i = \arg\min_{w_j \in \Xi_M} \Lambda(V_n, W_i)$
    Set $W_i = \text{span}\{W_{i-1}, w_i\}$
return W_m

Algorithm 2.2.5 Op $\|Q\|$ greedy algorithm

input: the reduced space V_n, the observation functional set Ξ_M
output: the optimal observation space W_m
Set $W_n = \text{span}\{w_1, \cdots, w_n\}$ with GEIM greedy algorithm from Ξ_M
for $i = n+1, \cdots, m$ **do**
    Set $W_i = \text{span}\{W_{i-1}, w_j\}$, where $w_j \in \Xi_M$
    With V_n, W_i, compute the stability factor with (2.2.47), i.e., $\|Q\|(V_n, W_i) = (s_n[\mathbb{M}_r])^{-1}$
    $w_i = \arg\min_{w_j \in \Xi_M} \|Q\|(V_n, W_i)$
    Set $W_i = \text{span}\{W_{i-1}, w_i\}$
return W_m

Numerical results and analysis

We compared the following three methods for the determination of the observation space W_m: by adding a new point (sensor center) that i) maximizes the minimum distance (Op Max-Min dist), ii) minimizes the Lebesgue constant $\Lambda = (s_n[\hat{\mathbb{M}}])^{-1}$ (Op Λ), iii) minimizes the stability factor $\|Q\| = (s_n[\mathbb{M}_r])^{-1}$ (Op $\|Q\|$). Once the space W_m is determined, we again use LC approach and LS approach for the field reconstruction. We list the variations of the two kinds of stability factor for a fixed reduced dimension $n = 10, 20$ and sensor width $r = 0.001, 0.005$ in Figure 2.2.5 – 2.2.8 respectively. We can draw several conclusions from the numerical results as follows.

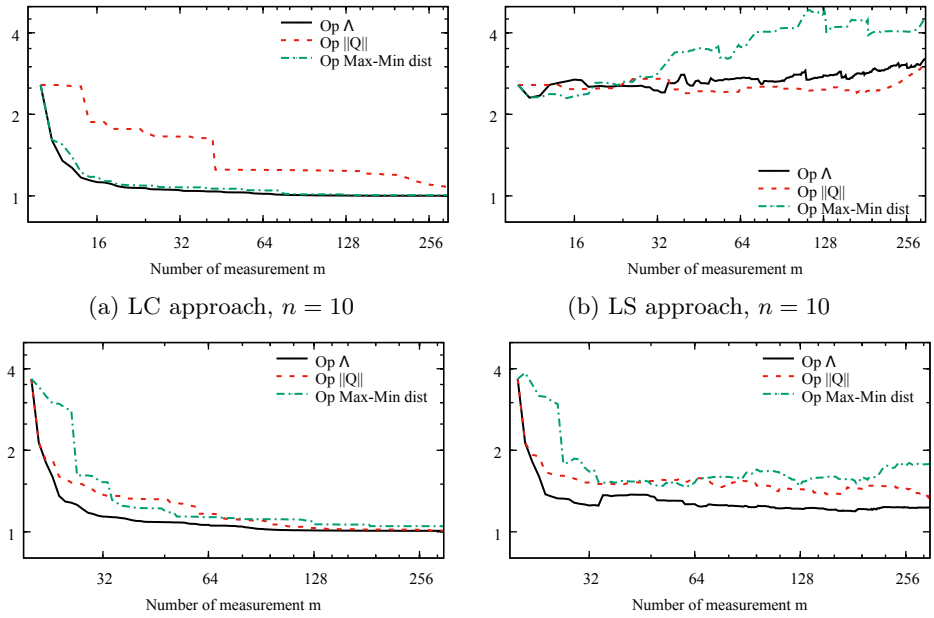
i) If the LC approach is selected for the recovery problem, the *Op* Λ *method* provides the optimal observation space W_m with respect to the Lebesgue constant Λ. On the contrary, *Op* Λ *method*

is unable to stabilize $\|Q\|$, especially when $r = 0.005$, where the observation functions are almost linear dependent.

ii) If the LS approach is selected for the recovery problem, the *Op* $\|Q\|$ *method* provides the optimal observation space W_m with respect to the stability factor $\|Q\|$. Furthermore, the *Op* Λ *method* doesn't explode the Lebesgue constant, i.e., there is not so much difference on Λ between the *Op* $\|Q\|$ *method* and the *Op* Λ *method*.

iii) Both the *Op* Λ *method* and the *Op* $\|Q\|$ *method* are able to catch the most interesting observations in the very beginning thanks to the greedy algorithm. These properties are verified by the rapid convergence of Λ for LC approach in Figure 2.2.5a, 2.2.5c, 2.2.6a, 2.2.6c, and the rapid convergence of $\|Q\|$ for LS approach in Figure 2.2.7b, 2.2.7d, 2.2.8b and 2.2.8d.

We observe that for LS approach, Figure 2.2.7b, 2.2.7d, 2.2.8b and 2.2.8d show that the convergences of $\|Q\|$ from *Op* $\|Q\|$ *method* behave as $m^{-\frac{1}{2}}$ when V_n is fixed. That is to say, we can expect the recovery error from noise observations to decrease with m and hence, converges to 0 in the limit of $m \to \infty$, but the convergence of the error with m is rather slow. We also note that in the beginning, the convergence of $\|Q\|$ doesn't follow $m^{-\frac{1}{2}}$, this is due to the greedy algorithm 2.2.5 providing the most important observations in the beginning.

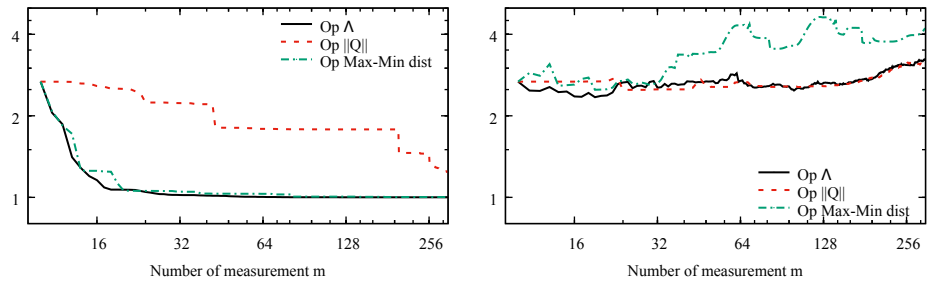


(a) LC approach, $n = 10$ (b) LS approach, $n = 10$

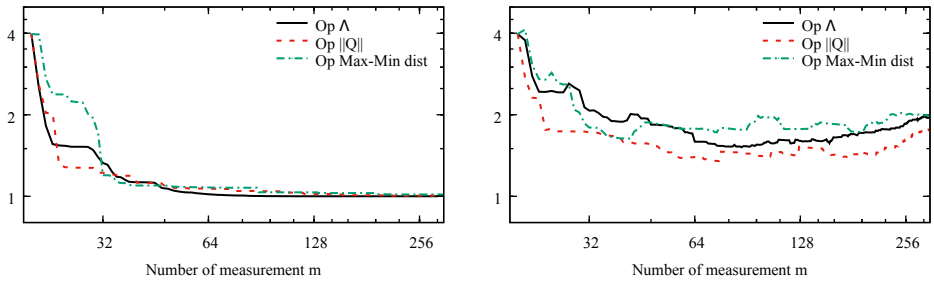
(c) LC approach, $n = 20$ (d) LS approach, $n = 20$

Figure 2.2.5: Variations of the Lebesgue constant Λ with respect to the number of observations m for LC approach and LS approach with sensor width $r = 0.001$, for three different optimization methods and two values of reduced dimension n.

We also note that, the two methods (Op $\|Q\|$ and Op Λ) — in practice mainly based on a greedy procedure — are however not robust enough with respect to noisy measurements for the reduced basis recovery problem 2.1.1, even though with a carefully selected observation space W_m, one is able to get a rather small Lebegue constant $\Lambda \to 1$ (e.g. see 'Op Λ' in Figure 2.2.5a, 2.2.5c, 2.2.6a, 2.2.6c), but the stability factor $\|Q\|$ is still relatively large, especially when the dimension of the reduced space V_n is relatively high (e.g. see Figure 2.2.3). Our question is: is there any

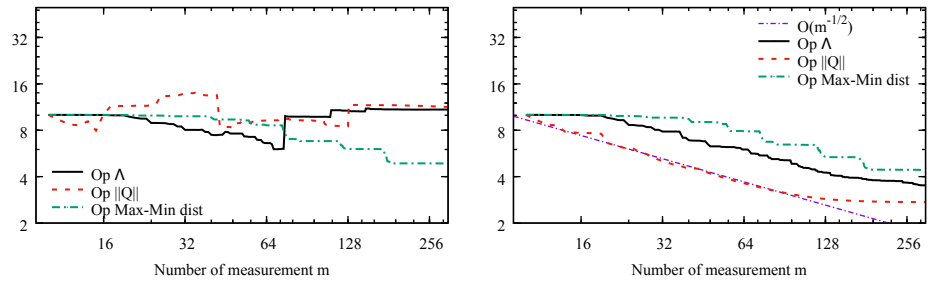


(a) LC approach, $n = 10$ (b) LS approach, $n = 10$

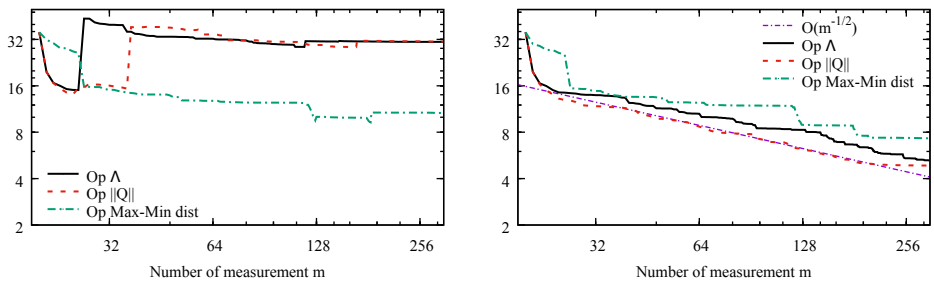


(c) LC approach, $n = 20$ (d) LS approach, $n = 20$

Figure 2.2.6: Variations of the Lebesgue constant Λ with respect to the number of observations m for LC approach and LS approach with sensor width $r = 0.005$, for three different optimization methods and two values of reduced dimension n.



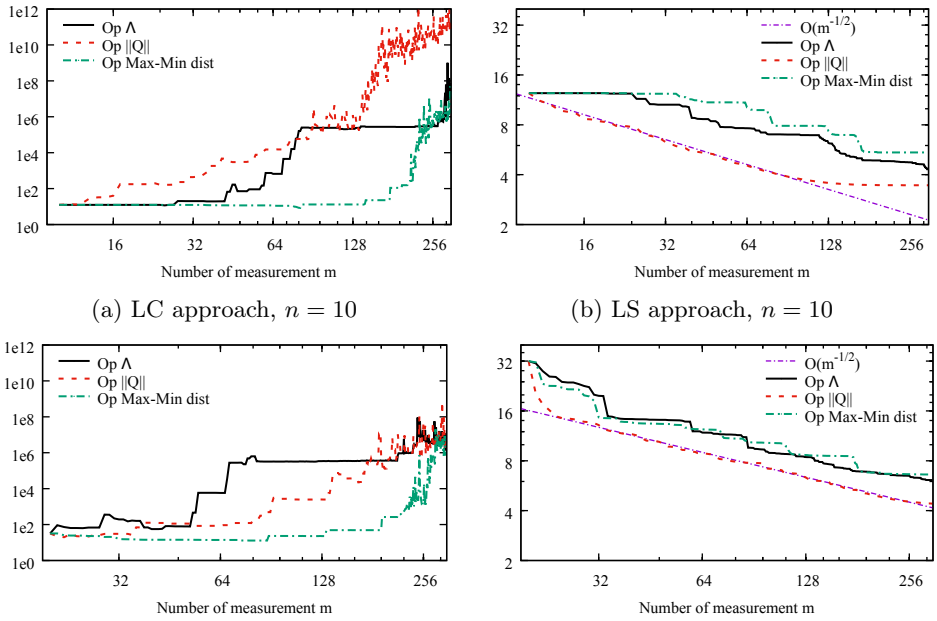
(a) LC approach, $n = 10$ (b) LS approach, $n = 10$



(c) LC approach, $n = 20$ (d) LS approach, $n = 20$

Figure 2.2.7: Variations of the stability factor $\|Q\|$ with respect to the number of observations m for LC approach and LS approach with sensor width $r = 0.001$, for three different optimization methods and two values of reduced dimension n.

algorithm which is able to further improve the stability with respect to measurement noise? This will be addressed in the following sections.



(a) LC approach, $n = 10$

(b) LS approach, $n = 10$

(c) LC approach, $n = 20$

(d) LS approach, $n = 20$

Figure 2.2.8: Variations of the stability factor $\|Q\|$ with respect to the number of observations m for LC approach and LS approach with sensor width $r = 0.005$, for three different optimization methods and two values of reduced dimension n.

2.3 Constrained stabilized recovery with reduced basis

As already stated in the previous sections, for those recovery methods with reduced basis, the stability with respect noisy observations is more sensitive than the stability with respect to model mismatch. By designing a 'high resolution' observation system and also increasing the amount of observations, both the stability with respect to model mismatch and the stability with respect noisy observations can be improved, but only up to some extent. Note that the stability analysis is based on the general assumption that there is no constraint on the field u to be reconstructed, i.e., $u \in V$. In practice, for our interest is the fields in the solution manifold $\mathcal{M} = \{u(\mu) : \mu \in \mathcal{D}\}$, which may be rather regular. Another way to stabilize the procedure is that, for the underlying problem, if the Kolmogorov n-width of the manifold $\mathcal{M}$ decays fast with n, it is possible to improve the stability by adding some constraints on the reduced basis, which is equivalent to add some physical constraints on the underlying problem.

2.3.1 Constraints on the reduced space

Let us assume that $\mathcal{S} = \overline{\text{span}(\mathcal{M})}$ (where the $\overline{X}$ denotes the closure in V of the set X) admits a basis $\{q_i\}_i$, i.e., for every $u \in \mathcal{S}$ there exists a unique sequence $\{\alpha_i(u)\}$ of scalars such that $\lim_{n\to\infty} \|u - \sum_{i=1}^{n} \alpha_i(u) q_i\| = 0$. For every $n \geq 1$, we define the n-dimensional subspace $V_n := \text{span}\{q_1, \ldots, q_n\}$. Let us formulate in a different manner the hypothesis made involving the Kolmogorov n-width of $\mathcal{M}$ in V: let us assume that the error in approximating the functions of $\mathcal{M}$

in V_n is

$$\max_{u\in\mathcal{M}} \operatorname{dist}(u, V_n) \leq \varepsilon_n, \tag{2.3.1}$$

where the sequence $(\varepsilon_n)_n$ decays fast with n. Note here that ε_n may not be the Kolmogorv n-width, as the best optimal spaces V_n for the Kolmogorov n-width may not be hierarchical (i.e. $V_n \subset V_{n+1}$). The greedy procedure however, allows to construct such a hierarchical series of spaces and is only *sub-obtimal* (see [30]).

Let us now focus on a class of linear operator $\mathcal{J}_n : V \to V_n$. By construction, for any $n \geq 1$ and any $u \in V$, $\mathcal{J}_n(u) = \mathcal{J}_{n-1}(u) + \alpha_n(u)q_n$, here we set $\mathcal{J}_0 = 0$ for notational coherence.

When we use the interpolation operator $\mathcal{J}_n$, which is from GEIM to approximate the functions of $\mathcal{M}$, yields the error bound

$$\max_{u\in\mathcal{M}} \|u - \mathcal{J}_n(u)\| \leq (1 + \Lambda_n)\,\varepsilon_n, \tag{2.3.2}$$

where

$$\Lambda_n \coloneqq \sup_{u\in V} \|\mathcal{J}_n(u)\|/\|u\| \tag{2.3.3}$$

is the Lebesgue constant. Note that [29, 145] narrow the error bound of inequation (2.3.2) to $\Lambda_n\varepsilon_n$. The value of Λ_n diverges at a certain rate so the behavior of $\max_{u\in\mathcal{M}} \|u - \mathcal{J}_n(u)\|$ with the dimension is dictated by the trade-off between the rate of divergence of (Λ_n) (that is generally slow) and the convergence of (ε_n) (that is generally very fast, one can refer to [146] for further analysis of the convergence rate). From the construction of $\mathcal{J}_n$, we have the following proposition on the coefficients.

Proposition 2.3.1 *If the basis $(q_i)_i$ are from GEIM greedy algorithm, and $(\mathcal{J}_i)_i$ are GEIM interpolants with functionals $\{\ell_i\}$ in the dual space V' of V of unity norm. Then, for any $u \in \mathcal{M}$,*

$$|\alpha_n(u)| \leq \tau(1 + \Lambda_{n-1})\varepsilon_{n-1}, \quad n \geq 1 \tag{2.3.4}$$

where $\tau \geq 1$ is a constant because of the weak greedy algorithm of GEIM *(see Algorithm 1.4.3 in Section 1.4), and for notational coherence we set $\Lambda_0 = 0$ and $\varepsilon_0 = \max_{u\in\mathcal{M}} \|u\|$.*

Proof: From the construction of GEIM (see Algorithm 1.3.2), we have the snapshots

$$\mu_n = \arg\max_{u\in\mathcal{M}} \ \|u - \mathcal{J}_{n-1}(u)\|$$

and the observation functionals

$$\ell_n = \arg\max_{\ell\in\Sigma} |\ell\,(u_n - \mathcal{J}_{n-1}(u_n))\,|$$

and the basis

$$q_n = (u_n - \mathcal{J}_{n-1}(u_n))/\ell_n\,(u_n - \mathcal{J}_{n-1}(u_n))\,.$$

We then have for every $n \geq 1$ and every $1 \leq j \leq n$, $\forall i,\ 1 \leq i \leq j \quad \ell_i(q_j) = \delta_{i,j}$. Thus we have for all $u \in \mathcal{M}$, we can write $\mathcal{J}_n(u) = \sum_{i=1}^{n} \alpha_i(u)q_i$ with $\mathcal{J}_n(u) = \mathcal{J}_{n-1}(u) + \alpha_n(u)q_n$. From the relation $\ell_n(\mathcal{J}_n(u)) = \ell_n(u)$ we further have $\alpha_n(u) = \ell_n(u - \mathcal{J}_{n-1}(u))$. Then we recall the unity norm of ℓ_n and immediately have $|\alpha_n(u)| = |\ell_n(u - \mathcal{J}_{n-1}(u))| \leq \|\ell_n\|_{V'}\|u - \mathcal{J}_{n-1}(u)\| = \|u - \mathcal{J}_{n-1}(u)\| \leq (1 + \Lambda_{n-1})\varepsilon_{n-1}$. In practice, the above *arg max* over $\mathcal{M}$ (and Σ) is replaced with an *arg max* over a very fine sample $\mathcal{M}^{(training)}$ (and $\Sigma^{(training)}$) (see Algorithm 1.4.3), thus we introduce a constant $\tau \geq 1$ to make the relation $|\alpha_n(u)| \leq \tau(1 + \Lambda_{n-1})\varepsilon_{n-1}$ holds for all $u \in \mathcal{M}$. □

Figure 2.3.9 shows the variations of $\alpha_n^0 = \max_{u\in\mathcal{M}} |\alpha_n(u)|$ for two different sensor width ($r = 0.001, 0.005$) for Test case 2.2.1, which confirms the statement in Proposition 2.3.1.

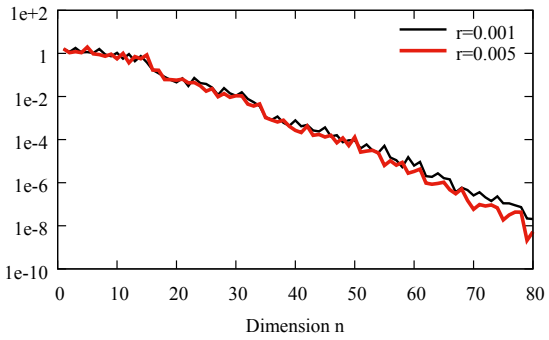


Figure 2.3.9: The decay of the coefficients α_n^0 with respect to the dimension n of the reduced space for Test case 2.2.1. The reduced basis are derived from GEIM greedy algorithm with sensor width $r = 0.001, 0.005$ respectively.

Proposition 2.3.2 *If the linear operator $\mathcal{J}_n$ is an orthogonal projector i.e. P_{V_n}, and $(q_i)_i$ are the basis resulting from an orthogonalization process with the standard greedy algorithm for the reduced basis method [48, 185], if the Kolmogorov n-width $d_i(\mathcal{S}; V) \leq ce^{-\lambda i}$ with $\lambda > log\ 2$, there exists $\beta > 0$ such that $|\alpha_i(u)| \leq Ce^{-\beta(i-1)}$.*

Proof: We recall the greedy algorithm (or the algorithm from [48, Algorithm 1]) for the choice of the elements that constitute the reduced basis is defined as follows [48, 185]:

i. $u_1 := \arg\max_{u\in\mathcal{S}} \|u\|$ and $q_1 = \frac{u_1}{\|u_1\|}$,

ii. assume $u_1, \cdots, u_{i-1}$ and $q_1, \cdots, q_{i-1}$ are defined, consider $V_{i-1} := \text{span}\{q_1, \cdots, q_{i-1}\}$,

iii. $u_i := \arg\max_{u\in\mathcal{S}} \|u - P_{V_{i-1}} u\|$ and $q_i = \frac{u_i - P_{V_{i-1}} u_i}{\|u_i - P_{V_{i-1}} u_i\|}$,

where $P_{V_{i-1}}$ denotes the orthogonal projection on V_{i-1} for the scalar product in V. We recall the following result from [48, Theorem 2.1, Theorem 3.1]: if the Kolmogorov n-width $d_i(\mathcal{S}; V) \leq ce^{-\lambda i}$ with $\lambda > \log 2$, there exists $\beta > 0$ such that

$$\forall u \in \mathcal{S},\ \|u - P_{V_i} u\| \leq Ce^{-\beta i}\ . \tag{2.3.5}$$

By recognizing $V_{i-1} \subset V_i$ and $q_i \perp V_{i-1}$, we can denote u by $u = P_{V_{i-1}} u + \alpha_i(u) q_i + P_{V_i^\perp} u$. Thus we have

$$\begin{aligned} Ce^{-\beta(i-1)} &\geq \|u - P_{V_{i-1}} u\| \\ &= \|P_{V_{i-1}} u + \alpha_i(u) q_i + P_{V_i^\perp} u - P_{V_{i-1}} u\| \\ &= \|\alpha_i(u) q_i + P_{V_i^\perp} u\| \\ &= \|\alpha_i(u) q_i\| + \|P_{V_i^\perp} u\| \\ &\geq \|\alpha_i(u) q_i\|\ . \end{aligned} \tag{2.3.6}$$

Thus we have $\|\alpha_i(u) q_i\| \leq Ce^{-\beta(i-1)}$. From the orthogonalization of q_i we deduce that $|\alpha_i(u)| \leq Ce^{-\beta(i-1)}$.

□

Proposition 2.3.1 and Proposition 2.3.2 imply that, if the Kolmogorov n-width of the manifold $\mathcal{M}$ decays fast with n, for any $u \in \mathcal{M}$, there exists an approximation $\mathcal{J}_n[u] = \sum_{i=1}^n \alpha_i(u) q_i$ of u

such that i) $\|\mathcal{J}_n[u] - u\| \leq \epsilon_n$, ii) $|\alpha_i(u)| \leq \mathcal{O}(\epsilon_{i-1})$, where $\epsilon_i \to 0$ with $i \to \infty$. That is to say, any function of the manifold can be approximated by an element from a 'long and narrow polyhedron' associated with some particular basis. Note that this statement can be easily extended to POD [128, 27, 172] case.

2.3.2 Constrained stabilized GEIM

Formulation

For a given series of linearly independent vectors $(q_i)_i \in V$ and a scalar $\tau > 0$, and also the series $\{\alpha_i^0\}_i$ being given, let us define a *polyhedron* in V corresponding to the basis $(q_i)_i$:

$$\mathcal{K}_n(\tau, \alpha^0) \coloneqq \{v \in V_n \ : \ v = \sum_{i=1}^{n} \alpha_i q_i \ \ |\alpha_i| \leq \tau \alpha_i^0\}. \tag{2.3.7}$$

We have for any $n \geq 1$ and any $u^t \in \mathcal{M}$, $\mathcal{J}_n(u^t) \in \mathcal{K}_n(1, \alpha^0)$, where α_i^0 represent the upper bound.

In presence of noisy measurements $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o$, reconstruction from these values yields an element in V_n denoted as $\mathcal{J}_n(u^t;\ \boldsymbol{e}^o)$. With noisy measurements, we propose to correct the recovery operator by using more the structure of the manifold $\mathcal{M}$ that, at the discrete level, is expressed in the fact that the approximation should belong to $\mathcal{K}_n(1, \alpha^0)$. Indeed, the belonging of $\mathcal{J}_n(u^t;\ \boldsymbol{e}^o)$ to $\mathcal{K}_n(1, \alpha^0)$ is not satisfied any more except if there exists $\tilde{u}$ in $\mathcal{M}$ such that $\ell_i(\tilde{u}) = y_i^o$ for any $i, 1 \leq i \leq m$ (which is rarely the case). In addition, in order to minimize the effect of the noise, we can increase the number of measurements and use m larger than n linear functional evaluations at a given dimension n. This leads to propose a least-squares projection onto $\mathcal{K}_n(\tau, \alpha^0)$ for $\tau \geq 1$.

Definition 2.3.1 (Constrained Stabilized GEIM) *If there exist a (generalized) projection operator $\mathcal{J}_n : V \to V_n$, and an associated series $\{\alpha_i^0\}_i$ such that for any $n \geq 1$ and $u^t \in \mathcal{M}$, $\mathcal{J}_n(u^t) = \mathcal{J}_{n-1}(u^t) + \alpha_n(u^t) q_n$, with $|\alpha_i| \leq \tau \alpha_i^0$. We now collect the values $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o$. There exists an operator $A_{m,n}$ that, to any $\boldsymbol{y}^0$ associates an approximation of u^t defined as follows*

$$A_{m,n}(u^t) = \underset{v \in \mathcal{K}_n(\tau, \alpha^0)}{\arg\min} \sum_{i=1}^{m} \left(\ell_i(v) - y_i^o\right)^2 = \underset{v \in \mathcal{K}_n(\tau, \alpha^0)}{\arg\min} \ \|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_2^2 \tag{2.3.8}$$

where $\mathcal{K}_n(\tau, \alpha^0)$ is a polyhedron defined in (2.3.7).

We consider the algebraic form of the CS-GEIM statement. We recall the notations of $\mathcal{B}, \mathcal{W}$ and $\mathbb{M} = \mathcal{W}^T \mathcal{B}$ introduced in the previous sections. The algebraic form of (2.3.8) is: find $\boldsymbol{\alpha}^* \in \mathbb{R}^n$ such that

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha} \in \mathbb{R}^n}{\arg\min} \|\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o\|_2^2 \quad \text{subject to: } |\alpha_i| \leq \tau \alpha_i^0, \quad i = 1, \cdots, n \tag{2.3.9}$$

For the above definition, we have several remarks:

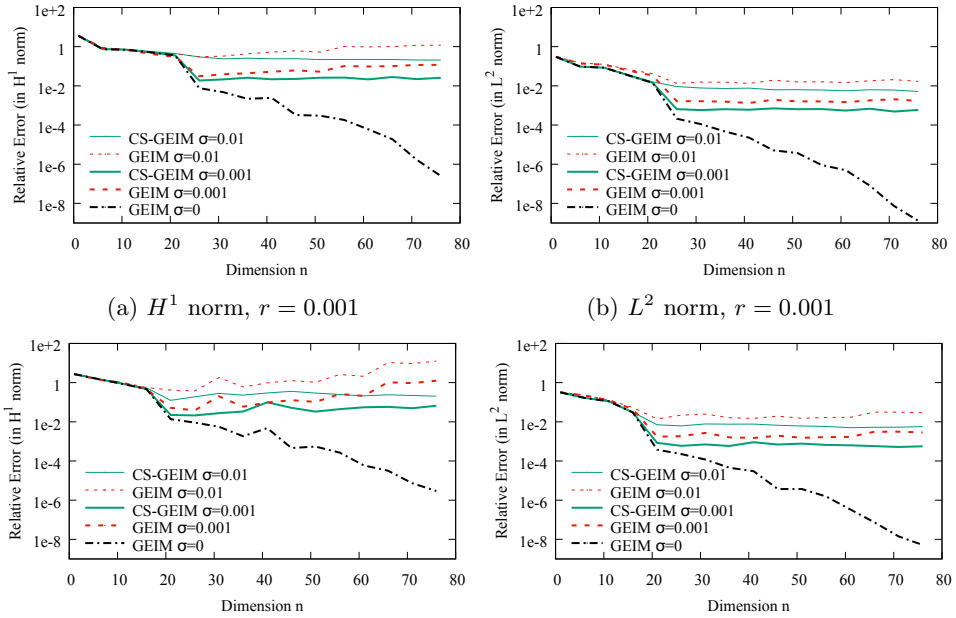
i) As already stated, the LS approach (2.3.9) provides the smallest possible operator norm among all perfect recovery operators with respect to noise. Here we choose this operator instead of (2.2.7) which provides the optimal Lebesgue constant with respect to model bias.

ii) The operator $A_{m,n}$ from (2.3.8) is not a linear operator any more. Indeed it is a constrained quadratic programming (QP [96]) problem aims at finding a function from the polyhedron $\mathcal{K}_n(\tau, \alpha^0)$ which makes the distance between the real and simulated observations minimal.

iii) In absence of noise or with $\tau = \infty$, when $m = n$, the operator $A_{m,n}$ from (2.3.8) boils down to GEIM.

Numerical results

The comparison between GEIM and CS-GEIM based on Test case 2.2.1 is shown in Figure 2.3.10. Assume the measurements suffer from Gaussian noise with standard deviations $\sigma = 0.01$ or 0.001. Here we work on $(\cdot,\cdot)_{H^1(\Omega)}$ inner product and induced norm when selecting the basis with GEIM greedy algorithm. There are two kinds of measurement functionals with sensor width $r = 0.001$ and 0.005. Numerical results show that, the noise is well controlled with coefficients constraints, especially in H^1 norm. Note that with the increasing of reduced dimension n, GEIM amplifies the noise but CS-GEIM provides a rather robust stability performance with respect to noise.



(a) H^1 norm, $r = 0.001$ (b) L^2 norm, $r = 0.001$

(c) H^1 norm, $r = 0.005$ (d) L^2 norm, $r = 0.005$

Figure 2.3.10: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for GEIM and CS-GEIM, for two different sensor width r and two standard deviations σ of Gaussian noise.

Note that there is no explicit solution to the constrained QP problem (2.3.9), one can e.g. use the *fmincon* solver from Optimization toolbox of MATLAB [67] to solve this problem numerically. *fmincon* finds the minimum of a problem specified by

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \text{ such that } \begin{cases} c(\boldsymbol{x}) \leq \mathbf{0} \\ ceq(\boldsymbol{x}) = \mathbf{0} \\ A \cdot \boldsymbol{x} \leq \mathbf{b} \\ Aeq \cdot \boldsymbol{x} = \mathbf{beq} \\ \mathbf{lb} \leq \boldsymbol{x} \leq \mathbf{ub} \end{cases} \tag{2.3.10}$$

where $\mathbf{b}$ and $\mathbf{beq}$ are vectors, A and Aeq are matrices, $c(\boldsymbol{x})$ and $ceq(\boldsymbol{x})$ are functions that return vectors, and $f(\boldsymbol{x})$ is a function that returns a scalar. $f(\boldsymbol{x})$, $c(\boldsymbol{x})$, and $ceq(\boldsymbol{x})$ can be nonlinear

functions, $\boldsymbol{x}$, **lb**, and **ub** can be vectors or matrices.

We also note that even with *fmincon* the iteration is need when solving (2.3.9), our advantage here is that, we work with relative small matrices thanks to the RB method. Later we will show that the regularized stabilized approach which provides simpler and closed form of (2.3.9), avoiding to solve this constrained QP problem. With the regularized approximation, one can directly obtain the algebraic solution for the underlying recovery problem.

2.4 Constrained stabilized PBDW

The CS-GEIM looks for a solution in the reduced space V_n, and doesn't take into account the update part $\eta \in W_m$, which belongs to the observation spaces W_m. Now we turn to the analysis of PBDW approach, which looks for a solution in a wilder space, i.e., $V_n + W_m$ for the underlying recovery problem.

2.4.1 General constraint on the reduced space

Formulation

Recall that in noisy case, [202] proposed an adapted PBDW (APBDW) approach based on (1.3.43) which reads:

$$u^* = \underset{\eta\in V, v\in V_n}{\arg\min}\ \xi_m\|\eta\|^2 + \|\boldsymbol{\ell}(\eta + v) - \boldsymbol{y}^o\|_2^2, \tag{2.4.1}$$

where $\xi_m > 0$ is a regularization coefficient that regulates the relative importance of the mismatch from the reduced space V_n. Theory and numerical results in the previous section both confirm that the least-squares term $\|\boldsymbol{\ell}(\eta + v) - \boldsymbol{y}^o\|_2^2$ provides a better stability factor with respect to noise.

Our approach is to add a constraint on v^*, which is similar to CS-GEIM to control the noise. Now we state the constrained stabilized PBDW (CS-PBDW) minimization statement with least-squares approach.

Proposition 2.4.1 (Constrained stabilized PBDW) *The constrained stabilized PBDW (CS-PBDW) minimization statement with least-squares approach is: find* $(u^* = v^* + \eta^* \in V, v^* \in V_n, \eta^* \in W_m)$ *such that*

$$(v^*, \eta^*) = \underset{\eta\in V, v\in\mathcal{K}_n(\tau,\alpha^0)}{\arg\min}\ \xi_m\|\eta\|^2 + \|\boldsymbol{\ell}(v + \eta) - \boldsymbol{y}^o\|_2^2, \tag{2.4.2}$$

where ξ_m is a trade-off factor between the complementary part $\|\eta\|^2$ and the observation misfit.

We first show that the constraint $\eta \in V$ in (2.4.2) can be weakened to $\eta \in W_m$ like in the plain PBDW method. We have the following remark.

Remark 2.4.1 *Let $u^* = v^* + \eta^* \in V$ be an optimal solution to* (2.4.2)*. Then we have $v^* \in V_n$ and $\eta^* \in W_m$.*

Proof: Let $\eta^* = P_{W_m}\eta^* + P_{W_m^\perp}\eta^*$ be the solution to (2.4.2), and recall that $\ell_i(u^*) = (u^*, w_i)$, thus we have $\ell_i(u^*) = (v^* + P_{W_m}\eta^*, w_i)$, then we deduce that

$$\xi_m\|\eta^*\|^2 + \|\boldsymbol{\ell}(u^*) - \boldsymbol{y}^o\|_2^2 = \xi_m\|P_{W_m}\eta^*\|^2 + \xi_m\|P_{W_m^\perp}\eta^*\|^2 + \|\boldsymbol{\ell}(v^* + P_{W_m}\eta^*) - \boldsymbol{y}^o\|_2^2.$$

Indeed η^* is supposed to be the minimum, for $P_{W_m}\eta^* \in V$ we should have

$$\xi_m\|\eta^*\|^2 + \|\boldsymbol{\ell}(u^*) - \boldsymbol{y}^o\|_2^2 \leq \xi_m\|P_{W_m}\eta^*\|^2 + \|\boldsymbol{\ell}(v^* + P_{W_m}\eta^*) - \boldsymbol{y}^o\|_2^2,$$

which means that $P_{W_m^\perp}\eta^* = 0$, and we get the conclusion.

Now we consider the algebraic form of (2.4.2). We recall the notations of $\mathbb{W} = \mathcal{W}^T\mathcal{W}$ and $\mathbb{M} = \mathcal{W}^T\mathcal{B}$ introduced in the previous sections and further denote $\mathbb{A}^{m\times(m+n)} = [\mathbb{W}, \mathbb{M}]$, $\mathbb{K}^{(m+n)\times(m+n)} = \begin{bmatrix}\mathbb{W} & 0\\ 0 & 0\end{bmatrix}$. Then the algebraic form of (2.4.2) is: find $\boldsymbol{\theta}^* = (\beta_1^*, \cdots, \beta_m^*,\ \alpha_1^*, \cdots, \alpha_n^*)^T \in \mathbb{R}^{m+n}$ (i.e. $u^* = \sum_{i=1}^m \beta_i^* w_i + \sum_{j=1}^n \alpha_j^* q_j \in V_n + W_m$) such that:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}\in\mathbb{R}^{m+n}}{\arg\min}\, \xi_m\boldsymbol{\theta}^T\mathbb{K}\boldsymbol{\theta} + (\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o)^T(\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o) \text{ s.t. } |\theta_i| \leq \theta_i^0,\quad i = 1, \cdots, m+n \tag{2.4.3}$$

where $\theta_i^0 = +\infty,\ i = 1, \cdots m; \theta_{i+m}^0 = \tau\alpha_i^0,\ i = 1, \cdots n.$

We now make several remarks:

i) The property $\eta^* \in W_m \cap V_n^\perp$ doesn't hold any more because of the constraint, i.e., we constrain the reduced space part v^*, but relax the constraint on the update part η^*;

ii) In absence of noise, the CS-PBDW approach (2.4.2) boils down to the classical PBDW approach (1.3.19);

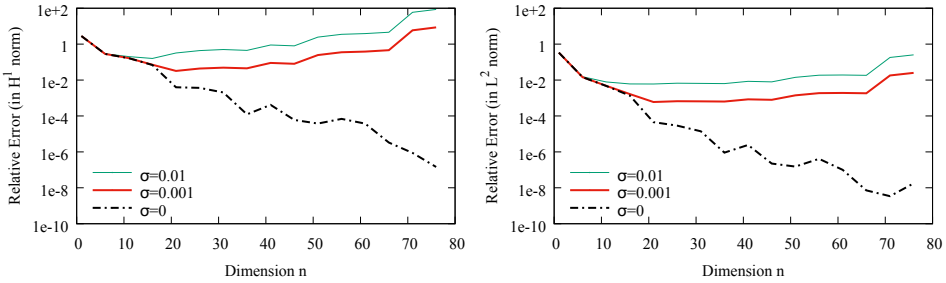
iii) In absence of model update part η^*, the CS-PBDW approach (2.4.2) boils down to the CS-GEIM approach (2.3.8).

Numerical results

We first illustrate the variations of the relative error versus the reduced dimension n for PBDW with uniformly distributed noise with noise level $\sigma = 0.01$ or 0.001 in Figure 2.4.11 based on Test case 2.2.1. The sensor width is $r = 0.005$. The number of observations $m = 2n$ with the first n observations selected with GEIM greedy algorithm and the left selected with Algorithm 2.2.5. As shown in Figure 2.4.11, even though PBDW is able to correct the model bias, it still gives no solution to address the noise. Then we present the recovery errors in Figure 2.4.12 for CS-PBDW approach (2.4.2). We observe that the noise is well controlled with the method we proposed.

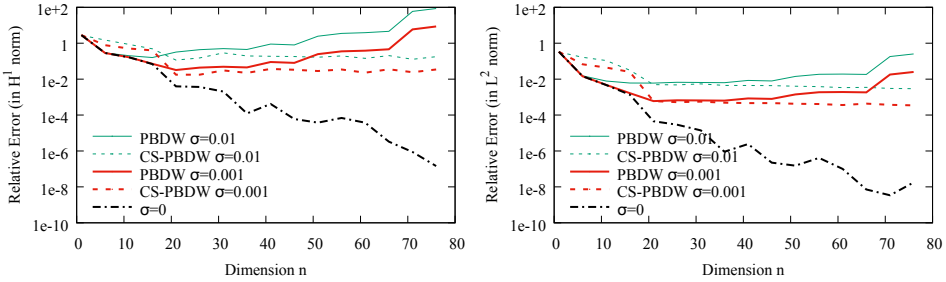
We notice here that the strange initial behavior of CS-PBDW comes from the improper factor ξ_m. Indeed, for a reduced space V_n of relatively low dimension, the error from the plain PBDW method itself (i.e. $(\eta^*, v^*) = \arg\min_{\eta\in W_m, v\in V_n}\|\eta\|$ s.t. $\|\boldsymbol{\ell}(v+\eta) - \boldsymbol{y}^o\| = 0$) is much larger than the error caused by the noise, in this case the plain PBDW method still provides an optimal solution, this corresponds to $\xi_m \to +\infty$ in (2.4.2). In Figure 2.4.11 we used a value of ξ_m which is optimal for a relatively higher reduced dimension, thus we get this strange behavior.

The CS-PBDW approach takes no priori noise information into consideration. By knowing more detail information of the noise, it's hopeful to improve the above reconstruction process, e.g., by taking the maximum a posteriori estimation (MAP) or maximum likelihood (ML) [19] to weight the least-squares problem, with the inverse of the covariance matrix $\mathbb{D}$ of the observations. We will analysis three typical noise distribution cases, i.e., uniform distribution, Gaussian distribution and Laplacian distribution, etc.



(a) H^1 norm (b) L^2 norm

Figure 2.4.11: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for PBDW with sensor width $r = 0.005$, for two different noise level σ of uniformly distributed noise.



(a) H^1 norm (b) L^2 norm

Figure 2.4.12: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for CS-PBDW and PBDW with sensor width $r = 0.005$, for two different noise level σ of uniformly distributed noise.

2.4.2 Noise with uniform distribution

Formulation

To simplify the analysis, we take the first assumption on the noise $\boldsymbol{e}^o$ that each term e_i^o is uniformly distributed on $[-e_{i,Max}, e_{i,Max}]$. In this case, from [41, page 352], if $\{e_i^o\}_{i=1}^m$ are uniformly distributed on $[-e_{i,Max}, e_{i,Max}]$, we have the corresponding probability distributions $p(e_i^o) = 1/(2e_{i,Max})$, and the maximum likelihood (ML) estimation of $u^t \in V$ is: find $u \in V$ s.t.

$$|\boldsymbol{\ell}(u) - \boldsymbol{y}^o| \leq \boldsymbol{e}_{Max} \ , \tag{2.4.4}$$

here the inequality is to vector, it represents component-wise inequality, i.e., $|\ell_i(u) - y_i| \leq e_{i,Max},\ i = 1, \cdots, m$.

Now we state the maximum likelihood constrained stabilized PBDW (MLCS-PBDW) approach with uniformly distributed noise by adding an maximum likelihood estimation constraint (2.4.4), which reads: find $(u^* = v^* + \eta^* \in V, v^* \in V_n, \eta^* \in V)$ such that

$$(v^*, \eta^*) = \underset{\substack{v \in \mathcal{K}_n(\tau, \alpha^0) \\ \eta \in V}}{\arg\min} \ \|\eta\|^2 \text{ s.t. } |(\ell_i(v^* + \eta^*) - y_i^o| \leq e_{i,Max} \ i = 1, \cdots, m \tag{2.4.5}$$

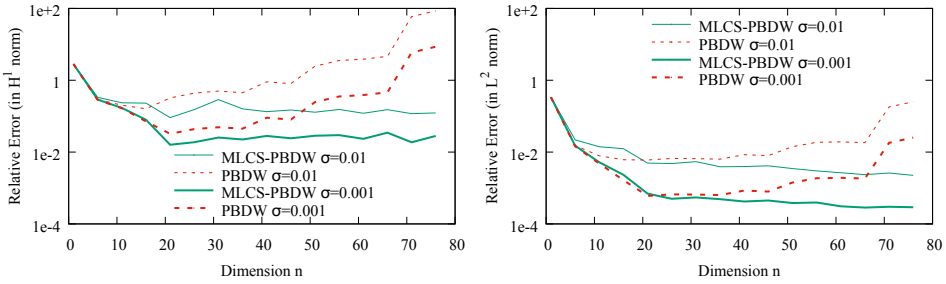
where $\{e_{i,Max}\}_{i=1}^m$ are the upper limits of the noise, and we call the latter constraint *ML constraint*. If $(e_i)_{i=1}^m$ are uniformly independent, identically distributed (IID) on $[-e_{Max}, e_{Max}]$, the ML constraint term reduces to

$$\|\boldsymbol{\ell}(v^* + \eta^*) - \boldsymbol{y}^o\|_\infty \leq e_{Max} \ . \tag{2.4.6}$$

Numerical results

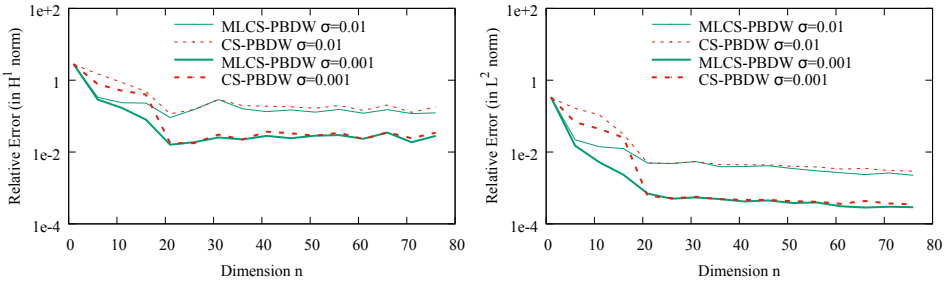
Figure 2.4.13 illustrates the relative errors from MLCS-PBDW approach and PBDW approach for different reduced dimension n, for sensor width $r = 0.005$ with uniformly distributed noise with noise level $\sigma = 0.01$ or 0.001. The number of observations $m = 2n$ with the first n observations selected with GEIM greedy algorithm and the left selected with Algorithm 2.2.5. All the calculations are based on Test case 2.2.1. It is easy to find that the noise is well controlled both in L^2 norm and in H^1 norm.

Furthermore, we illustrate the comparison between general CS-PBDW approach (2.4.2) and MLCS-PBDW approach (2.4.5) in Figure 2.4.14, which confirms that in uniformly distributed noisy case, the MLCS-PBDW approach is better than the general CS-PBDW approach, especially when the noise is smaller than the best estimate error (or model bias) with out noise, i.e., the cases when $n < 20$ in Figure 2.4.14.



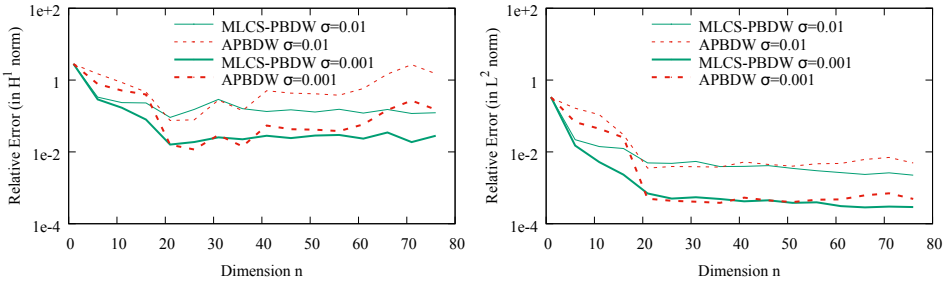
(a) H^1 norm (b) L^2 norm

Figure 2.4.13: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for MLCS-PBDW and PBDW with sensor width $r = 0.005$, for two different noise level σ of uniformly distributed noise.



(a) H^1 norm (b) L^2 norm

Figure 2.4.14: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for general CS-PBDW approach (2.4.2) and MLCS-PBDW approach (2.4.5) with sensor width $r = 0.005$, for two different noise level σ of uniformly distributed noise.



(a) H^1 norm (b) L^2 norm

Figure 2.4.15: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for MLCS-PBDW approach (2.4.5) and APBDW approach (2.4.1) with sensor width $r = 0.005$, for two different noise level σ of uniformly distributed noise.

We also illustrate the recovery errors with APBDW approach (2.4.1) in Figure 2.4.15. From this figure we observe that i) APBDW is still unable to balance the noise and model bias when the noise is relatively small with respect to the model bias, ii) APBDW is unable to stabilize the noise when the reduced space dimension n is relatively large, the cases when PBDW/APBDW are able to provide very high accuracy. On the contrary, MLCS-PBDW stabilizes the noise in all cases.

2.4.3 Noise with Gaussian distribution

Formulation

We then analyze the Gaussian noise. Assume each e_i is Gaussian noise with zero mean and variance σ_i^2, the density is $p_i(z) = (2\pi\sigma_i^2)^{-1/2}e^{-z^2/2\sigma_i^2}$, and the maximum likelihood function [41] of u that provides the noisy observations $\boldsymbol{y}^o$ is

$$l(u) = -\sum_{i=1}^{m}\frac{1}{2}\log(2\pi\sigma_i^2) - \sum_{i=1}^{m}\frac{1}{2\sigma_i^2}(\ell_i(u) - y_i^o)^2. \tag{2.4.7}$$

Therefore the ML estimation of u^t is equivalent to solve

$$(u)_{ML} = \arg\min_{u\in V}\sum_{i=1}^{m}\frac{1}{\sigma_i^2}(\ell_i(u) - y_i^o)^2. \tag{2.4.8}$$

Now we state the maximum likelihood constrained stabilized PBDW (MLCS-PBDW) approach with Gaussian noise, which reads: find $(u^* = v^* + \eta^* \in V, v^* \in V_n, \eta^* \in V)$ such that

$$(v^*, \eta^*) = \underset{\substack{v\in\mathcal{K}_n(\tau,\alpha^0)\\ \eta\in V}}{\arg\min}\ \xi_m\|\eta\|^2 + \sum_{1}^{m}\frac{1}{\sigma_i^2}(\ell_i(v+\eta) - y_i^o)^2, \tag{2.4.9}$$

where ξ_m is a trade-off factor between the complementary part $\|\eta\|^2$ and the maximum likelihood estimation part $\sum_1^m \frac{1}{\sigma_i^2}(\ell_i(v+\eta) - y_i^o)^2$, because for each term we look for minimum value.

