A new decomposition of Gaussian random elements in Banach spaces with application to Bayesian inversion.

Jean-Charles Croix

N° d'ordre NNT : 2018LYSEM019

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
(MINES Saint-Etienne)

Ecole Doctorale N° 488
(Sciences, Ingénierie, Santé)

Spécialité de doctorat : Mathématiques Appliquées

Soutenue publiquement le 09/10/2018, par:
Jean-Charles Croix

A new decomposition of Gaussian random elements in Banach spaces with application to Bayesian inversion.

Devant le jury composé de :

Nouy, Anthony	Professeur	Ecole Centrale de Nantes	Président du Jury
Ayache, Antoine	Professeur	Université Lille 1	Rapporteur
Garnier, Josselin	Professeur	Ecole Polytechnique	Rapporteur
Lifshits, Mikhail	Professeur	Université de Saint-Petersbourg	Examinateur
Batton-Hubert, Mireille	Professeure	MINES Saint-Etienne	Directrice de thèse
Bay, Xavier	Maître-assistant	MINES Saint-Etienne	Co-encadrant
Touboul, Eric	Ingénieur	MINES Saint-Etienne	Co-encadrant

Spécialités doctorales	Responsables :	Spécialités doctorales	Responsables
SCIENCES ET GENIE DES MATERIAUX	K. Wolski Directeur de recherche	MATHEMATIQUES APPLIQUEES	O. Roustant, Maître-assistant
MECANIQUE ET INGENIERIE	S. Drapier, professeur	INFORMATIQUE	O. Boissier, Professeur
GENIE DES PROCEDES	F. Gruy, Maître de recherche	IMAGE, VISION, SIGNAL	JC. Pinoli, Professeur
SCIENCES DE LA TERRE	B. Guy, Directeur de recherche	GENIE INDUSTRIEL	X. Delorme, Maître assistant
SCIENCES ET GENIE DE L'ENVIRONNEMENT	D. Graillot, Directeur de recherche	MICROELECTRONIQUE	Ph. Lalevée, Professeur

EMSE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'Etat ou d'une HDR)

ABSI	Nabil	CR	Génie industriel	CMP
AUGUSTO	Vincent	CR	Image, Vision, Signal	CIS
AVRIL	Stéphane	PR2	Mécanique et ingénierie	CIS
BADEL	Pierre	MA(MDC)	Mécanique et ingénierie	CIS
BALBO	Flavien	PR2	Informatique	FAYOL
BASSEREAU	Jean-François	PR	Sciences et génie des matériaux	SMS
BATTON-HUBERT	Mireille	PR2	Sciences et génie de l'environnement	FAYOL
BEIGBEDER	Michel	MA(MDC)	Informatique	FAYOL
BLAYAC	Sylvain	MA(MDC)	Microélectronique	CMP
BOISSIER	Olivier	PR1	Informatique	FAYOL
BONNEFOY	Olivier	MA(MDC)	Génie des Procédés	SPIN
BORBELY	Andras	MR(DR2)	Sciences et génie des matériaux	SMS
BOUCHER	Xavier	PR2	Génie Industriel	FAYOL
BRODHAG	Christian	DR	Sciences et génie de l'environnement	FAYOL
BRUCHON	Julien	MA(MDC)	Mécanique et ingénierie	SMS
BURLAT	Patrick	PR1	Génie Industriel	FAYOL
CHRISTIEN	Frédéric	PR	Science et génie des matériaux	SMS
DAUZERE-PERES	Stéphane	PR1	Génie Industriel	CMP
DEBAYLE	Johan	CR	Image Vision Signal	CIS
DELAFOSSE	David	PR0	Sciences et génie des matériaux	SMS
DELORME	Xavier	MA(MDC)	Génie industriel	FAYOL
DESRAYAUD	Christophe	PR1	Mécanique et ingénierie	SMS
DJENIZIAN	Thierry	PR	Science et génie des matériaux	CMP
DOUCE	Sandrine	PR2	Sciences de gestion	FAYOL
DRAPIER	Sylvain	PR1	Mécanique et ingénierie	SMS
FAVERGEON	Loïc	CR	Génie des Procédés	SPIN
FEILLET	Dominique	PR1	Génie Industriel	CMP
FOREST	Valérie	MA(MDC)	Génie des Procédés	CIS
FOURNIER	Jacques	Ingénieur chercheur CEA	Microélectronique	CMP
FRACZKIEWICZ	Anna	DR	Sciences et génie des matériaux	SMS
GARCIA	Daniel	MR(DR2)	Génie des Procédés	SPIN
GAVET	Yann	MA(MDC)	Image Vision Signal	CIS
GERINGER	Jean	MA(MDC)	Sciences et génie des matériaux	CIS
GOEURIOT	Dominique	DR	Sciences et génie des matériaux	SMS
GONDRAN	Natacha	MA(MDC)	Sciences et génie de l'environnement	FAYOL
GRAILLOT	Didier	DR	Sciences et génie de l'environnement	SPIN
GROSSEAU	Philippe	DR	Génie des Procédés	SPIN
GRUY	Frédéric	PR1	Génie des Procédés	SPIN
GUY	Bernard	DR	Sciences de la Terre	SPIN
HAN	Woo-Suck	MR	Mécanique et ingénierie	SMS
HERRI	Jean Michel	PR1	Génie des Procédés	SPIN
KERMOUCHE	Guillaume	PR2	Mécanique et Ingénierie	SMS
KLOCKER	Helmut	DR	Sciences et génie des matériaux	SMS
LAFOREST	Valérie	MR(DR2)	Sciences et génie de l'environnement	FAYOL
LERICHE	Rodolphe	CR	Mécanique et ingénierie	FAYOL
MALLIARAS	Georges	PR1	Microélectronique	CMP
MOLIMARD	Jérôme	PR2	Mécanique et ingénierie	CIS
MOUTTE	Jacques	CR	Génie des Procédés	SPIN
NIKOLOVSKI	Jean-Pierre	Ingénieur de recherche	Mécanique et ingénierie	CMP
NORTIER	Patrice	PR1		SPIN
OWENS	Rosin	MA(MDC)	Microélectronique	CMP
PERES	Véronique	MR	Génie des Procédés	SPIN
PICARD	Gauthier	MA(MDC)	Informatique	FAYOL
PIJOLAT	Christophe	PR0	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR1	Génie des Procédés	SPIN
PINOLI	Jean Charles	PR0	Image Vision Signal	CIS
POURCHEZ	Jérémy	MR	Génie des Procédés	CIS
ROBISSON	Bruno	Ingénieur de recherche	Microélectronique	CMP
ROUSSY	Agnès	MA(MDC)	Génie industriel	CMP
ROUSTANT	Olivier	MA(MDC)	Mathématiques appliquées	FAYOL
STOLARZ	Jacques	CR	Sciences et génie des matériaux	SMS
TRIA	Assia	Ingénieur de recherche	Microélectronique	CMP
VALDIVIESO	François	PR2	Sciences et génie des matériaux	SMS
VIRICELLE	Jean Paul	DR	Génie des Procédés	SPIN
WOLSKI	Krzystof	DR	Sciences et génie des matériaux	SMS
XIE	Xiaolan	PR1	Génie industriel	CIS
YUGMA	Gallian	CR	Génie industriel	CMP