We consider the algebraic form of MLCS-PBDW approach (2.4.9). We recall the notations of $\mathbb{W}, \mathbb{M}, \mathbb{A}, \mathbb{D}, \mathbb{K}$, then the algebraic form of (2.4.9) is: find $\boldsymbol{\theta}^* = (\beta_1^*, \cdots, \beta_m^*, \alpha_1^*, \cdots, \alpha_n^*)^T \in \mathbb{R}^{m+n}$

(or $u^* = \sum_{I=1}^{m} \beta_i^* w_i + \sum_{j=1}^{n} \alpha_j^* q_j \in V_n + W_m$) such that:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}\in\mathbb{R}^{m+n}}{\arg\min}\, \xi_m \boldsymbol{\theta}^T \mathbb{K}\boldsymbol{\theta} + (\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o)^T \mathbb{D}^{-1} (\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o) \text{ s.t. } |\theta_i| \leq \theta_i^0, \quad i = 1, \cdots, m+n \tag{2.4.10}$$

where $\theta_i^0 = +\infty,\ i = 1, \cdots m; \theta_{i+m}^0 = \tau\alpha_i^0,\ i = 1, \cdots n.$

Remark 2.4.2 *When the noise on each observation is independent and identically distributed (IID) gaussian noise,* (2.4.10) *is equivalent to the general CS-PBDW approach* (2.4.3)*.*

Remark 2.4.3 (Noise with Laplacian distribution) *When e_i is Laplacian distributed noise, i.e., $p_i(z) = (1/2a)e^{-|z|/a}$, where $a > 0$, the ML estimation of u^t is equivalent to solve*

$$(u)_{ML} = arg\, \underset{u\in V}{min} \|\boldsymbol{\ell}(u) - \boldsymbol{y}^o\|_1 \tag{2.4.11}$$

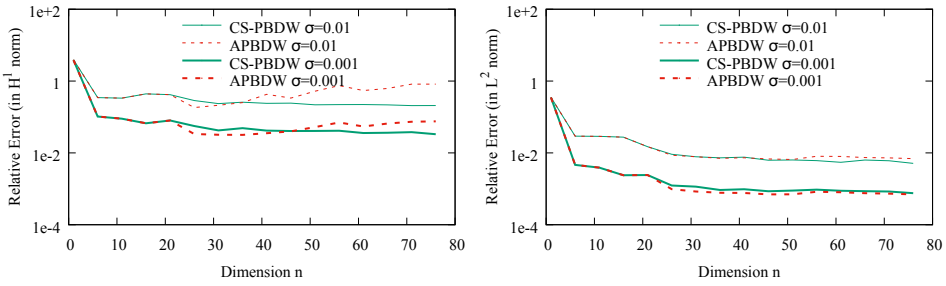
Now we state the maximum likelihood constrained stabilized PBDW (MLCS-PBDW) minimization statement with Laplacian distributed noise, which reads: find $(u^* = v^* + \eta^* \in V, v^* \in V_n, \eta^* \in V)$ such that

$$(v^*, \eta^*) = \underset{\substack{v\in\mathcal{K}_n(\tau,\alpha^0)\\ \eta\in V}}{\arg\min}\ \xi_m\|\eta\|^2 + \|\boldsymbol{\ell}(v+\eta) - \boldsymbol{y}^o\|_1 \tag{2.4.12}$$

where ξ_m is a trade-off factor between the complementary part $\|\eta\|^2$ and the maximum likelihood (ML) constraint $\|\boldsymbol{\ell}(v+\eta) - \boldsymbol{y}^o\|_1$.

Numerical results

The numerical results are based on Test case 2.2.1. We list the variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for MLCS-PBDW and APBDW approaches in Figure 2.4.16 for Gaussian noise with standard deviation $\sigma = 0.01$ or 0.001. The sensor width is $r = 0.001$. The number of observations $m = 2n$ with the first n observations selected with GEIM greedy algorithm and the left selected with Algorithm 2.2.5. The numerical results again confirm that the CS version improves the stability with respect to PBDW/APBDW.



(a) H^1 norm (b) L^2 norm

Figure 2.4.16: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for MLCS-PBDW and APBDW with sensor width $r = 0.001$, for two different standard deviations σ of Gaussian noise.

2.5 Regularized stabilized recovery

2.5.1 Formulation

The constrained stabilized recovery problem, e.g., (2.3.9) or (2.4.3) is a constrained QP problem, that aims at finding a function from a polyhedron $\mathcal{K}_n(\tau, \alpha^0)$ making the distance between the real and simulated observations minimal. Successful approaches to solving bound-constrained optimization problems for general linear or nonlinear objective functions could be found in [112, 159, 138, 160] and the Matlab function *fmincon*. Approaches which are specific to least-squares problem are described in [28, 142, 161] and the Matlab function *lsqlin*. The shortcoming of these algorithms is time-consuming because of iteration process. To meet online calculation for monitoring purpose, a simpler and closed form to solve these problems is necessary.

Here we introduce an approach whereby the bound constrained problem is written with n quadratic inequality constraints, (2.3.9) becomes

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min} \|\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o\|_2^2 \quad \text{subject to: } \alpha_i^2 \leq (\tau\alpha_i^0)^2, \quad i = 1, \cdots, n \tag{2.5.1}$$

The necessary and sufficient KKT conditions for a feasible point $\boldsymbol{\alpha}^*$ to be a solution to (2.5.1) are

$$\begin{aligned} &(\mathbb{M}^T\mathbb{M} + \boldsymbol{\lambda}^*)\boldsymbol{\alpha}^* = \mathbb{M}^T\boldsymbol{y}^o \\ &\lambda_i^* \geq 0 \quad i = 1, \cdots, n, \\ &\lambda_i^*[(\tau\alpha_i^0)^2 - \alpha_i^2] = 0 \quad i = 1, \cdots, n, \\ &(\tau\alpha_i^0)^2 - \alpha_i^2 \geq 0 \quad i = 1, \cdots, n, \end{aligned} \tag{2.5.2}$$

where $\boldsymbol{\lambda}^* := diag(\lambda_i^*)$ is a diagonal matrix. Reformulating the box constraints $|\alpha_i| \leq \tau\alpha_i^0$ as quadratic constraints $\alpha_i^2 \leq (\tau\alpha_i^0)^2, i = 1, \cdots, n$ effectively circumscribes an ellipsoid constraint around the original box constraint. In other words, we replace $\mathcal{K}_n(\tau, \alpha^0)$ with the ellipsoid one $\varepsilon_n(\tau, \alpha^0)$ where $\varepsilon_n(\tau, \alpha^0) := \{x \in \mathbb{R}^n \ : \ \sum_{i=1}^n \frac{\alpha_i^2}{(\tau\alpha_i^0)^2} \leq 1\}$. In [171] box constraints were reformulated in exactly the same manner, in which parameters were found which ensure that there is a convex combination of the objective function and the constraints. Another possibility to solve this optimization problem is with the penalty or weighted approach, similar work can be found in [156, 125].

Here, we apply the penalty or weighted approach to the quadratic, inequality constrained problem to(2.5.1), then Tikhonov regularization [11] can be viewed as a quadratically constrained least squared problem when the constraint in (2.5.1) is considered as a regularized term. In this case, a penalty term on $\boldsymbol{\alpha}$ is added to the objective function, and multiplied by a matrix that contains the bounds of the inequality constraints i.e. α_i^0. This matrix comes from the first KKT condition (2.5.2), which is the solution of the least-squares problem:

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min} \|\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o\|_2^2 + \|(\boldsymbol{\lambda}^*)^{1/2}\boldsymbol{\alpha}\|_2^2 \tag{2.5.3}$$

We view the inequality constraints as a penalty term by replacing λ^* with $\mathbb{C} = diag((\alpha_i^0)^2)$, and taking the maximum a posteriori estimation (MAP) [19] to weight the least-squares problem with the inverse of the covariance matrix $\mathbb{D}$ for the mismatch between the observation data and the simulated ones, thus we have the following closed form:

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min} \|\mathbb{D}^{-1/2}(\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o)\|_2^2 + \zeta_m\|\mathbb{C}^{-1/2}\boldsymbol{\alpha}\|_2^2 \tag{2.5.4}$$

where ζ_m is the penalty or regularization factor. In our case, the penalty factor reflects the relative

weight of each basis to the function u we want to approximate. (2.5.4) is a typical way to regularize a problem with Tikhonov regularization, the L-curve [33] approach is used to find the factor ζ_m. This is done by solving (2.5.4) multiple times with various ζ_m to get multiple solutions $\boldsymbol{\alpha}$. Once these solutions are obtained, a log-log plot of $J_1 = \|\mathbb{D}^{-1/2}(\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o)\|_2^2$ versus $J_2 = \|\mathbb{C}^{-1/2}\boldsymbol{\alpha}\|_2^2$ will typically be in the shape of an L, and the optimal value of ζ_m is the one at the corner.

We now state the regularized stabilized recovery with reduced basis.

Proposition 2.5.1 (Regularized stabilized GEIM, R-GEIM) *The regularized stabilized GEIM in algebraic form is: find $\boldsymbol{\alpha}^*$ (or $u^* = \mathcal{B}\boldsymbol{\alpha}^*$) s.t.*

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min} \quad (\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o)^T\mathbb{D}^{-1}(\mathbb{M}\boldsymbol{\alpha} - \boldsymbol{y}^o) + \zeta_m\boldsymbol{\alpha}^T\mathbb{C}^{-1}\boldsymbol{\alpha} \tag{2.5.5}$$

with solution $\boldsymbol{\alpha}^* = [\mathbb{M}^T\mathbb{D}^{-1}\mathbb{M} + \zeta_m\mathbb{C}^{-1}]^{-1}\mathbb{M}^T\mathbb{D}^{-1}\boldsymbol{y}^o$, thus $u^* = \mathcal{B}\boldsymbol{\alpha}^* = \sum_{i=1}^n \alpha_i q_i$, where ζ_m is determined with L-curve method.

Apply the same idea to the constrained stabilized PBDW problem (2.4.3), by defining a matrix $\mathbb{P}^{(m+n)\times(m+n)} = diag((\theta_i^0)^2)$, and following all the notations in (2.4.3), we have the following proposition:

Proposition 2.5.2 (Regularized stabilized PBDW, R-PBDW) *The regularized stabilized PBDW in algebraic form is: find $\boldsymbol{\theta}^* = (\beta_1^*, \cdots, \beta_m^*, \alpha_1^*, \cdots, \alpha_n^*)^T \in \mathbb{R}^{m+n}$ (or $u^* = \sum_{I=1}^m \beta_i w_i + \sum_{j=1}^n \alpha_j q_j \in V_n \times W_m$) such that:*

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}\in\mathbb{R}^{m+n}}{arg\ inf} \quad \xi_m\boldsymbol{\theta}^T\mathbb{K}\boldsymbol{\theta} + \zeta_m\boldsymbol{\theta}^T\mathbb{P}^{-1}\boldsymbol{\theta} + (\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o)^T\mathbb{D}^{-1}(\mathbb{A}\boldsymbol{\theta} - \boldsymbol{y}^o) \tag{2.5.6}$$

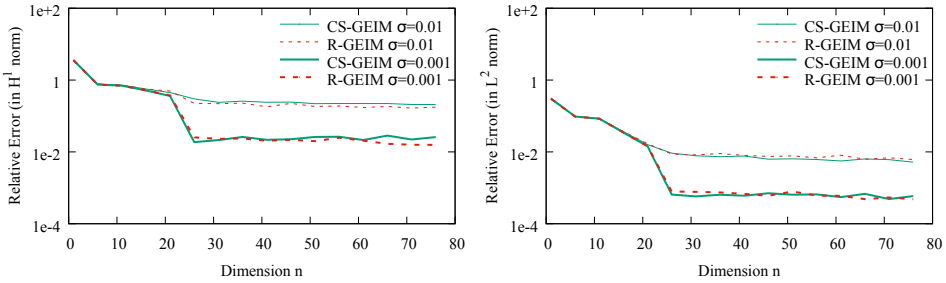
the solution is $\boldsymbol{\theta}^* = [\mathbb{A}^T\mathbb{D}^{-1}\mathbb{A} + \zeta_m\mathbb{P}^{-1} + \xi_m\mathbb{K}^{-1}]^{-1}\mathbb{A}^T\mathbb{D}^{-1}\boldsymbol{y}^o$, where ζ_m is the regularization factor for the reduced space constraint, and ξ_m is a trade-off factor for model update/bias.

2.5.2 Numerical results

We illustrate the performance of R-GEIM and R-PBDW based on Test case 2.2.1 as stated before.

We first analyze R-GEIM. Let us select three different reduced dimension cases with $n = 40, 60, 80$ and two observation systems with sensor width $r = 0.001, 0.005$ respectively. In all cases, we set the number of observations $m = 2n$. Then we solve the recovery problem with (2.5.5), for two different standard derivations $\sigma = 0.01$ or 0.001 of Gaussian noise with different regularization factor ζ_m. The L-curves of all the cases are presented in Figure 2.5.18, which shows that i) the regularization factors ζ_m are almost the same for different reduced dimension, but ii) for different noise level or sensor width, they are different.

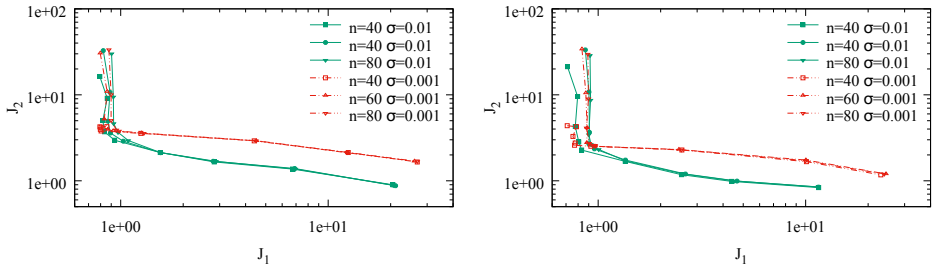
In Figure 2.5.17, we show the variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for R-GEIM, for sensor width $r = 0.001, 0.005$ with Gaussian noise with standard deviation $\sigma = 0.01$ or 0.001. For each case, the regularization factor is selected corresponding to the one at the corner of the L-curve. It is easy to see that, with a proper regularization factor, R-GEIM shows a better stability performance than CS-GEIM.



(a) H^1 norm, $r = 0.001$ (b) L^2 norm, $r = 0.001$

(c) H^1 norm, $r = 0.005$ (d) L^2 norm, $r = 0.005$

Figure 2.5.17: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for CS-GEIM and R-GEIM, for two values of the sensor width r and two standard deviations σ of Gaussian noise.



(a) $r = 0.001$ (b) $r = 0.005$

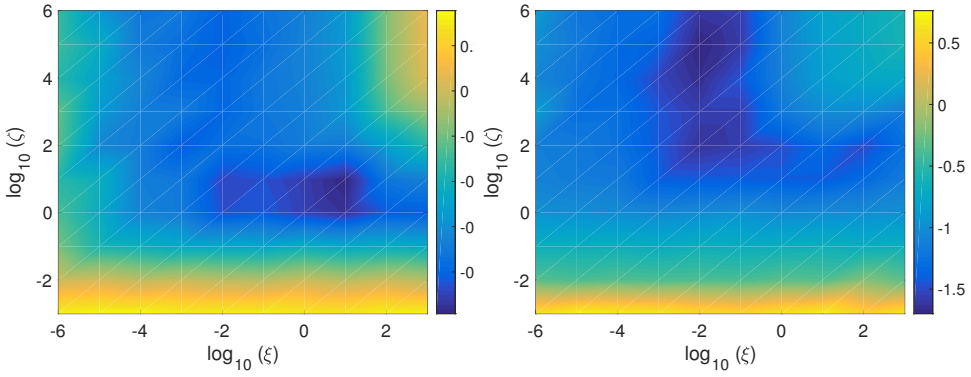
Figure 2.5.18: Variations of the L-curves for R-GEIM for Test case 2.2.1, for three different values of reduced dimension n, two different values of sensor width r and two different standard derivations σ of Gaussian noise.

Now let us turn to analyze R-PBDW. From (2.5.6), we observe that there are two regularization factors, ξ and ζ, the classical L-curve method is thus not feasible, the way to choose the proper regularization factor pair (ξ, ζ) is a little complicated. A proposed method is that, for a fixed reduced space dimension n, the number of observations m and the noise level σ, one can simulate the recovery problem multiple times with different regularization factor pairs, and choose the pair which provides the smallest recovery error. Figure 2.5.19 shows the variations of the H^1 relative errors (in *log* scale) for different regularization factor pairs (ξ, ζ) and two different standard deviations $\sigma = 0.01$ or 0.001 of Gaussian noise, for $n = 60$, $m = 120$ and $r = 0.005$.

Then we show the variations of the L^2 and H^1 relative errors for R-PBDW approach in Figure

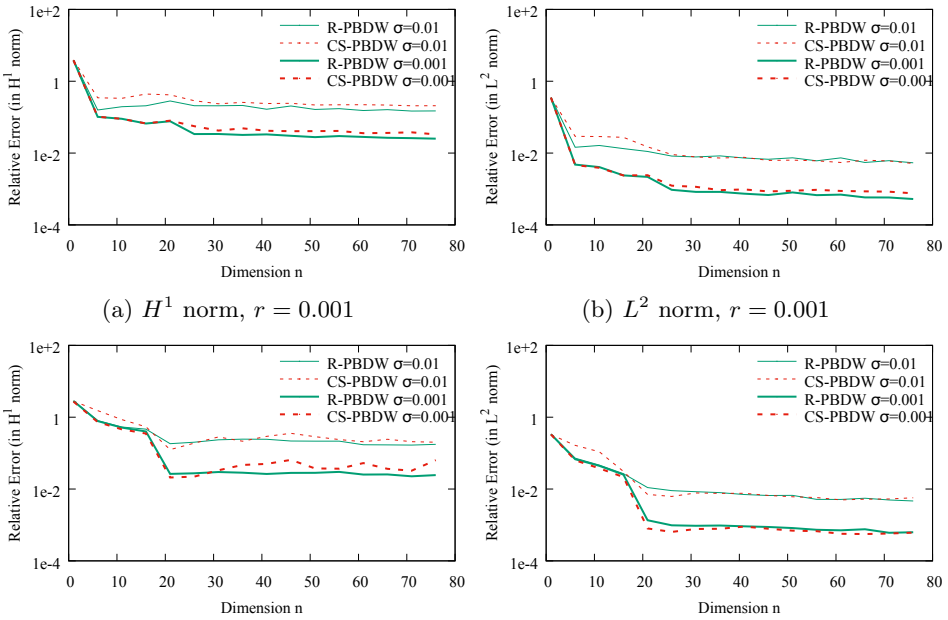
2.5.20, for two sensor width values $r = 0.001, 0.005$ and two standard deviations $\sigma = 0.01$ or 0.001. The numerical results again confirm that, R-PBDW shows almost the same stability performance with respect to CS-PBDW.

We conclude that, the linear form (2.5.5) and (2.5.6) of the regularized stabilized GEIM and PBDW provide a very high efficiency computing framework, meanwhile, keep the same accuracy with respect to the CS versions, the latter are relatively time-consuming as constrained QP problems. Thus the regularized versions are more feasible for online purpose.



(a) $\sigma = 0.01$ (b) $\sigma = 0.001$

Figure 2.5.19: Variations of the H^1 relative errors (in *log* scale) for different regularization factor pairs (ξ, ζ), for two different standard deviations σ of Gaussian noise, and $n = 60$, $m = 120$, $r = 0.005$.



(a) H^1 norm, $r = 0.001$ (b) L^2 norm, $r = 0.001$

(c) H^1 norm, $r = 0.005$ (d) L^2 norm, $r = 0.005$

Figure 2.5.20: Variations of the L^2 and H^1 relative errors with respect to the reduced dimension n for R-PBDW and CS-PBDW, for two values of the sensor width r and two standard deviations σ of Gaussian noise.

2.6 Conclusions

In this chapter, we first analyzed the stability performance for the recovery problem with reduced basis. We presented two stability factors in order to measure the stability, i) with respect to the observation noise, the stability factor $\|Q\|$ is used, ii) with respect to model mismatch, the Lebesgue constant Λ is used. With the stability analysis, one is able to i) calculate the two factors for the underlying practical problem, and evaluate the performance of the recovery methods e.g. GEIM or PBDW, etc., ii) optimize the sensor placements or the amount of sensors needed for different practical goal, particularly in controlling the model mismatch or measurement noise.

Furthermore, stability analysis also showed that, for those recovery methods with reduced basis, the stability with respect to noisy observations is more sensitive than the stability with respect to model mismatch. By designing a 'high resolution' observation system and also increasing the amount of observations, both the stability with respect to model mismatch and the stability with respect noisy observations can be improved, but only up to some extent. Another way to stabilize the procedure is, for the underlying problem, if the Kolmogorov n-width of the manifold $\mathcal{M}$ decays fast with n, it is possible to improve the stability by adding some constraints on the reduced basis, which is equivalent to add some physical constraints on the underlying problem. In this spirit, we proposed a constrained stabilized GEIM and PBDW respectively. Numerical results showed that in presence of measurement noise, CS-GEIM and CS-PBDW are able to control the noise amplification. Furthermore, by adding more measurements, the recovery errors with CS version are even lower than the noise level.

In the end of this chapter, we proposed the regularized stabilized GEIM and PBDW, i.e., R-GEIM and R-PBDW, which provide a very high efficiency computing framework, meanwhile, keep the same accuracy with respect to the CS versions, the latter are relatively time-consuming as constrained QP problems. Thus the regularized versions are more feasible for online purpose.

In the next chapter, we will show more industrial applications for these stabilized recovery methods in nuclear reactor physics domain.

Part II

Applications of the methodologies to neutronics

Chapter 3

Data assimilation framework with GEIM

Contents

In this chapter, we introduce the data assimilation framework with the generalized empirical interpolation method. We shall remind that some basics may have been explained in the previous chapters but for the sake of convenience we remind in this part so that the reading of the chapters can be done separately. First in Section 3.1, we give a general background for the application of GEIM to nuclear reactor domain. Then in Section 3.2, we present the physical context of neutronic modeling and simulation. In Section 3.3, we remind the GEIM methodology, then we adapt this methodology to neutronic field reconstruction problem by exploiting a greedy algorithm within measurements from thermal flux and reconstructing the fast/thermal flux and power distribution in Section 3.4. In Section 3.5 we present the numerical tests of the adapted GEIM based on two benchmark problems in nuclear domain, i.e., i) a 1D homogeneous slab reactor, and ii) a 2D IAEA benchmark problem. Finally, in Section 3.6 we draw a short conclusion for this chapter.

We indicate here that part of the chapter (e.g. from Section 3.3 to Section 3.5) has been presented in a published paper with J.-P. Argaud, B. Bouriquet, F. de Casoa, Y. Maday and O. Mula. Its reference in the manuscript is [14].

3.1 Introduction

Nowadays, the production of nuclear energy is done under very high safety standards where tight criteria must be satisfied both at design and operational levels. What is essentially required is the accurate knowledge of significant quantities like temperature, neutron flux, power, irradiation or fluence. The quantities can be global outputs like the maximum or average temperature or the total generated power but the knowledge of more detailed information like temperature, flux and/or power maps in the whole reactor may also be required. The knowledge of any of these quantities is accessed either through the study of parametrized models and/or through measurement data collected from the reactor itself. In very general lines, the usual way to work with models and data at design and operation levels is the following:

- At design stages, the goal is essentially to first find the most realistic model for the physics' core and then optimize the parameters of the model to find the safest core configurations. At this stage, if measured data from previous experiments are at reach, they are only used to find the model.
- At the operational level, the information is primarily obtained via sensor measurements. Their placement has to be carefully optimized in order to retrieve as much information as possible while the reactor is running.

In all these modeling and optimization steps, the experience of engineers plays a crucial role in order to find the best acceptable configurations.

Even more, due to the complexity of the physics, it is sometimes necessary to combine the expertise of engineers from different fields and the modeling/optimization process might require several iterations between experts before satisfying all the desired criteria. In this context, this chapter is a contribution to making these tasks become more *systematic*. For this, we apply the method that we have been presenting and analyzing in the previous chapters for data assimilation and model order reduction to the field of reactor simulations. For a given quantity of interest, the method provides a quick approximation of it by combining the measurement data and the knowledge of a parametrized model. The key idea is to do the reconstruction on a well chosen finite dimensional space of *reduced* dimension. The basis functions of this space are solutions to the model problem for appropriately chosen parameter values. This idea, which is the root of reduced order modeling techniques such as reduced basis, constitutes the reduced modeling part of our approach.

An overview of reduced bases can be found in the reference books [107, 24] and a summary on some relevant applications to nuclear engineering is given in Chapter 1. With respect to these works, the new main ingredient that we bring is the combination of this idea with data assimilation. This is done via the Empirical Interpolation Method (EIM, [21]) which we will use in its generalized version (GEIM, see [144, 145]). In this approach, at least the basic step, the approximation in the space of reduced dimension is defined in such a way that the measures on the approximant coincide with the measured data. Two features of this methodology which might be of interest to the community of nuclear engineering are the following:

- The information from the model and the measurements is incorporated *simultaneously* and not in a sequential manner like in the classical procedures in nuclear engineering.
- Since the approximation is done on a space of small dimension (dimension between 30 and 50 is usually enough in many cases, in the real application it is more about 50), its computation is very quick. This could be helpful to accelerate some steps at design and operation stages.

We would like to emphasize that GEIM belongs to a broader class of recovery methods which gathers other approaches like the PBDW methodology of [148] where the use of data allows not only

to reconstruct the quantities of interest, but also to correct the possible bias in the mathematical model. Here we shall not use this feature and assume that the mathematical model is perfect. The whole class is subject of current active research in the community of applied mathematics (see, e.g., [169, 29, 77, 146]) since it carries potential to address in a unified methodology different types of inverse or uncertainty quantification problems arising in a large variety of physical systems. Among the possible applications of the method to the field of reactor physics stand

(i) the search for optimal sensor locations to measure certain quantities of interest during the operation of the core,

(ii) the acceleration in the search for optimally safe and/or efficient core configurations since GEIM gives a quick reconstruction of the quantities of interest.

Much more ambitious is the possibility to take into account the accuracy of the sensors in the placement selection. Indeed one could be interested in using few, very accurate, sensors and more, less accurate ones. The natural questions are then to place them in an optimal way. Another related question is: given a certain budget, what is the best location/quality/number of sensors to recover the best approximation. This question is difficult to answer in general but a recent investigation on this topic can be found in [31]. At any rate, we emphasize that the method cannot completely replace the experience of experts of the field. It should be seen as a tool to assist them in doing these tasks more efficiently and especially in a more *systematic* way.

3.2 Physical context

The purpose of this section is to present the physical context of neutronic modeling and simulation. Some notions of neutronic physics related to our problem will be introduced. We refer the reader to the books [200, 101, 153] for more details of neutronic physics and nuclear reactors.

3.2.1 Objective of reactor physics

Nuclear reactors are engineering devices in which controlled nuclear fission chain reactions are maintained and from which the produced nuclear energy is extracted for useful uses, such as generation of electricity. In such a device, neutrons induce nuclear fission reactions with heavy nuclei called nuclear fuel. The constituent materials of a reactor are generally fuel, coolant/moderator, structural materials, and fission control material. We illustrate a two-dimensional schematic of a typical configuration of a power reactor core and fuel assembly in Figure 3.2.1. In general, these materials are arranged very heterogeneously due to neutronics, thermal-hydraulics, and structural considerations, etc. In addition, these structural arrangement and the constituents may change depending on the life-cycle of the fuel or on the operational mode of the reactor, including accident conditions.

Although the discipline of reactor physics that deals with the design and analysis of such reactors encompasses several areas in science and engineering, the reactor physics has matured on its own and established a unique field; and thus in particular, reactor analysis and methods development may be characterized as a discipline concerning determination and prediction of the states of a reactor that sustains chain reaction by balancing neutron production by fission and loss by capture and leakage.

More specifically and summarily, we can say that the objective of the reactor analysis is to determine:

i) neutron multiplication factors for various configurations of a reactor, and

ii) neutron flux distributions (hence, power distributions that are generated), spatial and temporal, under various operating (including accident) conditions.

Thus, the results of reactor design and analysis become the base or springboard to other activities necessary in the realization of a nuclear power plant. More details on this process may be found in any of the numerous textbooks written on nuclear physics [200, 153, 101].

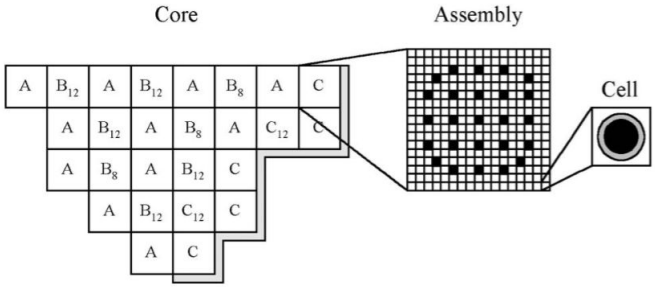


Figure 3.2.1: Typical configuration of 1/8 ractor core and fuel assembly.

3.2.2 Governing equations

In this thesis, we shall focus on the second objective above: neutron flux/power distributions. For that, we have to take into account the motion of the neutrons and their interactions with the host nuclei of various kinds. Thus, we need a mathematical model or theory to describe this particle transport phenomena.

As a high-level model that describes the distribution of neutrons in a medium such as a reactor, we usually consider the following Boltzmann transport equation [55, 74]:

$$\begin{aligned}\frac{1}{v}\frac{\partial}{\partial t}\varphi(r,E,\vec{w},t) + \vec{w}\nabla\varphi(r,E,\vec{w},t) + \Sigma_t(r,E)\varphi(r,E,\vec{w},t) \\ = \int_{4\pi} d\vec{w}' \int_0^\infty dE'\Sigma_s(r,E'\to E,\vec{w}'\cdot\vec{w})\varphi(r,E',\vec{w}',t) \\ + \frac{\chi(E)}{4\pi}\int_{4\pi} d\vec{w}' \int_0^\infty dE'\nu\Sigma_f(r,E')\varphi(r,E',\vec{w}',t) + q_{ex}(r,E,\vec{w},t)\end{aligned} \tag{3.2.1}$$

with appropriate initial and boundary conditions provided. In full generality, this equation has seven independent variables: three spatial variables r, two direction-of-flight (or angular) variables $\vec{w}$, energy E and time t. In (3.2.1), the angular neutron flux φ is defined as

$$\varphi(r,E,\vec{w},t) = vn(r,E,\vec{w},t) \tag{3.2.2}$$

where $v = \left(\frac{2E}{m}\right)^{1/2}$ with energy E in direction $\vec{w}$, and the angular neutron density n has the following meaning: $n(r,E,\vec{w},t)drdEd\vec{w}$ is the expected number of neutrons in $dr, dE, d\vec{w}$ around the phase space in point $r, E, \vec{w}$ at time t. The upper case Σ stands for macroscopic cross sections, other notations are standard [200, 153].

It is customary to first represent the differential scattering cross section in Legendre components:

$$\Sigma_s(r, E' \to E, \mu_0) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \Sigma_{sl}(r, E' \to E) P_l(\mu_0) \tag{3.2.3}$$

where $\mu_0 = \vec{w}' \cdot \vec{w}$. In the case of time-independent or steady-state situation, (3.2.1) becomes

$$\begin{aligned} &\vec{w}\nabla\varphi(r, E, \vec{w}) + \Sigma_t(r, E)\varphi(r, E, \vec{w}) \\ &\quad = \int_{4\pi} d\vec{w}' \int_0^{\infty} dE' \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \Sigma_{sl}(r, E' \to E) P_l(\mu_0) \varphi(r, E', \vec{w}') \\ &\quad + \frac{\chi(E)}{4\pi} \int_{4\pi} d\vec{w}' \int_0^{\infty} dE' \nu\Sigma_f(r, E')\varphi(r, E', \vec{w}') + q_{ex}(r, E, \vec{w}) \end{aligned} \tag{3.2.4}$$

Because the solutions to the transport equation for realistic reactor problems are difficult to obtain, and also due to the fact that knowledge of the neutron density or the neutron flux

$$\phi(r, E) = \int \varphi(r, E, \vec{w}) d\vec{w} \tag{3.2.5}$$

is sufficient for most applications such as the fission rate distribution, one is interested in obtaining a governing equation with $\phi(r, E)$ as unknown instead of $\varphi(r, E, \vec{w})$. We begin by defining, in addition to (3.2.5), also neutron current density

$$\vec{J}(r, E) = \int_{4\pi} \varphi(r, E, \vec{w}) \vec{w} d\vec{w} \tag{3.2.6}$$

and integrate (3.2.4) over the angular variable, resulting in P_1 equations [123]. If we assume that the angular flux is only weakly dependent on the angle, i.e., linearly anisotropic (which would not be good in a highly absorbing medium or near the boundary or in a medium of rapid variation of cross sections):

$$P_1 \text{ approximation: } \varphi(r, E, \vec{w}) = \frac{1}{4\pi}\phi(r, E) + \frac{3}{4\pi}\vec{w} \cdot J(r, E), \tag{3.2.7}$$

then we define the so called diffusion coefficient,

$$D(r, E) = \frac{1}{3}\Big[\Sigma_t(r, E) - \frac{\int_0^{\infty} \Sigma_{sl}(r, E' \to E) J_u(r, E')}{J_u(r, E')}\Big]^{-1}, \tag{3.2.8}$$

where $u = x, y, z$. We can write formally as

$$J(r, E) = -D(r, E)\nabla\phi(r, E), \tag{3.2.9}$$

which is the so-called Fick's law [81]. Finally, the two P_1 equations are combined to give the so called continuous-energy diffusion equation:

$$\begin{aligned} -D(r, E)\nabla\phi(r, E) + \Sigma_t(r, E)\phi(r, E) &= \int_0^{\infty} \Sigma_{s0}(r, E' \to E)\phi(r, E')dE' \\ &+ \chi(E) \int_0^{\infty} \nu\Sigma_f(r, E' \to E)\phi(r, E')dE' + Q_{ex}(r, E') \end{aligned} \tag{3.2.10}$$

The transport equation (3.2.1) and the diffusion equation (3.2.10) are of continuous form in independent variables. Except for extremely simple cases, it is not feasible to find exact solutions for them. We need to call for variety of methods by which the governing equations are discretized

and solved numerically. The first discretization we consider is the energy variable, called multigroup approximation.

The multigroup discretisation splits up the energy into a number of so-called energy groups determined by a discrete set of values $E_G < E_{G-1} < \cdots < E_1 < E_0$. Each group has an associated scalar flux solution, ϕ_g, and a set of material parameters. The behaviour of the neutrons is described by the macroscopic cross-sections corresponding to fission, scattering and absorption, denoted $\Sigma_g^f, \Sigma_{g'\to g}^s, \Sigma_g^a$ respectively, where the subscripts g and g' indicate a particular group. The cross-sections represent the probabilities that fission, scattering or absorption will occur. The parameter D_g which controls how the neutrons diffuse through the domain, depends on the cross-sections according to $D_g(r) = \left(s(\Sigma_g^a(r) + \Sigma_g^s(r))\right)^{-1}$. The criticality is a measure of the growth or decay of the neutron population in a system over successive neutron generations in a fission chain reaction. In order to study criticality, the governing equation is cast in the form of an eigenvalue problem by introducing the effective multiplication factor, k_{eff}. If neutrons removed from the system exactly balance those created, the system is adjudged critical ($k_{eff} = 1$). If more neutrons are produced than removed then the system is super-critical ($k_{eff} > 1$). If fewer neutrons are produced than removed the system is sub-critical ($k_{eff} < 1$).

For group g, the criticality eigenvalue problem of the multigroup neutron diffusion equation is

$$-\nabla \cdot (D_g(r)\nabla\phi_g(r)) + \Sigma_g^a(r)\phi_g(r) - \sum_{\substack{g'=1 \\ g'\neq g}}^{G} \Sigma_{g'\to g}^s(r)\phi_{g'}(r) = \frac{\chi_g}{k_{eff}} \sum_{g'=1}^{G} \nu_{g'}\Sigma_{g'}^f(r)\phi_{g'}(r) \ , \qquad (3.2.11)$$

where χ_g is the probability that fission will result in a neutron being born in group g, ν_g is the average number of neutrons produced per fission event and k_{eff} is the effective multiplication factor. There are two typical boundary conditions considered: reflective boundary conditions $D_g\nabla\phi_g(r)\cdot n = 0$, and void boundary conditions $-\frac{1}{2}D_g\nabla\phi_g(r)\cdot n = \frac{1}{4}\phi_g(r)$, where n is the outer normal vector on the boundary.

Note that the numerical scheme to compute what can be interpreted as an eigenvalue $1/k_{eff}$ is based on the well-known power method (see e.g. [101]). Even though for most practical problems, there is no analytical solution for (3.2.11), we will show the special one dimension case in which the analytical solution can be derived in Appendix A. Later the simple analytical case will be used for the very beginning verification of the methodology.

3.2.3 Neutronics modeling and simulation at EDF R&D

As shown in Figure 3.2.2, the EDF neutronic calculation scheme is based on a classical two-step calculation approach:

i) The *micro-level* calculation, also referred to as *lattice calculation*, determines the neutron density over a two-dimensional (2D) radial slice of a nuclear reactor fuel assembly based on assumed boundary conditions (e.g. reflective boundary conditions, specular reflection boundary conditions, periodic boundary conditions, white boundary conditions, vacuum boundary conditions, etc. [120]). The lattice calculation often comprises relatively accurate mathematical and/or numerical models, e.g., 1D spatial ultra-fine group neutron transport calculations for each fuel pin in the lattice, 2D spatial multi-group neutron transport calculations for each fuel lattice. From these calculations, effective few-group cross sections – representing the macro-level parameters, i.e., few energy groups and homogenized materials – appropriately averaged over angular direction, energy and space are determined so as to preserve neutron reactions rates, and correction factors, e.g., discontinuity factors, are introduced to preserve neutron leakage. The few-group cross sections are then input to a macro-level model, denoted

by the core simulator. There are approximately 10^6 few-group cross sections that are fed to the core simulator model as input data.

The APOLLO2 package [188, 213] is used at this level. The APOLLO2.8 / REL2005 / CEA2005v.4 package is a state-of-the-art lattice calculation scheme developed at CEA. There are currently two flux solvers that are used in APOLLO2, one is based on a multi-cell collision probability approach (CP) and the other is a full heterogeneous transport solver based on the method of characteristics (MOC). The cross sections depend on a few feedback parameters such as boron concentration, fuel temperature, fuel burnup and so on which span a phase space. A *calculation point* is an APOLLO2 calculation for a given point in this phase space. Since there are thousands of calculation points to perform for a given assembly, an application software converts these data in one object, the XSLIB. XSLIB contains all the input data needed to perform calculations such as technological data (geometry of components, assemblies, pins or grids, positions of burnable poisons, instrumentation types, clad constituents, etc.) or depletion chain.

ii) The *macro-level* calculation, also referred to as *core simulation*, uses the few-group cross sections to calculate approximately 10^5 macroscopic attributes, including neutron multiplication factor, safety margins, i.e., distance to design limits, and core-wide flux/power distributions, etc.

At EDF, all the core simulations are performed with the core code COCAGNE [50] which is still under development at EDF R&D. This code is part of the new EDF calculation chain, ANDROMÈDE, which is a major step forward for an industrial core code. It has state-of-the-art flux solvers, i.e., simplified transport solver (SPn) and Discrete Ordinates (Sn) transport solver, an efficient microscopic depletion solver, and different levels of cross section homogenizations (including pin-homogenized or pin-by-pin data) and relies on the APOLLO2 package.

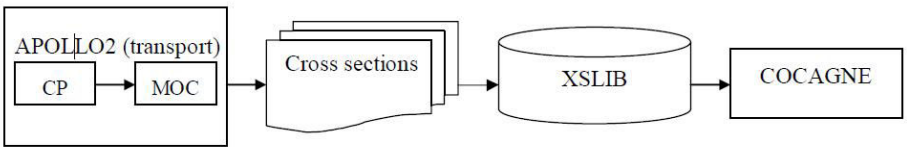


Figure 3.2.2: EDF new neutronic calculation scheme.

3.3 The Generalized Empirical Interpolation Method

3.3.1 Rationale of the GEIM

We first introduce the mathematical notations used throughout this chapter. Let V be a Hilbert space defined over a physical domain $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ and let $(\cdot, \cdot)$ be the inner product, and $\|\cdot\|$ be the associated norm. In our case Ω will be the reactor domain and V will be either $L^2(\Omega)$, $L^\infty(\Omega)$ or a product of these spaces.

In order to reconstruct the state $f \in V$ of a physical quantity, we assume that we have access to two types of information:

i) Information coming from measurements collected directly from the physical system. In the following, we model sensors with linear functionals $\ell \in V'$ parametrized by the position $x \in V$

and the quantity $\ell(f, x)$ denotes the measurement value given from a sensor located in $x \in \Omega$. The linear functionals can be chosen as Dirac masses, in which case the measurements are pointwise evaluations of f at some points $x \in \Omega$,

$$\ell(f, x) = \delta_x(f) = f(x).$$

A perhaps more realistic option is to work with local averages centered around some point x. If $D(x)$ is a neighborhood of x and g is a radial function, one can work with

$$\ell(f, x) = \int_{D(x)} f(t)g(x - t)\ \mathrm{d}t.$$

In our numerical experiments, we work either with Dirac masses or local averages depending on the test case.

ii) An information coming from a physical model, which reads in general in the form of a parametric partial differential equation (PDE). We assume that f is the solution of

$$\mathcal{F}(f, \mu) = 0, \tag{3.3.1}$$

where $\mathcal{F}$ is a differential operator defined in V and $\mu \in \mathbb{R}^p$ is a vector of $p \geq 1$ parameters. In our case, we assume that for any μ, there exists a unique solution $f(\cdot, \mu) \in V$ to problem (3.3.1). However, the particular parameter μ that best describes the system under consideration is in general not well known. Instead, one usually knows that the parameters may lie in some range $\mathcal{D} \subset \mathbb{R}^p$ so that the information from the parametrized model is that our function f of interest belongs to the set

$$\mathcal{M}_f \coloneqq \{f(\cdot, \mu)\ : \mu \in \mathcal{D}\}, \tag{3.3.2}$$

which is called the manifold of states. For certain classes of PDEs and for a given target accuracy $\varepsilon > 0$, it is possible to approximate the elements of $\mathcal{M}_f$ with n-dimensional linear spaces V_n such that

$$\operatorname{dist}(\mathcal{M}_f, V_n) \coloneqq \max_{u \in \mathcal{M}_f} \min_{v \in V_n} \|u - v\| \leq \varepsilon,$$

and where the dimension $n = n(\varepsilon)$ increases moderately with the accuracy ε going to zero (see [124]). In other words, one can find spaces V_n of reduced dimension which approximate $\mathcal{M}_f$ at high accuracy. The underlying reason for having this property is the regularity in μ of $f(\cdot, \mu)$ for certain PDEs like elliptic ones (see [150, 64, 63]). One technique to build these spaces are reduced basis methods, which use a greedy algorithm to select functions $f(\cdot, \mu_i)$ for appropriately chosen parameters μ_i and then define $V_n \coloneqq \operatorname{span}\{f(\cdot, \mu_i)\ : i = 1, \ldots, n\}$.

GEIM combines measured data and a reduced space as follows. We first run a greedy algorithm which selects

- a set of functions $\{f(\cdot, \mu_1), \cdots, f(\cdot, \mu_n)\}$ from the manifold $\mathcal{M}_f$. These functions will span our reduced basis $V_n \coloneqq \operatorname{span}\{f(\cdot, \mu_i)\ : i = 1, \ldots, n\}$. For convenience, we will sometimes work with other functions $q_i \in V_n$ which are built from linear combinations of the $f(\cdot, \mu_i)$. Their exact form is given in section 3.3.2.

- a set of locations $x_1, \ldots, x_n \in \Omega$ for the sensors. The locations are searched among all admissible positions $x \in \Omega_{\text{admissible}}$. Mathematically, this is expressed by saying that we search for sensors $\sigma(\cdot, x)$ among a dictionary $\Sigma \coloneqq \{\sigma(\cdot, x)\ x \in \Omega_{\text{admissible}}\}$ of potential candidates. One technical condition which is required is the following *unisolvence property*: if $f \in \operatorname{span}\{\mathcal{M}_f\}$ is such that $\sigma(f) = 0, \forall \sigma \in \Sigma$, then $f = 0$.

Once V_n and the sensor locations are selected, we approximate any function $f(\cdot,\mu) \in \mathcal{M}_f$ with

$$\mathcal{J}_n[f](\mu) \coloneqq \sum_{j=1}^{n} c_j(\mu) q_j \in V_n \tag{3.3.3}$$

The $c_j(\mu)$ are coefficients which depend on the parameters μ. They are computed using measurement information by imposing the interpolating conditions

$$\ell\left(\mathcal{J}_n[f](\cdot,\mu), x_k\right) = \ell\left(f(\cdot,\mu), x_k\right), \ k \in \{1,\dots,n\}, \tag{3.3.4}$$

where we see that the coefficients $(c_1,\dots,c_n)$ are the solution of an $n \times n$ linear system of equations. The reconstruction gives enough accuracy for a small dimension n of V_n. We refer to [145, 146] for the mathematical analysis of the approach.

The next section explains how the greedy algorithm is exactly defined. Before moving to it, we would like to emphasize that, in this approach, the mathematical model is assumed to be perfect in the sense indicated above. Other approaches like the PBDW methodology of [148] allow to correct the possible model bias.

3.3.2 GEIM greedy algorithm

We start by finding a parameter μ_1 in $\mathcal{D}$ such that

$$\|f(\cdot,\mu_1)\| = \max_{\mu\in\mathcal{D}} \|f(\cdot,\mu)\|. \tag{3.3.5}$$

Note that several parameters μ might maximize the function $\mu \to \|f(\cdot,\mu)\|$. In this case, μ_1 is picked among the set of maximizers. The state $f(\cdot,\mu_1)$ defines $V_1 = \text{span}\{f(\cdot,\mu_1)\}$. The first sensor location x_1 is one of possibly multiple maximizers such that

$$|\ell\left(f(\cdot,\mu_1), x_1\right)| = \max_{x\in\Omega} |\ell\left(f(\cdot,\mu_1), x\right)|. \tag{3.3.6}$$

To facilitate the practical computation of the generalized interpolant, we do a change of basis in V_1. Instead of working with $f(\cdot,\mu_1)$ as basis function, we use

$$q_1 = \frac{f(\cdot,\mu_1)}{\ell(f(\cdot,\mu_1), x_1)}.$$

For any $\mu \in \mathcal{D}$, the generalized interpolant of $f(\cdot,\mu)$ is

$$\mathcal{J}_1[f](\cdot,\mu) = c_1(\mu) q_1 \tag{3.3.7}$$

and $c_1(\mu)$ is found with the interpolating conditions (3.3.4) for $n = 1$. We then proceed by induction. Assume that, for a given $M \geq 1$, we have selected a set of states $\{f(\cdot,\mu_j)\}_{j=1}^{M}$ and the associated basis functions $\{q_1, q_2, \dots, q_M\}$ that span V_M. Assume also that we have chosen positions $x_1,\dots,x_M$ to locate the first M sensors. The generalized interpolant is assumed to be well defined by (3.3.3) for $n = M$, i.e.,

$$\mathcal{J}_M[f](\cdot,\mu) \coloneqq \sum_{j=1}^{M} c_j(\mu) q_j.$$

The coefficients $c_j(\mu),\ j \in \{1,\dots,M\}$, are given by the interpolation problem (3.3.4) for $n = M$, i.e.,

$$\begin{cases} \text{Find } \{c_j(\mu)\}_{j=1}^{M} \text{ such that:} \\ \sum_{j=1}^{M} c_j(\mu) B_{k,j} = \ell\left(f(\cdot,\mu), x_k\right),\ \forall k \in \{1,\dots,M\}. \end{cases}$$

where $B_{k,j}$ are the coefficients of the $M \times M$ matrix $B := \left(\ell(q_j, x_k)\right)_{1 \le k,j \le M}$. We now define $f(\cdot, \mu_{M+1})$ such that

$$\|(f - \mathcal{J}_M[f])(\cdot,\mu_{M+1})\| = \max_{\mu \in \mathcal{D}} \|(f - \mathcal{J}_M[f])(\cdot,\mu)\|, \tag{3.3.8}$$

and x_{M+1} such that

$$|\ell\left((f - \mathcal{J}_M[f])(\cdot,\mu_{M+1}), x_{M+1}\right)| = \max_{x \in \Omega} |\ell\left((f - \mathcal{J}_M[f])(\cdot,\mu_{M+1}), x\right)|. \tag{3.3.9}$$

Like in the case $M = 1$, we might have several maximizers and, in this case, we just pick μ_{M+1} and x_{M+1} as one of them. The next basis function is then

$$q_{M+1} = \frac{(f - \mathcal{J}_M[f])(\cdot,\mu_{M+1})}{\ell\left((f - \mathcal{J}_M[f])(\cdot,\mu_{M+1}), x_{M+1}\right)}$$

We finally set $V_{M+1} = \operatorname{span}\{f(\cdot,\mu_j)\}_{j=1}^{M+1} = \operatorname{span}\{q_j\}_{j=1}^{M+1}$ and the generalized interpolant of $f(\cdot,\mu)$ at dimension $M+1$ is defined by formula (3.3.3) with $n = M+1$. It satisfies the interpolating conditions (3.3.4) for the sensors located at the $M+1$ positions given by the algorithm.

It has been proven in [146] that for any $n \ge 1$, the set $\{q_1,\dots,q_n\}$ is linearly independent and that this interpolation procedure is well-posed in V. This follows from the fact that the matrix B is lower triangular with diagonal entries equal to 1.

Let us now make several remarks.

i) The particular case where $V = L^\infty(\Omega)$ and the dictionary Σ is composed of Dirac masses is called EIM (and not Generalized EIM). This variant is the first one presented in the literature (see [21]).

ii) $\mathcal{D}$ is a set containing parameters in a continuous range so, in practice, it is not possible to compute maximum values over $\mathcal{D}$ as required in formulas (3.3.5) and (3.3.8). The same applies for the computation of the maximum over $x \in \Omega$ in (3.3.6) and (3.3.9). This is the reason why it is necessary to work with discrete subsets $\mathcal{D}^{(\text{training})}$, $\Omega^{(\text{training})}$ of $\mathcal{D}$ and Ω. They have to be fine enough so that the maximum over $\mathcal{D}^{(\text{training})}$ (resp. $\Omega^{(\text{training})}$) is representative of the maximum over $\mathcal{D}$ (resp. Ω).

iii) In practice, problem (3.3.1) is solved with a numerical scheme that we denote by SOLVE and which yields an approximation $\bar{f}(\cdot,\mu)$ of $f(\cdot,\mu)$,

$$\bar{f}(\cdot,\mu) = \text{SOLVE}\left(\mathcal{F}\left(f,\mu\right) = 0\right).$$

For a given $\mu \in \mathcal{D}^{(\text{training})}$, note that $f(\cdot,\mu)$ is the *exact* solution of the PDE (3.3.1). So $f(\cdot,\mu)$ is not known exactly but only via an approximation $\bar{f}(\cdot,\mu)$ coming from $\text{SOLVE}\left(\mathcal{F}\left(f,\mu\right) = 0\right)$. $\bar{f}(\cdot,\mu)$ is the quantity that is considered in the practical implementation of the algorithm. For any $\bar{\mu} \in \mathcal{D}^{(\text{training})}$, $\bar{f}(\cdot,\mu)$ is called a snapshot and

$$\mathcal{M}_{\mathcal{D}^{(\text{training})}} \coloneqq \{\bar{f}(\cdot,\mu)\ :\ \mu \in \mathcal{D}^{(\text{training})}\}$$

is called the set of snapshots. It is intended to be representative enough of the set $\mathcal{M}_f$ defined

in (3.3.2).

The discretizations $(\mathcal{D}^{(\text{training})},\ \Omega^{(\text{training})},\ \bar{f})$ yield an implementable version of the greedy algorithm. If they are fine enough, the algorithm yields, up to a factor, a convergence rate which is the same as the greedy algorithm with $(\mathcal{D},\ \Omega,\ f)$, see [146]. However, quantifying how fine these discretizations should be is a difficult task. Even more difficult is finding a systematic way of building the discretizations. One result in this direction is lemma 1 of [146], which quantifies how much one needs to refine $\mathcal{D}^{(\text{training})}$ and $\Omega^{(\text{training})}$ at each step M of the algorithm. However, the result is difficult to use for practical purposes. Another relevant reference is [151] where the authors propose an approach to refine $\mathcal{D}^{(\text{training})}$ by adapting the local approximation spaces to the local anisotropic behavior in the parameter space, thus expected to be representative enough of the set $\mathcal{M}_f$.

3.4 Implementation of GEIM to a nuclear reactor core

3.4.1 Physical model and remarks on how to apply GEIM

In this work, the neutron flux ϕ is modeled with the two-group neutron diffusion equation with null flux boundary conditions, see (3.2.11). So ϕ has two energy groups $\phi = (\phi_1, \phi_2)$. Index 1 denotes the fast energy group and 2 the thermal energy one. The flux is the solution of the following eigenvalue problem:

$$\text{Find } (\lambda, \phi) \in \mathbb{C} \times (L^\infty(\Omega) \times L^\infty(\Omega)) \text{ such that for all } x \in \Omega,$$

$$\begin{cases} -\nabla\left(D_1 \nabla \phi_1\right) + (\Sigma_{a,1} + \Sigma_{s,1\to 2})\phi_1 = \frac{1}{k_{eff}}\left(\chi_1 \nu \Sigma_{f,1}\phi_1 + \chi_1 \nu \Sigma_{f,2}\phi_2\right) \\ -\nabla\left(D_2 \nabla \phi_2\right) + \Sigma_{a,2}\phi_2 - \Sigma_{s,1\to 2}\phi_1 = \frac{1}{k_{eff}}\left(\chi_2 \nu \Sigma_{f,1}\phi_1 + \chi_2 \nu \Sigma_{f,2}\phi_2\right), \end{cases} \tag{3.4.1}$$

with

$$\phi_i|_{\partial\Omega} = 0 \quad \text{for } i = 1,\, 2.$$

The coefficients involved are the following:

- D_i is the diffusion coefficient of group i with $i \in \{1, 2\}$.
- $\Sigma_{a,i}$ is the macroscopic absorption cross section of group i.
- $\Sigma_{s,1\to 2}$ is the macroscopic scattering cross section from group 1 to 2.
- $\Sigma_{f,i}$ is the macroscopic fission cross section of group i.
- ν is the average number of neutrons emitted per fission.
- χ_i is the fission spectrum of group i.

The generated power is

$$P \coloneqq \nu\Sigma_{f,1}\phi_1 + \nu\Sigma_{f,2}\phi_2. \tag{3.4.2}$$

We next make some comments on the coefficients and recall well-posedness results of the eigenvalue problem (3.4.1). First of all, the first four coefficients (D_i, $\Sigma_{a,i}$, $\Sigma_{s,1\to 2}$ and $\Sigma_{f,i}$) might depend on the spatial variable. In the following, we assume that they are either constant or piecewise constant so that our set of parameters is

$$\mu = \{D_1, D_2, \Sigma_{a,1}, \Sigma_{s,1\to 2}, \nu, \Sigma_{f,1}, \Sigma_{f,2}, \chi_1, \chi_2\}. \tag{3.4.3}$$

By abuse of notation, in (3.4.3) we have written D_i to denote the set of values that this coefficient might take in space and similarly for the other parameters.

Note that k_{eff} is not a parameter in our setting because, for each value of the parameters μ, k_{eff} is determined by the solution of the eigenvalue problem. In the computation part, the numerical scheme SOLVE to compute the eigenvalue $1/k_{\text{eff}}$ is based on the well-known power method (see, e.g., [101]). The spatial approximation uses $\mathbb{P}_1$ finite elements of a grid of size h (this value will be specified later in the following section).

If the parameters of our diffusion model range in, say,

$$D_1 \in [D_{1,\min}, D_{1,\max}],\ D_2 \in [D_{2,\min}, D_{2,\max}], \dots, \chi_2 \in [\chi_{2,\min}, \chi_{2,\max}],$$

then

$$\mathcal{D} \coloneqq [D_{1,\min}, D_{1,\max}] \times \cdots \times [\chi_{2,\min}, \chi_{2,\max}] \tag{3.4.4}$$

and the set of all possible states of the flux and power is given by

$$\mathcal{M}_{\phi_1,\phi_2,P} \coloneqq \{(\phi_1, \phi_2, P)(\mu)\ :\ \mu \in \mathcal{D}\}, \tag{3.4.5}$$

which is the manifold of solutions of our problem. It is composed of vectorial quantities $(\phi_1, \phi_2, P)(\mu)$. Running GEIM in this case leads to various possibilities. First we could work on a product space V^3, with measurements on the triplet (ϕ_1, ϕ_2, P), or on ϕ_1, ϕ_2, P separately, letting the algorithm select which quantity is best to include at each step n of the greedy algorithm. One could also view the problem differently and define three independent manifolds

$$\begin{cases} \mathcal{M}_{\phi_1} \coloneqq \{\phi_1(\mu)\ :\ \mu \in \mathcal{D}\} \\ \mathcal{M}_{\phi_2} \coloneqq \{\phi_2(\mu)\ :\ \mu \in \mathcal{D}\} \\ \mathcal{M}_{P} \coloneqq \{P(\mu)\ :\ \mu \in \mathcal{D}\} \end{cases} \tag{3.4.6}$$

for which we run three separate GEIM algorithms. To go either one of these two above ways we need to have access to sensor measurements. For this reason, it is preferable to consider the manifold (3.4.5) and devise a reconstruction strategy where only thermal flux measurements are taken. We will follow the approach of [145] where the authors reconstruct the velocity and pressure of a fluid by using pressure measurements only. We describe how to adapt the strategy to the current neutronics problem in the next section. We also take into account that there are usually restrictions on the locations to place the sensors in the reactor Ω. A typical situation is that they can only be placed in the subdomain of Ω corresponding to the core Ω_{core} but there are no sensors in the reflector Ω_{refl}.

3.4.2 A GEIM algorithm for the neutronics problem

In the following $\mathcal{C}$ denotes a subdomain of the reactor Ω which is admissible for sensor placement. To work with the manifold (3.4.5), we modify the greedy algorithm to approximate the flux and power of $\mathcal{M}_{\phi_1,\phi_2,P}$ when:

- we only use thermal flux measurements (related to ϕ_2),
- the sensors can only be placed in a partial region $\mathcal{C}$ of the reactor, e.g., $\mathcal{C} = \Omega_{\text{core}}$ or $\mathcal{C} = \Omega$. We insist here that the norms that are used (e.g in (3.4.7)) are norms on Ω.

Let $\Psi(\mu)$ be the triplet $(\phi_1(\mu), \phi_2(\mu), P(\mu))$, and denote by $\Psi_1(\mu)$ the $\phi_1(\mu)$, $\Psi_2(\mu)$ the $\phi_2(\mu)$ and $\Psi_3(\mu)$ the $P(\mu)$. We have measurements $\ell_1(\Psi) = \ell(\phi_1)$, or $\ell_2(\Psi) = \ell(\phi_2)$ or again $\ell_3(\Psi) = \ell(P)$, but we use only ℓ_2. Then we assume that:

- the Kolmogorov n-width of the manifold of all $\Psi(\mu)$ decays fast with n,
- the thermal flux sensors satisfy the *unisolvence property*,
- for any given $\mu \in \mathcal{D}$, the knowledge of $\Psi_2(\mu)$ allows to uniquely reconstruct $\Psi_1(\mu)$ such that $\phi = (\Psi_1, \Psi_2)(\mu)$ is the unique positive eigenfunction of the eigenvalue problem (3.4.2) for the given parameter μ. This property is satisfied if we assume that the mapping $\mu \mapsto \Psi_2(\mu)$ is one to one. This is a strong hypothesis and, in the following, we assume that we are in a regime $\mathcal{D}$ of parameters where this is true (we will be able to confirm this in our numerical examples). We explain why this assumption is necessary at the end of the section.

We start by defining μ_1 as the quantity maximizing

$$\max_{\mu\in\mathcal{D}} \max_{i=1,2,3} \|\Psi_i(\mu)\|. \tag{3.4.7}$$

The first sensor location x_1 is now the one such that

$$|\ell_2(\Psi(\mu_1), x_1)| = \max_{x\in\mathcal{C}} |\ell_2(\Psi(\mu_1), x)|. \tag{3.4.8}$$

Instead of working with $\Psi(\mu_1)$ as basis function, we use

$$q_1 := \left(\frac{\Psi_1(\mu_1)}{|\ell_2(\Psi(\mu_1), x_1)|}, \frac{\Psi_2(\mu_1)}{|\ell_2(\Psi(\mu_1), x_1)|}, \frac{\Psi_3(\mu_1)}{|\ell_2(\Psi(\mu_1), x_1)|}\right). \tag{3.4.9}$$

For any $\mu \in \mathcal{D}$, we can then define $\mathcal{J}_1[\Psi](\mu)$ as in the previous algorithm, i.e.

$$\mathcal{J}_1[\Psi](\mu) := c_1(\mu) q_1. \tag{3.4.10}$$

For subsequent dimensions, we proceed by induction. We assume that, for a given $M > 1$, we have selected $\{\mu_1, \ldots, \mu_M\}$ and we have the set of locations $\{x_1, \ldots, x_M\}$ for the sensors of ϕ_2. Then the approximation of $\Psi(\mu)$ reads

$$\mathcal{J}_M[\Psi](\mu) := \sum_{j=1}^{M} c_j(\mu) q_j, \tag{3.4.11}$$

and is well defined. We then define μ_{M+1} as the parameter which maximizes

$$\max_{\mu\in\mathcal{D}} \max_{i=1,2,3} \frac{\|((\mathcal{J}_M[\Psi])_i - \Psi_i)(\mu)\|}{\|\Psi_i(\mu)\|}\ , \tag{3.4.12}$$

where we use relative errors in order to deal with possible differences in the magnitude orders of $\Psi_i,\ i = 1, 2, 3$. The next sensor location x_{M+1} satisfies

$$|\ell_2\left((\Psi - \mathcal{J}_M[\Psi])(\mu_{M+1}), x_{M+1}\right)| = \max_{x\in\mathcal{C}} |\ell_2\left((\Psi - \mathcal{J}_M[\Psi])(\mu_{M+1}), x\right)|. \tag{3.4.13}$$

Finally, the $(M+1)$-th basis function is

$$q_{M+1} := \left(\frac{((\mathcal{J}_M[\Psi])_1 - \Psi_1)(\mu_{M+1})}{|\ell_2(\Psi(\mu_{M+1}), x_{M+1})|}, \frac{((\mathcal{J}_M[\Psi])_2 - \Psi_2)(\mu_{M+1})}{|\ell_2(\Psi(\mu_{M+1}), x_{M+1})|}, \frac{((\mathcal{J}_M[\Psi])_3 - \Psi_3)(\mu_{M+1})}{|\ell_2(\Psi(\mu_{M+1}), x_{M+1})|}\right), \tag{3.4.14}$$

and the inductive step is completed.

As we will see later, the interpretation (3.4.11) that is nice (i.e. being able to reconstruct Ψ_1 from measurements on Ψ_2!). The above approach itself, provided that the set of all $\Psi(\mu)$ is indeed

with small dimension, is standard. The other approach we speak about is to consider indeed three independent GEIM approaches: one from $(\phi_1(\mu), \ell_1)$, one from $(\phi_2(\mu), \ell_2)$, one from $(P(\mu), \ell_3)$. We again emphasize that the same approach was used in [145] in a Stokes problem with a good accuracy in the reconstruction. The reference might be of interest to the community studying thermal-hydraulics in reactor cores.

Before giving some numerical results, a remark on the hypothesis about the bijectivity of the mapping between $\mu \mapsto \Psi_2(\mu)$ is in order. Classical results from spectral theory ensure that, for $\mu \in \mathcal{D}$, the mapping $\mu \mapsto (\Psi_1, \Psi_2)(\mu)$ is one-to-one. However, they do not ensure bijectivity of the mapping $\mu \mapsto \Psi_2(\mu)$. We assume that we are in a parameter regime where this holds because we reconstruct $(\Psi_1, \Psi_2, \Psi_3)(\mu)$ from measurements only acquired from thermal flux $\Psi_2(\mu)$. If $\mu \mapsto (\Psi_1, \Psi_2)(\mu)$ is bijective and $\mu \mapsto \Psi_2(\mu)$ was not, then we would have existence of $\mu_1 \neq \mu_2$ in $\mathcal{D}$ such that $\Psi_2(\mu_1) = \Psi_2(\mu_2)$ but $\Psi_1(\mu_1) \neq \Psi_1(\mu_2)$. This would imply that $(\mathcal{J}_M[\Psi](\mu_1))_2 = (\mathcal{J}_M[\Psi](\mu_2))_2$ and therefore $\mathcal{J}_M[\Psi](\mu_1) = \mathcal{J}_M[\Psi](\mu_2)$. As a result, we would have reconstructed $\Psi_1(\mu_1)$ and $\Psi_1(\mu_2)$ with the same function and the quality of approximation would no longer be ensured.

We further note that the above greedy algorithm can be extended to the general case, i.e., to recovery the multi-physics field $\Psi = (\Psi_1(\cdot, \mu), \cdots, \Psi_m(\cdot, \mu)) \in \mathbb{R}^m,\ m \in \mathbb{N}$ with the measurements from the i-th component $\Psi_i,\ 1 \leq i \leq m$, where Ψ is the solution to the parameter-dependent problem $\mathcal{F}(\Psi, \mu) = 0$. The bijectivity hypothesis about the mapping between $\mu \mapsto \Psi_i(\cdot, \mu)$ is necessary.

3.5 Numerical tests on benchmark problems

In this section, we reconstruct $\phi_1,\ \phi_2$ and P with the only knowledge of thermal flux measurements for two benchmark problems. The examples are in one and two dimensions and use the methodology of Section 3.4.2. For the sensor selection and placement, we consider two cases:

- Case I: the sensors can be placed at any point, in Section 3.4.2, the partial region $\mathcal{C} = \Omega$.
- Case II: the admissible sensor locations are restricted to the core Ω_{core}, which corresponds to setting $\mathcal{C} = \Omega_{\text{core}}$ in the algorithm of Section 3.4.2.

With these two cases, we aim at showing how the GEIM algorithm can be adapted to explore different restrictions in the positioning of the sensors.

3.5.1 1D homogeneous slab reactor

We consider the classical one-dimensional test case presented in [200, Chapter 4]. The reactor domain is $\Omega = [0, 30\text{cm}]$. The core and the reflector are $\Omega_{\text{core}} = [0, 25\,\text{cm}]$ and $\Omega_{\text{refl}} = [25, 30\,\text{cm}]$ respectively. The parametrized model is the one given in (3.4.3) with periodic boundary conditions. The discretization with $\mathbb{P}_1$ finite elements uses a mesh of size $h = 0.01\,\text{cm}$. We consider only the value of $D_1|_{\Omega_{\text{refl}}}$ in the reflector Ω_{refl} as a parameter (so $p = 1$ and $\mu = D_1|_{\Omega_{\text{refl}}}$). We assume that $D_1|_{\Omega_{\text{refl}}} \in [0.5, 2.0]$. The rest of the coefficients of the diffusion model (3.4.1) (including $D_1|_{\Omega_{\text{core}}}$) are fixed to the values indicated in Table 3.1. In principle, one could also consider these coefficients as parameters but we have decided to focus only on $D_1|_{\Omega_{\text{refl}}}$ because of its crucial role in the physical state of the core: its variation can be understood as a change in the boundary conditions in Ω_{core} which, up to a certain extent, allows to compensate the bias of the diffusion model with respect to reality.

	Core		Reflector	
Energy group	1	2	1	2
χ	1.0	0.0	0.0	0.0
$\nu\Sigma_f(\text{cm}^{-1})$	0.0085	0.1851	0.0	0.0
$\Sigma_a(\text{cm}^{-1})$	0.0121	0.121	0.0004	0.020
$\Sigma_{s,1\to2}(\text{cm}^{-1})$	0.0241	–	0.0493	–
$D(\text{cm})$	1.267	0.354	$\in$ [0.5,2.0]	0.166

Table 3.1: Coefficient values.

In Figure 3.5.3, we show some examples of the behavior of (ϕ_1, ϕ_2, P) for different values of the parameter $D_1|_{\Omega_{\text{refl}}}$. We confirm visually that $\mu \mapsto \phi_2(\mu)$ is a bijective mapping. In addition, we remind that it is possible if more values of the parameter would have been given, the bijectivity (i.e. the unisolvence) would not be true anymore.

For any $D_1|_{\Omega_{\text{refl}}} \in [0.5, 2.0]$, we reconstruct $\Psi = (\phi_1, \phi_2, P)$ as outlined at the end of Section 3.4.2, that is, we approximate Ψ with its generalized interpolant $\mathcal{J}_M[\Psi]$ (see (3.4.11)). For this, we run a weak greedy algorithm over a set of 300 snapshots of ϕ_2 (solutions of the PDE for a discrete grid $\mathcal{D}^{(\text{training})} \subset \mathcal{D}$ of 300 parameters). We use pointwise evaluations as a model for the sensors. The admissible domain for the search of the interpolating points is either $\mathcal{C} = \Omega$ (case I) or $\mathcal{C} = \Omega_{\text{core}}$ (case II).

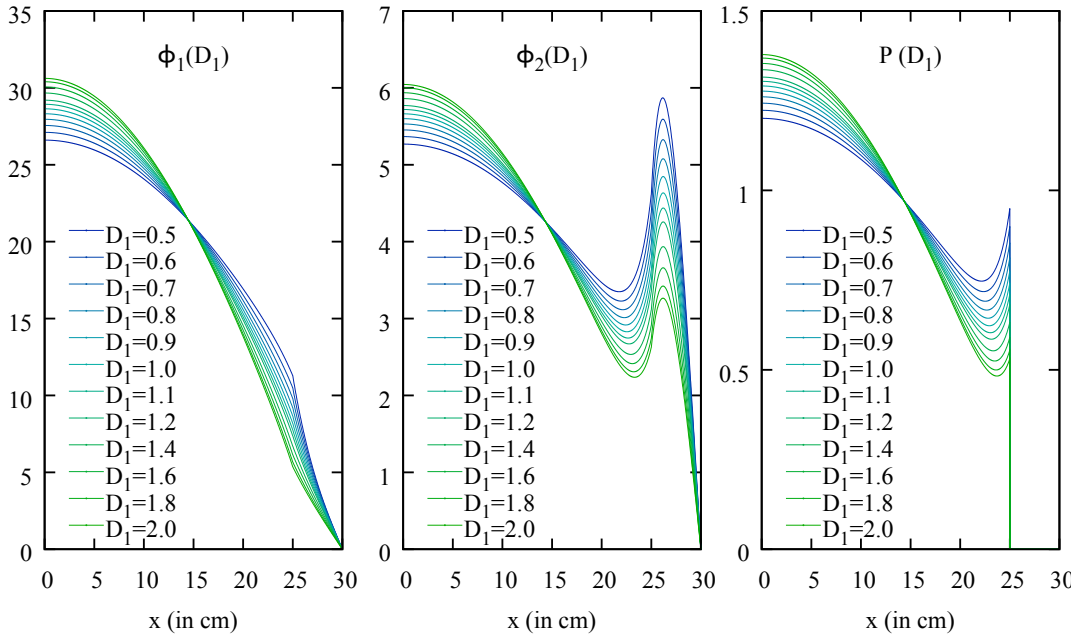


Figure 3.5.3: (ϕ_1, ϕ_2, P) for different values of $D_1|_{\Omega_{\text{refl}}}$. The values have been normalized to a reference quantity.

Let us now turn to the analysis of the results. We study the performance of the reconstruction

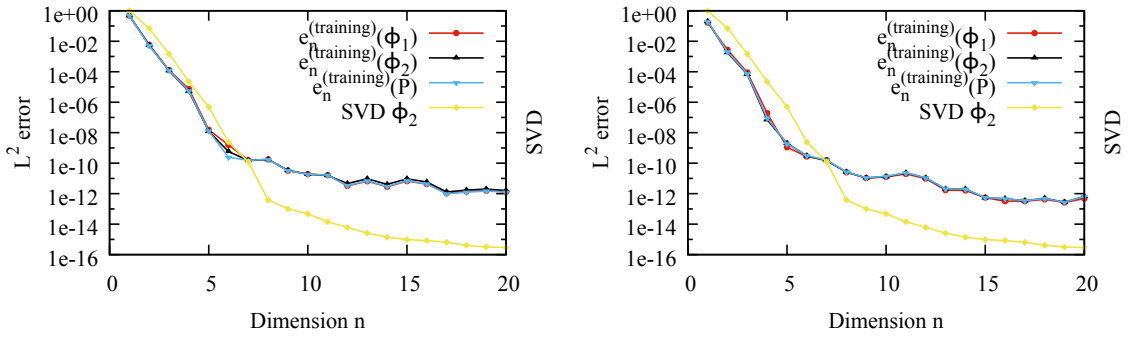
strategy by considering first of all the decay of the relative errors

$$\begin{cases} e_n^{(\text{training})}(\phi_1) & := \max_{\mu\in\mathcal{D}^{(\text{training})}} \|\phi_1(\mu) - (\mathcal{J}_M[\Psi])_1(\mu)\|_{L^2(\Omega)}/\|\phi_1(\mu)\|_{L^2(\Omega)} \\ e_n^{(\text{training})}(\phi_2) & := \max_{\mu\in\mathcal{D}^{(\text{training})}} \|\phi_2(\mu) - (\mathcal{J}_M[\Psi])_2(\mu)\|_{L^2(\Omega)}/\|\phi_2(\mu)\|_{L^2(\Omega)} \\ e_n^{(\text{training})}(P) & := \max_{\mu\in\mathcal{D}^{(\text{training})}} \|P(\mu) - (\mathcal{J}_M[\Psi])_3(\mu)\|_{L^2(\Omega)}/\|P(\mu)\|_{L^2(\Omega)} \end{cases} \tag{3.5.1}$$

in the greedy algorithm. In Figures 3.5.4a and 3.5.4b we plot the decay for case I and II respectively and observe that both yield very similar results. This is possibly due to the simplicity of the example. The decay is compared to an indicator of the optimal performance in $L^2(\Omega)$ which is obtained by a singular value decomposition of the snapshots $\phi_2(\mu)$, $\forall\mu \in \mathcal{D}^{(\text{training})}$. We see that $e_n^{(\text{training})}(\phi_2)$ decays at a similar rate as the SVD which suggests that GEIM behaves in a quasi-optimal way (see [146]). We now estimate the accuracy to reconstruct $(\phi_1, \phi_2, P)(D_1|_{\Omega_{\text{refl}}})$ for parameter values of $D_1|_{\Omega_{\text{refl}}} \in [0.5, 2.0]$ which do not necessary belong to the training set $\mathcal{D}^{(\text{training})}$. For this, we consider a test set of 300 parameters $\mathcal{D}^{(\text{test})}$ different from $\mathcal{D}^{(\text{training})}$ and compute the relative errors

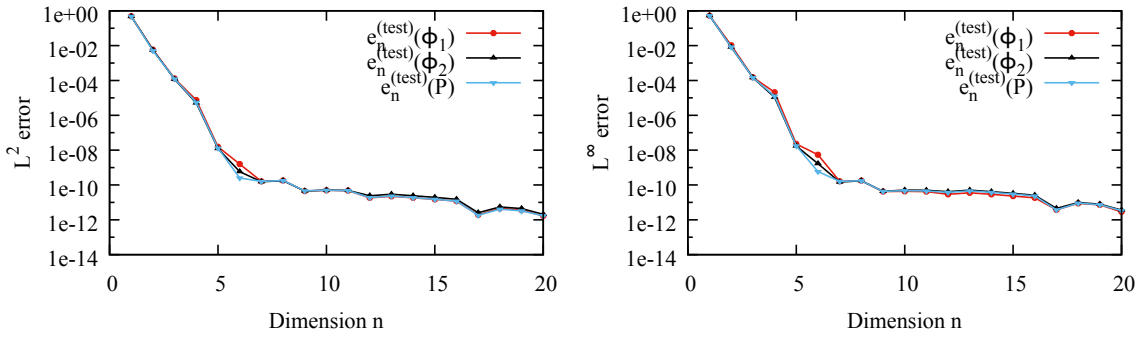
$$\begin{cases} e_n^{(\text{test})}(\phi_1) & := \max_{\mu\in\mathcal{D}^{(\text{test})}} \|\phi_1(\mu) - (\mathcal{J}_M[\Psi])_1(\mu)\|_{L^2(\Omega)}/\|\phi_1(\mu)\|_{L^2(\Omega)} \\ e_n^{(\text{test})}(\phi_2) & := \max_{\mu\in\mathcal{D}^{(\text{test})}} \|\phi_2(\mu) - (\mathcal{J}_M[\Psi])_2(\mu)\|_{L^2(\Omega)}/\|\phi_2(\mu)\|_{L^2(\Omega)} \\ e_n^{(\text{test})}(P) & := \max_{\mu\in\mathcal{D}^{(\text{test})}} \|P(\mu) - (\mathcal{J}_M[\Psi])_3(\mu)\|_{L^2(\Omega)}/\|P(\mu)\|_{L^2(\Omega)}. \end{cases} \tag{3.5.2}$$

Figure 3.5.5a shows the decay of these quantities for the first case where $\mathcal{C} = \Omega$. The behavior is almost identical to the case with $\mathcal{D}^{(\text{training})}$, which suggests that the training set was large enough for this example (otherwise the decay of the errors on $\mathcal{D}^{(\text{test})}$ might not have followed the same trend). In addition to this, the very fast decay rate of the reconstruction errors for ϕ_1 and P shows that, in this simple example, it is possible to reconstruct these quantities with measurements of ϕ_2 only. For certain safety studies, one might be interested in the behavior of the algorithm in $L^\infty(\Omega)$. In this case, very similar results to the $L^2(\Omega)$ norm are observed as Figure 3.5.5b shows. Also, the exact same conclusions can be drawn from the results on the second case where $\mathcal{C} = \Omega_{\text{core}}$ and the plots are not given for the sake of brevity.



(a) Case I (b) Case II

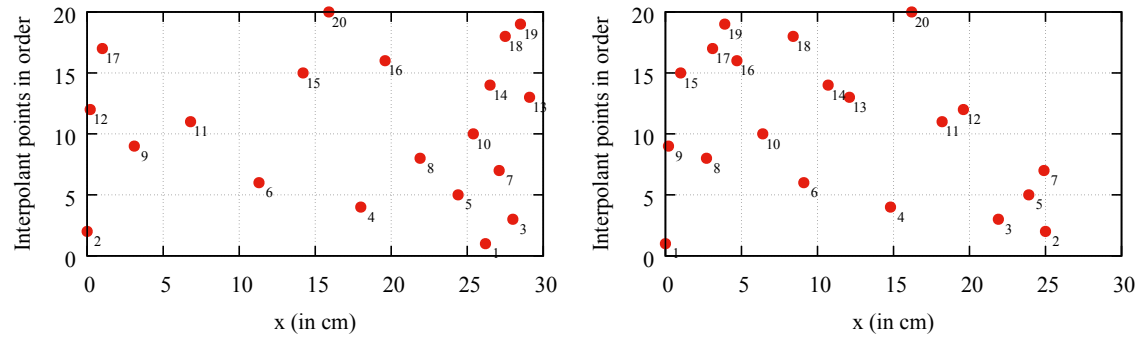
Figure 3.5.4: Relative errors $e_n^{(\text{training})}(\phi_1)$, $e_n^{(\text{training})}(\phi_2)$ and $e_n^{(\text{training})}(P)$ (Cases I and II, $L^2(\Omega)$ norm).



(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 3.5.5: Reconstruction errors $e_n^{(\text{test})}(\phi_1)$, $e_n^{(\text{test})}(\phi_2)$, $e_n^{(\text{test})}(P)$ (Case I, $L^2(\Omega)$ and $L^\infty(\Omega)$ norms).

As already brought up, the greedy algorithm finds sensor locations that can be used as a systematic tool for sensor placement. The locations selected in this simple test are given in Figures 3.5.6a and 3.5.6b for both cases I and II. We may note that the first locations (until more or less ten) tend to arrange themselves uniformly in the interval $\mathcal{C}$. This could be seen as an indication that the selection is of good quality.



(a) Case I ($\mathcal{C} = \Omega$) (b) Case II ($\mathcal{C} = \Omega_{\text{core}}$)

Figure 3.5.6: Sensor locations chosen by the greedy algorithm.

3.5.2 2D IAEA benchmark problem

We consider the classical 2D IAEA benchmark problem (see page 437 of [3] for its official definition and [1, 208] for implementations with different neutronic codes). The reactor geometry is shown in Figure 3.5.7. Only one quarter is given because the rest can be inferred by symmetry along the x and y axis.

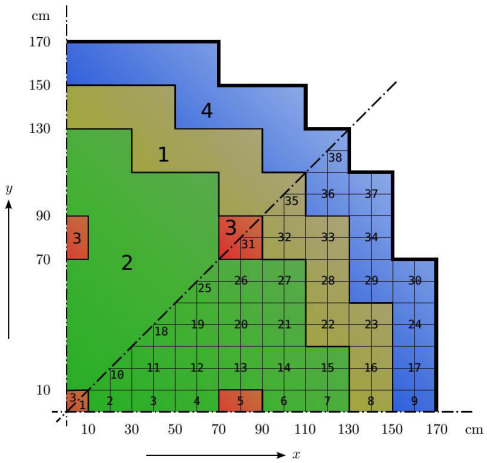


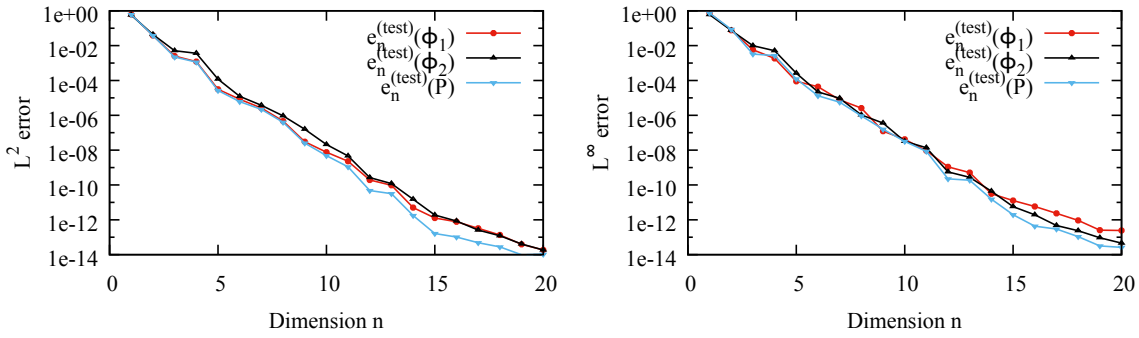
Figure 3.5.7: Geometry of 2D IAEA benchmark, upper octant: region assignments, lower octant: fuel assembly identification (from reference [208]).

Region	D_1	D_2	$\Sigma_{1\to 2}$	Σ_{a1}	Σ_{a2}	$\nu\Sigma_{f2}$	Material
1	1.5	0.4	0.02	0.01	0.080	0.135	Fuel 1
2	1.5	0.4	0.02	0.01	0.085	0.135	Fuel 2
3	1.5	0.4	0.02	0.01	0.130	0.135	Fuel 2 + Rod
4	[1.0, 3.0]	0.3	0.04	0	0.010	0	Reflector

Table 3.2: Parameter values: diffusion coefficients D_i (in cm) and macroscopic cross sections (in cm^{-1}).

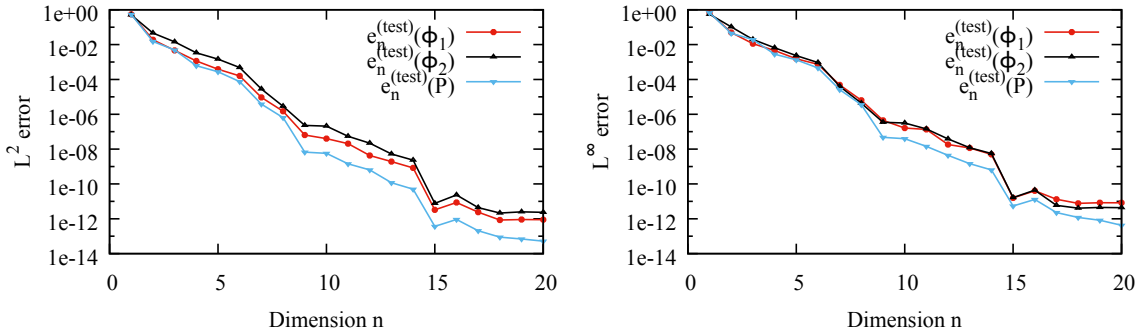
As in the previous example, the parametrized model is the one given in (3.4.3). The computational domain Ω was not the whole reactor but only the lower octant of Figure 3.5.7. Note that four regions are numbered: the first three correspond to the core domain Ω_{core}, and the fourth being the reflector domain Ω_{refl}. To account for the symmetries boundary conditions were enforced in the $x = 0$ and $y = x$ axis and the external border has zero boundary conditions. The discretization with $\mathbb{P}_1$ finite elements uses a mesh of size $h = 1$ cm. Like in the previous example, the only parameter is $D_1|_{\Omega_{\text{refl}}}$ for which we assume that it ranges in $[1.0\,\text{cm}, 3.0\,\text{cm}]$ (the value in the original benchmark being 2.0 cm). The rest of the coefficients of the diffusion model (3.4.1) (including $D_1|_{\Omega_{\text{core}}}$) are in accordance with the original benchmark problem. Their values are given in Table 3.2.

Using pointwise evaluations as a model for the sensors and working with a test set of 300 parameters for $\mathcal{D}^{\text{test}}$, we proceed in the same way as in the previous example to obtain convergence errors for $e_n^{(\text{test})}(\phi_1)$, $e_n^{(\text{test})}(\phi_2)$ and $e_n^{(\text{test})}(P)$. They are given in Figures 3.5.8 and 3.5.9 for both cases I and II and in norms $L^2(\Omega)$ and $L^\infty(\Omega)$. As we can see, the errors decays quickly in all cases, confirming the efficiency of the methodology in this two dimensional case. However, in comparison with the previous 1D example, in order to reach the same reconstruction accuracy the dimension of the reduced space and the number sensors is more or less doubled. The fact that the geometry is more complex and two dimensional is probably the most important factor to explain this observation. We finish this example by giving in Figure 3.5.10 the corresponding sensor placement for both cases I and II. We observe that in both cases the locations tend to cluster around the Dirichlet boundary.



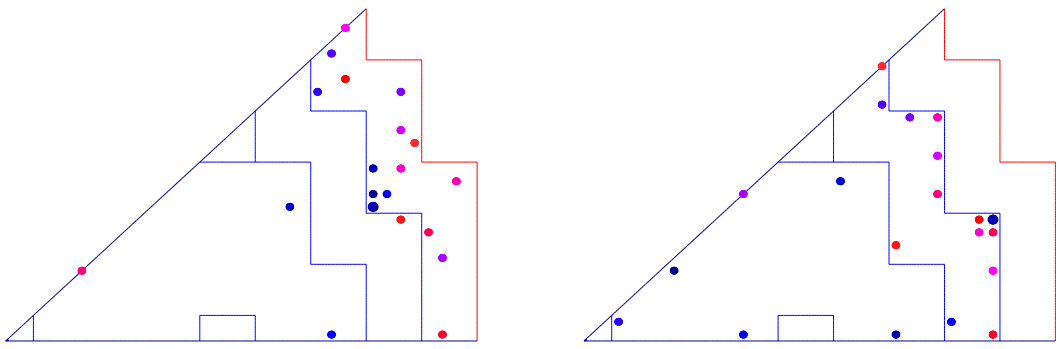
(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 3.5.8: Case I, error convergence for the reconstruction of $(\phi_1, \phi_2, P)(\mu)$ with $J_M[\phi_1, \phi_2, P](\mu)$ in 2D.



(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 3.5.9: Case II, error convergence for the reconstruction of $(\phi_1, \phi_2, P)(\mu)$ with $J_M[\phi_1, \phi_2, P](\mu)$ in 2D.



(a) Case I. (b) Case II.

Figure 3.5.10: Sensor locations chosen by the greedy algorithm in the 2D example.

3.6 Conclusions

In this chapter, we introduced a data assimilation framework with reduced basis for sensor placement and field reconstruction applications in nuclear domain. The framework of GEIM makes it possible for us to focus on the need for effective, many-query design evaluation in nuclear reactor domain: i) either real-time scenarios (control, parameter estimation, monitoring), where the solution of the problem needs to be known very quickly under limited resources for a previously unknown parameter, ii) or multi-query scenarios (design and optimization, multi-model/scale simulation), where the problem has to be solved repeatedly for many different parameters.

We set the goal to reconstruct the thermal/fast flux and power distributions with GEIM for a wide range of parameters variation domain, instead of solving the original neutronic governing equations with high resolution. To reach the goal, we first adapt the GEIM greedy algorithm, to be able to reconstruct the flux and power fields only from the thermal flux measurements. We then showed its applications to sensor placement and field reconstruction based on 1D and 2D benchmark problems, numerical results confirmed the feasibility of the methodology.

Chapter 4

Applications to sensor placement and field reconstruction

Contents

In this chapter, we show the applications of GEIM methodology on industrial cases, either based on the realistic nuclear reactors operated by EDF or based on the real physical problem. In Section 4.1, we apply the GEIM method to the determination of the optimal sensor locations over all the geometries of reactors operated by EDF: PWR 900 MWe, PWR 1300 MWe and PWR 1450 MWe.

In Section 4.2, the methodology is proposed to reconstruct the flux and power fields of the core during control rods movements based on the IAEA 2D benchmark. In Section 4.3, we apply GEIM to the neutron flux/power reconstruction based on EDF reactors in Pin-by-Pin framework, and the noisy measurements will be addressed in Section 4.4.

We indicate here that Section 4.1 has been presented in a published paper with J.-P. Argaud, B. Bouriquet, F. de Caso, Y. Maday and O. Mula. Its reference in the manuscript is [14]. Section 4.2 has been presented in a published paper with J.-P. Argaud, B. Bouriquet, Y. Maday and O. Mula. Its reference in the manuscript is [95]. Section 4.3 is a collaborative work with Francesco Silva, it is a part of his internship work at EDF R&D.

4.1 Sensor placement in a nuclear reactor core

4.1.1 Introduction

Knowledge of the best locations of measurements in the core is of great interest for industrial purposes. Of course, the main one is for new core design. But it can be also exploited to address the question of the sensitivity of the core with respect to the instrumentation failures [37]. In order to be sure that those two main issues could be addressed, we choose to work on real cases instead of benchmarks as already verified in previous chapter. Such an approach allows to demonstrate that the method can be applied in practical cases.

All referenced studies are originally driven by some effects that have been seen in data assimilation studies [36, 37], where it appears that all the instruments in the core do not have the same impact on the quality of the field reconstruction. As a consequence, the question is: where should we put the instruments to get the best results? This kind of questions have started getting answers in previous studies [35], in that work, the position optimization is realized using a Simulated Annealing algorithm, based on the Metropolis-Hastings proposition [157]. The results show that there are some locations that are better suited for instruments positioning, thus we can optimize a full instrumentation network at the design stage. However, no clear rule or systematic work has been found in making a good instrumentation network.

In this section, we apply the GEIM method to the determination of the optimal localizations over all the geometries of reactors operated by EDF: PWR 900 MWe, PWR 1300 MWe and PWR 1450 MWe. Several strategies are developed in order to underline the advantages and limitations of the determination of the instrumental setup. It is shown that GEIM gives a rather systematic instrument network that is consistent with what is expected from a physical point of view. Moreover, we demonstrate that, with respect to the random case, the results of the optimal network with GEIM are clearly better. Finally, we conclude that GEIM is able to provide a physically coherent sensor placement with the structure of the neutronic flux in the reactor core.

4.1.2 Problem setting

We consider a realistic core geometry of Pressured Water Reactors of 900, 1300 and 1450 MWe operated by EDF. The geometry for the PWR 900 and 1300 MWe is the one used in reference [39]. For the PWR1450 MWe, we use the one from [59] and an eighth of the core is presented in Figure 4.1.1a. Our goal is to find flux sensor locations which are well adapted for the whole life cycle of each reactor. To each time of the life cycle is associated a state of the flux, which is modelled by the solution to problem (3.4.1) with specific values of the coefficients

$$\mu = \{D_1, D_2, \Sigma_{a,1}, \Sigma_{a,2}, \Sigma_{s,1\to 2}, \nu, \Sigma_{f,1}, \Sigma_{f,2}, \chi_1, \chi_2\}.$$

Here we work with boundary values defined by engineering rules which we do not explain here for the sake of brevity. The entries of μ vary in space depending on the materials contained in each fuel assembly. Their values also depend on three macroscopic parameters which give the stage of the life cycle:

- Pw: the power level of the reactor core. It ranges between 0.3 and 1.
- Bu: the average burnup of the fuel in the whole core. It is a measure of how much energy is extracted from the fuel so it is an increasing function in time. It ranges between 0 (for the beginning of the life cycle) and $Bu_{\max}$ (the end of the life cycle) and its exact evolution depends on the operating history of the reactor.
- $D_1|_R$: the diffusion coefficient of the reflector (labelled with an R in Figure 4.1.1a). The coefficient follows a classical Lefebvre-Lebigot modelling [153] of the reflector and ranges between 1.0 and 1.6.

Therefore, in this case, μ depends on the vector of "general" parameters

$$g \coloneqq (Pw, Bu, D_1|_R)$$

that range in

$$\mathcal{G} := [0.3, 1] \times [0, Bu_{max}] \times [1.0, 1.6].$$

We remind that $\Psi(\mu(g)) := (\phi_1, \phi_2, P)(\mu(g))$, so the manifold of solutions is

$$\mathcal{M}_\Psi \coloneqq \{\Psi(\mu(g)) \ \ ; \ \ g \in \mathcal{G}\}. \tag{4.1.1}$$

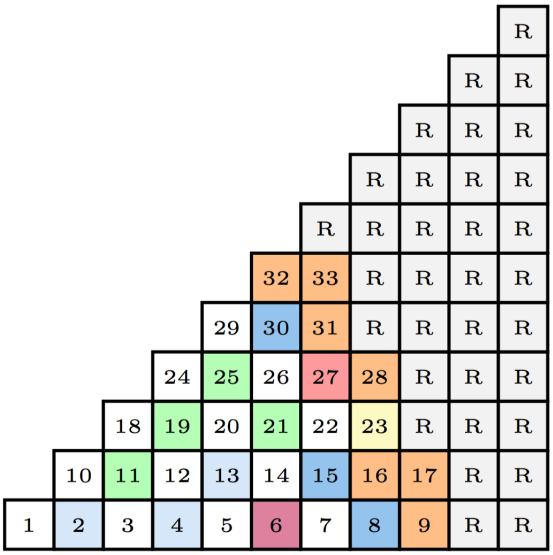
Note that the loading pattern of fuel assembly is octant-symmetry, in practice, the octant power tilt increases from zero to some extent with the burn-up of the fuel. Thus the simulation is based on the whole core symmetry, in order to take the possible power tilt into consideration when designing the sensor placement.

4.1.3 Selection of assemblies to place the sensors for ϕ_2

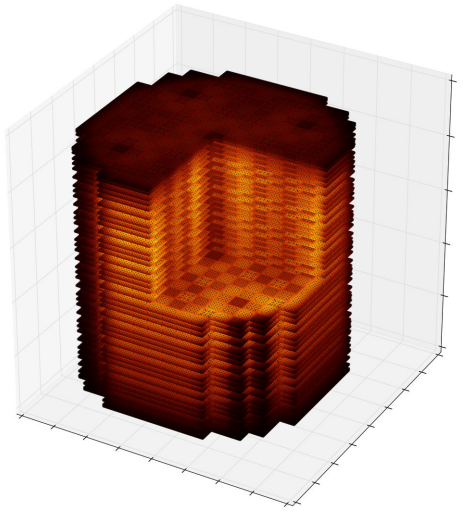
We have explored two approaches involving GEIM to find sensor locations for ϕ_2 in each PWR. For technological reasons, their placement is driven by the search of the most convenient assembly to take measurements so the output of our algorithm will be the assemblies where we locate the sensors. Therefore, even if the flux and power are three dimensional, our task can be seen as a 2D sensor placement problem and motivates to explore different strategies which we next explain. In all cases, we have run the greedy algorithms with a training set of 1000 snapshots. They are the solutions associated to a $5 \times 5 \times 40$ uniform grid $\mathcal{G}^{(\text{training})}$ of the parameter set $\mathcal{G}$. Note that we make a finer sampling on $D_1|_R$ because we know that the solutions are more sensitive to this parameter. The snapshots have been generated with COCAGNE [50], an R&D code developped at EDF R&D. We give in Figure 4.1.1b an example of the 3D power distribution in the core of the PWR 1450 MWe reactor.

The 3D2D method: We run the greedy algorithm of Section 3.4.2 with:

- the 1000 snapshots in 3D computed with COCAGNE,
- sensors that give local averages of ϕ_2 over portions of the assemblies that are called *assembly nodes*. Denoting by A the index set of the assemblies and, for each $a \in A$, defining $N(a)$ as



(a) Eighth of core (1-33: fuel assembly; R: reflector).



(b) Example of the 3D power distribution over the core from COCAGNE calculation.

Figure 4.1.1: The fuel assembly loading scheme and an example of 3D power distribution over the core in a realistic PWR 1450 MWe reactor at EDF.

the set of assembly nodes in assembly a, the measure on an assembly node $n \in N(a)$ is

$$\ell_n(\phi_2) := \frac{\int_{\mathrm{Vol(n)}} \phi_2 \, \mathrm{d}\Omega}{\mathrm{Vol}(n)}$$

As an example, we give the first 20 locations for the case of the PWR 1450 MWe in Figure 4.1.2.

Note that, with this approach, we may select more than one sensor per assembly. As a result, we need to post-process the output of the greedy algorithm in order to decide what are the most suitable assemblies. Our strategy goes as follows. We first compute

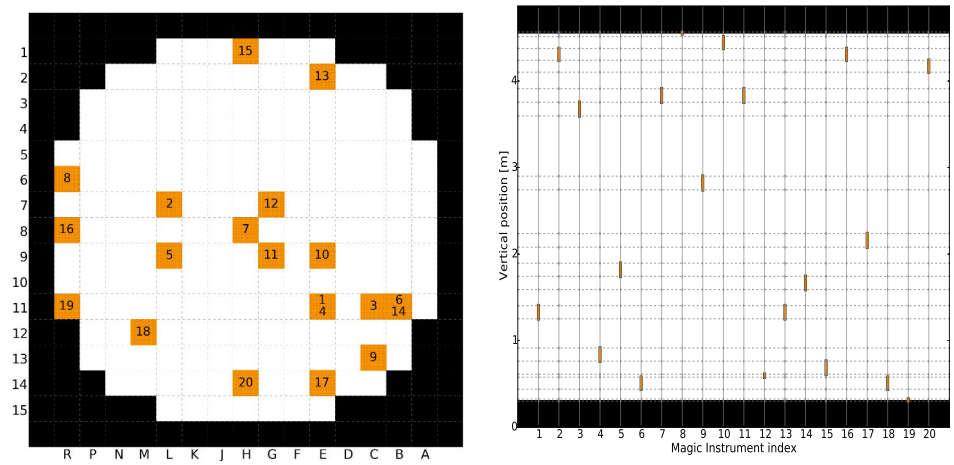
$$w_a = \sum_{n \in N(a)} w_n, \quad \forall a \in A,$$

where

$$w_n := \max_{g \in \mathcal{G}^{(\mathrm{training})}} |\ell_n \left(\phi_2(\mu(g)) - (\mathcal{J}_M[\Psi])_2(\mu(g))\right)|,$$

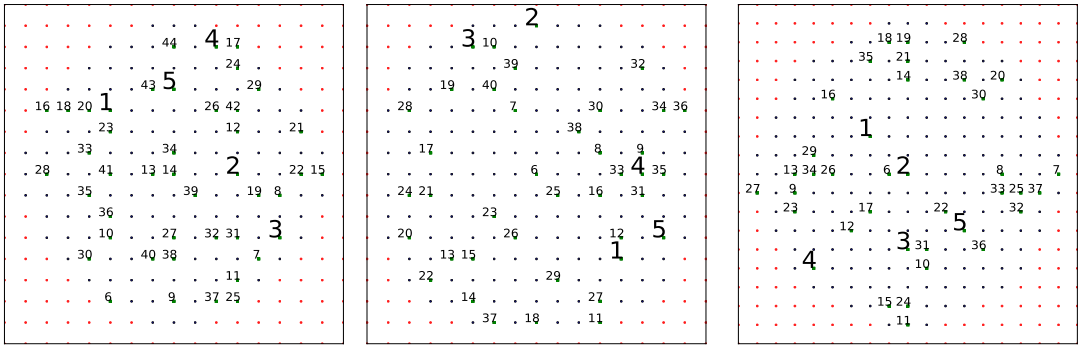
and $\mathcal{J}_M[\Psi]$ is defined in (3.4.11). We finally arrange the values $\{w_a\}_{a \in A}$ in decreasing order and choose the assemblies according to this ordering. Figure 4.1.3 shows the first 50 assemblies where to place the sensors for the three types of PWR's.

The 2D2D method: In this second approach, we integrate the snapshots over the vertical axis and run the greedy algorithm in the resulting 2D fluxes and power distributions. The measurements are averages over the assemblies. This approach gives directly a choice of the assemblies to select without a post-processing phase. Figure 4.1.4 shows the first 50 assemblies where to place the sensors for the three types of PWR's.



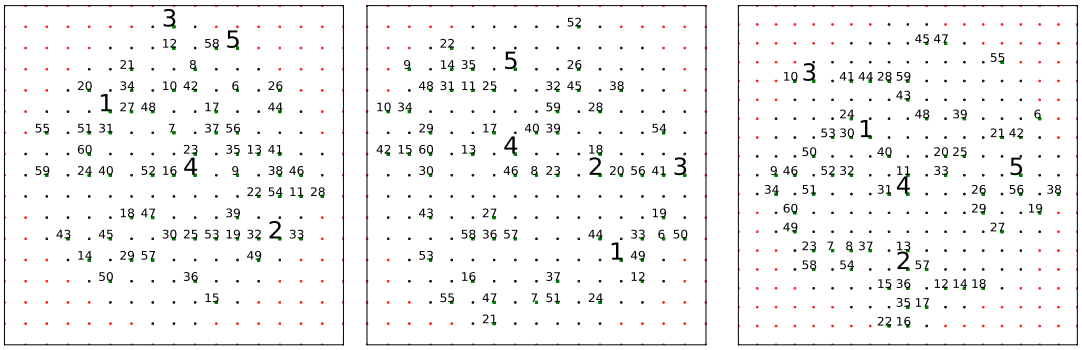
(a) Radial (b) Axial

Figure 4.1.2: The first 20 sensors placements, a realistic 1450 MWe reactor.



(a) PWR 900 (b) PWR 1300 (c) PWR 1450

Figure 4.1.3: 3D2D: The first 50 ordered optimized positioning of the instruments for PWR 900/1300/1450 MWe reactor cores.



(a) PWR 900 (b) PWR 1300 (c) PWR 1450

Figure 4.1.4: 2D2D: The first 50 ordered optimized positioning of the instruments for PWR 900/1300/1450 MWe reactor cores.

4.1.4 Analysis of the results

We first analyze the 3D2D case. From Figure 4.1.3, we notice that the mecanisms of sensor placement are pretty similar on the three types of PWR's. Globally speaking, two main areas are instrumented: one at the center of the core and another one at the limit between the core and the reflector. As a result, around the center of the core is a "ring" that is rather free of sensors in the three cases. If we focus on the order in which the first five assemblies are selected, we observe that the first ones are located at the border for the PWR 900 and 1300 MWe and at the center for PWR 1450 MWe.

The selected locations seem physically coherent with the structure of the neutronic flux in the core. The main information to be collected by sensors is the flux level inside the core. To get such information, the best option is to measure close to the center. This gives a good insight of the overall behavior of the core. Then the second global step is to characterize the behavior at the border which informs about the balance effect between the production (at the center) and the leakage (at the border). This explains the placement of the sensors at the border. It can also be linked to the importance, for global neutronic equilibrium in the core, of properly knowing the neutronic behavior at the transition between the reflector and the core. Moreover, these effects are also consistent with the burnup decrease during the whole life cycle of the reactor core, leading to changes in the power distribution and requiring information on each type of neutronic field distribution from the beginning to the end of the cycle. Thus the behavior of the method is in agreement with the neutronic physical expectations.

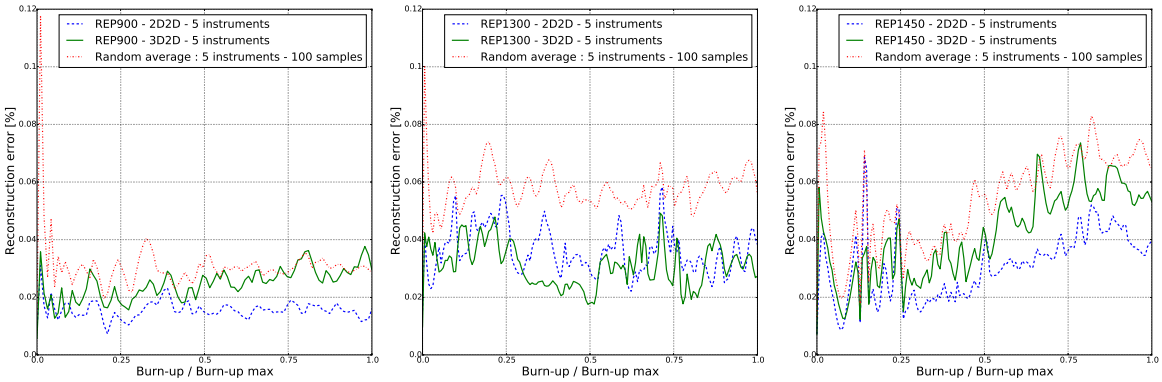
If we compare globally the sensor placement with the 3D2D and the 2D2D methods, both seem to be rather close. This confirms that working with the axially integrated flux and power is enough to determine the optimal positioning of the instrumental network.

We next study the quality of the reconstruction. From an engineering point of view, the most interesting parameter is the burnup Bu so we focus on its impact in the following. The study is carried out on examples with 5 and 15 instruments. This may seem to be a small number, but if we increase the number of sensors to be similar to real cores, the results become too close because the amount of information provided is enough. This saturation effect of mandatory information for field reconstruction has already been demonstrated in [37]. Thus, a setup with few sensors is chosen. In order to better evaluate the quality of the optimized sensor locations, we also analyze 100 random instrumental setups. In these random cases, the instrumental setup does not use any knowledge about the interest of each measurement for the overall field quality. This complement is given for all kinds of cores that we are taking into account in this study.

Figure 4.1.5 shows the mean reconstruction error over the whole core as a function of the burnup with 5 sensors for the 2D2D and 3D2D methods. The average of the error over the 100 random configurations is also plotted. We notice that the optimized sensor locations based on the greedy algorithm are clearly better than the averaged random ones. Moreover, this effect is the same for all cores. It demonstrates that the greedy algorithm of GEIM is also efficient in practice. We notice that the 3D2D results are a little worse than the 2D2D ones for the three cores. This can be understood in terms of dimension according to the instrumental network. As we look for a 2D network of instrumentation, it seems that the very accurate information we obtain with the 3D case is not fully useful.

We follow the same lines to analyze the case with 15 sensors. We give results on the field reconstruction in Figure 4.1.6. From this figure, we can draw a similar conclusion as with the 5 sensors. The sensor placement proposed with GEIM is better than a random selection. However, when comparing in detail the figures by pairs for the same core type, we notice that the difference between the random placement and the optimal one for 15 instruments is less obvious. Since more information is provided with 15 sensors than with 5, the differences between the random and optimal

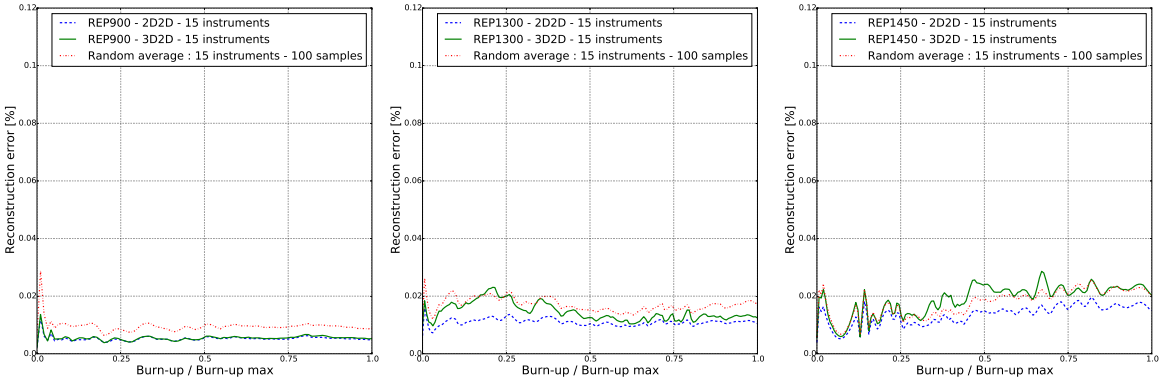
is less clear. This is even clearer if we compare the 3D2D and 2D2D cases, which are closer with 15 sensors than with only 5. The 2D2D method is still more efficient than the 3D2D method as it fits better with the design of the final instrumental network we aim to build.



(a) PWR 900 (b) PWR 1300 (c) PWR 1450

Figure 4.1.5: Quality of the reconstruction as a function of the burnup for 5 instruments, for PWR 900/1300/1450 MWe reactor cores.

Finally, whichever the strategies (3D2D method or 2D2D method) we choose to condensate information into 2D framework, the optimal networks are far better than the averaged random ones. It confirms the efficiency of the method for neutronic instrumental network analysis. Moreover, the 2D2D global method seems to be better suited than the 3D2D one.



(a) PWR 900 (b) PWR 1300 (c) PWR 1450

Figure 4.1.6: Quality of the reconstruction as a function of the burnup for 15 instruments, for PWR 900/1300/1450 MWe reactor cores.

4.1.5 Conclusions and future works

We have shown how GEIM can be applied to find sensor locations in a real reactor core in a systematic way. The example dealing with PWR illustrates the quality of the selection: the locations given by the algorithm yield better reconstruction results than the mean value of a random sampling. This proves that GEIM has the ability to learn the essential mechanisms of the physical system in order to provide good indications on sensor locations. In future works, we will include the deviation

of the model with respect to reality by working with PBDW [148]. In this approach, we work with a number of measurements $m \geq n$ and approximate with a least squares projection over the reduced model V_n. The greedy algorithm recently analyzed in [31] could be implemented in order to find optimal locations in this setting.

We currently work with synthetic measurements and it would be desirable to consider real data but there are still some obstructions to address in the methodology. Current research efforts go into this direction. One of the most important problems to address is that measurements are usually noisy. Some works in this topic whose results could be easily applied to nuclear engineering problems are [149, 16, 202].

4.2 Monitoring flux and power during control rods movements

4.2.1 Introduction

Control rods are an important mechanism for maintaining the desired state of fission reactions within a nuclear reactor. They constitute a real-time control of the fission process, which is crucial for both keeping the fission chain reaction active and preventing it from accelerating beyond control. During control rods movements, the power distribution is strongly affected locally, and any unanticipated rod misalignment from an electrical or mechanical failure may lead a reactor core shutdown. For safety reason to avoid such an issue, the power distribution monitoring system should be able to detect the abnormality of the control rods.

Many kinds of on-line core monitoring systems, such as BEACON [42], GNF-ARGOS [211] and SCORPIO-VVER [167] have been developed to estimate in-core power distributions during operations using several detectors. These detectors provide results at certain locations which reflect the actual reactor flux or power. Most of those systems also take advantage of the relationship between control rods position and the power distribution in the reactor [42, 86]. This relationship between flux shapes and rod positions is rather implicit, so for example neural network techniques are utilized in [90, 199] to unfold the rod position from the axial flux shape which is measured by movable in-core detectors. Recently, reduced basis methods have been applied to study the influence of effects of control rods movements [162, 193].

From the use of information point of view, the methods mentioned above neither take the physical governing equation/physical field information into consideration sufficiently, nor give an outline to optimize the sensor deployment in order to maximize the use of measurement information from a restricted number of sensors. On the other hand, from the on-line monitoring point of view, we need a simulation tool with real-time simulation capability, which is able to capture the features of the input/output behavior of a system in a rapid and reliable way, and the computational burden should be considerably low in terms of computational costs and times.

The GEIM method, which provides the possibility by taking the advantage of reduced basis method and optimized measurements, is thus proposed to reconstruct the flux and power fields of the core during control rods movements for the above considerations.

4.2.2 Problem setting

We still consider the classical 2D IAEA Benchmark Problem [3], the core geometry can be found in figure 3.5.7. The physical model is the one given in (3.4.1). The main effect of control rods movements is the variety of neutron flux/power shape, caused by the neutron absorption variety of the control rod assemblies. Note that, the control rods movements are always accompanied with boron concentration variations, but this doesn't affect the shape of the flux and power. Thus the

variable parameter in the diffusion model (3.4.1) is set to be $\Sigma_{a,2}|_{\Omega_{control}} \in \mathcal{D} := [0.080, 0.150]$, the rest of the coefficients of the model are in accordance with the original benchmark problem, see Table 3.2. We remind again that $\Psi(\mu) := (\phi_1, \phi_2, P)$, so the manifold of solution is

$$\mathcal{M}_\Psi := \{\Psi(\mu) \ : \ \mu \in \mathcal{D}\}. \tag{4.2.1}$$

We keep the same computing settings as in Section 3.5.2, i.e., to account for the symmetries boundary conditions were enforced in the $x = 0$ and $y = x$ axis and the external border has zero boundary conditions. The discretization with $\mathbb{P}_1$ finite elements uses a mesh of size $h = 1$ cm, and the pointwise evaluations are used as a model for the sensors.

4.2.3 Implementation of GEIM with an offline-online framework

We implement the greedy algorithm of Section 3.4.2 to reconstruct ϕ_1, ϕ_2 and P with the knowledge from thermal flux measurements. For the sensor selection and placement, we consider two cases:

- Case I: the sensors can be placed at any point, in Section 3.4.2, the partial region $\mathcal{C} = \Omega$.
- Case II: the admissible sensor locations are restricted to the core except the control rod domain i.e. $\Omega_{cote/control}$, which corresponds to setting $\mathcal{C} = \Omega_{cote/control}$ in the algorithm of Section 3.4.2.

We emphasize that, the GEIM greedy algorithm makes it possible to reconstruct the flux and power during control rods movements based on the following offline-online computation stages:

- **Offline phase:** select basis functions and measurement functionals with high computational cost.
  - Build a discrete solution set $\mathcal{M}^{(training)}$ which is intended to be representative enough of the set $\mathcal{M}_\Psi$ based on the given physical model i.e. (3.4.1).
  - Run the greedy algorithm of Section 3.4.2 to determine the basis functions $\{q_i := (\phi_1, \phi_2, P)(\mu_i)\}_{i=1}^n$ and measurement functionals $\{\ell_i\}_{i=1}^n$ where n is determined to meet the given accuracy.
- **Online phase:** reconstruct flux/power fields with low computational cost.
  - Acquire measurement data $\{\ell_i(\phi_2)\}_{i=1}^n$ of the current state (ϕ_1, ϕ_2, P).
  - Approximate (ϕ_1, ϕ_2, P) with $\mathcal{J}_n(\Psi)$ of (3.4.11) that we recall:

$$\mathcal{J}_n[\Psi] = \sum_{i=1}^n c_i q_i \tag{4.2.2}$$

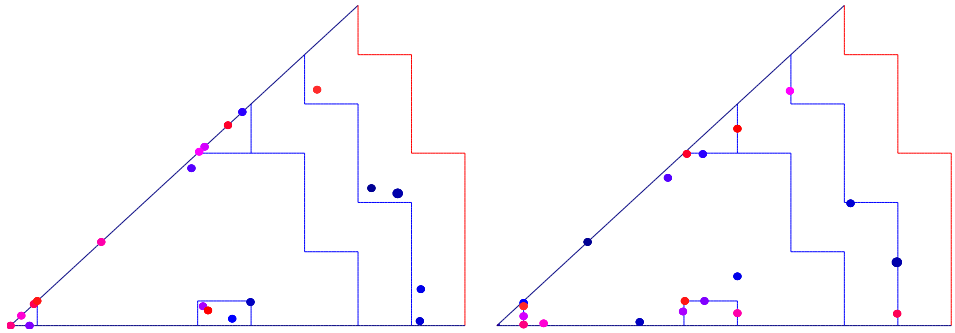
    where $\{c_i\}_{i=1}^n$ is determined by solving the following equations:

$$\ell_i((\mathcal{J}_n[\Psi])_2) = \ell_i(\phi_2), \quad i = 1, \cdots, n. \tag{4.2.3}$$

4.2.4 Numerical results and analysis

Let us now turn to the analysis of the results. We first give the first twenty sensor locations for case I and II in Figure 4.2.7. We observe that, because of the strong effect of the control rods, many sensors are deployed in the control rods domain (see Figure 4.2.7a) in order to capture the

behavior of control rods movements. When the control rod assemblies are not available for sensor placement, which is the normal case, the sensors have to be placed around the control rod assemblies (see Figure 4.2.7b). Furthermore, the effect of reflector is also considerable, thus several sensors are placed around the interface of fuel/reflector. That is to say, GEIM has the ability to deal with the essential mechanisms of the control rods movements based on a learning process from the manifold, and provides coherent indications on sensor locations.



(a) Case I ($\mathcal{C} = \Omega$) (b) Case II ($\mathcal{C} = \Omega_{core/control}$)

Figure 4.2.7: Sensor locations chosen by the greedy algorithm.

We study the performance of the reconstruction strategy by considering first of all the decay of the relative errors $(e_n^{(training)}(\phi_1), e_n^{(training)}(\phi_2), e_n^{(training)}(P))$ as defined in (3.5.1) for the training set $\mathcal{M}^{(training)}$ of 300 snaptshots. We also build a test set of 300 snaptshots $\mathcal{M}^{(test)}$ which is different from $\mathcal{M}^{(training)}$, and compute the errors $(e_n^{(test)}(\phi_1), e_n^{(test)}(\phi_2), e_n^{(test)}(P))$ as defined in (3.5.2). To evaluate the error convergence, we propose two measures: L^2 norm and L^∞ norm, the former is an indicator of the average/total effect of the reconstruction error, and latter reflects the maximum error in the domain, which is essential for safety analysis.

We give the errors from Figure 4.2.8 to Figure 4.2.11 for Case I and II. From these figures, we can draw the conclusions as follows:

i) Even with the measurements only from the thermal flux, the GEIM algorithm proposed in Section 3.4 is able to reconstruct the fast flux, thermal flux and the power distribution with almost the same accuracy, which proves, a posteriori, the bijectivity of the application $\mu \mapsto \phi_2(\mu)$.

ii) The errors on the test set $\mathcal{M}^{(test)}$ are almost identical to the errors on the training set $\mathcal{M}^{(training)}$ suggests that the training set is large enough to present the control rods movements process(for the specified IAEA 2D benchmark).

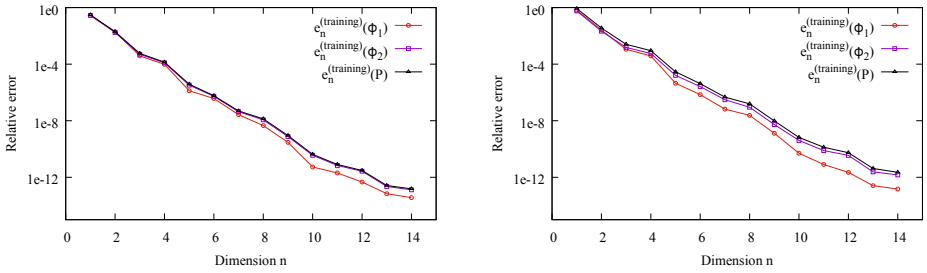
iii) The very fast decay rate of the reconstruction errors shows that the Kolmogorov n-width of the manifold decays rather fast with n. From this point of view, the flux/power fields set produced by the control rods movements process are rather regular, and only a few number of basis and measurements can provide very high accurate reconstructions.

iv) The very fast decay rate for Case II also confirms that, this methodology is of engineering feasibility for control rods movements problem.

4.2.5 Conclusions

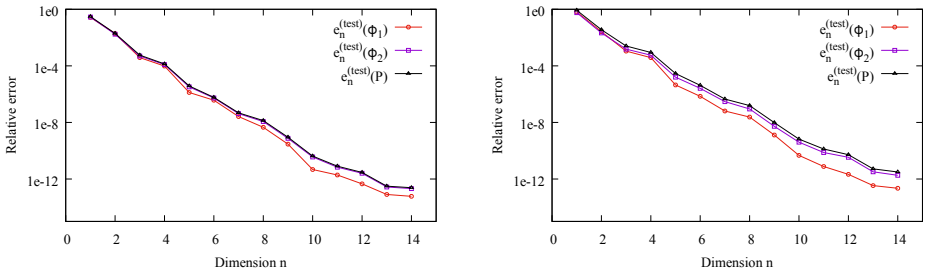
To summarize, we applied the GEIM to the control rods movements problem to improve the flux and the power distribution knowledge in the whole domain, without knowing the control rods

position (step) parameters thanks to the measurements. To reach this goal, we first construct a set of snapshots by solving large amounts of PDEs for the underlying physical problem. Then we select the sensors and the basis functions in a greedy way with GEIM greedy algorithm. With the optimized measurements, we are able to reconstruct the flux and power distribution in high accuracy. Numerical results confirm that the proposed GEIM methodology is hopefully acceptable for control rods movements problem, additionally, the online-offline computational framework is good for on-line monitoring purpose.



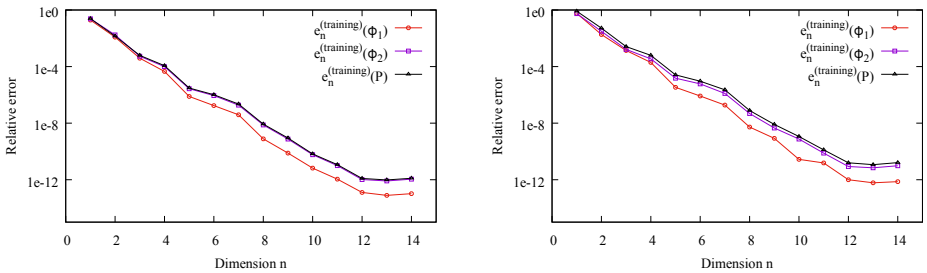
(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 4.2.8: Cases I: Relative errors $e_n^{(\text{training})}(\phi_1)$, $e_n^{(\text{training})}(\phi_2)$ and $e_n^{(\text{training})}(P)$.



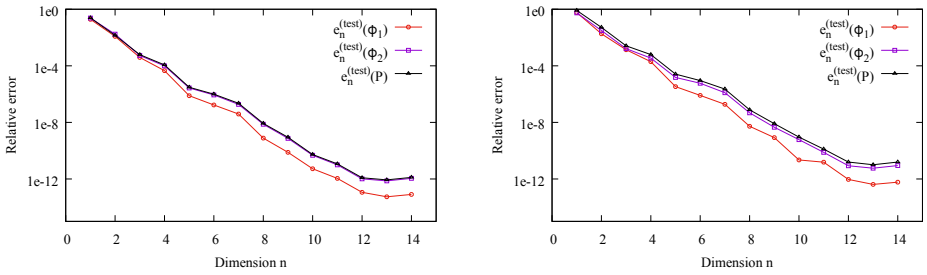
(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 4.2.9: Cases I: Relative errors $e_n^{(\text{test})}(\phi_1)$, $e_n^{(\text{test})}(\phi_2)$ and $e_n^{(\text{test})}(P)$.



(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 4.2.10: Cases II: Relative errors $e_n^{(\text{training})}(\phi_1)$, $e_n^{(\text{training})}(\phi_2)$ and $e_n^{(\text{training})}(P)$.



(a) $L^2(\Omega)$ norm (b) $L^\infty(\Omega)$ norm

Figure 4.2.11: Cases II: Relative errors $e_n^{(\text{test})}(\phi_1)$, $e_n^{(\text{test})}(\phi_2)$ and $e_n^{(\text{test})}(P)$.

4.3 Field reconstruction on EDF nuclear reactor cores

4.3.1 Introduction

Nowadays, the two-group and few-group neutron diffusion theory continues to be the workhorse for practical LWR core calculations because of its relatively low computational cost with respect to the transport theory. Increasingly coarse mesh nodal methods, such as Analytic Nodal Method (ANM), Semi-Analytic Nodal Method (SANM), or Nodal Expansion Method (NEM) have been widely used in the analysis of nuclear reactor cores [197, 196, 214, 181, 88]. These methods provide only averaged values on the relatively large mesh nodes, while for the safety analysis of the reactor, it is necessary to know the distribution of the power density in a Pin-by-Pin wise way. In order to obtain the detailed distributions, the average values of the nodal calculations are used to reconstruct the pin power distribution [122]. The reconstruction process involves a fundamental assumption, that is *the detailed Pin-by-Pin distributions inside a fuel assembly can be estimated by the product of a homogeneous flux distribution by local heterogeneous form function.* While the homogeneous flux distribution takes into account the effects of fuel assembly on the vicinity of the core, the form function loads the heterogeneity of fuel assembly (pin fuel, water holes, burnable poison pins, etc.).

It has only been possible recently with the availability of petascale computing that LWR researchers have focused considerable efforts on performing whole-core LWR calculations using higher order transport methods [127, 116, 154, 50]. Although several 3D deterministic transport methods [135, 126, 134, 198] have been developed, it is not easy to apply thenm to realistic whole-core problems due to accuracy and memory problems. Among them, the Discrete Ordinates (Sn) method [139] is the most widely used to solve the 3D transport equation. However, this method has difficulty in treating curvilinear geometry arising from heterogeneity. During the last several years, the Method of Characteristics (MOC) has been revisited and refined to apply to 2D/1D whole core calculations [135, 126, 134, 198, 58, 57], that is a synergistic combination of the method of characteristics for radial 2D calculation and the Sn-like method for axial 1D calculation. More recently, a direct 3D MOC approach has been studied [121], and the computational burden is considerable for practical LWR applications. Additionally, the EDF core code COCAGNE [50] also implemented the state-of-the-art flux solvers i.e. simplified transport solver (SPn) and Discrete Ordinates transport solver.

Even though the direct 3D transport approach is possible, the real-time computation — especially for on-line monitoring purpose, or many-queries computation for design optimization or safety evaluation purposes — is still full of challenges because of tremendous amounts of memory and long computing time for the whole core calculation. The data assimilation with reduced basis (e.g. GEIM) provides another possibility to estimate the state in Pin-by-Pin wise instead of running

the transport code directly in this scenario. At any rate, we emphasize that the methodology cannot completely replace the high performance computing, but provides a computational paradigm that combines the high performance computing and advanced reduced order modeling techniques.

4.3.2 Problem setting

For the practical reactor core application, we chose an operating PWR 1450 MWe reactor core that have been used in [59]. Our goal is to reconstruct the pin flux/power distribution for the whole life cycle of the reactor. To each time of the life cycle is associated a state of the fields, which is modeled by the solution to the following transport problem

$$\mathcal{F}(\phi_1, \phi_2, P, \mu) = 0, \tag{4.3.1}$$

with engineering boundary conditions and specified macroscopic parameters μ, where $\mathcal{F}$ is the transport operator, one can refer to (3.2.4) in Section 3.2.2 for more information. We consider two main macroscopic parameters:

- Pw: the power level of the reactor core. It ranges between 0.3 and 1.
- Bu: the average burnup of the fuel in the whole core. It ranges between 0 (for the beginning of the life cycle) and $Bu_{\max}$ (the end of the life cycle) and its exact evolution depends on the operating history of the reactor.

Therefore, in this case, μ depends on the vector of "general" parameters

$$\mu \coloneqq (Pw, Bu)$$

that range in

$$\mathcal{D} := [0.3, 1] \times [0, Bu_{max}].$$

So the manifold of solutions is

$$\mathcal{M}_\Psi \coloneqq \{\Psi(\mu) = (\phi_1, \phi_2, P)(\mu) \ : \mu \in \mathcal{D}\}. \tag{4.3.2}$$

The sampling of parameters over the space $\mathcal{D}$ is done on a regular mesh, and $50 \times 50 = 2500$ snapshots are generated with COCAGNE transport solver (Pin-by-Pin calculation) to be the training set $\mathcal{M}^{(training)}$. Then the reduced basis and sensor locations are selected with the greedy algorithm in Section 3.4.2. In order to evaluate the quality of reconstructions, a randomly sampled set of snapshots $\mathcal{M}^{(test)}$ are also generated.

For comparison purposes, we design the following approaches:

- 2D Pin-by-Pin case. We first integrate the snapshots over the vertical axis and run the greedy algorithm in the resulting 2D fluxes and power distributions.
- 3D Pin-by-Pin case. We work on the 3D Pin-by-Pin snapshots directly.

In both cases, the measurements can be averages over the assembly nodes (e.g. on node can be one whole assembly or a quarter of an assembly, etc.)

$$\ell_n(\phi_2) \coloneqq \frac{\int_{\mathrm{Vol}^{ass}(n)} \phi_2 \, \mathrm{d}\Omega}{\mathrm{Vol}^{ass}(n)}, \tag{4.3.3}$$

or averages over the pin-nodes

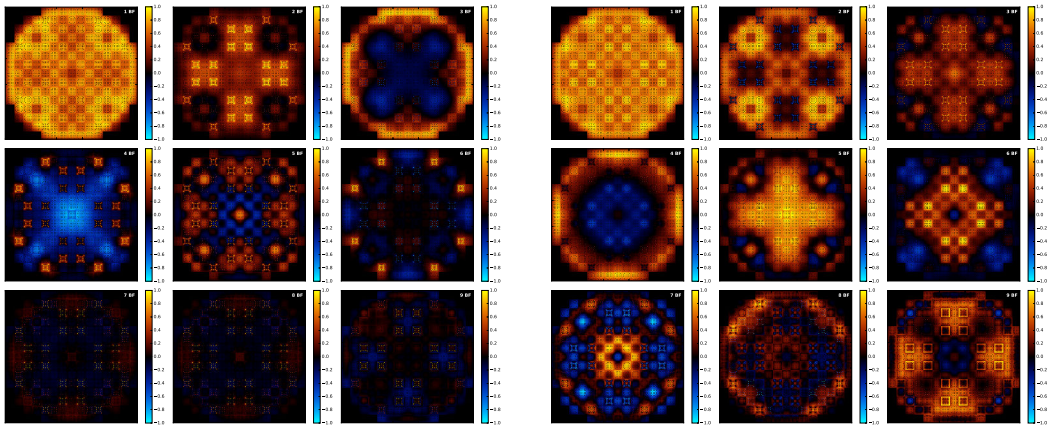
$$\ell_n(\phi_2) := \frac{\int_{\mathrm{Vol}^{pin}(n)} \phi_2 \, \mathrm{d}\Omega}{\mathrm{Vol}^{pin}(n)}, \tag{4.3.4}$$

where $\mathrm{Vol}^{ass}(n)$ and $\mathrm{Vol}^{pin}(n)$ denote the volume of the n-th assembly node and pin node in the reactor respectively. For comparison reason, we denote the approach based on the assembly node wise measurements by *GEIM*, and denote the approach based on the pin node wise measurements by *EIM* though the latter is different from EIM as in [21].

4.3.3 Numerical results and analysis

The 2D case

Let us now turn to the analysis of the results. We first give in Figure 4.3.12 the first nine basis in 2D case, corresponding to the pin node wise and assembly node wise cases. For the assembly node wise case, the measurements are averages over each assembly as a whole. We notice here that, there is no big difference for the first seven basis functions between the above two cases. With Pin-by-Pin measurements, the reduced basis functions are able to keep the high order information of the fields, and the assembly wise approach can not.



(a) Assembly wise (b) Pin wise

Figure 4.3.12: The first nine basis functions for pin node wise measurements and assembly node wise measurements in 2D case.

We apply the relative errors ($e_n^{(\text{training})}(\cdot)$ and $e_n^{(\text{test})}(\cdot)$ in L^2 and L^∞ norms defined in Section 3.5) to the training and test sets in order to evaluate the performance of the reconstruction.

We show in Figure 4.3.13 the L^2 errors on the training set for pin node wise (EIM) and assembly node wise (GEIM) cases. For comparison purposes, we also list the best estimation case, i.e., the orthogonal projection of the field on the reduced space of the same dimension which is built with POD method (denoted by POD). We observe that, the manifold — even in Pin-by-Pin wise — which reflects the burnup and power evolution of the reactor core of the whole life is rather regular. This makes it possible to approximate the field with only a few basis functions and measurements, e.g., with only 20 basis functions, one can reach an accuracy of 10^{-4}, even with assembly wise measurements (see 'GEIM' in Figure 4.3.13). The reconstruction with GEIM is a little worse than the best estimation with POD method (the difference is less than one order). Additionally, a smaller size measurement system (EIM) is better than a large size one, but the difference only appears when

the number of the reduced basis increases up to some extent (e.g. the reduced dimension $n \geq 20$ in our case).

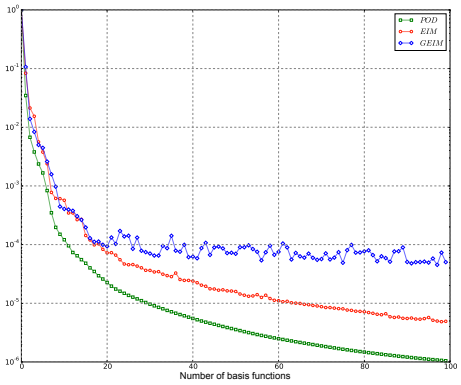
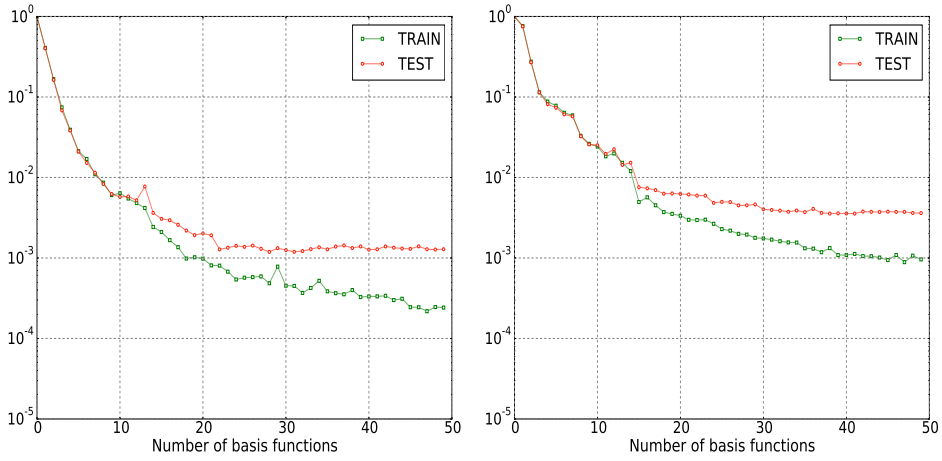


Figure 4.3.13: Variations of the L^2 relative errors with respect to the reduced dimension n for EIM, GEIM and POD in 2D case.



(a) L^2 norm (b) L^∞ norm

Figure 4.3.14: Variations of the L^2 and L^∞ relative errors with respect to the reduced dimension n with pin node wise measurements in 3D case.

The 3D case

We now turn to the analysis of the 3D case. We first study the case where the measurements the averages on the pin nodes. We first present the L^2 and L^∞ errors on the training and test set in Figure 4.3.14, the measurements are based on the pin node values. We notice that, within only 20 basis and measurements, we reach an accuracy of 0.1%, which is one order worse than the 2D case as shown before. The difference between test errors and training errors also reflects that the snapshots set $\mathcal{M}^{(training)}$ which is used to derive the reduced basis is not fine enough, i.e., for practical usage of GEIM for the underlying problem, a finer snapshot is necessary. Though the 3D case is a little worse than the 2D case, the 0.1% relative error is still acceptable for engineering applications.

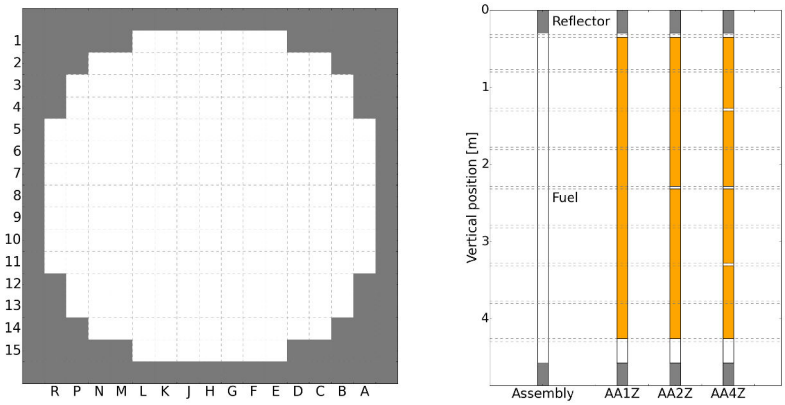


Figure 4.3.15: The instrument frame inside the core. The sensor can be placed in any assembly of the core with different axial segment number (AA1Z for 1, AA2Z for 2 and AA4Z for 4).

$$\ell_n(\phi_2) \coloneqq \frac{\int_{\mathrm{Vol}^{ass}(n)} \phi_2 \, \mathrm{d}\Omega}{\mathrm{Vol}^{ass}(n)}, \tag{4.3.5}$$

In reality, the measurements are acquired from assembly node wise sensors rather than pin node wise sensors. Thus we design the measurements with assembly node $\mathrm{Vol}^{ass}(n)$ in (4.3.5) as follows:

i) $\mathrm{Vol}^{ass}(n) \sim \mathrm{Vol}^{assambly}$, i.e., the measurement is the average over the whole assembly.

ii) $\mathrm{Vol}^{ass}(n) \sim \frac{1}{2}\mathrm{Vol}^{assambly}$, i.e., the measurement is the average over half of an assembly.

iii) $\mathrm{Vol}^{ass}(n) \sim \frac{1}{4}\mathrm{Vol}^{assambly}$, i.e., the measurement is the average over a quarter of an assembly.

We show the admissible sensor placement framework of the above three sizes in Figure 4.3.15. The left shows the radial distribution of the assemblies and the right shows the axial distribution of the three kinds of sensors, which are denoted by *AA1Z*, *AA2Z* and *AA4Z* respectively. Note that we already take the grids of the fuel assemblies into consideration, thus there is larger blank in the bottom than in the top of each assembly.

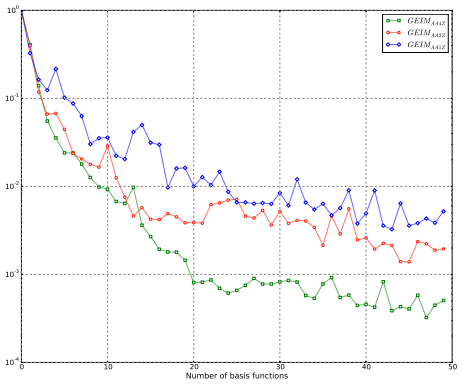


Figure 4.3.16: Variations of the L^2 relative errors with respect to the reduced dimension n for different measurements ($\mathrm{Vol}^{ass}(n) \sim \mathrm{Vol}^{assambly}$, $\frac{1}{2}\mathrm{Vol}^{assambly}$, $\frac{1}{4}\mathrm{Vol}^{assambly}$) in 3D case.

We show in Figure 4.3.16 the L^2 relative errors on the training set for the three cases. As is expected, the reconstruction with 4 measurements in the axial direction provides the highest accuracy than others. We also list the sensor locations for this case in Figure 4.3.17. Note that Figure 4.3.17 can be a guide for sensor placement in Pin-by-Pin framework, and it's very close to the real engineering case.

The comparison of the L^2 errors on the training set between GEIM (with assembly node wise measurements, the AA4Z case) and EIM (with pin node wise measurements) is shown in Figure 4.3.18. We observe that the AA4Z case is rather close to the EIM case, which also shows an accuracy of 0.1%. Recall that the AA4Z case is very close to the real case for sensor placement, i.e., we do not lose too much accuracy by considering the engineering constraints.

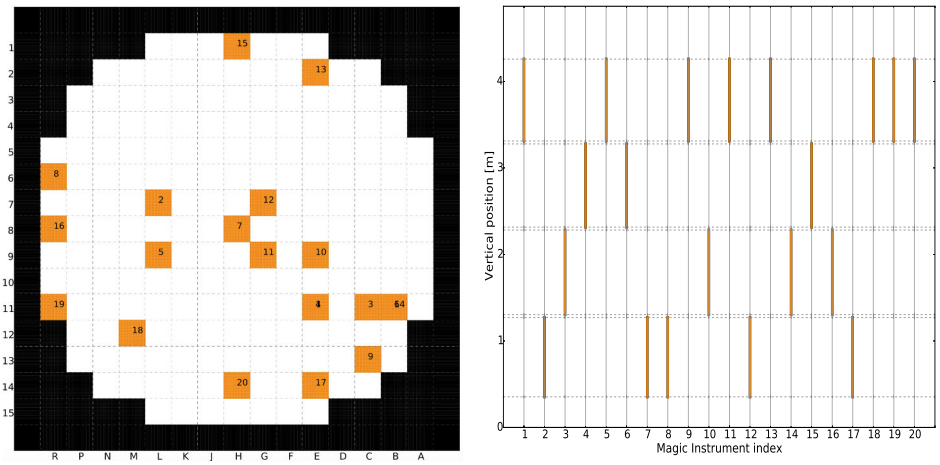
In addition, we show the comparison of the L^2 errors on the training set and test set for GEIM in Figure 4.3.19. Even though the training set is not fine enough, the relative error of the reconstruction on the test set is still below 1%, this 'methodology error' is negligible with respect to the realistic measurement error which ranges between 1% and 5%. Indeed, for the underlying problem, enrich the training set is necessary in order to improvement the quality of the reconstruction.

4.3.4 Computational cost

Table 4.1 gives some computational times to illustrate the cost of the greedy algorithm (computed in an offline stage) and the reconstruct with the reduced basis (online stage).

Items	Off-line	On-line
One snapshot calculation	46	
Prepare $\mathcal{M}^{(training)}$, 2500 snapshots	115000	
Select 50 basis functions and sensors with GEIM greedy algorithm	5328	
Field reconstruction with 5 basis functions		0.05
Field reconstruction with 15 basis functions		0.25

Table 4.1: Computational cost (in seconds) of GEIM with PWR 1450 MWe reactor core.



(a) Radial (b) Axial

Figure 4.3.17: Sensor locations of GEIM for AA4Z case.

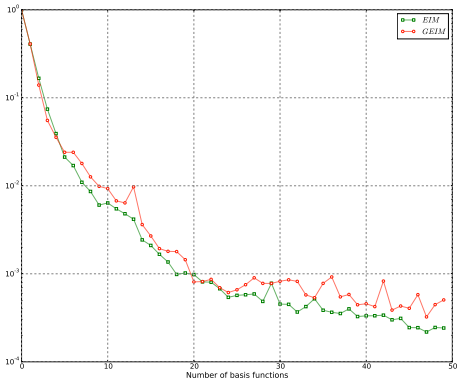


Figure 4.3.18: Variations of the L^2 errors on the training set for GEIM (with assembly node wise measurements i.e. the AA4Z case) and EIM (with pin node wise measurements).

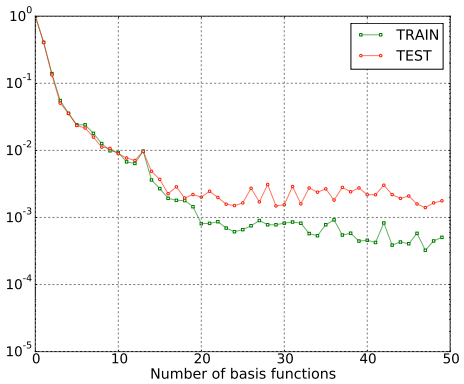


Figure 4.3.19: Variations of the L^2 errors of GEIM (with assembly node wise measurements i.e. the AA4Z case) for the training and test sets.

4.3.5 Conclusions

In this section, we applied the GEIM method adapted in Section 3.4.2 to reconstruct the Pin-by-Pin flux/power distribution during the life cycle of a nuclear reactor core. All the simulations are based on the PWR 1450 MWe nuclear reactor operated by EDF. Numerical result shows that, the manifold of the underlying problem (even in Pin-by-Pin wise), which reflects the burnup and power evolution of the reactor core of the whole life, is rather regular. This makes it possible to reconstruct the fields with only a few basis functions and measurements. Furthermore, by considering the measurement system which is close to the real engineering case, the proposed method still provides an acceptable accuracy.

The extension of the method is to take into account the effect of the noise and qualify its impact when few instruments remains and also to control the noise amplification through various technique, either data pre-processing or mathematical correction of the induced error, will be studied in the next section.

4.4 Reconstruction with noisy measurements

4.4.1 Introduction and problem setting

The examples dealing with PWR illustrate the quality of the field reconstruction given by GEIM. This proves that GEIM has the ability to learn the essential mechanisms of the physical system in order to provide good indications on sensor locations and good reconstructions. In previous sections, we work with synthetic measurements and it would be desirable to consider real data but there are still some obstructions to address for the methodology. One of the most important problems to address is that measurements are usually noisy.

In this section, we will study the noisy measurements case with respect to GEIM, PBDW, for nuclear applications, namely, the control rods movements problem. The problem setting can be found in Section 4.2. For the measurement noise, we consider two kinds of independent and identically distributed (IID) noise:

i) With respect to the upper limits of the reconstruction error, we assume that the observations suffer from a uniform distributed noise $e \in [\sigma_{min}, \sigma_{max}]$.

ii) With respect to general statistical performance of the reconstruction, we assume that the observations suffer from a Gaussian noise with probability density function $p = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e-\bar{e})^2}{2\sigma^2}}$, with mean $\bar{e}$ and standard deviation σ.

Besides, the localized observation is of particular interest in this work, for the sake of convenience, we set Gaussian convolution as the measurement functionals:

$$\ell_i(u) = \text{Gauss}(u, x_i^c, s_i) \equiv \int_\Omega \Big\{ (2\pi s_i^2)^{-d/2} \exp\Big(- \frac{\|x - x_i^c\|^2_{\ell^2(\mathbb{R}^d)}}{2s_i^2} \Big) \Big\} u(x) dx, \tag{4.4.1}$$

where $x_i^c \in \mathbb{R}^d$, $d = 2$ is the center of the detector in 2D case, and $s_i \in \mathbb{R}_{>0}$ is the filter width of the detector. We work with $L^2(\Omega) = \{u \mid \|u\|_{L^2(\Omega)} < \infty\}$ where the norm $\|\cdot\|_{L^2(\Omega)}$ is induced by the inner product $(w, v)_{L^2(\Omega)} = \int_\Omega w(x)v(x)dx$. In this setting, examples of the Riesz representation w_i of the functional ℓ_i can be found in Figure 4.4.20, for three different values of s.

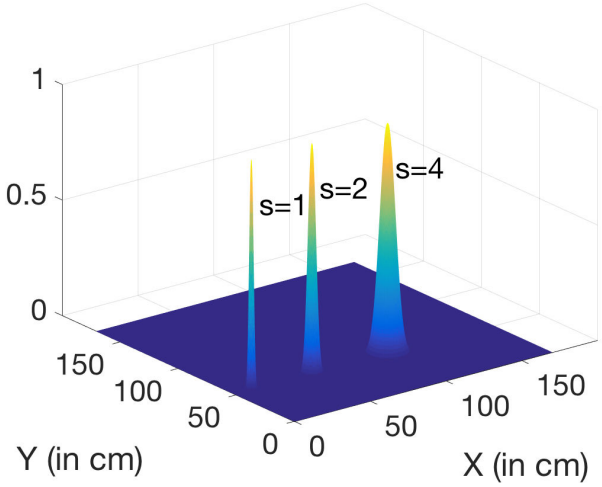


Figure 4.4.20: Example (4.4.1): plots of the Gaussian convolution as the measurement functionals for three different values of sensor width $s = 1, 2, 4$ cm.

To this end, we recall some notations introduced in the previous chapter: V is a Hilbert space over a domain $\Omega \subset \mathbb{R}^d$ ($d \geq 1$) endowed with inner product $(\cdot, \cdot)$ and induced norm $\|\cdot\| = \sqrt{(\cdot, \cdot)}$, $V_n \in V$ stands for the reduced space of dimension n, and $W_m \in V$ stands for the observation space spanned by the Riesz representations of the m functionals $\{\ell_i\}_{i=1}^m$.

4.4.2 Stability analysis

Our interest is the stability of GEIM for the control rods movements problem. More generally, based on the knowledge of Chapter 2, we hope to stabilize the reconstruction and also increase the accuracy by adding more measurements. Thus we analyze the general recovery problem with reduced basis and noisy measurements as described in Problem 2.1.1, we recall it here for the sake of convenience:

Problem 4.4.1 (Recovery with reduced basis and noisy measurements, RRB) *Let* $\mathcal{M} := \{u^t(\mu) : \mu \in \mathcal{D}\}$ *be the solution manifold of a physical problem, where* $\mathcal{D}$ *is the parameter set. Assume that we are given a series of linearly independent vectors* $(q_i)_{i\in\mathbb{N}} \in V$ *which span a sequence of reduced space* $V_n = \text{span}\{q_1, \cdots, q_n\} \in V$ *of dimension* n*, and linear measurements* $\boldsymbol{y}^o = \boldsymbol{\ell}(u^t) + \boldsymbol{e}^o = (\ell_1(u^t) + e_1, \cdots, \ell_m(u^t) + e_m)^T$ *of* $u^t \in \mathcal{M}$*. The recovery problem is: find* $v \in V_n$ *from observations* $\boldsymbol{y}^o$ *to approximate the true state* u^t*.*

The field u^t above can be ϕ_1, ϕ_2 or P, as already shown in Section 4.2, there is not so much difference for the reconstructions of the three fields with the measurements from thermal flux. Later we will show that in noisy case, the reconstruction errors of ϕ_1, ϕ_2 or P also follow the same behavior, thus we will only study the reconstruction of thermal flux in this section for the sake of brevity. Note that there are two different approaches introduced in Chapter 2 for the above recovery problem, we recall them here also for the sake of convenience:

i) The LC approach (2.2.7): find $v^*_{LC} \in V_n$ s.t.

$$v^*_{LC} = \arg\min_{v\in V_n} \|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_{\mathbb{W}} , \tag{4.4.2}$$

where the Gramian matrix $\mathbb{W}$ is defined with entries $\mathbb{W}_{i,j} = \ell_i(w_j) = (w_i, w_j),\ i, j = 1, \cdots m$.

ii) The LS approach (2.2.44): find $v^*_{LS} \in V_n$ s.t.

$$v^*_{LS} = \arg\min_{v\in V_n} \|\boldsymbol{\ell}(v) - \boldsymbol{y}^o\|_2 . \tag{4.4.3}$$

To evaluate the stability, we will use the stability factor $\|Q\|$ defined in (2.2.1) as a measure of stability with respect to the measurement noise, and use the Lebesgue constant Λ defined in (2.2.2) as a measure of stability with respect to the model bias. Many factors may affect the performance of the field reconstruction, such as i) the size of the sensors, namely the width s in (4.4.1); ii) the amount of reduced basis functions used to reconstruct the fields; iii) the amount of measurements used to reconstruct the fields; iv) the placement of the sensors, etc. Thus, the stability analysis is carried out to understand the effects of each factors mentioned above.

Now we work with spaces V_n and measurements $\{\ell_i\}_{i=1}^m$ which are determined as follows:

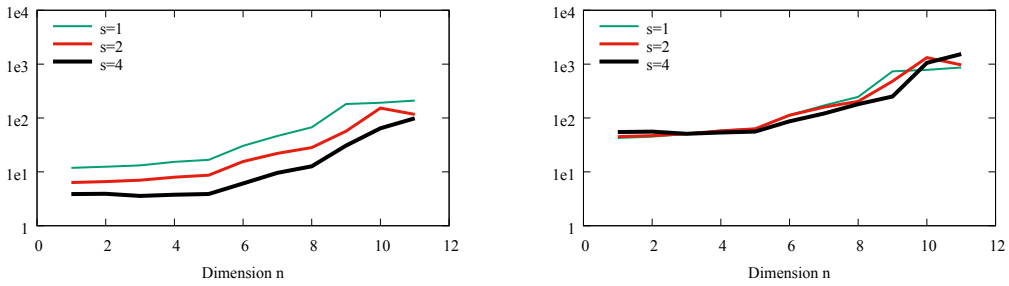
- V_n is spanned by the first n basis which are built with GEIM.
- The first n measurement functionals are from GEIM, and the remaining $n+1$ to m functionals are selected with *Op* $\|Q\|$ *method* which aims at minimizing the stability factor $\|Q\|$, see Algorithm 2.2.5.

Note that one can also use *Op* Λ *method* as stated in Algorithm 2.2.5 to select the $n+1$ to m functionals in order to get a better Lebesgue constant. Numerical results already confirmed that the stability factor $\|Q\|$ is more sensitive than the Lebesgue constant, thus *Op* $\|Q\|$ *method* is of interest. Furthermore, we will work with three typical sensors with different values of s as follows:

- $s = 1$ *cm* which corresponds to pin wise measurement.
- $s = 2$ *cm* which corresponds to a medium size measurement.
- $s = 4$ *cm* which corresponds to sub-assembly wise measurement.

Stability performance of GEIM

We first show in Figure 4.4.21 the variations of the Lebesgue constant Λ and the stability factor $\|Q\|$ with respect to the reduced dimension n for different values of sensor width, i.e., $s = 1, 2, 4\ cm$. We observe that, the stability factor $\|Q\|$ is one order higher in magnitude than the Lebesgue constant Λ, which means that GEIM algorithm is more sensitive with respect to the noisy measurements than with respect to the mode error, i.e., the error caused by using an element from the reduced spave V_n to approximate the true field. Furthermore, we observe that both the Lebesgue constant and stability factor explosion with the increase of the reduced dimension, though with higher dimension of reduced space one is able to reach a very high accuracy with noise-free measurements, as already shown in Section 4.2. That means the recovery algorithms will be not robust with respect to i) the model mismatch because of Λ diverges, and ii) the noisy measurements because of $\|Q\|$ diverges. At last, we note that with a larger values of sensor width, we get a better Lebesgue constant.



(a) Lebesgue constant Λ.

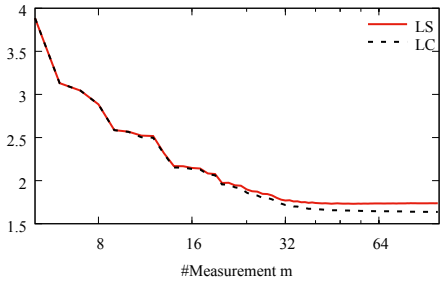
(b) Stability factor $\|Q\|$.

Figure 4.4.21: Variations of Lebesgue constant Λ and stability factor $\|Q\|$ of GEIM with respect to the reduced dimension n for different values of sensor width s.

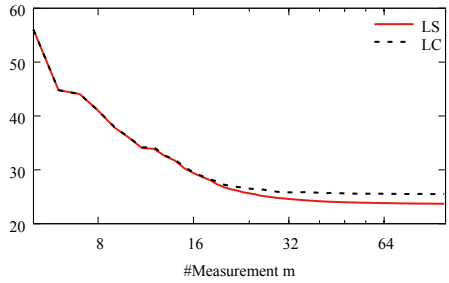
Stability with respect to LC approach and LS approach

We now study the effects of different reconstruction approaches, namely, LC approach (4.4.2) and LS approach (4.4.3). From Figure 4.4.21, we observe that when the reduced dimension n is no greater than 5, the behavior of Lebesgue constant and stability factor are rather stable, the diverging start when $n > 5$. Thus we select two typical reduced space with V_5 and V_{10} for further analysis.

We show the reconstruction errors with respect the number of measurements in Figure 4.4.22 and Figure 4.4.23 for two reduced spaces V_5, V_{10} for the two approaches with sensor width $s = 4$. We observe that, with the increase of the number of measurements, both the Lebesgue constant and stability factor decay very fast and stabilize to some extent, for the two recovery approaches. The LC approach shows better performance in controlling the Lebesgue constant, on the other hand, the LS approach shows better performance in controlling the stability factor. Even though there is difference between LC approach and LS approach, for the underlying control rods movements problem, the effect by adding more measurements to improve the quality of the reconstruction is much more significant than by selecting a suitable reconstruction method.

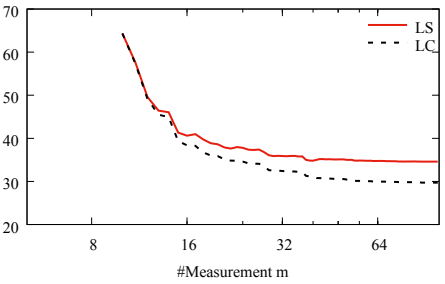


(a) Lebesgue constant Λ with V_5

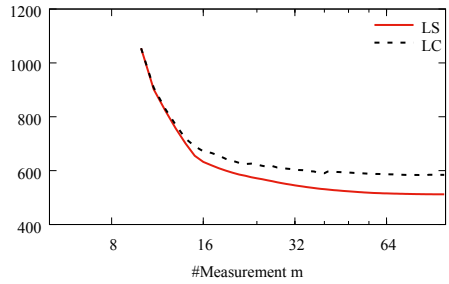


(b) Stability factor $\|Q\|$, with V_5

Figure 4.4.22: Variations of Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the number of measurements, for LC and LS approaches with sensor width $s = 4$ for V_5.



(a) Lebesgue constant Λ, with V_{10}



(b) Stability factor $\|Q\|$, with V_{10}

Figure 4.4.23: Variations of Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the number of measurements, for LC and LS approaches with sensor width $s = 4$ for V_{10}.

Stability with respect to the measurements

As already known from Figure 4.4.22 and Figure 4.4.23 that by adding more measurements, the Lebesgue constant and stability factor are improved. We show in Figure 4.4.24 and 4.4.25 the variations of the two factors for sensor width $s = 1, 2, 4$ *cm* for LS approach. We observe that a large s leads a better Lebesgue constant but a worse stability factor, vice versa.

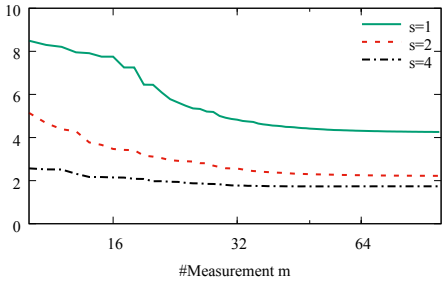
Finally, we draw the following conclusions that, in order to control the noise for the field reconstruction with noisy observations, one would

i) use sensors with relatively small sensor width,

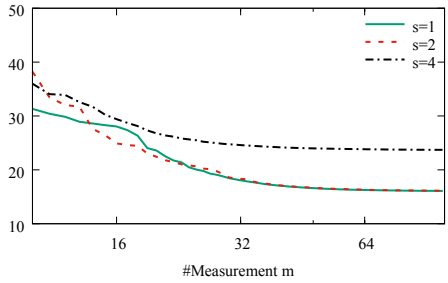
ii) add more measurements but to some extent because of the fast decay thanks to the *Op* $\|Q\|$ *method*,

iii) and use the least-square approach,

in order to control the noise amplification. Besides, the field reconstruction for a large dimensional reduced space V_n, $n > 5$ is still unsatisfactory (e.g. even for the best case in Figure 4.4.23b, the stability factor is of order 10^2 in magnitude). We will address this issue in the next section with CS-GEIM proposed in Chapter 2 (see Definition 2.3.1).

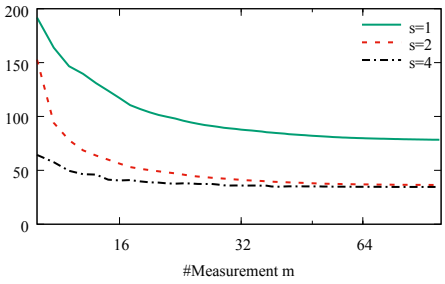


(a) Lebesgue constant Λ with V_5

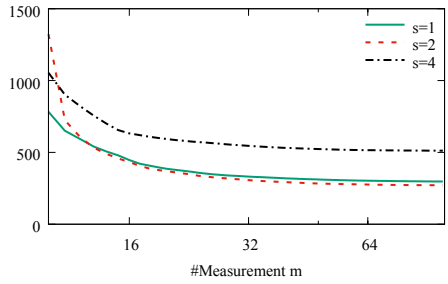


(b) Stability factor $\|Q\|$, with V_5

Figure 4.4.24: Variations of Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the number of measurements for LS approach, for three different values of sensor width s for V_5.



(a) Lebesgue constant Λ, with V_{10}



(b) Stability factor $\|Q\|$, with V_{10}

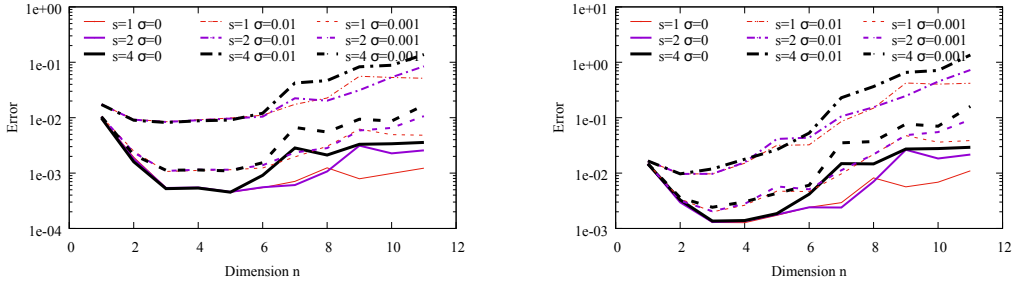
Figure 4.4.25: Variations of Lebesgue constant Λ and stability factor $\|Q\|$ with respect to the number of measurements for LS approach, for three different values of sensor width s for V_{10}.

4.4.3 CS-GEIM

As already shown in previous section, Λ and $\|Q\|$ diverge with the increase of the reduced dimension n. In order to study the effects of Λ and $\|Q\|$ to noisy measurement and model mismatch, we set the goal to reconstruct the biased field, i.e., $u \in \mathcal{M}_b$, where b stands for *bias*. The biased solution set $\mathcal{M}_b$ is built with parameter $\Sigma_{a,2}|_{\Omega_{control}}$ in the domain $\mathcal{D} := [0.080, 0.150]$, and other coefficients are given in Table 3.2, but the diffusion coefficient in reflector region (Region 4) is changed to $D_1|_{ref} = 1.995$ in order to make the biased snapshots. The reduced basis and sensor locations are derived from the original manifold $\mathcal{M}$ where $D_1|_{ref} = 2.0$ with GEIM algorithm.

We first study the case with uniformly distributed noise bounded by $\sigma = 0.01$ or 0.001 for each sensor, and apply the LS approach (4.4.3) to reconstruct the fields (we denote here 'GEIM' to highlight the difference with CS-GEIM, when $m = n$, the LS approach is exact GEIM). We emphasize that the first n sensor locations are selected with GEIM greedy algorithm, and the left $n+1$ to m one are selected with *Op* $\|Q\|$ *method*. When the sensor suffers from noise, we simulate the reconstruction many times for different noise values of the same distribution and mark the maximum relative error in L^2 and L^∞ norms.

In Figure 4.4.26, we present the variations of the errors with respect to the reduced dimension n for sensor width $s = 1, 2, 4$ (dotted line). For all cases, we set the number of measurement to $m = 10n$. For comparison purposes, we also present the errors for noise-free measurements (solid line).

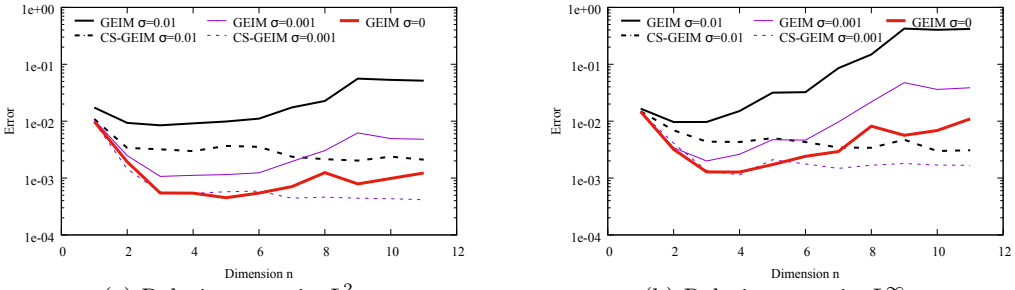


(a) L^2 norm (b) L^∞ norm

Figure 4.4.26: Variations of the error of GEIM with respect to the reduced dimension n for different values of sensor width s for LS approach with $m = 10n$.

From Figure 4.4.26, we observe that, in noise-free case, the quality of the reconstruction is acceptable only when the reduced dimension is small, i.e., $n < 6$. The model bias is amplified with the Lebesgue constant diverges when n is relatively large. The dotted lines in Figure 4.4.26 show that, if the measurements suffer from noise, the reconstructions are much more pessimistic because of the diverging of the stability factor $\|Q\|$. Furthermore, by changing the sensor size or adding more measurements doesn't improve the stability performance so much.

Thus the CS-GEIM is proposed in order to further control the noise. The method is stated in Definition 2.3.1. Figure 4.4.27 shows the comparison between CS-GEIM and GEIM. Here we only list the case with sensor width $s = 1$. From this figure we observe that, with CS-GEIM, the noise and the model mismatch are both well controlled. If the noise level (e.g. $\sigma = 0.01$) is larger than the model mismatch , the error with CS-GEIM can be lower than the noise level by adding more measurements. If the noise level (e.g. $\sigma = 0.001$) is lower than the model bias, the error with CS-GEIM can reach the same level to the model bias, but it is unable to correct the model bias. In a word, the reconstruction with CS-GEIM can reach an accuracy to the noise level or model bias.

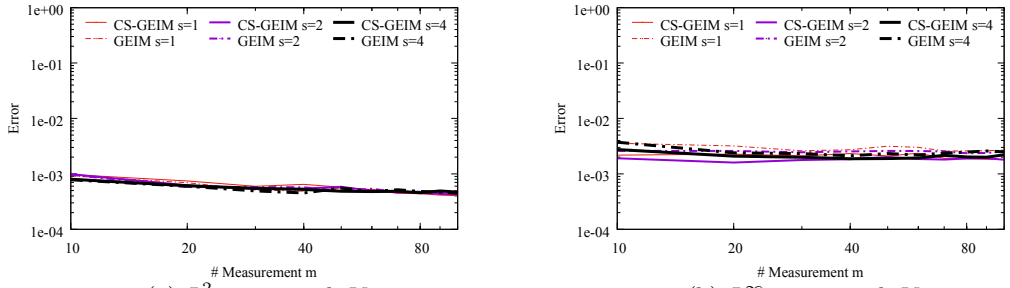


(a) Relative error in L^2 norm (b) Relative error in L^∞ norm

Figure 4.4.27: Variations of the error of GEIM and CS-GEIM with respect to the reduced dimension n for different noise levels with $m = 10n$.

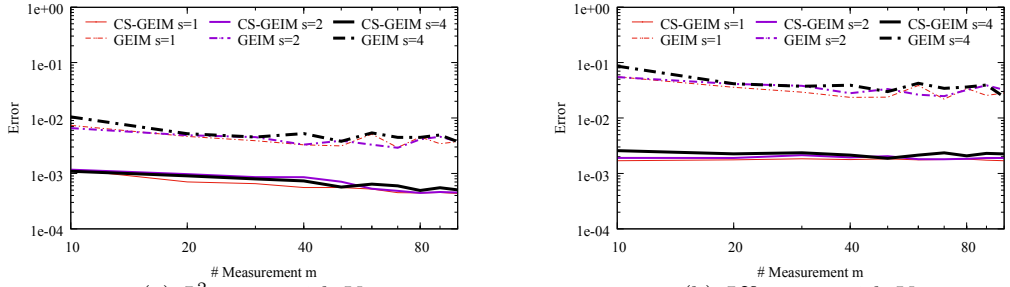
We now study the two typical cases: i) reconstruction based on V_5, which corresponds the stable Lebesgue constant and stability factor, and ii) reconstruction based on V_{10} which corresponds the unstable Lebesgue constant and stability factor. We first show in Figure 4.4.28 the errors of GEIM and CS-GEIM with respect to the number of measurement m for different values of sensor width, i.e., $s = 1, 2, 3\ cm$, the noise level $\sigma = 0.001$ and the reduced space is V_5. We confirm that, when the Lebesgue constant and stability factor are in a relatively 'good' state, there is not so much difference between CS-GEIM and GEIM. In this case, the reconstruction can be improved by adding more measurements. But Figure 4.4.29 shows that, when the Lebesgue constant and stability factor are

rather 'bad', the constrained stabilized version of GEIM is much better, in this case, by adding more measurements does improve the quality, but CS gains more.



(a) L^2 norm, with V_5. (b) L^∞ norm, with V_5.

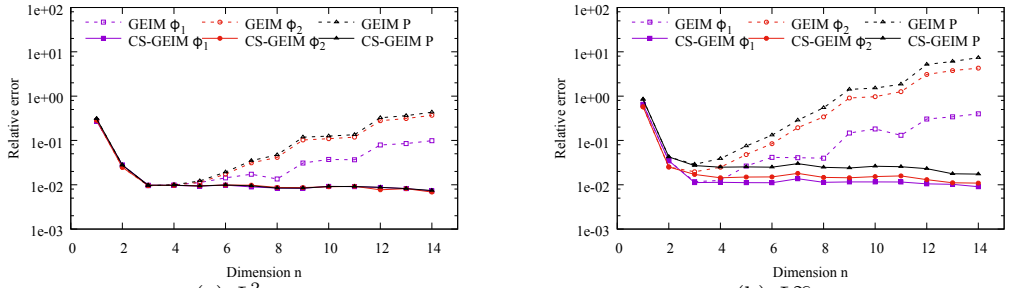
Figure 4.4.28: Variations of the errors with respect to the number of measurement m for GEIM and CS-GEIM for different values of sensor width s with noise level $\sigma = 0.001$ and reduced space V_5.



(a) L^2 norm, with V_{10}. (b) L^∞ norm, with V_{10}.

Figure 4.4.29: Variations of the errors with respect to the number of measurement m for GEIM and CS-GEIM for different values of sensor width s with noise level $\sigma = 0.001$ and reduced space V_{10}.

To this end, we show in Figure 4.4.30 the relative L^2 and L^∞ errors of ϕ_1, ϕ_2 and P with respect to the reduced dimension n for GEIM and CS-GEIM, for sensor width $s = 1$ and noise level $\sigma = 0.01$ (for CS-GEIM, the number of measurements $m = n$). From the figure we confirm that, in the noisy case, the performances of the noise amplification for the reconstructions of ϕ_1, ϕ_2 and P for both GEIM and CS-GEIM are coincident. Thus, we only study the reconstruction of ϕ_2 for brevity.



(a) L^2 norm. (b) L^∞ norm.

Figure 4.4.30: Variations of the recovery errors of ϕ_1, ϕ_2 and P with respect to the reduced dimension n for GEIM and CS-GEIM, for sensor width $s = 1$ and noise level $\sigma = 0.01$.

4.4.4 CS-PBDW

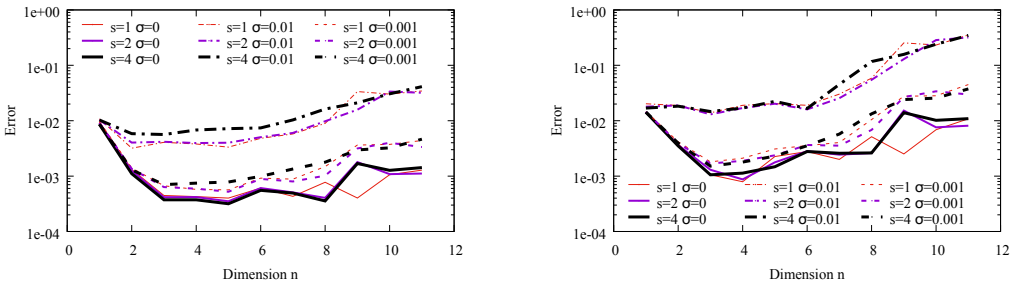
We already illustrated the benefits of CS-GEIM in controlling the noise and the model bias amplification in Section 4.4.3. Recall that the PBDW method is able to correct the model bias by adding an update term which belongs to the observation space W_m. In this section, we will show the noise amplification of PBDW, and the benefits of CS-PBDW in controlling the noise. The detailed formulation of PBDW (see (1.3.19)) and CS-PBDW (see Proposition 2.4.1) can be found in Chapter 1 and Chapter 2.

We still take the same problem setting as stated in Section 4.4.3, and we still assume that each sensor suffers from uniform distributed noise $e \in [-\sigma, \sigma]$, thus we use the CS-PBDW formulation given in (2.4.5) in Chapter 2.

We first show in Figure 4.4.31 the errors of PBDW for sensor width $s = 1, 2, 4$ for different noise levels. For comparison purposes, we also list the errors in noise-free case. We notice that, the error performance of PBDW is almost the same to GEIM, i.e., the mode bias or the measurement noise is amplified by the Lebesgue constant and stability factor.

For comparison purposes, we also studied the APBDW [202] approach for the noise control. The formulation of APBDW can be found in (2.4.1). Note that, the regularization factor ξ in APBDW formulation needs to be carefully selected, usually the L-curve [33] approach is used to determine ξ. The errors of APBDW with different ξ are shown in Figure 4.4.32, the sensor width is set to $s = 4$, and the noise level is $\sigma = 0.01$. For comparison purposes, we also list the errors of PBDW and CS-PBDW. From Figure 4.4.32a and Figure 4.4.32b we observe that, with a proper regularization factor, APBDW does improve the reconstruction when the reduced dimension n is relatively low (e.g. V_5), especially when more measurements are considered. In [202], the author shows the similar result, the improvement is shown with a low dimensional reduced space. Figure 4.4.32c and Figure 4.4.32d also show that, for a higher dimensional reduced space (e.g. V_{10}), we have not been able to control the noise, even by adding more measurements. On the other hand, CS-PBDW always improve the performance of the reconstruction. Furthermore, it follows the rule that more measurements, more accuracy.

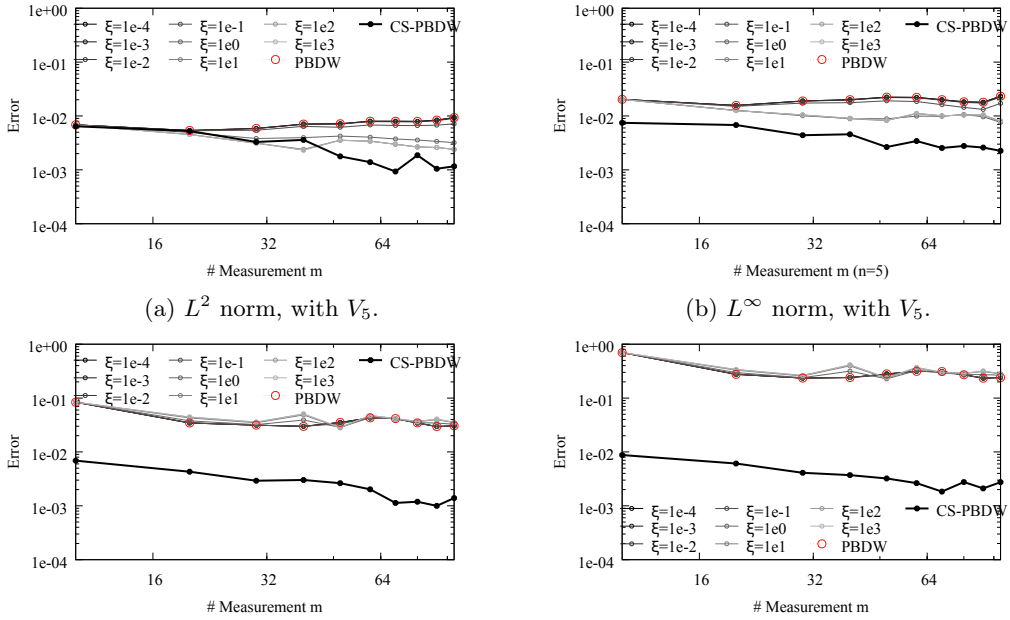
We further compare the two cases i.e. V_5 and V_{10} for sensor width $s = 1, 2, 4$ with noise level $\sigma = 0.01$. The errors of PBDW and CS-PBDW are shown in Figure 4.4.33. We again confirm that, in any case, CS-PBDW are able to control the noise with respect to PBDW.



(a) Relative error in L^2 norm

(b) Relative error in L^∞ norm

Figure 4.4.31: Variations of the errors with respect to the reduced dimension n for PBDW for different values of sensor width s with $m = 10n$



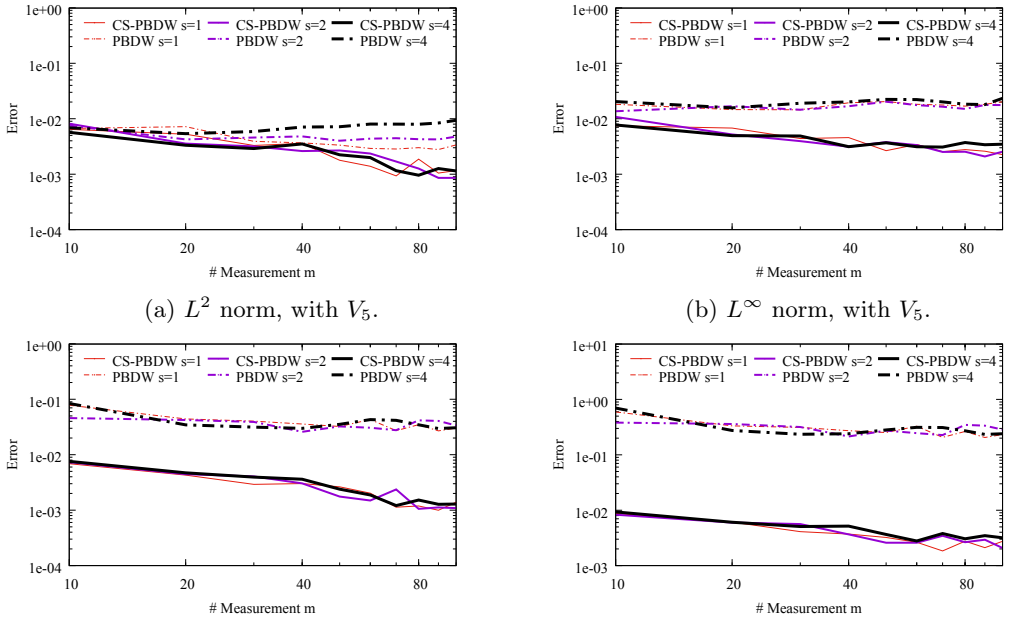
(a) L^2 norm, with V_5.

(b) L^∞ norm, with V_5.

(c) L^2 norm, with V_{10}.

(d) L^∞ norm, with V_{10}.

Figure 4.4.32: Variations of the errors with respect to the number of measurements m for different regularization factors for APBDW for two different reduced space with sensor width $s = 4$ and noise level $\sigma = 0.01$.



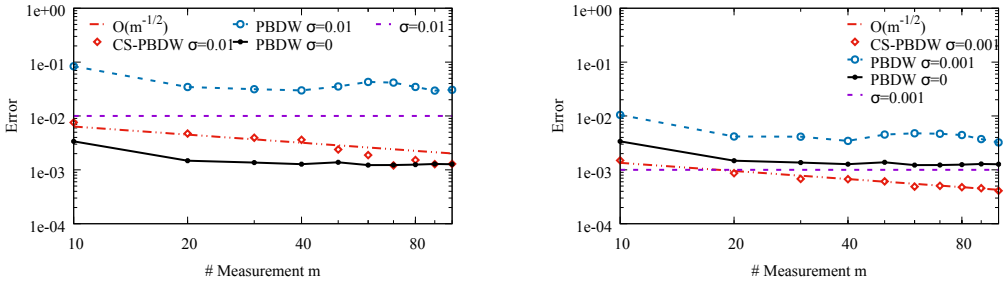
(a) L^2 norm, with V_5.

(b) L^∞ norm, with V_5.

(c) L^2 norm, with V_{10}.

(d) L^∞ norm, with V_{10}.

Figure 4.4.33: Variations of the errors with respect to the number of measurements m for CS-PBDW and PBDW for two different reduced space and three different values of sensor width s with noise level $\sigma = 0.01$.



(a) The noise level $\sigma = 0.01$.

(b) The noise level $\sigma = 0.001$.

Figure 4.4.34: Variations of the L^2 errors with respect to the number of measurements m for CS-PBDW and PBDW for noise level $\sigma = 0.01$ or 0.001 for V_{10} with sensor width $s = 4$.

We now study the case with sensor width $s = 4$ and the reduced space V_{10}. We show the errors of PBDW and CS-PBDW in Figure 4.4.34 with noise level $\sigma = 0.01$ or 0.001. For comparison purposes, the errors of PBDW in noise-free case are also shown. We again find that, with more measurements, CS-PBDW is able to not only control the noise, but also improve the results that are even better than the noise level. Furthermore, Figure 4.4.34b shows that CS-PBDW in noise case is even better than PBDW in noise-free case. Recall that for the underlying problem, if the Lebesgue constant Λ and stability factor $\|Q\|$ diverge, even without noise, PBDW may amplify the model bias, but the bias amplification is controlled by CS-PBDW. Additionally, we observe that, the red dotted lines in Figure 4.4.34a and Figure 4.4.34b show that the error of CS-PBDW converges with $m^{-1/2}$, which is a slow but rather expected convergence rate.

4.4.5 Gaussian noise and regularization

We emphasize again that knowing the behavior of the reconstruction with the uniformly distributed noise, one is able to estimate the upper bound of the reconstruction error, which is of great important for nuclear safety analysis. In section 4.4.3 and Figure 4.4.4, we studied the uniform noise case. In this section, we study the performance of the reconstruction with CS-GEIM and CS-PBDW with Gaussian noise, which is more general for measurements. We assume that the noise e of each sensor is IID, with probability density function $p(e) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(e-\bar{e})^2}{2\sigma^2}}$, with zero mean $\bar{e} = 0$ and standard deviation σ.

With no loss of generality, we only analyze two typical cases with reduced space V_5 and V_{10} for sensor width $s = 4$ and the standard deviation $\sigma = 0.01$. Furthermore, we simulate the reconstruction many times for different noise of the same distribution and mark the average error in L^2 and L^∞ norms (note that for uniformly distributed noise, we mark the maximum error among all simulations).

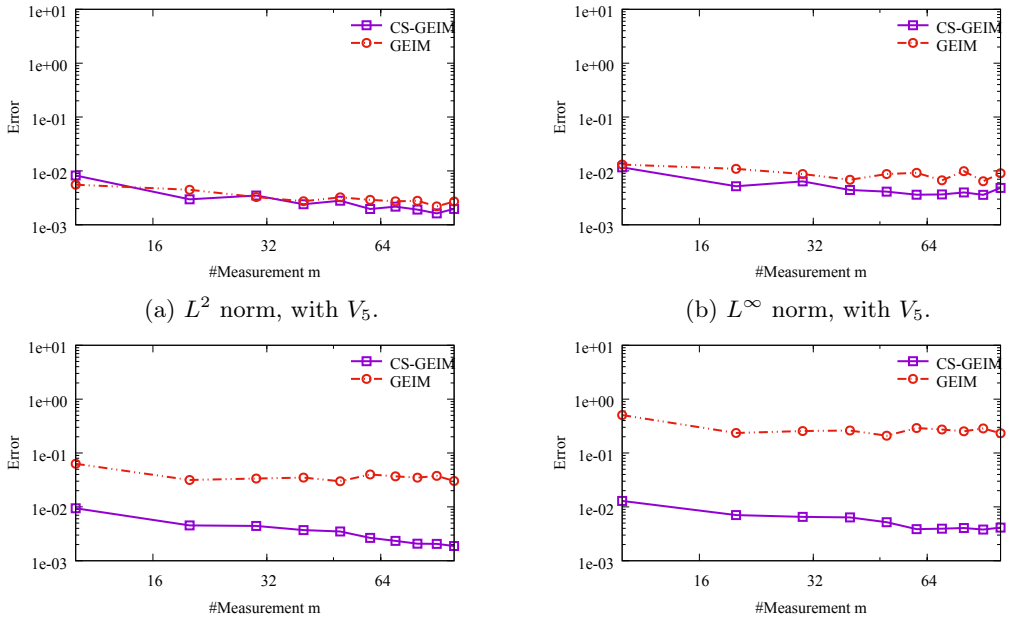
The reconstruction with CS-GEIM is based on the formulation (2.3.9) of Chapter 2. Note that this formulation does not depend on the specified noise distribution. We show in Figure 4.4.35 the L^2 and L^∞ relative errors of CS-GEIM and GEIM for different number of measurements, for V_5 and V_{10}. We immediately observe that, the performance of CS-GEIM with Gaussian noise is almost the same to the performance with uniform noise, this again confirms that the constraint stabilized GEIM approach is robust with respect to the noise.

The CS-PBDW formulation for Gaussian distributed noise is stated in (2.4.10) in Section 2.5. The L^2 and L^∞ relative errors of CS-PBDW and APBDW are shown in Figure 4.4.36. We again find that, in any case, CS-PBDW is robust with respect to the noise.

As explained in Section 2.5, CS-GEIM and CS-PBDW are constrained quadratic programming (QP) problems. Successful approaches to solve the bound-constrained optimization problems for general linear or nonlinear objective functions can be found in [112, 159, 138, 160]. The shortcoming of these algorithms is that it is time-consuming because of iterations. To meet online calculation for monitoring purpose, a simpler and closed form for these problems is proposed, see the regularized stabilized GEIM (R-GEIM, Proposition 2.5.5) and the regularized stabilized PBDW (R-PBDW, Proposition 2.5.2).

The comparison between R-GEIM and CS-GEIM is shown in Figure 4.4.37. We observe that, with a proper regularization factor, the performance of R-GEIM is the same to CS-GEIM, but the computational cost of R-GEIM is as low as solving an $m \times n$ matrix equation. Thus, R-GEIM is more suitable for on-line calculation.

We show the L^2 and L^∞ errors with respect to the number of measurement m for CS-PBDW and R-PBDW in Figure 4.4.38, from which we find that, R-PBDW sometimes provides even a little better reconstruction than CS-PBDW does. This again confirms that, the regularized version is more suitable for on-line calculation, if a proper regularization factor is selected.



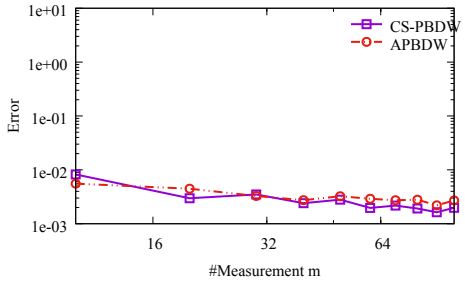
(a) L^2 norm, with V_5.

(b) L^∞ norm, with V_5.

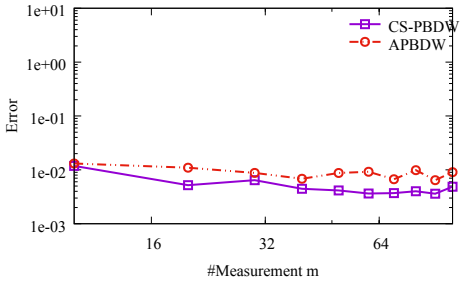
(c) L^2 norm, with V_{10}.

(d) L^∞ norm, with V_{10}.

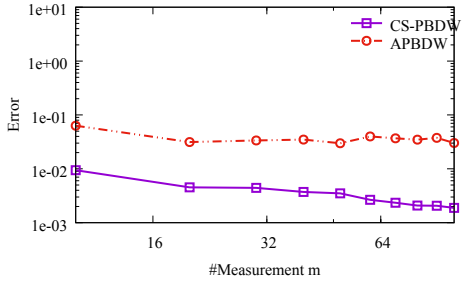
Figure 4.4.35: Variations of the L^2 and L^∞ errors with respect to the number of measurement m for GEIM and CS-GEIM for Gaussion noise with standard deviation $\sigma = 0.01$ and sensor width $s = 4$.



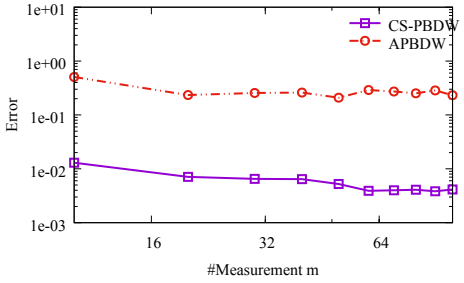
(a) Relative error in L^2 norm, with V_5.



(b) Relative error in L^∞ norm, with V_5.

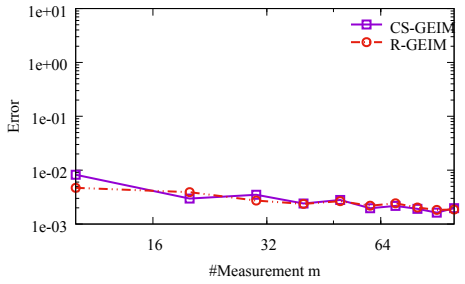


(c) Relative error in L^2 norm, with V_{10}.

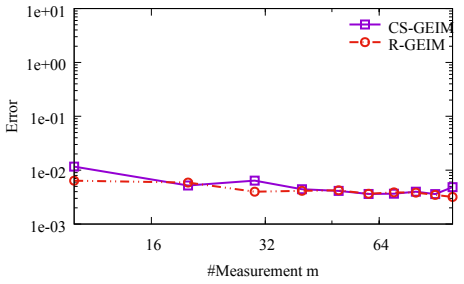


(d) Relative error in L^∞ norm, with V_{10}.

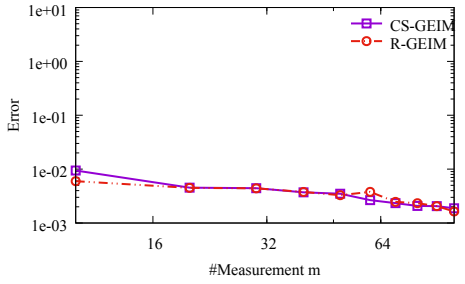
Figure 4.4.36: Variations of the L^2 and L^∞ errors with respect to the number of measurement m for CS-PBDW and APBDW for Gaussion noise with standard deviation $\sigma = 0.01$ and sensor width $s = 4$.



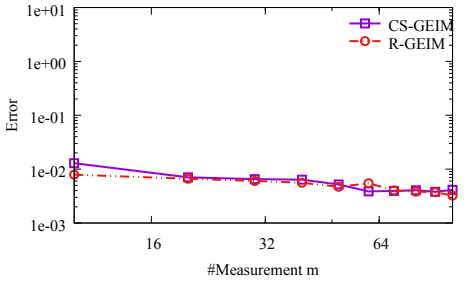
(a) Relative error in L^2 norm, with V_5.



(b) Relative error in L^∞ norm, with V_5.

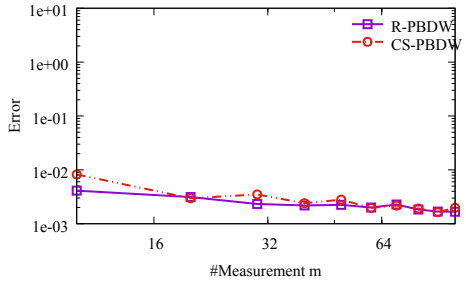


(c) Relative error in L^2 norm, with V_{10}.

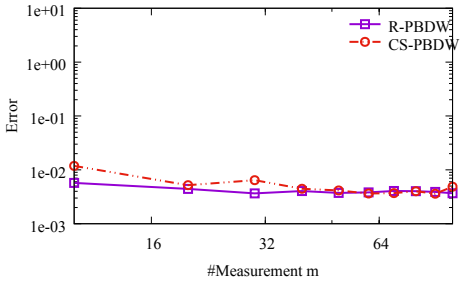


(d) Relative error in L^∞ norm, with V_{10}.

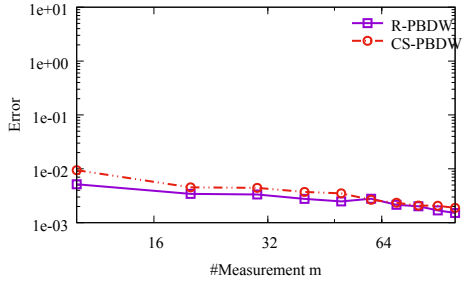
Figure 4.4.37: Variations of the L^2 and L^∞ errors with respect to the number of measurement m for R-GEIM and CS-GEIM for Gaussion noise with standard deviation $\sigma = 0.01$ and sensor width $s = 4$.



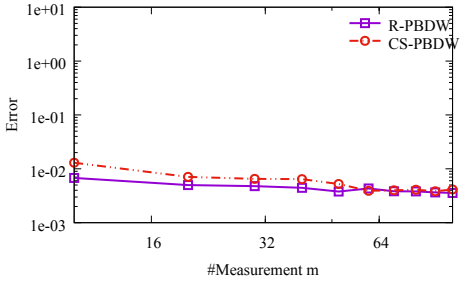
(a) Relative error in L^2 norm, with V_5.



(b) Relative error in L^∞ norm, with V_5.



(c) Relative error in L^2 norm, with V_{10}.



(d) Relative error in L^∞ norm, with V_{10}.

Figure 4.4.38: Variations of the L^2 and L^∞ errors with respect to the number of measurement m for CS-PBDW and R-PBDW for Gaussion noise with standard deviation $\sigma = 0.01$ and sensor width $s = 4$.

4.5 Conclusions

We have showed the industrial applications of GEIM either based on the realistic nuclear reactors operated by EDF or based on the real physical problems in this chapter.

This determination of the optimal sensor locations is done over all the geometries of reactors operated by EDF: PWR 900 MWe, PWR 1300 MWe and PWR 1450 MWe. Several strategies are developed in order to emphasize the advantages and limitations of the determination of the instrumental setup. It is shown that the method gives some instrumental setup that are consistent with what is expected from a physical point of view. Moreover, we demonstrate that, with respect to the random case, the results of optimal networks are clearly better. Finally, we conclude that GEIM is able to provide a physically coherent sensor placement with the structure of the neutronic flux in the reactor core.

We then applied the GEIM to control rods movements problem to improve the flux and the power distribution knowledge during control rods movements. Numerical results confirmed that, with GEIM, the fields reconstruction is well performed from measurements, which is hopefully acceptable in engineering usage, especially the realization meets the needs of online monitoring purpose based on the online-offline computational framework of GEIM.

The field reconstruction based on EDF reactors is also optimistic, especially the reconstruction is based Pin-by-Pin distribution framework. The important aspect is that we are providing a very accurate field reconstruction method that moreover follows the physical property of the core. On the case using the calculation code COCAGNE, the accuracy can reach up to 0.01% if a lot of basis functions (50) are used and it is already 0.1% with only 5 basis functions.

The extension of the method is to take into account the effect of the noise to qualify its im-

pact when few instruments remains and also to control it through various technique, either data pre-processing or mathematical correction of the induced error. In this respect, we propose the methodologies so called CS-GEIM and CS-PBDW to control the noise and also the model bias. Based on the stability analysis, we are able to optimize the sensor size and location. With the stabilized algorithms, we are able to control the measurement noise amplification, furthermore, by adding more measurements the proposed methods are able to provide even better accuracy than the noise level. The algorithms we proposed are robust with respect to the measurement noise and model bias. Finally, we emphasize that the regularized versions (R-GEIM, R-PBDW) with low computational cost are more suitable for on-line calculation in case the shape of the noise is known, i.e., Gaussion noise.

Part III

Supplementary developments for industry applications of the methodologies

Chapter 5

Stabilization for sensor constraints and failures

Contents

5.1 Introduction

In the framework of state estimation with reduced basis methods e.g. GEIM and PBDW, where the measurement is needed, there are at least two issues on the measurement in real engineering applications: i) there are constraints on the amount or the location for the sensors, and ii) the measurement always suffers from noise, let alone the measurement failures.

In presence of noise, the good properties of the reduced modeling approach may be blurred in the sense that the approximation error no longer converges but even diverges. So special denoising and stabilizing approaches are necessary for this class of recovery methods, this issue has already been addressed in Chapter 2.

Indeed, for the recovery methods with reduced basis, the number of measurements needed and the corresponding locations largely rely on the basis choosing procedure and the desired error threshold ε_{tol} on the training set. The fact is that, the measurements may be constrained in some specified sub-domain, e.g., in a nuclear reactor core, the control rod assemblies are unavailable for sensor deployment (see Figure 5.3.1). In addition, at operation stage, measurement failures may happen for the sensors. All these cases mentioned above may lead an unstable recovery, thus one has to reconsider the sensor deployment or the choice of the reduced basis to stabilize the recovery.

5.2 Measurement failure tolerant interpolation

5.2.1 Problem setting

Recall the notations we introduced in previous chapters : V is a Hilbert space over a domain $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ endowed with inner product $(\cdot,\cdot)$ and induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$, e.g. $V \subset L^2(\Omega)$ or $H^1(\Omega)$. $u(x;\mu) \in V$ is the solution of a parameter dependent PDE set on Ω and on a closed parametric domain $\mathcal{D} \subset \mathbb{R}^p$ $p \in \mathbb{N}$. Then we get the manifold $\mathcal{M} \equiv \{u(x;\mu); \mu \in \mathcal{D}\}$ of all parameter dependent solutions. We further denote by $\mathcal{D}^{(training)}$ the discrete parameter set of $\mathcal{D}$, $\mathcal{M}^{(training)}$ the corresponding discrete manifold of $\mathcal{M}$ and Ω_d is the discrete of Ω.

For the sake of brevity, the measurement failure tolerant interpolation in the frame of EIM is of particular interest in this section. The empirical interpolation method (EIM [21, 97, 147]) has been introduced to extend the reduced basis technique [150, 143, 185] to nonaffine and nonlinear partial differential equations (PDEs). The essential components of EIM procedure are (i) a good hierarchy of collateral reduced basis approximation spaces $(V_m)_m$ with $\dim(V_m) = m$ and $V_m \subset V_{m+1}$, (ii) a series of well selected interpolation points (also called 'magic point') $\{x_i\}_i$ on a domain $\Omega \in \mathbb{R}^d$, and (iii) an effective a posteriori estimator to quantify the interpolation errors.

The implantable EIM greedy algorithm [21] based on the discrete parameter training set $\mathcal{D}^{(training)}$ is shown in Algorithm 5.2.6. Once the basis functions and interpolation points are determined (off-line), the interpolant (on-line) of any $u \in \mathcal{M}$ is defined as

$$\begin{aligned} &\mathcal{I}_M[u(\cdot;\mu)] = \sum_{m=1}^{M} \beta_m(\mu) q_m \\ \text{subject to:} \quad &\mathcal{I}_M[u(\cdot;\mu)](x_m) = u(x_m;\mu), \quad m = 1, \cdots, M. \end{aligned} \tag{5.2.1}$$

Algorithm 5.2.6 Weak Greedy EIM

Input: $\mathcal{D}^{(training)}$, the parameter train set; M, the number of basis functions.
Output: $\{x_m\}_{m=1}^M, \{\mu_m\}_{m=1}^M, \{q_m\}_{m=1}^M$ ▷ q_m is the basis function
$\mu_1 = \arg\max_{\mu \in \mathcal{D}^{(training)}} \|u(\cdot;\mu)\|_{L^\infty(\Omega)}$
$x_1 = \arg\max_{x \in \Omega_d} |u(\cdot;\mu_1)|$ ▷ Ω_d is the discrete of Ω
$q_1 = u(\cdot;\mu_1)/u(x_1;\mu_1)$
for $m = 2 : M$ **do**
  $\mu_m = \arg\max_{\mu \in \mathcal{D}^{(training)}} \|u(\cdot;\mu) - \mathcal{I}_{m-1}[u(\cdot;\mu)]\|_{L^\infty(\Omega)}$
  $x_m = \arg\max_{x \in \Omega_d} |u(x;\mu_m) - \mathcal{I}_{m-1}[u(\cdot;\mu_m)](x)|$
  $q_m = \frac{u(\cdot;\mu_m) - \mathcal{I}_{m-1}[u(\cdot;\mu_m)]}{u(x_m,\mu_m) - \mathcal{I}_{m-1}[u(\cdot;\mu_m)](x_m)}$

If there are measurement failures (always randomly), the interpolant in (5.2.1) will not holds any more. In case of measurement failures, the problem becomes:

Problem 5.2.1 *For a given set of basis function $\{q_m^\circ\}_{m=1}^M$ (or $\{u(\cdot;\mu_m^\circ)\}_{m=1}^M$ equivalently, which span V_M) and the associated interpolation point set $I_M := \{x_m^\circ\}_{m=1}^M$, assume that n $(1 \leq n < M)$ measurements corresponding to the point set $I_n = \{x_{m_i}^\circ\}_{m_i=1}^n$ faile. What is the best definition of the new interpolation $\mathcal{I}_M^n[u] \in V_M$ based on the left observations $u(x_m)$, $x_m \in I_M/I_n$?*

The superscript '$\circ$' of $\mu_m^\circ, x_m^\circ, q_m^\circ$ quoted above denotes they are the initial sampling parameters, interpolation points and basis functions, e.g., from EIM greedy algorithm 5.2.6. Let us further denote by U_M^n the set of indices for valid points. Normally, for any $u \in \mathcal{M}$, there are infinite $v \in V_M$ which

satisfies $v(x_m) = u(x_m),\ m \in U_M^n$. To make Problem 5.2.1 feasible, one must select no more than $M - n$ basis functions, in this case, one can

i) use the first $(M - n)$ basis functions $\{q_1, \cdots, q_{M-n}\}$;

ii) use $(M - n)$ snapshots $u(\cdot; \mu_m)$ with $m \in U_M^n$ corresponding to the valid measurements;

iii) re-select $(M - n)$ basis functions to set-up a new interpolant.

5.2.2 Interpolation aiming to optimize the basis functions and interpolation points

A more general way to address the measurement failures is to design a new algorithm to constitute the basis functions and points instead of EIM greedy algorithm. For the sake of brevity, we first study the case with only one measurement failed. Suppose the first $m-1, m \geq 2$ sampling snapshots $\{u(\cdot; \mu_i)\}_{i=1}^{m-1}$, basis functions $\{q_i\}_{i=1}^{m-1}$ and point $\{x_i\}_{i=1}^{m-1}$ are already selected. Recall that with EIM greedy algorithm, we select the m-th snapshot $u(\cdot; \mu_m)$ where the L^∞ norm of the $(m-1)$-th order interpolant error

$$e_{m-1}[u(\cdot; \mu)] = \mathcal{I}_{m-1}[u(\cdot; \mu)] - u(\cdot; \mu) \tag{5.2.2}$$

is maximal, i.e.,

$$\mu_m = \underset{\mu \in \mathcal{D}^{(training)}}{\arg\max} \ \|e_{m-1}[u(\cdot; \mu)]\|_{L^\infty(\Omega)} \ , \tag{5.2.3}$$

and denote

$$\zeta_m = \|e_{m-1}[u(\cdot; \mu_m)]\|_{L^\infty(\Omega)}. \tag{5.2.4}$$

Then we select the point x_m where the maximum value (in absolute) of the error $e_{m-1}(u(\cdot; \mu_m))$ happens, i.e.

$$x_m = \underset{x \in \Omega_d}{\arg\max} \ |e_{m-1}[u(x; \mu_m)]|. \tag{5.2.5}$$

Note that several parameters μ might maximize the function $\mu \to \|e_{m-1}[u(\cdot; \mu)]\|_{L^\infty(\Omega)}$ and several x might maximize the function $x \to |e_{m-1}[u(x; \mu_m)]|$. In this case, μ_m and x_m are picked among the set of maximizers. For the general algorithm with measurement failure tolerance, the idea is that, at the m-th step, instead of selecting the m-th snapshot where the $(m-1)$-th order interpolant error is maximal, we first select all the candidates where the $(m-1)$-th order interpolation errors are large than a threshold $\eta\zeta_m$, where η with $0 < \eta < 1$ is a tolerance, i.e.,

$$\mathcal{D}_m := \{\mu \in \mathcal{D}^{(training)} \ ; \ \|\mathcal{I}_{m-1}[u(\cdot; \mu)] - u(\cdot; \mu)\|_{L^\infty(\Omega)} \geq \eta\zeta_m\}. \tag{5.2.6}$$

Assume among the $(m-1)$ points, the i-th measurement failed, the new $(m-2)$-th order interpolant is thus defined as

$$\begin{aligned} &\mathcal{I}'_{m-1,i}[u(\cdot; \mu)] := \sum_{j=1, j\neq i}^{m-1} \beta_j(\mu) u(\cdot; \mu_j), \\ \text{subject to:} \quad &\mathcal{I}'_{m-1,i}[u(\cdot; \mu)](x_k) = u(x_k; \mu),\ k = 1, ..., m-1,\ k \neq i. \end{aligned} \tag{5.2.7}$$

The interpolation error is

$$e'_{m-1,i}[u(\cdot; \mu)] := \mathcal{I}'_{m-1,i}[u(\cdot; \mu)] - u(\cdot; \mu) \ . \tag{5.2.8}$$

Then the next snapshot $u(\cdot; \mu_m)$ is determined which maximize the interpolation error $e'_{m-1,i}[u(\cdot; \mu)]$

for all $1 \leq i \leq m-1$ and all snapshots $u(\cdot;\mu) \in \mathcal{D}_m$:

$$\mu_m = \arg\max_{\mu \in \mathcal{D}_m} \max_{i=1,\dots m-1} \|\mathcal{I}'_{m-1,i}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)}. \tag{5.2.9}$$

In order to determine the m-th interpolation point, we first define a set of the error indicators:

$$\rho^m_{m-1,i} := \|\mathcal{I}'_{m-1,i}[u(\cdot;\mu_m)] - u(\cdot;\mu_m)\|_{L^\infty(\Omega)}, \quad i = 1, \cdots, m-1, \tag{5.2.10}$$

for a given tolerance α with $0 < \alpha < 1$, we select the point x with which the interpolation error $e'_{m-1,i}[u(\cdot;\mu_m)](x)$ is no smaller than a threshold $\alpha\rho^m_{m-1,i}$, i.e.,

$$L^m_{m-1,i,\alpha} := \{x \in \Omega_d \ ; \ |\mathcal{I}'_{m-1,i}[u(\cdot;\mu_m)](x) - u(x;\mu_m)| \geq \alpha\rho^m_{m-1,i}\} \ . \tag{5.2.11}$$

Similarly, we have a set $L^m_{m,\alpha}$ based on the $(m-1)$-th order interpolant:

$$L^m_{m,\alpha} := \{x \in \Omega_d \ ; \ |\mathcal{I}_{m-1}[u(\cdot;\mu_m)](x) - u(x;\mu_m)| \geq \alpha\zeta_m\} \ . \tag{5.2.12}$$

In this setting, we deduce a set L^m_α defined as

$$L^m_\alpha := \underset{i=1,\dots,m-1}{\cap} L^m_{m-1,i,\alpha} \cap L^m_\alpha \ , \tag{5.2.13}$$

in which the interpolation error on these points are relatively large in the sense of a given tolerance α. We then find the maximum α which satisfies $L^m_\alpha \neq \varnothing$, i.e.,

$$\alpha_{max} = \arg\max_{0<\alpha<1} \{\alpha \ ; \ L^m_\alpha \neq \varnothing\} \ . \tag{5.2.14}$$

The m-th interpolation point is determined:

$$x_m \in L^m_{\alpha_{max}} \ . \tag{5.2.15}$$

The algorithm is summarized in Algorithm 5.2.7. Note that the interpolation function $\mathcal{I}'_{m-1,i}, i = 1, ..., m-1$ is well defined. For any $u(\cdot;\mu) \in \mathcal{M}^{(training)}$, we have

$$\mathcal{I}'_{m-1,i}[u(\cdot;\mu)] = \sum_{j=1,j\neq i}^{m-1} \beta_j(\mu) u(\cdot;\mu_j), \tag{5.2.16}$$

the coefficients $\beta_j(\mu), \ j = 1, ..., m-1, \ j \neq i$ satisfy

$$\sum_{j=1,j\neq i}^{m-1} \beta_j(\mu) q_j(x_k) = u(x_k;\mu), \ k = 1, ..., m-1, \ k \neq i \ . \tag{5.2.17}$$

The above equation (5.2.17) has solution if and only if

$$u(x_k;\mu_k) - \mathcal{I}'_{k-1,i}[u(x_k;\mu_k)](x_k) \neq 0, \ k = i+1, ..., m-1,$$

and this holds evidently: when we select x_k, we have $x_k \in \{\underset{i=1,\dots,k-1}{\cap} L^k_{k-1,i,\alpha_{max}} \cap L^k_{k,\alpha_{max}}\}$, thus we have $x_k \in L^k_{k-1,i,\alpha_{max}}$, based on the definition of $L^k_{k-1,i,\alpha_{max}} := \{x \in \Omega_d \mid |u(x;\mu_k) - \mathcal{I}'_{k-1,i}[u(\cdot;\mu_k)](x)| \geq \alpha_{max}\rho^k_{k-1,i}\}$, we can easily find that $|u_k(x_k) - \mathcal{I}'_{k-1,i}[u_k](x_k)| \geq \alpha_{max}\rho^k_{k-1,i} \neq 0$.

Note that Algorithm 5.2.7 proposes a general way to address the measurement failures. Unlike EIM, we take the selected snapshot $u(\cdot;\mu_m)$ as the m-th basis function directly instead of the error

Algorithm 5.2.7 One measurement failed

Set $\eta, 0 < \eta < 1$
$\mu_1 = \arg\max_{\mu\in\mathcal{D}^{(training)}} \|u(\cdot;\mu)\|_{L^\infty(\Omega)}$
$x_1 = \arg\max_{x\in\Omega_d} |u(\cdot;\mu_1)|$
$q_1 = u(\cdot;\mu_1)$
$m = 2$
$\mu_m = \arg\max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)}$
$x_m = \arg\max_{x\in\Omega_d} |\mathcal{I}_{m-1}[u(\cdot;\mu_m)](x) - u(x;\mu_m)|$
$q_m = u(\cdot;\mu_m)$
for m = 3:M **do**
    $\zeta_m := \max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)}$
    $\mathcal{D}_m = \{\mu \in \mathcal{D}^{(training)} \ ; \ \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \geq \eta\zeta_m\}$
    $\mu_m = \arg\max_{\mu\in\mathcal{D}_m} \max_{i=1,\dots m-1} \|u(\cdot;\mu) - \mathcal{I}'_{m-1,i}[u(\cdot;\mu)]\|_{L^\infty(\Omega)}$
    Calculate $\rho^m_{m-1,i} := \|\mathcal{I}'_{m-1,i}[u(\cdot;\mu_m)]u(\cdot;\mu_m)\|_{L^\infty(\Omega)}, \ 1 \leq i \leq m-1$
    $L^m_{m-1,i,\alpha} := \{x \in \Omega_d \ ; \ |\mathcal{I}'_{m-1,i}[u(\cdot;\mu_m)](x) - u(x;\mu_m)| \geq \alpha\rho^m_{m-1,i}\}$
    $L^m_{m,\alpha} := \{x \in \Omega_d \ ; \ |\mathcal{I}_{m-1}[u(\cdot;\mu_m)](x) - u(x;\mu_m)| \geq \alpha\zeta_m\}$
    $L^m_\alpha := \cap_{i=1,\dots,m-1} L^m_{m-1,i,\alpha} \cap L^m_{m,\alpha}$
    $\alpha_{max} = \arg\max_{0<\alpha<1} \{\alpha \ ; \ L^m_\alpha \neq \varnothing\}$
    The m-th interpolation point: $x_m \in \{L^m_{\alpha_{max}}\}$
    The m-th basis function: $q_m = u(\cdot;\mu_m)$

function by an normalization process.

From the construction of the above algorithm, when there is one measurement failed, for any $\mu \in \mathcal{D}$, one can use the $(M-1)$ measurements and $(M-1)$ snapshots $u(\cdot;\mu_m)$ with $m \in U^1_M$ corresponding to the valid measurements to do the interpolation:

$$\begin{aligned} &\mathcal{I}'_{M,i}[u(\cdot;\mu)] := \sum_{j=1,j\neq i}^{M} \beta_j(\mu)u(\cdot;\mu_j), \\ \text{subject to:} \quad &\mathcal{I}'_{M,i}[u(\cdot;\mu)](x_k) = u(x_k;\mu), \ k = 1,...,M, \ k \neq i. \end{aligned} \tag{5.2.18}$$

5.2.3 Interpolation aiming to optimize the basis functions

As already known, the basis functions and points from EIM greedy algorithm are already quasi-optimal in some sense. Another possible way is to keep the original EIM magic points and re-select some basis functions. In this setting, the problem becomes: with fixed points $I_{M\setminus n} := I_M \setminus I_n$, where $I_M := \{x^\circ_m\}_{m=1}^M$ is the initial M interpolation points set and $I_n := \{x^\circ_{m_i}\}_{i=1}^n, n \geq 1$ is the failed n interpolation points set, we re-select the parameter sampling $\{\mu_m\}_{m=1}^{M-n}$ and the corresponding snapshots $\{u(\cdot;\mu_m)\}_{m=1}^{M-n}$ to do the interpolation.

Assume the index of the first failed point is $m_1, \ 1 \leq m_1 \leq M$, we keep the first (m_1-1) well selected points and basis functions (deduced from EIM) unchanged, i.e.,

$$\begin{cases} x_m = x^\circ_m \\ \mu_m = \mu^\circ_m \ , \quad m = 1, \cdots, m_1 - 1 \\ u(\cdot;\mu_m) = u(\cdot;\mu^\circ_m) \end{cases} \tag{5.2.19}$$

and re-select the following $(M - 1 - (m_1 - 1))$ points and basis functions in a greedy way. Suppose

the first $m-1$ points and basis functions are already determined, the $(m-1)$-th order interpolant is

$$\mathcal{I}_{m-1}[u(\cdot;\mu)] = \sum_{j=1}^{m-1} \beta_j(\mu) u(\cdot;u_j) \ . \tag{5.2.20}$$

We again define the error indicator

$$\zeta_m := \max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \ , \tag{5.2.21}$$

and select all the candidates where the $(m-1)$-th order interpolant errors are large than a threshold $\eta\zeta_m$, where η is a tolerance with $0<\eta<1$, i.e.,

$$\mathcal{D}_m := \{\mu \in \mathcal{D}^{(training)} \mid \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \geq \eta\zeta_m\}. \tag{5.2.22}$$

Our interest is the (μ, x) pairs which maximize the error $|\mathcal{I}_{m-1}[u(x;\mu)] - u(x;\mu)|$ for all the candidates in $\mathcal{D}_m$ on all the points $I_M \setminus I_n$, i.e.,

$$(\mu_m, x_m) = \arg\max_{\mu\in\mathcal{D}_m} \max_{x\in I_M\setminus I_n} |\mathcal{I}_{m-1}[u(\cdot;\mu)](x) - u(x;\mu)| \ . \tag{5.2.23}$$

The algorithm is summarized in Algorithm 5.2.8. We make the following remarks for this approach.

- If we use M basis functions and points to do the interpolation, and the system is required to be designed with no more than n sensor failures, then we have to pre-computed $\sum_{m=1}^{n} \frac{M!}{(M-m)!m!}$ sets of basis functions for all possible sensor failures cases in the off-line phase, this may lead computing and storage issues if M and n are relatively large.

- In practice engineering application, once the n measurement failures are detected, one can re-calculate the new basis functions with Algorithm 5.2.8 immediately, then use the new interpolation frame in the on-line phase until the failed measurements are recovered.

Algorithm 5.2.8 Measurement failure tolerant EIM (mftEIM) algorithm

Input: the points $\{x_m^\circ\}_{m=1}^M$ and parameters $\{\mu_m^\circ\}_{m=1}^M$ (e.g., from EIM algorithm).
Input: $I_n = \{x_{m_i}^\circ\}_{i=1}^n$, the failed measurement points.
Input: a train set $\mathcal{D}^{(training)}$, a tolerance η, with $1 \geq \eta \geq 0,\ M \geq n+1$.
Output: $\{x_m\}_{m=1}^{M-n}, \{\mu_m\}_{m=1}^{M-n}, \{u(\cdot;\mu_m\}_{m=1}^{M-n}$
for $m = 1 : m_1 - 1$ **do**
    $x_m = x_m^\circ$, $\mu_m = \mu_m^\circ$
for $m = m_1 : M - n$ **do** ▷ for notational coherence, $\mathcal{I}_0 = 0$
    $\zeta_m := \max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)}$
    $\mathcal{D}_m := \{\mu \in \mathcal{D}^{(training)} \ ; \ |\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)|_{L^\infty(\Omega)} \geq \eta\zeta_m\}$
    $(\mu_m, x_m) = \arg\max_{\mu\in\mathcal{D}_m} \max_{x\in I_M\setminus I_n} |\mathcal{I}_{m-1}[u(\cdot;\mu)](x) - u(x;\mu)|$

From the construction of the above algorithm, when there is n measurements failed, for any $\mu \in \mathcal{D}$, one can use the $(M-n)$ measurements and $(M-n)$ snapshots $u(\cdot;\mu_m)$ with $m \in U_M^n$

corresponding to the valid measurements to do the interpolation:

$$\begin{aligned} &\mathcal{I}_M^n[u(\cdot;\mu)] := \sum_{j\in U_M^n} \beta_j(\mu) u(\cdot;\mu_j), \\ \text{subject to:} \quad &\mathcal{I}_M^n[u(\cdot;\mu)](x_j) = u(x_j;\mu),\ j \in U_M^n. \end{aligned} \tag{5.2.24}$$

5.3 Location constraints

In the framework of data assimilation with reduced basis, the number of measurement needed and the corresponding locations largely rely on basis choosing procedure and the desired error threshold ε_{tol} on the training set $\mathcal{M}^{(training)}$. The fact is that, the measurements may be constrained in some specified sub-domain, e.g., in a nuclear reactor core, the control rod assemblies are not available for sensor deployment, see Figure 5.3.1 the EPR in-core instrumentation for example [68], thus, only part of the assemblies are able to implement the sensors.

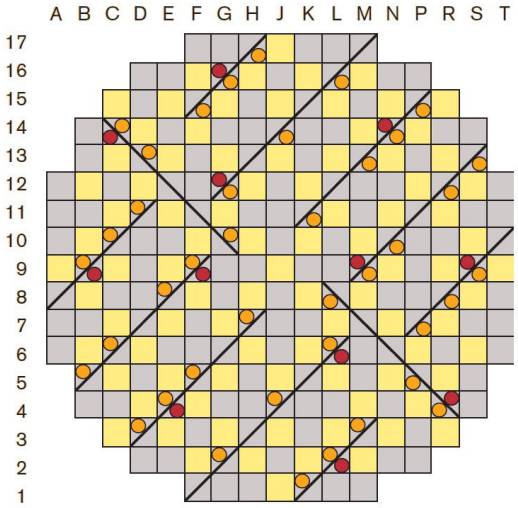


Figure 5.3.1: EPR in-core instrumentation. Purple rectangle: fuel assemblies, yellow rectangle: control rods. The black lines are instrumentation lances, red circles mark the position of in core sensors, while the orange ones mark the position of the aeroball systems. Source: `http://www.areva-np.com/us/liblocal/docs/EPR/U.S.EPRbrochure_1.07_FINAL.pdf`

Let's consider EIM greedy algorithm again. The domain for measurements is constrained in a sub domain $\Omega_{lc} \in \Omega$. Our goal is to determine the basis functions and the corresponding interpolation points with this domain constraint. Suppose the first $m-1$ points $\{x_i\}$ and basis functions $\{q_i\}$ are already determined, the $(m-1)$-th order interpolant is

$$\mathcal{I}_{m-1}[u(\cdot;\mu)] = \sum_{j=1}^{m-1} \beta_j(\mu) q_j \ . \tag{5.3.1}$$

Note that here for notational coherence, we set $\mathcal{I}_0 = 0$, and $m \geq 1$. We again define the error indicator

$$\zeta_m := \max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \ , \tag{5.3.2}$$

and select all the candidates where the $(m-1)$-th order interpolant errors are large than a threshold $\eta\zeta_m$, where η with $0 < \eta < 1$ is a tolerance, i.e.,

$$\mathcal{D}_m := \{\mu \in \mathcal{D}^{(training)} \ ; \ \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \geq \eta\zeta_m\}. \tag{5.3.3}$$

Our interest is the (μ, x) pairs which maximize the error $|\mathcal{I}_{m-1}[u(x;\mu)] - u(x;\mu)|$ for all the candidates in $\mathcal{D}_m$ and on all the points $x \in \Omega_{lc}$, i.e.,

$$(\mu_m, x_m) = \arg\max_{\mu\in\mathcal{D}_m} \max_{x\in\Omega_{lc}} |\mathcal{I}_{m-1}[u(\cdot;\mu)](x) - u(x;\mu)| \ . \tag{5.3.4}$$

The algorithm is summarized in Algorithm 5.3.9. Indeed, this idea is an natural expansion of Algorithm 5.2.8, we call it location constrained EIM (lcEIM), which constrains the possible interpolation points in to a set $\Omega_{lc} \subset \Omega$, where Ω is the whole space domain of our problem. An special case is that when $\Omega_{lc} = \{x_m^\circ\}_{m=1}^M \setminus I_n$, the algorithm becomes Algorithm 5.2.8. If $\Omega_{lc} = \Omega$, and the tolerance $\eta = 1$, this algorithm degenerates to classical EIM greedy algorithm.

Algorithm 5.3.9 Location Constrained EIM Greedy algorithm (lcEIM)

1: **Input:** a train set $\mathcal{D}^{(training)}$, a constrained set Ω_{lc}, a tolerance η, with $0 < \eta < 1$.
2: **Output:** $\{x_m\}_{m=1}^M, \{\mu_m\}_{m=1}^M, \{u(\cdot;\mu_m)\}_{m=1}^M$
3: **for** $m = 1 : M$ **do** ▷ here for notational coherence, $\mathcal{I}_0 = 0$
4: $\quad \zeta_m := \max_{\mu\in\mathcal{D}^{(training)}} \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)}$
5: $\quad \mathcal{D}_m := \{\mu \in \mathcal{D}^{(training)} \ ; \ \|\mathcal{I}_{m-1}[u(\cdot;\mu)] - u(\cdot;\mu)\|_{L^\infty(\Omega)} \geq \eta\zeta_m\}$
6: $\quad (\mu_m, x_m) = \arg\max_{\mu\in\mathcal{D}_m} \max_{x\in\Omega_{lc}} |\mathcal{I}_{m-1}[u(\cdot;\mu)](x) - u(x;\mu)|$

5.4 Numerical results

We first compare different algorithms for two test problems. Test case 5.4.1 is a pure numerical test and Test case 5.4.2 is based on the IAEA 2D benchmark problem described in Section 3.5.2.

Test case 5.4.1 *Consider the family of functions from [145],*

$$g(x,\mu) \equiv \frac{1}{\sqrt{1 + (25 + \mu_1 cos(\mu_2 x))x^2}} \ , \tag{5.4.1}$$

where $\mu = (\mu_1, \mu_2) \in \mathcal{D}, x \in [0,1]$*, and the parameter domain is modified from the original one* $\mathcal{D}^o = [0.01, 24.9] \times [0.01, 15]$.

Test case 5.4.2 *The IAEA 2D benchmark problem described in Section 3.5.2.*

Recall that the interpolation points $I_M := \{x_m^\circ\}_{m=1}^M$, snapshots $\{u(\cdot;\mu_m^\circ)\}_{m=1}^M$ corresponding to the basis functions $\{q_m^\circ\}_{m=1}^M$ are selected from EIM algorithm. There are n measurements on $I_n := \{x_{m_i}^\circ\}_{i=1}^n, n \geq 1$ failed among the I_M points. U_M^n is the set of indices for valid points. We set-up new interpolation with different basis functions as follows:

i) the first $(M-n)$ basis functions $q_1^\circ, \cdots, q_{M-n}^\circ$;

ii) the $(M-n)$ snapshots $\{u(\cdot;\mu_m^\circ) \ ; \ m \in U_M^n\}$;

iii) the $(M-n)$ basis functions selected with Algorithm 5.2.8; or

iv) the $(M-n)$ snapshots $\{u(\cdot;\mu_m^1)\ ;\ m \in U_M^n\}$, note that $(\mu_i^1)_i, (x_i^1)_i$ are selected with Algorithm 5.2.7 as the initial ones;

v) the first $(M-n)$ basis functions selected with POD method,

vi) the first $M-n$ basis functions $q_1^\circ, \cdots, q_{M-n}^\circ$, use the constrained stabilized EIM method (CS-EIM, see Section 2.3) to do the interpolation.

Note that for comparison purposes, we also tried the constrained stabilized EIM, introduced in Chapter 2. All these interpolation point and basis function combinations are summarized in Table 5.1.

Table 5.1: Different points and basis functions combinations

Case	Points	Basis	# Failed (n)	Comment
1	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{u(\cdot;\mu_m^\circ)\ ;\ m \in U_M^n\}$	0,1,2	EIM
2	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{q_m^\circ\}_{m=1}^{M-n}$	1,2	EIM
3	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{q_m^\circ\}_{m=1}^{M-n}$	1	CS-EIM
4	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{q_{m,POD}^\circ\}_{m=1}^{M-n}$	1	POD-EIM [1]
5	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{u(\cdot;\mu_m^1)\}_{m=1}^{M-n}$	1	Algorithm 5.2.7 [2]
6	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{u(\cdot;\mu_m^1)\ ;\ m \in U_M^n\}$	1	Algorithm 5.2.7
7	$\{x_m^\circ\ ;\ m \in U_M^n\}$	$\{u(\cdot;\mu_m^2)\}_{m=1}^{M-n}$	1,2	mftEIM, Algorithm5.2.8 [3]

[1] Here we use the first $(M-n)$ POD basis functions $\{q_{m,POD}^\circ\}_{m=1}^{M-n}$, but the interpolation points are from EIM algorithm.
[2] $(\mu_i^1)_i, (x_i^1)_i$ are selected with Algorithm 5.2.7 to be the initial ones.
[3] $(\mu_i^2)_i$ are selected with Algorithm 5.2.8.

To evaluate the qualities of the seven combinations shown in Table 5.1, we fist construct the discrete manifold $\mathcal{M}^{(training)}$ with cardinality $M_{max} = |\mathcal{M}^{(training)}|$, which is fine enough to be representative of the manifold $\mathcal{M}$. Then we select the first $M \leq M_{max}$ basis functions $\{q_m^\circ\}_{m=1}^M$ and interpolation points $\{x_m^\circ\}_{m=1}^M$ with EIM greedy algorithm. Then we compute the errors in $\mathcal{X} = L^2(\Omega)$ norm as in (5.4.2):

$$e_{M,n}^{(training)} \coloneqq \max_{\mu \in \mathcal{D}^{(training)}} \|\mathcal{I}_M^n[u(\cdot;\mu)](\cdot) - u(\cdot;\mu)\|_{\mathcal{X}} / \|u(\cdot;\mu)\|_{\mathcal{X}}, \tag{5.4.2}$$

where $\mathcal{I}_M$ is the M-th order interpolant with one of the combinations in Table 5.1. The measure with $e_m^{(training)}$ can be used as an indicator to evaluate how good the method is, in case the training set is fine enough to be representative for the manifold. In contrast to the training set, we construct this test set with enough sampling point $\mathcal{D}^{(test)}$ to be representative of the manifold. Then we compute the errors as in (5.4.3):

$$e_{M,n}^{(test)} \coloneqq \max_{\mu \in \mathcal{D}^{(test)}} \|\mathcal{I}_M^n[u(\cdot;\mu)](\cdot) - u(\cdot;\mu)\|_{\mathcal{X}} / \|u(\cdot;\mu)\|_{\mathcal{X}}. \tag{5.4.3}$$

The measure with $e_{M,n}^{(test)}$ can be used as an indicator to evaluate how robust the method is, in case the training set is not enough representative for the manifold.

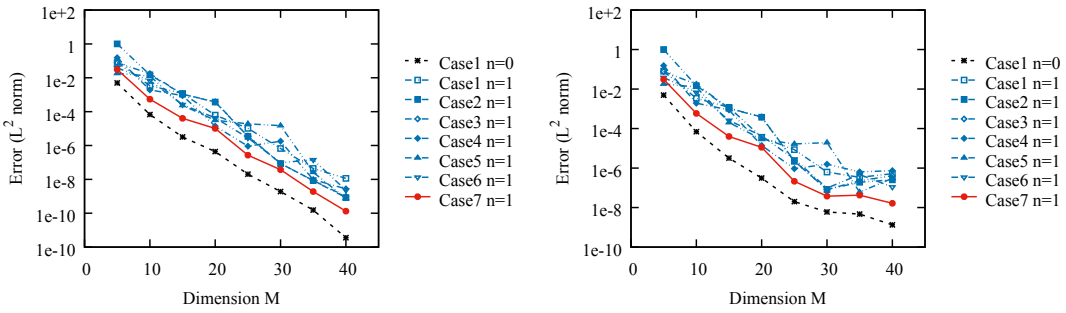
We first test all the seven combinations shown in Table 5.1 for Test case 5.4.1, but with a relative small parameter domain i.e. $\mathcal{D} = [15.0, 24.9] \times [14.9, 15]$. The reason why we start with a small parameter domain is that, in this setting, the Kolmogorov n-width of the manifold $\mathcal{M}$ of the problem decays fast, and easily for us to find the effect of measurement failures. The errors $e_{M,n}^{(training)}$ and $e_{M,n}^{(test)}$ are shown in Figure 5.4.2. We notice that, in case of one measurement failed,

all the combinations loose at least one order accuracy, which is normal because we loose one basis functions and measurement. Furthermore, we find that, mftEIM is always the best, and much regular than others, no matter for the test set or the training set.

We then enlarge the parameter domain to $\mathcal{D} = [15.0, 24.9] \times [4.9, 15]$. The interpolation errors for the seven combinations are shown in Figure 5.4.3, we can find the same result. Besides, with a relative large parameter domain, the situation for other combination is much worse and irregular. This means that, in this case, the measurement failures is really an issue, the mftEIM approach is necessary to replace the original EIM process.

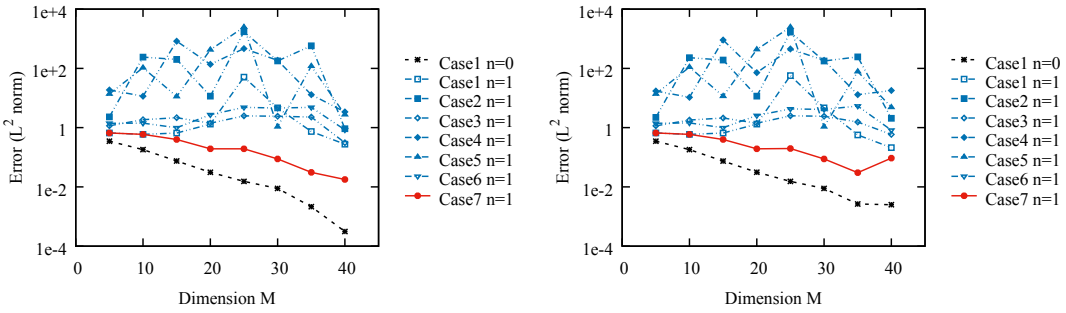
Figure 5.4.4 shows the interpolation errors for two sensors failures for the small parameter domain case, and the large parameter domain case is shown in Figure 5.4.5. In these figures, we also list the case only one sensor failed for mftEIM approach. This again confirms the robust of mftEIM approach in addressing the measurement failures.

The tests based on the IAEA 2D benchmark problem are shown in Figure 5.4.6 for one sensor failed, and in Figure 5.4.7 for two sensors failures respectively. For this problem, the Kolmogorov n-width of the manifold $\mathcal{M}$ decays fast, the results are similar to Test case 5.4.1 with a small parameter domain.



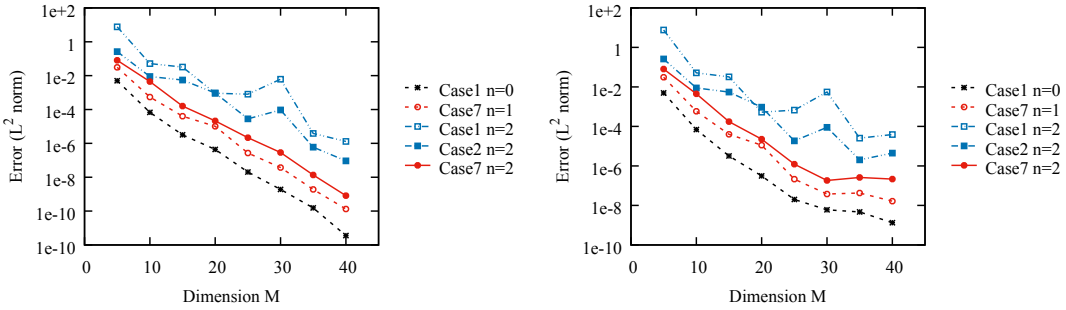
(a) Errors for the training set. (b) Errors for the test set.

Figure 5.4.2: Test case 5.4.1: The L^2 errors with respect reduced dimension M for one measurement failed case. $\mathcal{D} = [15.0, 24.9] \times [14.9, 15]$, n is the number of failures.



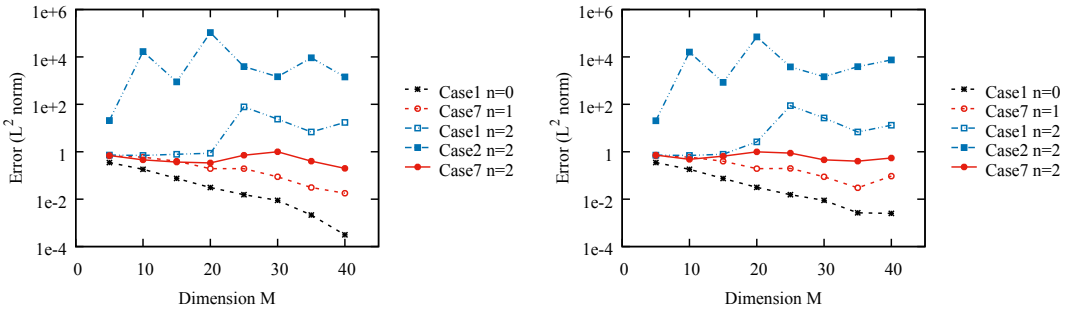
(a) Errors for the training set. (b) Errors for the test set.

Figure 5.4.3: Test case 5.4.1: The L^2 errors with respect reduced dimension M. $\mathcal{D} = [15.0, 24.9] \times [5.9, 15]$, n is the number of failures.



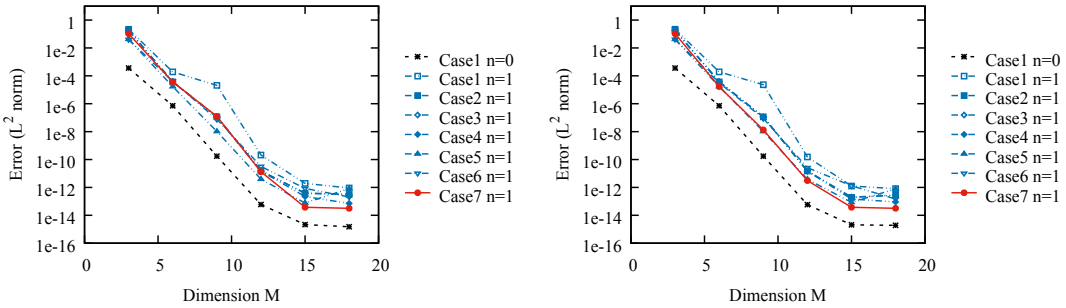
(a) Errors for the training set. (b) Errors for the test set.

Figure 5.4.4: Test case 5.4.1: The L^2 errors with respect reduced dimension M. $\mathcal{D} = [15.0, 24.9] \times [14.9, 15]$, n is the number of failures.



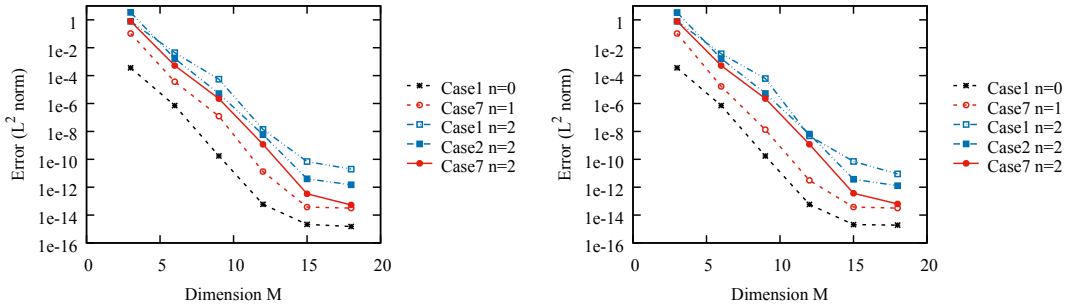
(a) Errors for the training set. (b) Errors for the test set.

Figure 5.4.5: Test case 5.4.1: The L^2 errors with respect reduced dimension M. $\mathcal{D} = [15.0, 24.9] \times [5.9, 15]$, n is the number of failures.



(a) Errors for the training set. (b) Errors for the test set.

Figure 5.4.6: Test case 5.4.2: The L^2 errors with respect reduced dimension M for IAEA 2D benchmark problem. One measurement failed case, n is the number of failures.



(a) Errors for the training set.

(b) Errors for the test set.

Figure 5.4.7: Test case 5.4.2: The L^2 errors with respect reduced dimension M for IAEA 2D benchmark problem. Two measurement failed case, n is the number of failures.

5.5 Conclusions and further works

Measurement failures for data assimilation with reduced basis, particularly, EIM, is a practical issue. We proposed a so called measurement failure tolerant EIM (mftEIM) approach to address this issue. Compared with other possible approaches, numerical results comfirmed that, the mftEIM approach provides the best interpolation in case of measurement failures. Need to point out that, the basis functions for mftEIM are not fixed, and depend on the detailed failure scheme. Thus, one can either calculate the specified new basis functions when the sensor failures happen, or pre-calculate the basis functions for all possible measurement failures cases, in the latter case, a relatively large storage is necessary.

More generally, if more measurements are available than the dimension of V_M, we consider the problem of reconstructing an approximation of f in the reduced basis space V_M from noiseless (or possibly noisy) samples of u at M' points $\{x_i\}_{i=1}^{M'}$, $M \leq M'$. Recently, the reconstruction with a weighted least-squares approximation in a given linear space V_M and m independent random samples has been studied in [62, 66]. The results show that stable results and optimal accuracy comparable to that of best approximation in V_M are achieved under the mild condition that m scales linearly with n up to an additional logarithmic factor and the points $\{x_i\}_{i=1}^{M'}$ are randomly chosen with respect to a sampling measure which depends on the space V_M : $k_M(x) := \sum_{i=1}^{M} |\tilde{q}_i(x)|^2$, where $\{\tilde{q}_i\}$ is an orthonormal basis of V_M.

Inspired by [66, 62], the on going work to stabilize EIM is based on a least-square framework with more points than originally required. In this framework, the essential issues is how to determine a 'relative optimal' set of the additional points under mild condition that M' is not so large with respect to the reduced dimension M, and the 'optimal' is in the sense that even loosing some points, the interpolation is still stable.

Chapter 6

Parameter sampling for multi-dimensional parameter problem

Contents

6.1 Introduction

Reduced basis methods (RBMs [150, 143, 185]) have nowadays become a widely accepted and used tools for realtime and/or multi-query simulations of parameterized partial differential equations (pPDEs). By using an off-line/on-line decomposition, the main idea is to use a high fidelity, detailed, but costly numerical solver off-line to compute approximations to the pPDEs for certain parameter values on-line.

The crucial ingredient of RBM is the choice of the basis functions $\{q_n\}_n$ which span the reduced basis space $V_n \in V$, typically based on the solution manifold $\mathcal{M} := \{u(\cdot,\mu)\ ;\ \mu \in \mathcal{D}\}$ over the parameter set $\mathcal{D} \subset \mathbb{R}^p$ with $p \in \mathbb{N}$, where V is a Hilbert space over a domain $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ endowed with inner product $(\cdot,\cdot)$ and induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$, e.g. $V \subset L^2(\Omega)$ or $H^1(\Omega)$. Due to practical implementation, the selection of these parameters is done e.g. by the generic greedy approach [175, 185, 48]. One typically needs a fine enough training set $\mathcal{D}^{(training)} \subset \mathcal{D}$ or $\mathcal{M}^{(training)}$ instead of the full set $\mathcal{D}$ or $\mathcal{M}$, over which an error estimator which is efficiently computable has to be evaluated, thus allows one to determine the 'worst' approximated snapshot, then let the corresponding parameter to be the next parameter sampling. Existing methods for basis construction by random parameter space sampling [98] or uniform parameter grids [99] reveal a breakdown of convergence of the global approximation error or high calculation times in case of very fine grids.

Note that in a high-dimensional parameter space case, the size of this training set quickly becomes considerable, rendering the computational cost substantial and perhaps even prohibitive. As consequence, since a fine enough training set is not realistic in practice, one is faced with the problem of ensuring the quality of the basis set under a non-rich enough training set, which leads to overfitting to the test set. It is worth noting that when dealing with certain high-dimensional problems, one may encounter the situation that the optimal basis set itself is of large size. This situation is, however, caused by the general complexity of the problems and we shall not discuss this further. In general, the two problems have to be addressed in the framework of RBM.

i) Overfitting: in case of too small training set, the error on $\mathcal{M}^{(training)}$ may be nicely decreasing or can ideally be reduced to 0 with growing basis dimension, but the error for independent test parameters remains large.

ii) Training times: in case of too large training set, the training time can be exorbitant, in particular in instationary scenarios without or with merely expensive error estimators.

Standard parameter sampling schemes such as uniform sampling (uniform gridding of the parameter space) or random sampling are another option for creating snapshots. However, if the dimension of the parameter space is large, uniform sampling will quickly become too computationally expensive due to the combinatorial explosion of samples needed to cover the parameter space. Random sampling, on the other hand, might fail to recognize important regions in the parameter space. One sampling strategy that provides a compromise between the uniformity and the size of the sample is the stratified sampling family, of which the popular Latin Hypercube Sampling (LHS) method is one example [155]. The LHS method is more efficient than uniform sampling and often more accurate than random sampling.

We would like to mention that there are existing literatures for RBM and various flavors of adaptivity, e.g., sampling set randomization, adaptive refinement of training sets, hp-RBM, time-partitioning etc., see, e.g., [5, 49, 54, 80, 108, 151], just to mention a few.

A recently proposed approach to address the challenge of sampling in a high-dimensional parameter space for building the reduced basis is the greedy algorithm. The basic idea of the greedy algorithm is to adaptively choose samples by placing the new sample point at the location in the parameter space where the estimated error with the reduced model is maximal. The approach is taken to determine a trial parameter set with n_{trial} parameters. At each point in this parameter set, the current reduced model is solved to determine the reduced states, which are then used to compute an error estimate. The parameter location in the trial set at which the error estimator is maximal is selected as the next sample point. Full model information is then generated via a full-scale system solve at this location, the basis is updated using the resulting snapshot information, and the reduced model is updated. These steps are then repeated, using the updated reduced model to compute the reduced states at each of the n_{trial} parameters.

It is known e.g. from [61] that adaptive methods show faster convergence rates if the Besov regularity of the solution in a certain scale exceeds the Sobolev regularity. For the off-line RB setting this means that the regularity of the solution with respect to the parameter is of crucial importance. If one single discretization is sufficient for approximating the solution $u(\cdot;\mu)$ well enough for all possible parameters μ, then adaptivity does not to make sense. On the other hand, if $u(\cdot;\mu)$ significantly differs with respect to μ, a joint discretization may be too fine.

6.2 Adaptive sampling for training set enrichment

6.2.1 Methodology

In this work we consider greedy adaptive algorithms and improvements of particular relevance to high-dimensional problems. While the ideas are of a general nature, we motivate and frame the discussion in the context of RBM and empirical interpolation method (EIM [21, 97, 147]) in which the greedy approximation approach plays a key role. The underlying reason for both overfitting and too long training times is the unknown ideal size and distribution of the training set. Hence, the main component of our first approach is an adaptive training set extension procedure to adapt the number and particular values of the training parameters for the underlying problem.

Assume we have the parameter domain $\mathcal{D} := [\mu_{min}^{(1)}, \mu_{max}^{(1)}] \times \cdots \times [\mu_{min}^{(p)}, \mu_{max}^{(p)}]$. For any parameter $\mu_i \in \mathcal{D}$, we denote by u_i the function $u(\cdot, \mu_i)$, then we define a grads like function (pseudo-grads) as follows:

$$d_\mu(u_i, u_j) = \frac{d[u](u_i, u_j)}{d[\mu](\mu_i, \mu_j)} \ , \tag{6.2.1}$$

where $d[u](u_i, u_j)$ is a distance function of u_i, u_j, defined as

$$d[u](u_i, u_j) := \| \frac{u_i}{\|u_i\|} - \frac{u_j}{\|u_j\|} \| \ , \tag{6.2.2}$$

or one can define

$$d[u](u_i, u_j) := \frac{|(u_i, u_j)|}{\|u_i\| \cdot \|u_j\|} \ , \tag{6.2.3}$$

where (u_i, u_j) is inner product of u_i and u_j. And $d[\mu](\mu_i, \mu_j)$ is the normalized distance of μ_i, μ_j in the parameter space, defined as

$$d[\mu](\mu_i, \mu_j) := \sqrt{\sum_{n=1}^{p} \big(\frac{\mu_i^{(n)} - \mu_j^{(n)}}{\mu_{max}^{(n)} - \mu_{min}^{(n)}}\big)^2}, \tag{6.2.4}$$

where $\mu_{max}^{(n)}$, $\mu_{min}^{(n)}$ are the maximal and minimal values of μ in the n-th dimension. When $\mu_{max}^{(n)} = \mu_{min}^{(n)}$, we use the definition $\frac{0}{0} = 0$ in (6.2.4).

To construct a global set of M sampling points in a multi-parameter space, we first initialize a set of initial uniform Cartesian parameter grid with $M_0 = m_1 \times m_2 \times ... \times m_p$ points, or we can initialize M_0 points with LHS method. Usually one can set $M_0 = 2^p$ initial sampling points on the corners of the parameter domain. For some particular problems e.g. the function $u(\cdot, \mu)$ doesn't change with the parameter on the boundary $\Gamma_{\mathcal{D}}$ of the parameter space, the LHS sampling method is proposed to set the initial points. With the initial parameter points we built a training set $\mathcal{M}_{M_0}^{(training)}$ of the manifold $\mathcal{M}$ which is not representative enough, then we enrich the training set based on the pseudo-grads of point-pairs: aways add the next parameter point which is the middle point of the point-pair (μ_i, μ_j) whose pseudo-grads is maximum among all the current $C_m^2 = m(m-1)$ pairs.

In more detail, at the m-th ($M_0 \leq m \leq M$) step, we have

$$\begin{aligned} \mathcal{D}_m^{(training)} &= \{\mu_1, \cdots, \mu_m\}, \\ \mathcal{M}_m^{(training)} &= \{u(\cdot; \mu_1), \cdots, u(\cdot; \mu_m)\}. \end{aligned} \tag{6.2.5}$$

Using (6.2.1), we are able to calculate the $m \times m$ grads-like matrix

$$D_\mu[\mathcal{M}_m^{(training)}] := (d_\mu(u_i, u_j)) \ , \ 1 \le i, j \le m \ . \tag{6.2.6}$$

Then the $m+1$-th parameter sampling is determined such that $u_{m+1} := (\mu_{i_1} + \mu_{j_1})/2$, where the parameter pair (μ_{i_1}, μ_{j_1}) satisfies

$$(i_1, j_1) = \underset{\substack{1 \le i,j \le m \\ d[\mu](\mu_{ij}, D_m^{(training)}) > r_{tol}}}{\arg\max} (D_\mu[\mathcal{M}_m^{(training)}])_{i,j}, \tag{6.2.7}$$

where $\mu_{ij} := (\mu_i + \mu_j)/2$ and $d[\mu](\mu_{ij}, D_m^{(training)}) := \min_{1 \le k \le m} d[\mu](\mu_{ij}, \mu_k)$ is the distance between μ_{ij} and $D_m^{(training)}$.

Indeed, the constraint in (6.2.7) is necessary. This can be understand as follows. We first select the pair (μ_{i_t}, μ_{j_t}) corresponding to the maximal element of $D_\mu[\mathcal{M}_m^{(training)}])$, then we evaluate the distance between $\mu_{i_t j_t}$ and $D_m^{(training)}$. If $d[\mu](\mu_{i_t j_t}, D_m^{(training)})$ is smaller than a given tolerance r_{tol}, that is to say, there already exists at least one sampling point μ_k in $D_m^{(training)}$ such that the candidate point $\mu_{i_t j_t}$ locates in the local domain $B(\mu_k, r_{tol})$ of μ_k, in this case we have reasons to believe that the difference between u_k and $u_{i_t j_t}$ is relatively small thus μ_k is of less interest. The tolerance r_{tol} is an indicator of the minimal admissible sampling density in the parameter domain $\mathcal{D}$.

Note also that several parameter pairs might maximize the element of $D_\mu[\mathcal{M}_m^{(training)}]$, in this case, (i_1, j_1) is picked among the set of maximizers. If $(D_\mu[\mathcal{M}_m^{(training)}])_{i_1,j_1}$ is smaller than a given tolerance ξ_{tol}, we would believe that the training set $\mathcal{D}_m^{(training)}$ is fine enough. The above process is summarized in Algorithm 6.2.10.

Algorithm 6.2.10 Adaptive sampling algorithm

Input: $(M_{max}, r_{tol}, \xi_{tol}, m_1, m_2, ..., m_p)$
Output: $(\mathcal{M}^{(training)}, \mathcal{D}^{(training)}, M^{(training)})$
Initialization: $\mathcal{D}_0$ is the initial mesh point set in the parameter space $\mathcal{D}$ e.g. with uniform Cartesian parameter grid with cardinality $|\mathcal{D}_0| = M_0 = m_1 \times m_2 \times ... \times m_p$.
$\mathcal{M}^{(training)} \leftarrow \{u(\cdot; \mu_0) \ ; \ \mu_0 \in \mathcal{D}_0\}$
$\mathcal{D}^{(training)} \leftarrow \mathcal{D}_0$, $M^{(training)} \leftarrow |\mathcal{D}_0|$
$D_\mu[\mathcal{M}^{(training)}] = (d_\mu(u_i, u_j))$, with $u_i, u_j \in \mathcal{M}^{(training)}$, $\mu_i, \mu_j \in \mathcal{D}^{(training)}$
while ($M^{(training)} < M$ and $\max(D_\mu[\mathcal{M}^{(training)}]) > \xi_{tol}$) **do**
    $(i_1, j_1) = \arg\max_{i,j} (D_\mu[\mathcal{M}^{(training)}])_{i,j}$
    $\mu_{i_1 j_1} = (\mu_{i_1} + \mu_{j_1})/2$
    if $d[\mu](\mu_{i_1 j_1}, \mathcal{D}^{(training)}) < r_{tol}$ **then**
        $(D_\mu[\mathcal{M}^{(training)}])_{i_1,j_1} \leftarrow 0$
    else
        $\mathcal{D}^{(training)} \leftarrow \mathcal{D}^{(training)} \cup \{\mu_{i_1 j_1}\}$
        $\mathcal{M}^{(training)} \leftarrow \mathcal{M}^{(training)} \cup \{u(\cdot; \mu_{i_1 j_1})\}$
        $M^{(training)} \leftarrow M^{(training)} + 1$
return $(\mathcal{M}^{(training)}, \mathcal{D}^{(training)}, M^{(training)})$

6.2.2 The determination of the adaptive parameters

The determination of the adaptive parameters ξ_{tol} and r_{tol} really depend on the underlying problem. Here we list some outlines.

For a fixed amount of the parameter sampling points M, we have the uniform sampling scale r_u

$$r_u = \sqrt[d]{1/M} \ , \tag{6.2.8}$$

where d is the dimension of the parameter space. Then r_{tol} satisfies

$$c_1 r_u \leq r_{tol} \leq r_u, \tag{6.2.9}$$

where $0 < c_1 < 1$, the left inequality constraints the minimal distance of sampling points, the right inequality constraints the sampling points departure from the uniform distribution. In a word, r_{tol} reflects the to what extent the sampling points can departure from the uniform distribution.

The determination of the lower limit ξ_{tol} of the grads-like function $d_\mu(u_i, u_j)$ is a little complicated. ξ_{tol} reflects the degree of the variation of the function with respect to the variation of the parameter. For a given initial sampling setting, $M_0 = m_1 \times \cdots \times m_p$, one can calculate the grads-like matrix D_μ and find the maximum element $d_0 = \max D_\mu$, then set

$$\xi_{tol} = c_2 d_0 \ , \tag{6.2.10}$$

where $0 < c_2 < 1$, for some good c_2. Note that ξ_{tol} can not be too large in order to make sure the loop in Algorithm 6.2.10 will not end until enough sampling points are selected.

6.3 Numerical results

We start by presenting numerical examples to illustrate the benefit of the adaptive sampling approach. Consider the following functions:

- Test Case 1, 2D parameter space, $\mathcal{D} \in \mathbb{R}^p$, $p = 2$, $u(x, \mu) = \frac{1}{\sqrt{(x_1+\mu_1)^2+(x_2+\mu_2)^2}}$, where $\mu \in \mathcal{D} = [0.01, 0.5]^2$, and $x \in \Omega := [0, 1]^2$ from [21].
- Test Case 2, 2D parameter space, $\mathcal{D} \in \mathbb{R}^p$, $p = 2$, $u(x, \mu) \equiv \frac{1}{\sqrt{1+(25+\mu_1 cos(\mu_2 x))x^2}}$, where $\mu = (\mu_1, \mu_2) \in \mathcal{D} = [0.01, 24.9] \times [0, 15]$, and $x \in \Omega := [0, 1]$ from [145].
- Test Case 3, 3D parameter space, $\mathcal{D} \in \mathbb{R}^p$, $p = 3$, $u(x, \mu) \equiv \frac{1}{\sqrt{1+\mu_3 x+(25+\mu_1 cos(\mu_2 x))x^2}}$, where $\mu = (\mu_1, \mu_2, \mu_3) \in \mathcal{D} = [0.01, 24.9] \times [0, 15] \times [0, 10]$, and $x \in \Omega := [0, 1]$.

To evaluate the quality of the sampling points for each sampling methods, we consider two model order reduction methods i.e. EIM and POD.

With EIM, we fist run EIM greedy algorithm based on the training set $\mathcal{M}^{(training)}$ with cardinality $m = |\mathcal{M}^{(training)}|$ to find the first $n \leq m$ basis functions $\{q_i\}_{i=1}^n$ and interpolation points $\{x_i\}_{i=1}^n$. With the basis functions and points, we reconstruct the field u in a test set $\mathcal{M}^{(test)}$ (corresponding to $\mathcal{D}^{(test)}$) which is different from the training set. In contrast to the training set, we construct this test set with enough sampling point such that $\mathcal{M}^{(test)}$ is representative of the manifold $\mathcal{M}$. Then we compute the errors as in (6.3.1):

$$e_n^{(test)}(u) \coloneqq \max_{\mu \in \mathcal{D}^{(test)}} \|u(\mu) - \mathcal{J}_n[u](\mu)\| / \|u(\mu)\|, \tag{6.3.1}$$

where $\mathcal{J}_n$ is the n-th order interpolant with EIM.

With POD, we can also construct the first n basis functions $\{q_i\}_{i=1}^n$, these basis functions span an n dimensional reduced space $V_n = \text{span}\{q_1, \cdots, q_n\}$. Then we evaluate the errors of the test set

with the following error indicator:

$$e_n^{(test)}(u) \coloneqq \max_{\mu \in \mathcal{D}^{(test)}} \|u(\mu) - P_{V_n} u(\mu)\| / \|u(\mu)\|, \tag{6.3.2}$$

where $P_{V_n} u$ is the orthogonal projection onto V_n.

We then compare the adaptive sampling approach with three different methods: uniform sampling, random sampling and Latin Hypercube Sampling. For each sampling method, we generate the same number of snapshots to construct the training set $\mathcal{M}^{(training)}$. To evaluate the quality of $\mathcal{M}^{(training)}$, we then apply POD and EIM to generate the reduced basis functions, and use (6.3.1) and (6.3.2) for the error estimation for the test set $\mathcal{M}^{(test)}$.

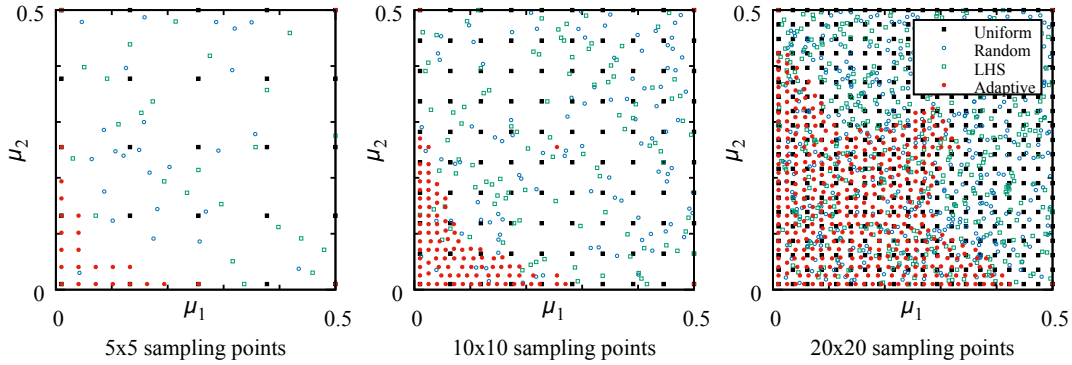
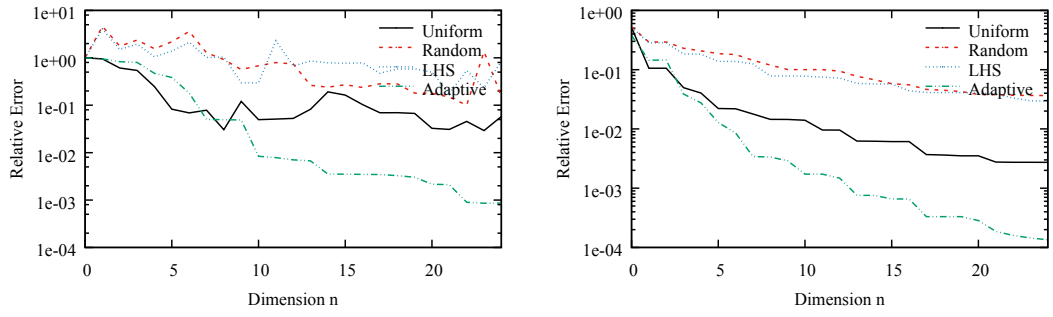


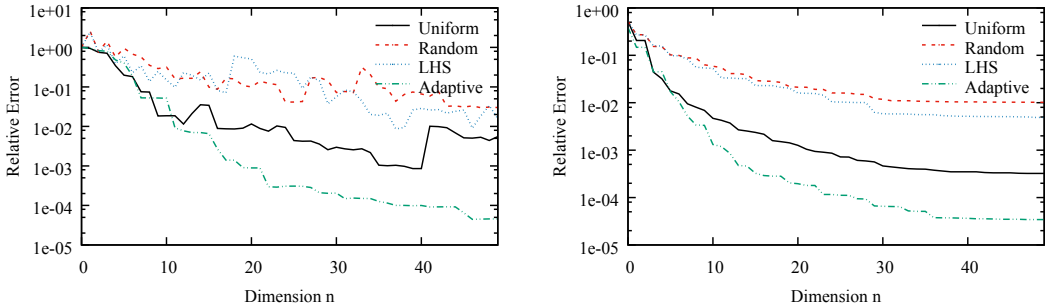
Figure 6.3.1: Test Case 1: Sampling points in the parameter space.

Figure 6.3.1 shows the sampling points distribution for Test Case 1 by using all the four methods (*Uniform, Random, LHS, Adaptive*) mentioned above, with a total sampling points $m = 25,\ 100,\ 400$. For the adaptive method, we set the initial points equal to $m_0 = 2^p$. We notice that, there is no clear difference for the distribution from *Uniform, Random, LHS* , but the adaptive one is rather different from others. For adaptive sampling, the points are symmetrical and gather around original point $(\mu_1 = 0, \mu_2 = 0)$, there are almost no points located in the region far from the original point. The distribution is consistent with the analytical expression of $u(x, \mu)$, which is symmetrical for μ_1 and μ_2 and also shows a divergence property at the original point.



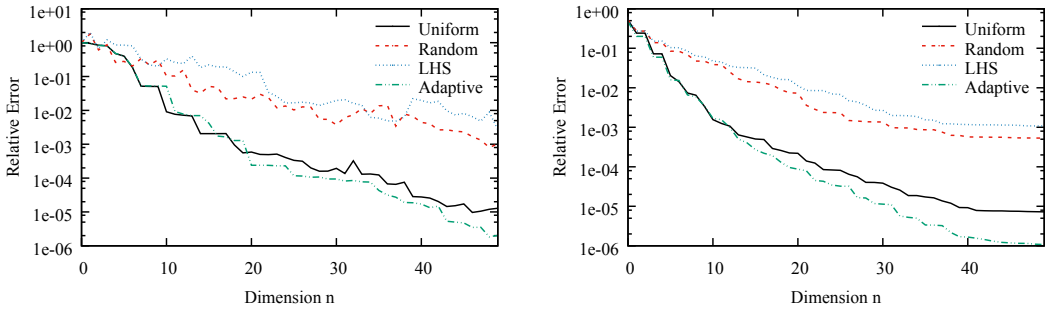
(a) EIM (b) POD

Figure 6.3.2: Test Case 1: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 25$.



(a) EIM (b) POD

Figure 6.3.3: Test Case 1: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 100$.



(a) EIM (b) POD

Figure 6.3.4: Test Case 1: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 400$.

We then show the different for the four sampling methods based on the test set which is different from the training set. In contrast to the training set, we construct the test set with enough snapshots to be representative to the manifold. The reconstruction errors for the test set are shown in Figure 6.3.2 - Figure 6.3.4.

In Figure 6.3.2, the total sampling points $m = 25$. With these points, we run EIM greedy algorithm to select the basis functions and interpolation points, then reconstruct the functions in the test set with different number of reduced basis for each sampling method, the maximum errors in L^2 norm is shown in Figure 6.3.2a. Besides, with the same training set, we use POD method to generate the basis functions and approximate the functions in the same test set. The errors are shown in Figure 6.3.2b. We observe that, in case of too small training set ($m = 25$), the overfitting happens, the error for independent test parameters remains large and also not stable. But in this case, the adaptive sampling method is much better than the others, both for EIM and POD methods.

We then enlarge the training set to $m = 100$, the results can be found in Figure 6.3.3. For a larger training set, the overfitting is improved a lot. In this case, the reconstruction for the test is also more stable for EIM and POD for all the four sampling methods. We notice that, the uniform sampling method is one order accurate than random sampling and LHS, but the adaptive sampling method behaviors much better than uniform case, even two order more accurate.

By enlarging the training set up to $m = 400$ snapshots, which is close to saturation. The difference between uniform sampling method and adaptive sampling method is narrowing. But the

latter is still the best.

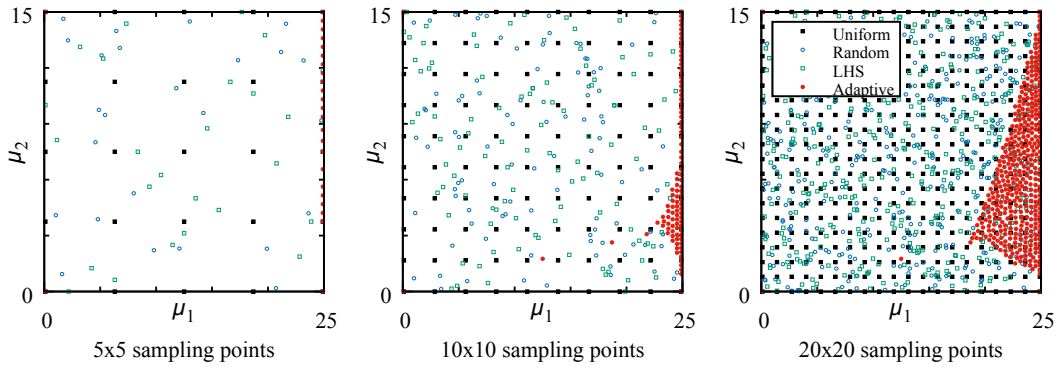
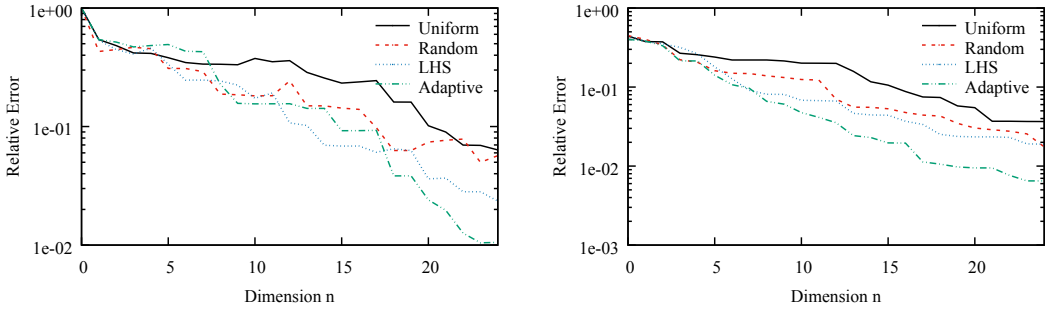
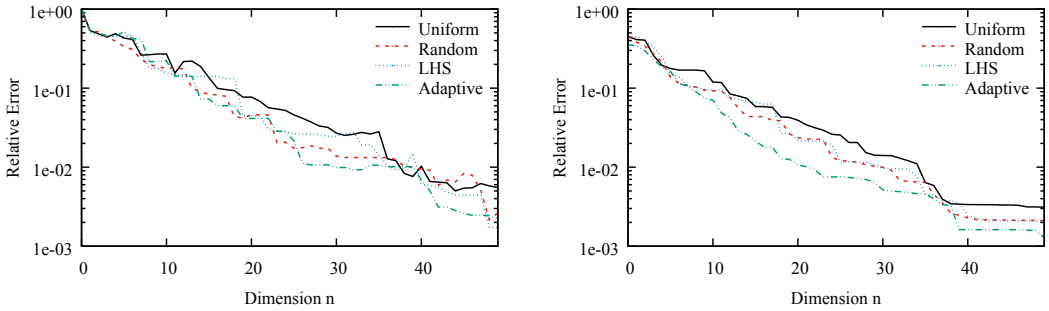


Figure 6.3.5: Test Case 2: Sampling points in the parameter space.



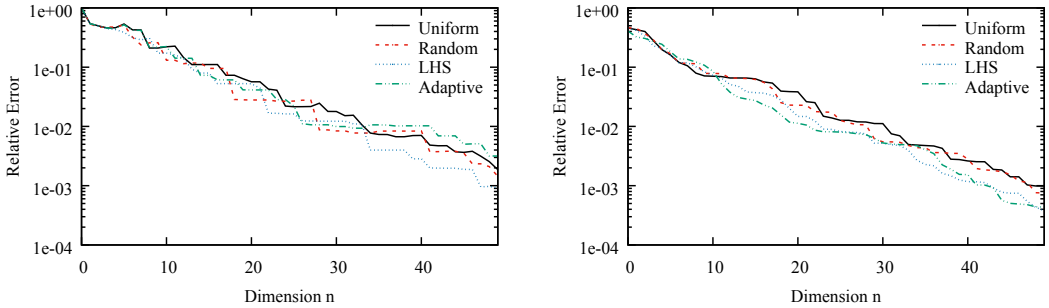
(a) EIM (b) POD

Figure 6.3.6: Test Case 2: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 25$.



(a) EIM (b) POD

Figure 6.3.7: Test Case 2: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 100$.



(a) EIM (b) POD

Figure 6.3.8: Test Case 2: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 400$.

We do the same test for Test Case 2. The sampling points distribution is shown in Figure 6.3.5. Notice that we still find a big difference for adaptive sampling method with respect to other three methods. The errors for each sampling case are shown in Figure 6.3.6 - Figure 6.3.8. We notice that, in this case, the uniform sampling method is the worst because of the specified parameter dependency. But the adaptive sampling method is still the best especially for the relatively small training set case, this method is able to take the most important information of the manifold in the beginning.

The benefit of adaptive sampling method is more obvious in higher dimensional parameter space case. In Figure 6.3.9 we show the sampling points distribution for the 3D parameter space case, Test Case 3. In this problem, we generate two level of the training set, with 125 and 1000 snapshots respectively. For adaptive sampling method, we set the initial points equal to $m_0 = 2^3$. The results are similar to 2D case, the sampling points for adaptived one is totally different with others. The errors show in Figure 6.3.10 and 6.3.11 again confirm that, the adaptive sampling method is the best when the total amount of sampling point is limited, this is usual for high dimensional parameter space problem.

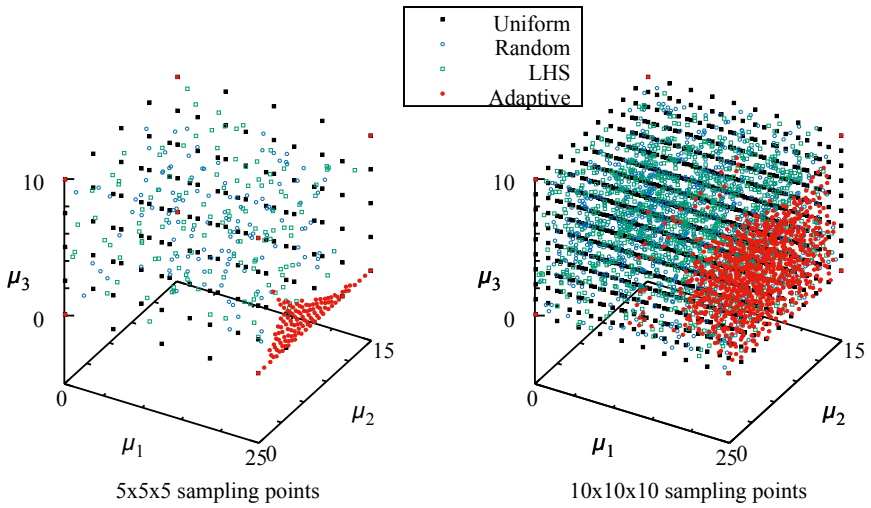
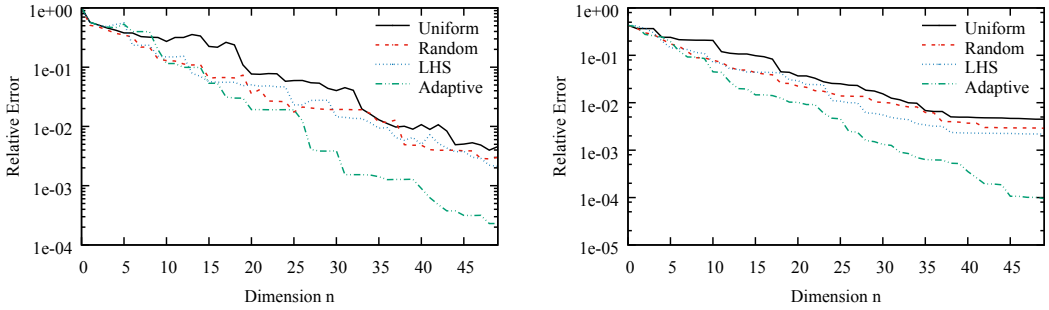
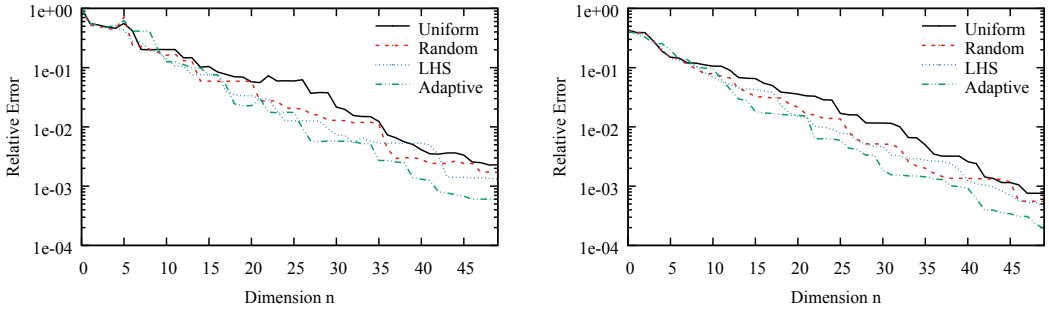


Figure 6.3.9: Test Case 3: Sampling points in the parameter space.



(a) EIM (b) POD

Figure 6.3.10: Test Case 3: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 125$.



(a) EIM (b) POD

Figure 6.3.11: Test Case 3: Relative error (in L^2 norm) for test set with respect to the reduced dimension n. The number of sampling points is $m = 1000$.

6.4 Application to EDF nuclear reactor cores

In this section, we test the sampling methods (uniform sampling, random sampling and adaptive sampling) based on real reactors at EDF. For this application we chose an operating PWR 1450 MWe reactor core that has been used in the publication [59]. On the basis of this core we test the method and see how good are the results. We will do those test in a simulation framework using the COCAGNE code for neutronic simulation that is developed by EDF [50]. The result is the comparison on the reconstruction done through the proposed method and the known calculation.

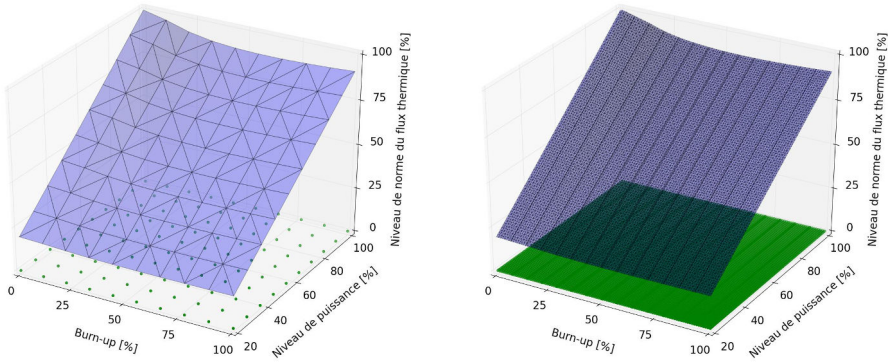
For the practical reactor core application, here we consider two-dimensional parameter space settings similar to Section 4.1 with $\mu = (Pw, Bu) \in \mathcal{D} := [0.3, 1] \times [0, Bu_{max}]$. We evaluate the quality of parameter sampling methods with GEIM, and the reduced basis are selected based on the greedy algorithm presented in Section 3.4.2, other settings are the same as in Section 4.3.

We list the norm of the thermal flux $\|\phi_2(\mu)\|$ as a function of parameter μ in Figure 6.4.12 for the total sampling points amount to $m = 125$ and 11300. As shown in this figure, the parameter dependence of the flux is more or less regular, the effect from power level is stronger than the burnup parameter. We construct the training set $\mathcal{M}^{(training)}$ with $m = 125$ snapshots and build a finer test set with $m = 11300$ snapshots.

We show the error distribution as a histogram in Figure 6.4.13 for the three sampling methods:

uniform sampling method, random sampling method and adaptive sampling method. In this figure, we use the first five basis functions for GEIM process to reconstruct the field in the test set. We notice that, from statistical viewpoint, the adaptive sampling method is the best among all the three methods.

Further more, we list the average and standard deviation of the error with different amount of basis functions in Figure 6.4.14. This again confirms that, in all case, the sampling points selected by adaptive sampling method is more representative than uniform and random sampling methods.



(a) Sampling points $m = 125$. (b) Sampling points $m = 11300$.

Figure 6.4.12: Variations of the norm of the thermal flux for each parameter sampling case.

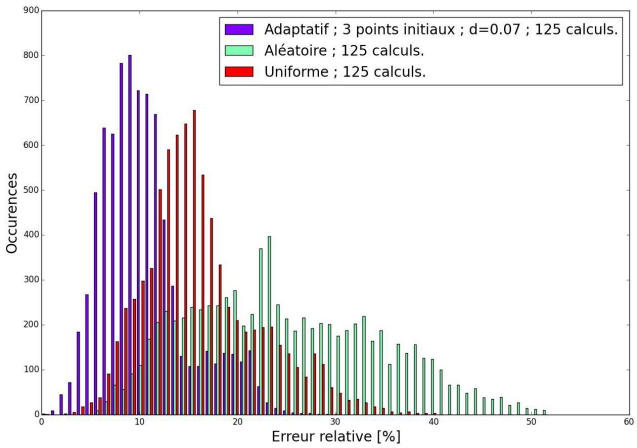


Figure 6.4.13: Reconstruction errors for the three methods (uniform sampling method: Uniforme; random sampling method: Aléatorie; Adaptive sampling method: Adaptatif) with the first five basis functions.

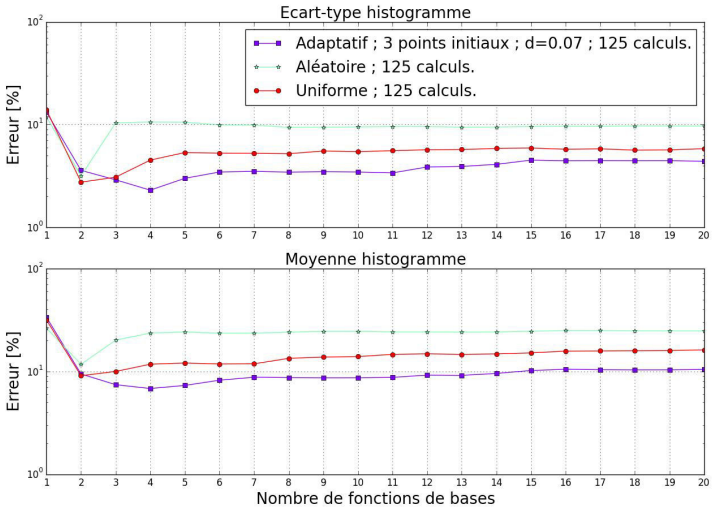


Figure 6.4.14: The average (Moyenne) and standard deviation (Escart-type) of the error with respect to the reduced dimension n.

6.5 Computational cost

Let us turn to analysis the computational cost of the adaptive sampling method. We denote by t_i the computing time for one inner product (u_i, u_j) evaluation, and t_s the one snapshot evaluation with COCAGNE code. From Algorithm 6.2.10, we notice that the evaluation of the grads-like matrix $D_\mu[\mathcal{M}_m^{(training)}]$ dominates the computational cost, to select m snapshots, we need $\mathcal{O}(m^2/2)$ inner product evaluations and $\mathcal{O}(m)$ snapshot evaluations.

With COCAGNE code, it costs around $t_s = 46$ seconds for one snapshot calculation in Pin-by-Pin case, and one inner product evaluation costs $t_i = 0.005$ second. Table 6.1 reports some computational times to illustrate the cost of the adaptive sampling method and the uniform sampling method. We notices that, for practical engineering problem, the off-line phase – preparing the training set – for the application of reduced order method is really time costly, especially for the multi-dimensional parameter space, the amount of the sampling points is rather limited. In this case, the adaptive sampling method can be used in the off-line phase to build the training set with only a slight increasing of the computational cost to obtain a higher accuracy than to other methods.

Items	Computational cost	
One snapshot calculation	46	
One inner product calculation	0.005	
	Adaptive	Uniform
Prepare 125 snapshots	5789	5750
Prepare 1000 snapshots	48500	46000
Prepare 11300 snapshots	838225	519000

Table 6.1: Computational cost (in seconds) for the adaptive sampling method and the uniform sampling method in a PWR 1450 MWe reactor core.

6.6 Conclusions

In a high-dimensional parameter space, the challenge for uniform sampling method is, unreachable, e.g. for $p = 10$, if we sample in each dimension by 10 points, the amount of sampling points is 10^{10}. On the other hand, with adaptive method, we first select 2^{10} initial sampling points, which is much smaller and also reachable, then with the greedy approach, we are able to pick the most information from the ideal manifold $\mathcal{M}$ with fixed number of snapshots.

The random sampling method and LHS method, are reachable for high-dimensional parameter case. But as the figures shown, some times may miss some important sampling points or with redundancy. Compared with other adaptive methods e.g. [5, 49, 54, 80, 108, 151], the approach as shown in Algorithm 6.2.10 saves computing cost, it reduces the number of function evaluation which might be very costly in the real engineering problem. This gives the adaptive sampling method more potential for engineering problem with high-dimensional parameter space. Nevertheless, we need to point out that, the adaptive sampling method may fail for some special case, e.g., the snapshots are from the the function in [151] with periodicity or symmetrical parameter space. The algorithm with pseudo-grads as an indicator may fail because of periodicity or symmetry. In this case, one can set the initial points with LHS method or one can refer to [151] with locally adaptive approach for anisotropic parameter space.

Chapter 7

Conclusions and perspectives

7.1 Summary and conclusions

This thesis presents work towards the improvement of the physical and numerical interpretation of the information involved in data assimilation with efficient model reduction strategies for systems modeled by PDEs. We specifically focused on the data assimilation task: state estimation for stationary problems, especially neutronic state estimation in nuclear reactor applications.

In the first part of the present thesis, we analyzed and adapted the generalized empirical interpolation method (GEIM) and the parametrized-background data-weak (PBDW) approach to the state estimation problem. These approaches aim at reducing the problem's complexity for state estimation based on the fact that the Kolmogorov n-width of the manifold of all solutions decays fast with n. The GEIM [144] that allows, from values of the measurements, to build a linear combination of some precomputed states associated to some well chosen parameters. The linear combination is determined in such a way that it takes the same values on the sensors as the original state. This concept generalizes the classical — e.g. polynomial or radial basis — interpolation procedure, the main difference is that the interpolating functions are not a priori known but are chosen as suitable state that, altogether, carry the information of the set of all states. The PBDW approach was originally proposed in [148] for perfect observations then the noisy observations case was analyzed in [149]. It is characterized by the following characteristics.

i) *Variational formulation*: the most important feature of PBDW is that, it is a special case of 3D-Var variational data assimilation for a parametrized background and a particular choice of (penalized-update) background covariance. Benefiting from the background space V_n, PBDW avoids modeling the covariance matrices in 3D-Var variational data assimilation formulation.

ii) *Background and update spaces*: PBDW provides the state estimate u as the sum of two contributions: $u = v + \eta$. The first contribution to u, v, is called "deduced background estimate" which represents anticipated model part; and V_n is interpreted as a background or prior space which approximates the best-knowledge manifold on which the true state resides. This second contribution to u, η is called "update estimate" which accommodates unanticipated update that the model couldn't represent.

We have provided the following contributions to the original GEIM and PBDW formulations in this thesis.

- *Stability analysis*: We formulated the posteriori error analysis with respect to measurement

noise for GEIM, RRB and PBDW. The estimation error is attributed to two distinct contributions: i) The *deterministic error*, which arises from the fact that modeling error is inevitable (i.e. $u^t \notin V_n$) and that the update space W_m and the background space V_n are finite dimensional; this error is present even if each observation is noise-free, though one is able to probe the true deterministic state. We use the so-called Lebesgue constant Λ to evaluate the deterministic error; ii) The *stochastic error*, which arises from the fact that each observation is noisy, corrupted by the random observation noise. We use the stability factor $\|Q\|$ to evaluate the stochastic error. We also proposed algorithms to optimize the update space W_m with respect to different measures, i.e., diminishing the deterministic error or diminishing the stochastic error.

- *Constrained Stabilized Reconstruction*: We proposed the so-called constrained stabilized generalized empirical interpolation method (CS-GEIM) and the constrained stabilized parametrized-background data-weak (CS-PBDW) approach to further improve the stability and robust performance for state estimation with data assimilation with noisy measurements. The constraints rely on the coefficients of the reduced basis functions, they are not the same order of quantity, on the contrary, they decay in certain rate with the reduced dimension n increases. We demonstrated that with the coefficients constraints, the recovery with reduced basis dramatically decreases the noise amplification, with good stability performance with respect to the reduced basis dimension n and the number of measurement m. To go further, we proposed the regularized versions (R-GEIM, R-PBDW) to diminish the computational cost in the on-line phase, while retaining the same accuracy as constrained stabilized formulations.

In the second part of the present thesis, we applied the methodologies for sensor placement and field reconstruction applications in nuclear domain. The framework of GEIM/PBDW makes it possible for us to focus on the need for effective, many-query design evaluation in nuclear reactor domain: i) either real-time scenarios (control, parameter estimation, monitoring), where the solution of the problem needs to be known very quickly under limited resources for a previously unknown parameter, ii) or multi-query scenarios (design and optimization, multi-model/scale simulation), where the problem has to be solved repeatedly for many different parameters. In the nuclear reactor applications, we set the goal to reconstruct the thermal/fast flux and power distributions for a wide range of parameters variation domain by using the measurements from sensors with combination of reduced basis methods, instead of solving the original neutronic governing equations with high resolution.

- *Sensor placement in a nuclear reactor core*: We first adapt the GEIM greedy algorithm, to be able to reconstruct the flux and power fields only from the thermal flux measurements. Then we show its application in optimizing the sensor placement in nuclear reactor cores. Numerical results on 1D/2D benchmarks confirm its feasibility. Furthermore, GEIM, can be applied efficiently to the analysis of real nuclear cores. With the GEIM approach and the sensor information, we are able to provide an optimized instrumental network to measure the neutronic flux inside the core. The locations of the instruments show a global coherence with respect to what is known from the physical behavior and process. This determination of the optimal localizations is done over all the geometries of reactors operated by EDF: PWR 900 MWe, PWR 1300 MWe and PWR 1450 MWe. Several strategies have been developed in order to emphasize the advantages and limitations of the determination of the instrumental setup. Numerical results show that the proposed method gives some instrumental setup that are consistent with what is expected from a physical point of view. Moreover, we demonstrate that, with respect to the random case, the results of optimal networks are clearly better. Finally, we conclude that the strategy which is the closest to the actual instrumentation setup gives the best result.

- *Neutronic field reconstruction*: We first apply the GEIM to control rods movements problem to improve the flux and the power distribution knowledge in the whole domain without knowing the control rods position (step) parameters thanks to the measurements. To reach this goal, we first construct a set of snapshots by solving large amounts of PDEs for the underlying physical problem. Then we select the sensor places (interpolation points) and the basis functions in a greedy way with GEIM greedy algorithm. With the selected measurements, we are able to reconstruct the flux and power distribution in high accuracy. Numerical results confirm that the proposed GEIM methodology is hopefully acceptable in engineering usage, especially the online-offline computational framework is good for on-line monitoring purpose.

  Then we applied the GEIM method to reconstruct the Pin-by-Pin flux/power distribution during the life cycle of a nuclear reactor core. All the simulations are based on the PWR 1450 MWe nuclear reactor operated by EDF. Numerical result shows that, the manifold of the underlying problem (even in Pin-by-Pin wise), which reflects the burnup and power evolution of the reactor core of the whole life, is rather regular. This makes it possible to reconstruct the fields with only a few basis functions and measurements. Further more, by considering the measurement system which is close to the real engineering case, the proposed method still provides an acceptable accuracy.

  Neutronic field reconstruction with noisy measurements: The extension of the method is to take into account the effect of the noise to qualify its impact when few instruments remains and also to control it through various technique, either data pre-processing or mathematical correction of the induced error. In this aspect, we propose the methodologies so called CS-GEIM and CS-PBDW to control the noise and also the model bias. Based on the stability analysis, we are able to optimize the sensor size and location. With the stabilized algorithms, we are able to control the measurement noise amplification. Further more, by adding more measurements, the proposed methods are able to provide even better accuracy than the noise level. The algorithms we proposed are robust with respect to the measurement noise and model bias. Finally, we emphasize that the regularized versions (R-GEIM, R-PBDW) with low computational cost are more suitable for on-line calculation in case the shape of the noise is known, i.e., Gaussion noise.

In the third part of the thesis, we present the supplementary developments for industrial applications of the methodologies.

- *Sensor failures and constraints on the sensor locations*: Measurement failures for data assimilation with reduced basis, particularly, EIM, is a practical issue. We proposed a so called measurement failure tolerant EIM (mftEIM) approach to address this issue. Compared with other possible approaches, numerical results comfirmed that, the mftEIM approach provides the best interpolation in case of measurement failures. Need to point out that, the basis functions for mftEIM are not fixed, and depend on the detailed failure scheme. Thus, one can either calculate the specified new basis functions when the sensor failures happen, or pre-calculate the basis functions for all possible measurement failures cases, in the latter case, a relatively large storage is necessary.

- *Parameter sampling for multi-dimensional parameter problems*: In the framework of reduced basis method, the crucial ingredient is the choice of reduced basis $\{q_n\}$ from the discrete manifold $\mathcal{M}^{(training)} := \{u(\mu) \ ; \ \mu \in \mathcal{D}^{(training)}\}$, which span the reduced basis space V_n . We proposed an adaptive sampling algorithm to sample the parameters in $\mathcal{D}$ to construct the discrete manifold. Compared with other adaptive methods, the method we proposed saves computing cost, it reduces the number of function evaluation which might be very costly in the real engineering problem. This gives the adaptive sampling method more potential for engineering problem with high-dimensional parameter space. Nevertheless, we need to point out that, the adaptive sampling method may fail for some special case, e.g., the snapshots are

from the the function in [151] with periodicity or symmetrical parameter space. In this case, one can set the initial points with LHS method or one can refer to [151] with locally adaptive approach for anisotropic parameter space.

Additionally, we shall point out that, the reduced order modeling techniques we have studied can make important contributions to the controlling or the analysis of the large scale industry systems. The modeling and controlling of the uncertainty for data assimilation with reduced basis is an issue, because the reduced order model is not robust with respect to the sensor locations and observation values. Similar works have been done in [6, 70, 79] etc. In [6, 70], the ill-conditioning because of the presence of the reduced basis is improved by using Tikhonov regularization, which is proved to be effective in terms of reduction of errors. But in those works, not measurement information has been taken into consideration. In [79], the authors present a technique for statistically modeling errors introduced by reduced order models, the method employs Gaussian-process regression to construct a mapping from a small number of computationally inexpensive 'error indicators' to a distribution over the true error. But in that work, the study in controlling the noise amplification because of the present of the reduced basis is rather limited. Our contribution in controlling the error of reduced order modeling is rather different from the previous, the constraints on the reduced basis has been taken in to consideration, which is of great important not only from the mathematic point of view, but also from the physical point of view. Besides, more theoretical studies in this aspect shall be carried out in the future.

7.2 Perspectives and future works

During the course of this work, we have identified several areas of future research. For purposes of presentation, we shall distinguish between the methodologies themselves and the benefits to industrial (e.g. nuclear power plant systems) design, operation, surveillance and maintenance, etc.

7.2.1 Further verification for engineering applications

Sensor placement designing at EDF: In Chapter 4, we demonstrated the sensor placement optimization framework with GEIM based on all types of nuclear reactors that are operating at EDF (PWR of 900 MWe, 1300 MWe and 1450 MWe). With this methodology, the locations of the instruments show a global coherence with respect to what is known from the physical behavior and process. But all those demonstrations were done without considering measurement noise. However, in real engineering case, quality and noise-free data is not easy to obtain – e.g., for MFC, noise is always exist because of the data acquisition procedure, signal processing, background noise, more especially, the delayed signal response for Self Powered Neutron Detector (SPND [94, 210]) which is fixed in the nuclear reactor core during operation, etc. Thus, the optimal instrument network designed with GEIM shall be tested in noise case, or even take the noise into consideration when designing the network.

Noise control in field reconstruction: Evaluating the neutronic state of the whole nuclear core is a very important topic that has strong implication for nuclear core management and for security monitoring. In Chapter 4, we demonstrated the benefits of GEIM/PBDW in field reconstruction. We tested the methodologies based on EDF reactors in assembly/Pin-by-Pin wise with noise-free measurements, the reconstruction errors stabilized at the order of 10^{-4} for assembly wise and 10^{-3} for Pin-by-Pin wise which are totally acceptable in engineering point of view. Usually, the noise level for sensors in nuclear reactor cores is around 1 - 5 10^{-2} [38], which is one to two order higher than the accuracy derived with noise-free measurements. We demonstrated the benefits of CS-GEIM/CS-PBDW based on the 2D IAEA benchmark for controlling the noise amplification.

Further demonstrations shall be done based on EDF reactors with real measurements.

Field reconstruction using heterogeneous instruments: In this thesis, we only studied the reconstruction with measurements from the Mobile Fission Chambers (MFC). However, the core state estimation should be more accurate when more measurements are collected in the core. There are four types of instrument that are usually used to monitor the nuclear power core:

- Mobile Fission Chambers, which are inside the active nuclear core,
- Thermocouples (TC), which are above the active nuclear core,
- fixed ex-core detectors,
- Self Powered Neutron Detectors (SPND [94, 210]) which are fixed inside the active nuclear core.

Using information from heterogeneous sources is a difficult task. Some works have been done in this respect at EDF [37, 38, 39] in the framework of data assimilation i.e. BLUE. Such a method allows to combine in a coherent framework the information coming from model and the one coming from various type of observations. Beyond the inner advantage to use heterogeneous instruments, this leads to obtain a significant increasing of the quality of neutronic global state reconstruction with respect to individual use of measures. The extending/adapting for GEIM/PBDW with heterogeneous instruments could be studied for neutronic applications.

One natural solution is that, in the off-line phase, select the reduced basis and locations of instruments using heterogeneous instruments with different class of measurement functionals; in the on-line phase, reconstruct the fields using heterogeneous instruments with different class of measurement functionals and different noise levels. This approach has been verified in [145], where the authors reconstruct the velocity and pressure of a fluid by using pressure measurements only.

7.2.2 Data-enabled, physics-constrained predictive modeling for nuclear power plant systems

Monitoring nuclear reactor core spatial kinetics: In the analysis of the nuclear reactor dynamics, which is governed by the neutronics, the most used approach is constituted by the point-kinetics equations, which are a set of coupled non-linear ordinary differential equations that describe both the time-dependence of the neutron population in the reactor and the decay of the delayed neutron precursors that allowing for the main feedback reactivity effects. For the point-kinetics equations, the strongest approximation regards the shape of the neutron flux, which is assumed to be represented by a single, time-independent spatial mode. Otherwise, if the reactors are characterized by complex geometries and asymmetric core configurations, more accurate modeling approach, namely, aims at solving the energy-, space-, and time-dependent kinetic equations is necessary to provide more detailed insights to the reactor behaviour during operational transients. Furthermore, in order to develop suitable control strategies for such reactors, the spatial effects induced by the movements of the control rods have to be taken into account as well. Thus, the Multi-Physics (MP) approach in which all the involved physics are simulated is necessary. The main drawback of the MP approach is that the computational burden is quite high, and simulating the entire core turns out to be very demanding in terms of computational costs and times. In addition, it is quite difficult to get the dynamics of the governing system and then set up a simulation tool that may assess and represent the dynamic response of the overall system at different operating conditions.

In this context, the ROM technique, such as RB methods, which can lead to a simulation tool with real time simulation, still solving a set of partial differential equations is proposed in

[191, 189, 190, 192, 193]. With ROM, one is able to efficiently reproduce the neutron flux distribution allowing to take into account the spatial effects induced by the dynamics – e.g., the movements of the control rods – with a computational speed-up of 30000 times, with respect to the full model. But all those approaches, are conducted only for simulations with correct parameters, and no additional ingredient of incorporating measurement information to the simulations. The specified GEIM/PBDW frameworks, are able to integrate additional measurement information, also benefit from RB methods, shows high potential for monitoring propose. In Chapter 3, the fields reconstruction during control rods movements were studied, but that study was based on the stationary governing equations. Thus, a new data assimilation framework based on GEIM/PBDW, with energy-, space-, and time-dependent kinetic equations, aims to monitor nuclear reactor core spatial kinetics could be on the agenda. For long period or wide range monitoring propose, the locally adaptive GEIM/PBDW could be a solution, similar works on locally or adaptive methods can be found in [151, 168].

Data-enabled, physics-constrained predictive modeling: Full scale nuclear reactor testing, numerical simulation of high dimensional (full-order) dynamic models or operation testing are some of the fundamental but complex steps in the various design phases of recent or next generation civil nuclear power plant. Current nuclear reactor designs have increased in complexity (multidisciplinary, multi-objective or multi-fidelity) and need to address the challenges posed by the nonlinearity of the objective functions and constraints, uncertainty quantification in neutronics, thermohydraulics, hydromechanics and materialogy problems or the restrained computational budgets. Even with recent advances in computing power and algorithms, a multitude of first-principles-based computations of physical problems remains out of reach, even on the most powerful supercomputers. With the aim to reduce the computational burden and generate low-cost but accurate models that mimic those full order models at different values of the design variables, recent progresses have witnessed the introduction, in real-time and many-query analyses, of surrogate-based approaches as rapid and cheaper to simulate models.

The reduced order modeling (ROM) methodology has found widespread applications in simplifying the physics model and reducing numerical complexity. The various ROM methods do this in general by exploiting similarity within an ensemble of high fidelity “snapshot" solutions which sample a certain parametric domain of interest. The number of degrees of freedom (DoF) is then reduced while retaining the problem's physical fidelity, thus allowing predictions of the physical data to be provided with lower evaluation time and storage than the original full scale model. Example can be found in [201], the authors proposes developing and incorporating a ROM server into the engineering design workflow. The ROM server stores all data associated with a given engineering model and automatically constructs a ROM every time a model is created or updated, thus maintaining a consistent version of information across multiple engineering teams.

A favoured application of ROM is in aerodynamic analyses [217, 212]. In these works, a comprehensive and state-of-the-art survey on common surrogate modeling techniques and surrogate-based optimization methods is given, with an emphasis on models selection and validation, dimensionality reduction, sensitivity analyses, constraints handling or infill and stopping criteria. Benefits, drawbacks and comparative discussions in applying those methods are described. In [212], the steady ROMs are used to predict the static aeroelastic loads in a multidisciplinary design and optimization (MDO) context, where the structural model is to be sized for the (aerodynamic) loads. The different ROM methods are applied to a 3D test case of a transonic wing-body transport aircraft configuration.

In nuclear domain, ROM is used for the improvement of the control-oriented modelling of the Gen-IV Lead-cooled Fast Reactor European Demonstrator (ALFRED [9]), in which the spatial dependence plays a relevant role for the respect of technological constraints and for the reactivity feedbacks. Thus, it has a different control strategies with respect to LWRs and SFRs. As reported in [141], a ROM-based components for object-oriented simulator in which a POD-Galerkin method

for Finite Volume approximation of Navier-Stokes and RANS equations (POD-FV-ROM) was developed. The new approach make it possible to provide the nuclear reactor control simulation tool with spatial information capabilities, enhancing the level of detail without a strong computational burden.

All those examples mentioned above, as a result, scientific and engineering community continue to rely on physical intuition and empiricism to derive approximate models for prediction and control. Advanced ROM techniques have allowed people to move from data to information, because in many systems, and despite the large amount of data, the hidden information was quite reduced thanks to ROM. One natural question is, can one extract cause-and-effect relationships and create reliable predictive models based on a large number of observations of physical phenomena? To develop improved predictive models of complex real-world problems, however, one may need to pursue a more balanced view. We refer to [82], a more informal article on the use of data in modeling complex physical problems. The main idea is that data cannot be an alternative for physical modeling, but when combined with – and informed by – a detailed knowledge of the physical problem and problem-specific constraints, it is likely to yield successful solutions. Nowadays, advanced clustering techniques not only help engineers and analysts, but are also crucial in many areas where models, approximation bases, parameters and so forth are adapted depending on the local state of the system thanks to the large amount of data. Machine and manifold learning also helps to extract the manifold in which the solutions of complex and coupled engineering problems are defined. Many works have been done to addresses construction of solution manifolds and its use for interpolating new solutions on the manifold, construction of parametric solutions on the just extracted manifold and definition of behavior manifolds to perform data-driven simulation while avoiding the use of usual constitutive equations. We refer to [179, 180, 12, 186, 8, 45] for further information in this aspect.

As already explained, the data assimilation frameworks based on GEIM/PBDW, are able to take the physics information and the observation information into consideration for field reconstruction. Thus, a data-enabled, physics-constrained predictive modeling framework based on GEIM/BPDW for nuclear power plant systems design, operation, surveillance and maintenance, etc., is of great interests in the further.

Appendix A

Analytical solution to the neutron diffusion equation for a homogeneous slab reactor

In this appendix, we use the one-dimensional multigroup neutron diffusion equation to reconstruct the neutron flux in a slab reactor from the nuclear parameters of the reactor, boundary and symmetry condition, initial flux and k^{eff}. The neutron multigroup difussion equation (3.2.11) in one dimension, two groups of energy and stationary state are expressed by the following equations:

$$-\frac{d}{dx}\Big(D_g(x)\frac{d}{dx}\phi_g(x)\Big)+\Sigma_g^a(x)\phi_g(x)=\sum_{\substack{g'=1\\g'\neq g}}^{2}\Sigma_{g'\to g}^s(x)\phi_{g'}(x)+\frac{\chi_g}{k^{eff}}\sum_{g'=1}^{2}\nu\Sigma_{g'}^f(x)\phi_{g'}(x),\quad g=1,2 \tag{A.0.1}$$

Due to the fact of the reactor to be homogeneous in the case of one single region (fuel) or homogeneous per part in the case of more than one region (fuel and reflector), these nuclear parameters are constant in each region, i.e.: $D_g(x)\equiv D_g,\ \Sigma_g^a(x)\equiv\Sigma_g^a,\ \Sigma_{g'\to g}^s(x)\equiv\Sigma_{g'\to g}^s$ and $\Sigma_{g'}^f(x)\equiv\Sigma_{g'}^f$. With this, (A.0.1) it becomes:

$$-D_g\frac{d^2}{dx^2}\phi_g(x)+\Sigma_g^a\phi_g(x)=\sum_{\substack{g'=1\\g'\neq g}}^{2}\Sigma_{g'\to g}^s\phi_{g'}(x)+\frac{\chi_g}{k^{eff}}\sum_{g'=1}^{2}\nu\Sigma_{g'}^f\phi_{g'}(x),\quad g=1,2 \tag{A.0.2}$$

We shall seek an analytical solution from (A.0.2) for the following cases: Homogeneous slab reactor consisting only of fuel and a heterogeneous slab reactor consisting of fuel and reflector as described.

The multigroup neutron diffusion equation defined as a eigenvalue problem can be written in the matrix form as follows:

$$\begin{bmatrix}-D_1\nabla^2+\Sigma_1^a-\frac{1}{k^{eff}}\nu\Sigma_1^f & -\frac{1}{k^{eff}}\nu\Sigma_2^f\\ -\Sigma_{1\to 2}^s & -D_2\nabla^2+\Sigma_2^a\end{bmatrix}\cdot\begin{pmatrix}\phi_1(x)\\ \phi_2(x)\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix} \tag{A.0.3}$$

where $\phi_1(x)$ and $\phi_2(x)$ represent respectively the fast and thermal neutron flux. If we study the equation above it is possible to see that the equations both for the fast flux and for the thermal flux satisfy Helmholtz's equation, given by:

$$\nabla^2\phi_g(x)+B^2\phi_g(x)=0,\quad g=1,2 \tag{A.0.4}$$

where B^2 denotes any one of the two roots of the equation characteristic of the second-order equation in B^2. Substituting (A.0.4) into (A.0.3), it results that:

$$\begin{bmatrix} -D_1B^2 + \Sigma_1^a - \frac{1}{k^{eff}}\nu\Sigma_1^f & -\frac{1}{k^{eff}}\nu\Sigma_2^f \\ -\Sigma_{1\to2}^s & -D_2B^2 + \Sigma_2^a \end{bmatrix} \cdot \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{A.0.5}$$

To solve (A.0.5), it is enough that the determinant of the matrix is null, i.e., we have a second-degree equation for B^2 whose solution is given by: $B_1 = \pm\sqrt{\frac{1}{2}\left(-1+\sqrt{1-\frac{4c}{b^2}}\right)}$ and $B_2 = \pm\sqrt{\frac{1}{2}\left(-1-\sqrt{1-\frac{4c}{b^2}}\right)}$ with $b = \frac{\Sigma_1^a}{D_1} + \frac{\Sigma_2^a}{D_2} - \frac{1}{k^{eff}}\frac{\nu\Sigma_1^f}{D_1}$ and $c = \frac{\Sigma_1^a}{D_1}\frac{\Sigma_2^a}{D_1} - \frac{1}{k^{eff}}\left(\frac{\nu\Sigma_1^f\Sigma_2^a + \nu\Sigma_2^f\Sigma_{1\to2}^s}{D_1D_2}\right)$.

Note that until now we only seek to find the roots of the characteristic equation. As we have not defined the type of solution and we know that the diffusion equation is a second-order differential equation, the types of solution that satisfy this equation are many, although we will consider that this solution has an exponential behaviour for both fast and thermal flux. Thus, the general solution for the thermal flux is given by:

$$\phi_2(x) = C_1e^{-B_1x} + C_2e^{B_1x} + C_3e^{-B_3x} + C_4e^{B_2x} \tag{A.0.6}$$

Due to the fact that the system provided by Eq (A.0.3) is coupled, the solution for the fast flux can be obtained by substituting (A.0.6) into (A.0.2) to $g = 2$, of which we have

$$\phi_1(x) = \left(\frac{\Sigma_2^a - D_2B_1^2}{\Sigma_{1\to2}^s}\right)\left(C_1e^{-B_1x} + C_2e^{B_1x}\right) + \left(\frac{\Sigma_2^a - D_2B_2^2}{\Sigma_{1\to2}^s}\right)\left(C_3e^{-B_2x} + C_4e^{B_2x}\right). \tag{A.0.7}$$

We will now seek some conditions that have to be applied to (A.0.6) and (A.0.7) in order to determine the coefficients of the equation. Note that the coefficients are the same for the fast and thermal flow, so one only needs to impose contour and symmetry conditions for only one of the equations. In the case of a slab reactor [200, 81] (plain slab in direction x) of dimension a as shown in Figure A.0.1, we can impose the following conditions:

i) Null flux in the boundary the left, such that, $\phi_g(-a/2) = 0$;

ii) Null flux in the boundary the right, such that, $\phi_g(a/2) = 0$;

iii) Maximum flux in the origin, i.e., $\phi_g(0) = \phi_{go}$;

iv) Null net current in the origin, $J_g(0) = 0$.

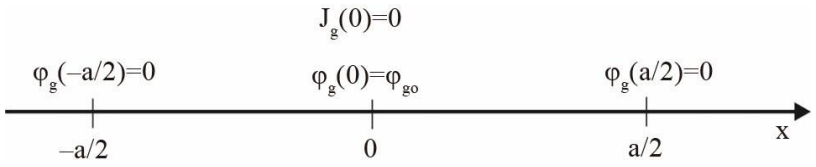


Figure A.0.1: Slab reactor with an a dimension.

The one-dimensional multigroup neutron diffusion equation for the fuel region was presented in the previous section, whose solution was given by (A.0.6) and (A.0.7) . For the reflector region this equation undergoes some modifications both in the nuclear parameters that will be different in relation to the nuclear parameters for the fuel, as well as for the form of the diffusion equation, given

that for this region there is no neutron fission. Therefore, this equation can be written as follows:

$$-D_1 \frac{d^2}{dx^2}\Psi_1(x) + \Sigma_1^a \Psi_1(x) = 0 \tag{A.0.8}$$

and

$$-D_2 \frac{d^2}{dx^2}\Psi_2(x) + \Sigma_2^a \Psi_2(x) = \Sigma_{1\to 2}^s \Psi_1(x) \tag{A.0.9}$$

The solutions of the diffusion equation for the fast and thermal group for the reflector region are, respectively,

$$\Psi_1(x) = R_1 e^{L_1 x} + R_2 e^{L_1 x} \tag{A.0.10}$$

and

$$\Psi_2(x) = \left(\frac{D_1 \Sigma_{1\to 2}^a}{\Sigma_2^a D_1 - \Sigma_1^a D_2}\right)\left(R_1 e^{L_1 x} + R_2 e^{L_1 x}\right) + R_3 e^{L_2 x} + R_4 e^{L_2 x} \tag{A.0.11}$$

where $L_1 = \sqrt{\frac{\Sigma_1^a}{D_1}}$ and $L_2 = \sqrt{\frac{\Sigma_2^a}{D_2}}$ are the respective diffusion lengths for the fast and thermal groups.

(A.0.11) and (A.0.11) represent the solutions for the neutron flux in the reflector region, while (A.0.6) and (A.0.7) represent the solutions for the neutron flux in the fuel region. These regions are shown in Figure A.0.2.

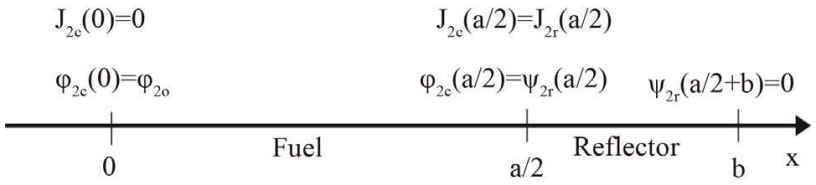


Figure A.0.2: Slab reactor with two regions of dimension $a + 2b$.

As we already have the solutions for the flux in the two regions of the reactor, we will now seek to find which condition we should impose to the flux in order to determine the coefficients of (A.0.6), (A.0.7), (A.0.11) and (A.0.11) . As the coefficients of (A.0.11) are present in (A.0.11) and the coefficients of (A.0.6) and (A.0.7) are identical, we will apply the conditions only for fast and thermal flux in the two regions, fuel and reflector, such that

i) Continuity of flux in the interface between regions to the left and right, such that, $\phi_{gc}(\pm a/2) = \Psi_{gr}(\pm a/2)$;

ii) Continuity of current in the interface between regions to the left and right, such that, $J_{gc}(\pm a/2) = J_{gr}(\pm a/2)$;

iii) Null flux in the boundary the left and right, such that, $\Psi_{gr}(-a/2-b) = 0$ and $\Psi_{gr}(a/2+b) = 0$;

iv) Maximum flux in the origin, i.e., $\phi_{gc}(0) = \phi_{go}$;

v) Null net current in the origin, $J_{gc}(0) = 0$.

Subscripts c and r represent respectively, the fuel and reflector regions.

Bibliography

[1] PARCS. IAEA 3D PWR problem. `https://engineering.purdue.edu/PARCS/Code/TestSuite/CalculationMode/StandAloneMode/Eigenvalue/IAEA3DPWR`.

[2] SALOME Platform Website. `http://www.salome-platform.org/`.

[3] Benchmark Problem Book. ANL-7416, Suppl.2. Argonne National Laboratory, 1977.

[4] H. S. Abdel-Khalik, P. Turinsky, M. Jessee, J. Elkins, T. Stover, and M. Iqbal. Uncertainty quantification, sensitivity analysis, and data assimilation for nuclear systems simulation. *Nuclear Data Sheets*, 109(12):2785–2790, 2008.

[5] A. Abdulle and Y. Bai. Adaptive reduced basis finite element heterogeneous multiscale method. *Computer Methods in Applied Mechanics and Engineering*, 257:203–220, 2013.

[6] B. Abou El Majd and L. Cordier. New regularization method for calibrated POD Reduced-Order Models. *Mathematical Modelling and Analysis*, 21(1):47–62, 2016.

[7] B. Adcock, A. C Hansen, and C. Poon. Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *SIAM Journal on Mathematical Analysis*, 45(5):3132–3167, 2013.

[8] J. V. Aguado, D. Borzacchiello, E. Lopez, E. Abisset-Chavanne, D. Gonzalez, E. Cueto, and F. Chinesta. New Trends in Computational Mechanics: Model Order Reduction, Manifold Learning and Data-Driven. *From Microstructure Investigations to Multiscale Modeling: Bridging the Gap*, pages 239–266, 2017.

[9] A. Alemberti, V. Smirnov, C. F. Smith, and M. Takahashi. Overview of lead-cooled fast reactor activities. *Progress in Nuclear Energy*, 77:300–307, 2014.

[10] M. Ali, K. Steih, and K. Urban. Reduced basis methods with adaptive snapshot computations. *Advances in Computational Mathematics*, 43(2):257–294, 2017.

[11] N. T. Andrey. Regularization of Incorrectly Posed Problems. *Soviet Mathematics Doklady*, 4(6):1624–1627, 1963.

[12] A. C. Antoulas and I. V. Gosea. Data-driven model reduction for weakly nonlinear systems: A summary. *IFAC-PapersOnLine*, 48(1):3–4, 2015.

[13] A. C. Antoulas, D. C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. *Contemporary mathematics*, 280:193–220, 2001.

[14] J-P. Argaud, B. Bouriquet, F. de Caso, H. Gong, Y. Maday, and O. Mula. Sensor placement in nuclear reactors based on the Generalized Empirical Interpolation Method. *Journal of Computational Physics*, 363(2018):354–370, 2018.

[15] J.-P. Argaud, B. Bouriquet, P. Erhard, S. Massart, and S. Ricci. Data assimilation in nuclear power plant core. *Progress in Industrial Mathematics at ECMI 2008*, pages 401–406, 2010.

[16] J.-P. Argaud, B. Bouriquet, H. Gong, Y. Maday, and O. Mula. Stabilization of (G)EIM in Presence of Measurement Noise: Application to Nuclear Reactor Physics. In Marco L. Bittencourt, Ney A. Dumont, and Jan S. Hesthaven, editors, *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2016: Selected Papers from the ICOSAHOM conference, June 27-July 1, 2016, Rio de Janeiro, Brazil*, pages 133–145. Springer International Publishing, 2017.

[17] G. Artana, A. Cammilleri, J. Carlier, and E. Memin. Strong and weak constraint variational assimilations for reduced order fluid flow modeling. *Journal of Computational Physics*, 231(8):3264–3288, 2012.

[18] M. Asch, M. Bocquet, and M. Nodet. *Data assimilation: methods, algorithms, and applications*. SIAM, 2016.

[19] R. C. Aster, C. H. Thurber, and B. Borchers. *Parameter Estimation and Inverse Problems*. Number v. 1 in International geophysics series. Elsevier Academic Press, 2005.

[20] R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, 2008.

[21] 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 Mathematique*, 339(9):667–672, 2004.

[22] J. Bartzis, J. Ehrhardt, S. French, J. Lochard, K. N. Morrey, M .and Papamichail, K. Sinkko, and A. Sohier. RODOS: Decision support for nuclear emergencies. In *Decision making: Recent developments and worldwide applications*, pages 381–395. Springer, 2000.

[23] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[24] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model Reduction and Approximation: Theory and Algorithms*, volume 15. SIAM, 2017.

[25] A. F. Bennett. *Inverse methods in physical oceanography*. Cambridge university press, 1992.

[26] P. Berger, K. Gröchenig, and G. Matz. Sampling and Reconstruction in Distinct Subspaces Using Oblique Projections. *arXiv preprint arXiv:1706.06444*, 2017.

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

[28] M. Bierlaire, P. L. Toint, and D. Tuyttens. On iterative algorithms for linear least squares problems with bound constraints. *Linear Algebra and its Applications*, 143:111–143, 1991.

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

[30] P. Binev, A. Cohen, W. Dahmen, Ro. 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] P. Binev, A. Cohen, O. Mula, and J. Nichols. Greedy algorithms for optimal measurements selection in state estimation using reduced models. *arXiv preprint arXiv:1602.06342*, 2017.

[32] Gary Bishop, Greg Welch, et al. An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8(27599-23175):41, 2001.

[33] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[34] B. Bouriquet and J.-P. Argaud. Best linear unbiased estimation of the nuclear masses. *Annals of Nuclear Energy*, 38(9):1863–1866, 2011.

[35] B. Bouriquet, J.-P. Argaud, and R. Cugnart. Optimal design of measurement network for neutronic activity field reconstruction by data assimilation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 664(1):117–126, 2012.

[36] B. Bouriquet, J.-P. Argaud, P. Erhard, S. Massart, A. Ponçot, S. Ricci, and O. Thual. Differential influence of instruments in nuclear core activity evaluation by data assimilation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 626:97–104, 2011.

[37] B. Bouriquet, J.-P. Argaud, P. Erhard, S. Massart, A. Ponçot, S. Ricci, and O. Thual. Robustness of nuclear core activity reconstruction by data assimilation. *Journal of Power and Energy Systems*, 6(2):289–301, 2012.

[38] B. Bouriquet, J.-P. Argaud, P. Erhard, and A. Ponçot. Unified nuclear core activity map reconstruction using heterogeneous instruments with data assimilation. *arXiv preprint arXiv:1108.5971*, 2011.

[39] B. Bouriquet, J.-P. Argaud, P. Erhard, and A. Ponçot. Nuclear core activity reconstruction using heterogeneous instruments with data assimilation. *EPJ Nuclear Sciences & Technologies*, 1:18, 2015.

[40] F. Bouttier and P. Courtier. Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*, page 59, 2002.

[41] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[42] W. A. Boyd and R. W. Miller. The BEACON on-line core monitoring system: functional upgrades and applications. In *Proc. Specialists' Meeting "In-core instrumentation and core assessment". Mito-shi, Japan*, pages 14–17, 1996.

[43] P. Brasseur and J. Verron. The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics*, 56(5):650–661, 2006.

[44] P. Brousseau, L. Berre, F. Bouttier, and G. Desroziers. Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Quarterly Journal of the Royal Meteorological Society*, 137(655):409–422, 2011.

[45] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[46] A. G. Buchan, A. A. Calloo, M. G. Goffin, S. Dargaville, F. Fang, C. C. Pain, and I. M. Navon. A POD reduced order model for resolving angular direction in neutron/photon transport problems. *Journal of Computational Physics*, 296:138–157, 2015.

[47] A. G. Buchan, C. C. Pain, F. Fang, and I. M. Navon. A POD reduced-order model for eigenvalue problems with application to reactor physics. *International Journal for Numerical Methods in Engineering*, 95(12):1011–1032, 2013.

[48] 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 method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(3):595–603, 2012.

[49] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.

[50] A. Calloo, D. Couyras, F. Févotte, and M. Guillo. Cocagne: EDF new neutronic core code for ANDROMEDE calculation chain. In *Proceedings of International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering (M&C), Jeju, Korea*, 2017.

[51] A. Cammi, L. Luzzi, and G. Rozza. A Reduced Basis Approach for Modeling the Movement of Nuclear Reactor Control Rods. *Journal of Nuclear Engineering and Radiation Science APRIL*, 2:021019–1, 2016.

[52] Y. Cao, J. Zhu, Z. Luo, and I. M. Navon. Reduced-order modeling of the upper tropical pacific ocean model using proper orthogonal decomposition. *Computers & Mathematics with Applications*, 52(8-9):1373–1386, 2006.

[53] Y. Cao, J. Zhu, I. M. Navon, and Z. Luo. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583, 2007.

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

[55] S. Chandrasekhar. Radiative transfer. *Oxford, Clarendon Press, 1950.*, 1950.

[56] F. Chinesta, A. Huerta, G. Rozza, and K. Willcox. Model Order Reduction: a survey. Wiley, 2016.

[57] N. Z. Cho. Fundamentals and recent developments of reactor physics methods. *Nuclear Engineering and Technology*, 37(1):25–78, 2005.

[58] N. Z. Cho, G. S. Lee, and C. J. Park. A Fusion Technique of 2-D/l-D Methods for Three-Dimensional Whole-Core Transport Calculations. In *Proceedings of the Korean Nuclear Society Conference*. Korean Nuclear Society.

[59] T. Clerc, A. Hébert, H. Leroyer, J.-P. Argaud, B. Bouriquet, and A. Ponçot. An advanced computational scheme for the optimization of 2D radial reflector calculations in pressurized water reactors. *Nuclear Engineering and Design*, 273:560–575, 2014.

[60] T. Clerc, A. Hebert, H. Leroyer, J.-P. Argaud, A. Poncot, and B. Bouriquet. Presentation of the MERC work-flow for the computation of a 2D radial reflector in a PWR. In *Proceedings of the 2013 International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering-M and C 2013*, 2013.

[61] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Mathematics of Computation*, 70(233):27–75, 2001.

[62] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Foundations of computational mathematics*, 13(5):819–834, 2013.

[63] A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 24:1–159, 2015.

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

[65] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Analysis and Applications*, 9(01):11–47, 2011.

[66] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *Détail*, 3:181–203, 2017.

[67] T. Coleman, M. A. Branch, and A. Grace. Optimization toolbox. *For Use with MATLAB. User's Guide for MATLAB 5, Version 2, Relaese II*, 1999.

[68] Wikimedia Commons. File:EPR in-core instrumentation.jpg — Wikimedia Commons, the free media repository, 2017.

[69] Wikipedia contributors. Uncertainty quantification — wikipedia, the free encyclopedia, 2018. [Online; accessed 2-March-2018].

[70] L. Cordier, B. Abou El Majd, and J. Favier. Calibration of POD reduced-order models using Tikhonov regularization. *International Journal for Numerical Methods in Fluids*, 63(2):269–296, 2010.

[71] P. Courtier and O. Talagrand. Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus A*, 42(5):531–549, 1990.

[72] N. Daget, A. T. Weaver, and M. A. Balmaseda. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, 135(641):1071–1094, 2009.

[73] R. Daley. *Atmospheric data analysis*. Number 2. Cambridge university press, 1993.

[74] B. Davison and J. B. Sykes. Neutron transport theory. 1957.

[75] G. Descombes, T. Auligné, Fr. Vandenberghe, D. M. Barker, and J. Barre. Generalized background error covariance matrix model (GEN_BE v2. 0). *Geoscientific Model Development*, 8(3):669–696, 2015.

[76] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in Banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.

[77] R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in Banach spaces. *arXiv preprint arXiv:1602.06342*, 2016.

[78] R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation and sampling in Banach spaces. *Calcolo*, 54(3):963–1007, 2017.

[79] M. Drohmann and K. Carlberg. The ROMES method for statistical modeling of reduced-order-model error. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):116–145, 2015.

[80] M. Drohmann, B. Haasdonk, and M. Ohlberger. Adaptive reduced basis methods for nonlinear convection-diffusion equations. *Finite Volumes for Complex Applications VI-Problems & Perspectives*, 1:369–377, 2011.

[81] J. J. Duderstadt and L. J. Hamilton. *Nuclear reactor analysis*, volume 1. Wiley New York, 1976.

[82] K. Duraisamy. Data-enabled, Physics-constrained Predictive Modeling of Complex Systems. 2017.

[83] S. Durbiano. *Vecteurs caractéristiques de modèles océaniques pour la réduction d'ordre en assimilation de données*. PhD thesis, Université Joseph Fourier (Grenoble), 2001.

[84] J. L. Eftang, A. T. Patera, and E. M. Rønquist. An "hp" certified reduced basis method for parametrized elliptic partial differential equations. *SIAM Journal on Scientific Computing*, 32(6):3170–3200, 2010.

[85] J. L. Eftang and B. Stamm. Parameter multi-domain 'hp' empirical interpolation. *International Journal for Numerical Methods in Engineering*, 90(4):412–428, 2012.

[86] M. Eissa, M. Naguib, and A. Badawi. PWR control rods position monitoring. *Annals of Nuclear Energy*, 81:106–116, 2015.

[87] H. Eleveld, Y. S. Kok, and C. J.W. Twenhofel. Data assimilation, sensitivity and uncertainty analyses in the Dutch nuclear emergency management system: a pilot study. *International Journal of Emergency Management*, 4(3):551–563, 2007.

[88] H. Finnemann, F. Bennewitz, and M. R. Wagner. Interface current techniques for multidimensional reactor calculations. *Atomkernenergie*, 30(2):123–128, 1977.

[89] M. Fisher. Background error covariance modelling. In *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pages 45–63, 2003.

[90] N. S. Garis, I. Pázsit, U. Sandberg, and T. Andersson. Determination of PWR control rod position by core physics and neural network methods. *Nuclear technology*, 123(3):278–295, 1998.

[91] G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

[92] I .Y. Gejadze, F.-X. L. Dimet, and V. Shutyaev. On analysis error covariances in variational data assimilation. *SIAM Journal on Scientific Computing*, 30(4):1847–1874, 2008.

[93] A. Gloria, T. Goudon, and S. Krell. Numerical homogenization of a nonlinearly coupled elliptic–parabolic system, reduced basis method, and application to nuclear waste storage. *Mathematical models and methods in applied sciences*, 23(13):2523–2560, 2013.

[94] N. P. Goldstein and W. H. Todt. A survey of Self-Powered Detectors-present and future. *IEEE Transactions on Nuclear Science*, 26(1):916–923, 1979.

[95] H. Gong, J.-P. Argaud, B. Bouriquet, Y. Maday, and O. Mula. Monitoring flux and power in nuclear reactors with data assimilation and reduced models. In *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2017), Jeju, Korea*, 2017.

[96] N. I. M. Gould and P. L. Toint. A quadratic programming bibliography. *Numerical Analysis Group Internal Report*, 1:32, 2000.

[97] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(3):575–605, 2007.

[98] M. A. Grepl and A. T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 39(1):157–181, 2005.

[99] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(2):277–302, 2008.

[100] C. E. Heaney, A. G. Buchan, C. C. Pain, and S. Jewer. A reduced order model for criticality problems in reactor physics varying control rod settings. In *Proceedings of the 24th UK Conference of the Association for Computational Mechanics in Engineering, Cardiff University, Cardiff*, 2016.

[101] A. Hebert. *Applied Reactor Physics*. Presses inter Polytechnique, 2009.

[102] F. Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[103] A. W. Heemink, M. Verlaan, and A. J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129(7):1718–1728, 2001.

[104] J. Heo, S.-W. Lee, and K. D. Kim. Implementation of data assimilation methodology for physical model uncertainty evaluation using post-CHF experimental data. *Nuclear Engineering and Technology*, 46(5):619–632, 2014.

[105] J. Heo, P. J Turinsky, and J. M. Doster. Design of nuclear power plant experiments for data assimilation to reduce modeling uncertainties. In *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering*, volume 2, page 15, 2011.

[106] J. Heo, P. J. Turinsky, and J. M. Doster. Optimization of Thermal-Hydraulic Reactor System for SMRs via Data Assimilation and Uncertainty Quantification. *Nuclear Science and Engineering*, 173(3):293–311, 2013.

[107] J. S. Hesthaven, G. Rozza, B. Stamm, et al. *Certified reduced basis methods for parametrized partial differential equations*. Springer, 2016.

[108] J. S. Hesthaven, B. Stamm, and S. Zhang. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(1):259–283, 2014.

[109] J. M. Hite, H. S. Abdel-Khalik, R. C. Smith, M. Wentworth, E. Prudencio, and B. Williams. Uncertainty Quantification and Data Assimilation (UQ/DA) Study on a VERA Core Simulator Component for CRUD Analysis CASL-I-2013-0184-000. *Milestone Report for L*, 2, 2013.

[110] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1990.

[111] P. L. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell. A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124(6):1225–1242, 1996.

[112] W. Huyer and A. Neumaier. Global optimization by multilevel coordinate search. *Journal of Global Optimization*, 14(4):331–355, 1999.

[113] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc. Unified Notation for Data Assimilation: Operational, Sequential and Variational (gtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):181–189, 1997.

[114] R. L. Iman and J. C. Helton. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk analysis*, 8(1):71–90, 1988.

[115] M. A. Jessee, P. J. Turinsky, and H. S. Abdel-Khalik. Many-group cross-section adjustment techniques for boiling water reactor adaptive simulation. *Nuclear Science and Engineering*, 169(1):40–55, 2011.

[116] H. G. Joo, J. Y. Cho, K. S. Kim, C. C. Lee, and S. Q. Zee. Methods and performance of a three-dimensional whole-core transport code DeCART. In *Proc. PHYSOR2004, CDROM*. Citeseer, 2004.

[117] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[118] G. Katata, H. Terada, H. Nagai, and M. Chino. Numerical reconstruction of high dose rate zones due to the Fukushima Dai-ichi Nuclear Power Plant accident. *Journal of environmental radioactivity*, 111:2–12, 2012.

[119] B. A. Khuwaileh and P. J. Turinsky. L2: VMA. P12. 01, Data Assimilation and Uncertainty Quantification Using VERA-CS for a Core Wide LWR Problem with Depletion.

[120] D. Knott and A. Yamamoto. Lattice physics computations. In *Handbook of nuclear engineering*, pages 913–1239. Springer, 2010.

[121] B. M. Kochunas. *A Hybrid Parallel Algorithm for the 3-D Method of Characteristics Solution of the Boltzmann Transport Equation on High Performance Compute Clusters*. PhD thesis, University of Michigan, 2013.

[122] K. Koebke and M. R. Wagner. The determination of the pin power distribution in a reactor core on the basis of nodal coarse mesh calculations. *Atomkernenergie*, 30(2):136–142, 1977.

[123] W. Kofink. Studies of the spherical harmonics method in neutron transport theory. *Il Nuovo Cimento (1955-1965)*, 9(2):497–541, 1958.

[124] A. Kolmogoroff. Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Annals of Mathematics*, 37:107–110, 1936.

[125] S. Konishi. *Introduction to multivariate analysis: Linear and nonlinear modeling*. CRC Press, 2014.

[126] S. Kosaka and E. Saji. Transport theory calculation for a heterogeneous multi-assembly problem by characteristics method with direct neutron path linking technique. *Journal of nuclear science and technology*, 37(12):1015–1023, 2000.

[127] D. B. Kothe. CASL: The Consortium for Advanced Simulation of Light Water Reactors. *Bulletin of the American Physical Society*, 55, 2010.

[128] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische mathematik*, 90(1):117–148, 2001.

[129] P. Ladevèze and L. Chamoin. On the verification of model reduction methods based on the proper generalized decomposition. *Computer Methods in Applied Mechanics and Engineering*, 200(23):2032–2047, 2011.

[130] W. Lahoz, B. Khattatov, and R. Menard. *Data assimilation: making sense of observations*. Springer Science & Business Media, 2010.

[131] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. Generalized reduced basis methods and n-width estimates for the approximation of the solution manifold of parametric PDEs. In *Analysis and numerics of partial differential equations*, pages 307–329. Springer, 2013.

[132] K. Law, A. Stuart, and K. Zygalakis. *Data assimilation: a mathematical introduction*, volume 62. Springer, 2015.

[133] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):97–110, 1986.

[134] G. S. Lee, N. Z. Cho, and S. G. Hong. Acceleration and parallelization of the method of characteristics for lattice and whole-core heterogeneous calculations. *Proc. Int. Mtg. Advances in Reactor Physics and Mathematics and Computations into the New Millennium (PHYSOR 2000)*, pages 7–11, 2000.

[135] G. S. Lee, N. Z. Cho, and S. G. Hong. Whole-core heterogeneous transport calculations and their comparison with diffusion results. *Trans. Am. Nucl. Soc*, 83:292, 2000.

[136] B. Legras and R. Vautard. A guide to Liapunov vectors. In *Proceedings 1995 ECMWF Seminar on Predictability*, volume 1, pages 143–156, 1996.

[137] J. M. Lewis and J. C. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A: Dynamic Meteorology and Oceanography*, 37(4):309–322, 1985.

[138] C.-J. Lin and J. J. Moré. Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9(4):1100–1127, 1999.

[139] K.-N. Liou. A numerical experiment on Chandrasekhar's discrete-ordinate method for radiative transfer: Applications to cloudy and hazy atmospheres. *Journal of the Atmospheric Sciences*, 30(7):1303–1326, 1973.

[140] S. Lorenzi, A. Cammi, and L. Luzzi. Spatial neutronics modelling to evaluate the temperature reactivity feedbacks in a lead-cooled fast reactor. *Proceedings of ICAPP 2015, Nice, France, May 03-06*, 2015.

[141] S. Lorenzi, A. Cammi, L. Luzzi, and G. Rozza. A reduced order model for investigating the dynamics of the Gen-IV LFR coolant pool. *Applied Mathematical Modelling*, 46:263–284, 2017.

[142] P. Lötstedt. Solving the minimal least squares problem subject to bounds on the variables. *BIT Numerical Mathematics*, 24(2):205–224, 1984.

[143] Y. Maday. Reduced basis method for the rapid and reliable solution of partial differential equations. In *in International Congress of Mathematicians. Vol. III, 1255–1270, Eur. Math. Soc., Zürich*. Citeseer, 2006.

[144] Y. Maday and O. Mula. A generalized empirical interpolation method: application of reduced basis techniques to data assimilation. In *Analysis and numerics of partial differential equations*, pages 221–235. Springer, 2013.

[145] Y. Maday, O. Mula, A. T. Patera, and M. Yano. The Generalized Empirical Interpolation Method: stability theory on Hilbert spaces with an application to the Stokes equation. *Computer Methods in Applied Mechanics and Engineering*, 287(0):310–334, 2015.

[146] Y. Maday, O. Mula, and G. Turinici. Convergence analysis of the Generalized Empirical Interpolation Method. *SIAM Journal on Numerical Analysis*, 54(3):1713–1731, 2016.

[147] Y. Maday, N. C. Nguyen, A. T. Patera, and S. H. Pau. A general multipurpose interpolation procedure: the magic points. *Communications on Pure & Applied Analysis*, 8(1):383–404, 2009.

[148] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A Parameterized-Background Data-Weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.

[149] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. PBDW State Estimation: Noisy Observations; Configuration-Adaptive Background Spaces; Physical Interpretations. *ESAIM: Proceedings and Surveys*, 50:144–168, 2015.

[150] Y. Maday, A. T. Patera, and G. Turinici. A priori convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations. *Journal of Scientific Computing*, 17(1-4):437–446, 2002.

[151] Y. Maday and B. Stamm. Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. *SIAM Journal on Scientific Computing*, 35(6):A2417–A2441, 2013.

[152] A. Manzoni, A. Quarteroni, and G. Rozza. Computational reduction for parametrized PDEs: strategies and applications. *Milan Journal of Mathematics*, pages 1–27, 2012.

[153] S. Marguet. *La physique des réacteurs nucléaires*. Lavoisier, 2011.

[154] W. R. Martin. Challenges and prospects for whole-core Monte Carlo analysis. *Nuclear Engineering and Technology*, 44(2):151–160, 2012.

[155] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

[156] J. L. Mead and R. A. Renaut. Least squares problems with inequality constraints as quadratic constraints. *Linear Algebra and its Applications*, 432(8):1936–1949, 2010.

[157] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[158] C. Meyer. *Matrix Analysis and Applied Linear Algebra Book and Solutions Manual*. Society for Industrial and Applied Mathematics, 2000.

[159] Z. Michalewicz and C. Z. Janikow. GENOCOP: a genetic algorithm for numerical optimization problems with linear constraints. *Communications of the ACM*, 39(12es):175, 1996.

[160] J. Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37. Springer Science & Business Media, 2012.

[161] S. Morigi, L. Reichel, F. Sgallari, and F. Zama. An iterative method for linear discrete ill-posed problems with box constraints. *Journal of Computational and Applied Mathematics*, 198(2):505–520, 2007.

[162] N. C. Nguyen, G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced Basis Approximation and a Posteriori Error Estimation for Parametrized Parabolic PDEs: Application to Real-Time Bayesian Parameter Estimation. *Large-Scale Inverse Problems and Quantification of Uncertainty*, pages 151–177, 2010.

[163] J. Nocedal and S. J. Wright. *Sequential quadratic programming*. Springer, 2006.

[164] M. Ohlberger and S. Rave. Reduced Basis Methods: Success, Limitations and Future Challenges. In *Proceedings of the Conference Algoritmy*, pages 1–12, 2016.

[165] G. Palmiotti, M. Salvatores, H. Hiruta, M. Herman, P. Oblozinsky, and M. Pigni. Use of Covariance matrices in a consistent (Multiscale) data assimilation for improvement of basic nuclear parameters in nuclear reactor applications: from meters to femtometers. Technical report, Idaho National Laboratory (INL), 2010.

[166] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[167] M. Pecka, J. Svarnay, and J. Kment. Some aspects of the new core monitoring system at NPP Dukovany and first experience. In *Core monitoring for commercial reactors: improvements in systems and methods*. 2000.

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

[169] B. Peherstorfer and K. Willcox. Dynamic data-driven reduced-order models. *Computer Methods in Applied Mechanics and Engineering*, 291:21–41, 2015.

[170] D. T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review*, 129(5):1194–1207, 2001.

[171] J. E. Pierce and B. W. Rust. Constrained least squares interval estimation. *SIAM Journal on Scientific and Statistical Computing*, 6(3):670–683, 1985.

[172] R. Pinnau. Model reduction via proper orthogonal decomposition. In *Model Order Reduction: Theory, Research Aspects and Applications*, pages 95–109. Springer, 2008.

[173] A. Ponçot, J.-P. Argaud, B. Bouriquet, P. Erhard, S. Gratton, and O. Thual. Variational assimilation for xenon dynamical forecasts in neutronic using advanced background error covariance matrix modelling. *Annals of Nuclear Energy*, 60:39–50, 2013.

[174] E. Privas, P. Archier, D. Bernard, C. De Saint Jean, C. Destouche, P. Leconte, G. Noguère, Y. Peneliau, and R. Capote. Generation of 238U Covariance Matrices by Using the Integral Data Assimilation Technique of the CONRAD Code. In *EPJ Web of Conferences*, volume 106, page 04015. EDP Sciences, 2016.

[175] C. Prud'Homme, D. V. Rovas, K. Veroy, L. Machiels, Yvon Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.

[176] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*, volume 92 of *UNITEXT*. Springer International Publishing, 2015.

[177] A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry*, 1(1):3, 2011.

[178] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics*, volume 37. Springer Science & Business Media, 2010.

[179] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations. *arXiv preprint arXiv:1711.10561*, 2017.

[180] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. *arXiv preprint arXiv:1711.10566*, 2017.

[181] H. L. Rajic and A. M. Ougouag. ILLICO: A nodal neutron diffusion method for modern computer architectures. *Nuclear Science and Engineering*, 103(4):392–408, 1989.

[182] A. Ribes and C. Caremoli. Salome platform component model for numerical simulation. In *Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International*, volume 2, pages 553–564. IEEE, 2007.

[183] C. Robert, S. Durbiano, E. Blayo, J. Verron, J. Blum, and F.-X. Le Dimet. A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine Systems*, 57(1):70–82, 2005.

[184] C. Rojas-Palma, J. Ehrhardt, and A. Sohier. Data assimilation for off site nuclear emergency management. Technical report, Technical report, SCK-CEN, DAONEM final report, RODOS (RA5)-RE (04)-01, 2005.

[185] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):229, 2008.

[186] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

[187] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989.

[188] R. Sanchez, I. Zmijarevi, M. Coste-Delclaux, E. Masiello, S. Santandrea, E. Martinolli, L. Villate, N. Schwartz, and N. Guler. APOLLO2 year 2010. *Nuclear Engineering and Technology*, 42(5):474–499, 2010.

[189] A. Sartori. *Reduced order methods: applications to nuclear reactor core spatial dynamics*. PhD thesis, Italy, 2015.

[190] A. Sartori, D. Baroli, A. Cammi, D. Chiesa, L. Luzzi, R. Ponciroli, E. Previtali, M. E. Ricotti, G. Rozza, and M. Sisti. Comparison of a Modal Method and a Proper Orthogonal Decomposition approach for multi-group time-dependent reactor spatial kinetics. *Annals of Nuclear Energy*, 71:217–229, 2014.

[191] A. Sartori, D. Baroli, A. Cammi, L. Luzzi, and G. Rozza. A reduced order model for multigroup time-dependent parametrized reactor spatial kinetics. In *2014 22nd International Conference on Nuclear Engineering*, pages V005T17A048–V005T17A048. American Society of Mechanical Engineers, 2014.

[192] A. Sartori, A. Cammi, L. Luzzi, M. E. Ricotti, G. Rozza, et al. Reduced Order Methods: Applications to Nuclear Reactor Core Spatial Dynamics. In *ICAPP 2015*, pages 3318–3327. Omnipress, 2015.

[193] A. Sartori, A. Cammi, L. Luzzi, and G. Rozza. Reduced basis approaches in time-dependent noncoercive settings for modelling the movement of nuclear reactor control rods. 2015.

[194] Y. Sasaki. Numerical variational analysis with weak constraint and application to surface analysis of severe storm gust. *Monthly Weather Review*, 98(12):899–910, 1970.

[195] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[196] R. A. Shober. Nodal method for fast reactor analysis. Technical report, Argonne National Lab., IL (USA), 1979.

[197] K. S. Smith. *An analytic nodal method for solving the two-group, multidimensional, static and transient neutron diffusion equations*. PhD thesis, Massachusetts Institute of Technology, 1979.

[198] M. A. Smith, N. Tsoulfanidis, E. E. Lewis, G. Palmiotti, and T. A. Taiwo. Whole-core neutron transport calculations without fuel-coolant homogenization. In *International Topical Meeting on Advances in Reactor Physics and Mathematics and Computation into the Next Millennium*, pages 7–12, 2000.

[199] R. M. G. P Souza and J. M. L. Moreira. Neural network correlation for power peak factor estimation. *Annals of Nuclear Energy*, 33(7):594–608, 2006.

[200] W. M. Stacey. *Nuclear Reactor Physics*. John Wiley & Sons, 2007.

[201] S. Suram and K. M. Bryden. Integrating a reduced-order model server into the engineering design process. *Advances in Engineering Software*, 90:169–182, 2015.

[202] T. Taddei. *Model order reduction methods for data assimilation; state estimation and structural health monitoring*. PhD thesis, 2016.

[203] T. Taddei. An adaptive parametrized-background data-weak approach to variational data assimilation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(5):1827–1858, 2017.

[204] O. Talagrand. Assimilation of Observations, an Introduction (gtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):191–209, 1997.

[205] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1311–1328, 1987.

[206] A. Tangborn and S. Q. Zhang. Wavelet transform adapted to an approximate Kalman filter system. *Applied Numerical Mathematics*, 33(1-4):307–316, 2000.

[207] V. N. Temlyakov. Greedy approximation. *Acta Numerica*, 17:235–409, 2008.

[208] G. Theler, F.J. Bonetto, and A. Clausse. Solution of the 2D IAEA PWR Benchmark with the neutronic code Milonga. *Actas de la Reunión Anual de la Asociación Argentina de Tecnología Nuclear, XXXVIII*, 2011.

[209] J.-N. Thépart, D. Vasiljevic, P. Courtier, and J. Pailleux. Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. *Quarterly Journal of the Royal Meteorological Society*, 119(509):153–186, 1993.

[210] W. H. Todt. Characteristics of self-powered neutron detectors used in power reactors. In *Proc. of a Specialists' Meeting on In-core Inst. and Reactor Core Assessment, NEA Nuclear Science Committee*, 1996.

[211] M. Tojo, T. Naka, T. Iwamoto, J. Koyama, and K. Harada. Application of the New Core Monitoring System,'GNF-ARGOS', to the Start-up Tracking Calculation of ABWR. 2008.

[212] M. J. Verveld, T. Kier, N. Karcher, T. Franz, M. Abu-Zurayk, M. Ripepi, and S. Görtz. Reduced Order Models for Aerodynamic Applications, Loads and MDO. 2016.

[213] J. F. Vidal, O. Litaize, D. Bernard, A. Santamarina, C. Vaglio-Gaudard, and R. Tran. New modelling of LWR assemblies using the APOLLO2 code package. *Proc. Joint Int. Top. Mtg. on Mathematics & Computation and Supercomputing in Nuclear Applications (M&C+ SNA 2007)*, pages 15–19, 2007.

[214] M. R. Wagner. Three-dimensional nodal diffusion and transport theory methods for hexagonal-z geometry. *Nuclear Science and Engineering*, 103(4):377–391, 1989.

[215] W. E. Walker, P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B.A. van Asselt, P. Janssen, and M. P. Krayer von Krauss. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, 4(1):5–17, 2003.

[216] F. Wols, D. Lathouwers, and W. Uyttenhove. Transient analyses of accelerator driven systems using modal expansion techniques. *Delft University of Technology*, 2010.

[217] R. Yondo, E. Andrés, and E. Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 2017.

[218] D. Q. Zheng, J. K. C. Leung, B. Y. Lee, and H. Y. Lam. Data assimilation in the atmospheric dispersion model for nuclear accident assessments. *Atmospheric Environment*, 41(11):2438–2446, 2007.