Mise à jour : 01/02/2016

Abstract

Inferring is a fundamental task in science and engineering: it gives the opportunity to compare theory to experimental data. Since measurements are finite by nature whereas parameters are often functional, it is necessary to offset this loss of information with external constraints, by regularization. The obtained solution then realizes a trade-off between proximity to data on one side and regularity on the other.

For more than fifteen years, these methods include a probabilistic thinking, taking uncertainty into account. Regularization now consists in choosing a prior probability measure on the parameter space and explicit a link between data and parameters to deduce an update of the initial belief. This posterior distribution informs on how likely is a set of parameters, even in infinite dimension.

In this thesis, approximation of such methods is studied. Indeed, infinite dimensional probability measures, while being attractive theoretically, often require a discretization to actually extract information (estimation, sampling). When they are Gaussian, the Karhunen-Loève decomposition gives a practical method to do so, and the main result of this thesis is to provide a generalization to Banach spaces, much more natural and less restrictive. The other contributions are related to its practical use with in a real-world example.

Résumé

L'inférence est une activité fondamentale en sciences et en ingénierie: elle permet de confronter et d'ajuster des modèles théoriques aux données issues de l'expérience. Ces mesures étant finies par nature et les paramètres des modèles souvent fonctionnels, il est nécessaire de compenser cette perte d'information par l'ajout de contraintes externes au problème, via les méthodes de régularisation. La solution ainsi associée satisfait alors un compromis entre d'une part sa proximité aux données, et d'autre part une forme de régularité.

Depuis une quinzaine d'années, ces méthodes intègrent un formalisme probabiliste, ce qui permet la prise en compte d'incertitudes. La régularisation consiste alors à choisir une mesure de probabilité sur les paramètres du modèle, expliciter le lien entre données et paramètres et déduire une mise-à-jour de la mesure initiale. Cette probabilité a posteriori, permet alors de déterminer un ensemble de paramètres compatibles avec les données tout en précisant leurs vraisemblances respectives, même en dimension infinie.

Dans le cadre de cette thèse, la question de l'approximation de tels problèmes est abordée. En effet, l'utilisation de lois infini-dimensionnelles, bien que théoriquement attrayante, nécessite souvent une discrétisation pour l'extraction d'information (calcul d'estimateurs, échantillonnage). Lorsque la mesure a priori est Gaussienne, la décomposition de Karhunen-Loève est une réponse à cette question. Le résultat principal de cette thèse est sa généralisation aux espaces de Banach, beaucoup plus naturels et moins restrictifs que les espaces de Hilbert. Les autres travaux développés concernent son utilisation dans des applications avec données réelles.

Remerciements

Pendant les trois années qui m'ont mené à rédiger ce manuscrit, j'ai conscience d'avoir bénéficié du soutien de nombreuses personnes, chacune contribuant à l'établissement des résultats qui sont retranscrits dans ce manuscrit à sa façon. Il m'est impossible d'être exhaustif, mais je souhaite en remercier ici un certain nombre d'entre elles.

Mes tout premiers remerciements s'adressent aux deux rapporteurs de cette thèse, Antoine Ayache et Josselin Garnier. Au travers de leurs rapports respectifs très détaillés, ils m'ont tous deux fait part d'un regard avisé sur les travaux présentés, ainsi que de précieuses remarques menant à de nouvelles pistes de recherche. Leur validation a été essentielle à mes yeux et me conforte dans ce choix de carrière.

En second lieu, je souhaite exprimer toute ma gratitude envers Anthony Nouy et Mikhail Lifshits qui ont acceptés d'être membres examinateurs du Jury. Conjointement avec celles des autres membres, je suis convaincu que leurs analyses et remarques me permettront d'envisager de nouvelles pistes fructueuses.

Les personnes que je vais maintenant remercier, forment l'équipe qui m'a encadré ces trois années : Mireille Batton-Hubert, Xavier Bay et Eric Touboul. J'ai de nombreuses raisons de leur être reconnaissant, mais je souhaiterai d'abord commencer par souligner les qualités humaines dont ils ont fait preuve tout au long de cette collaboration. Merci de m'avoir constamment fait confiance, d'avoir été aussi disponibles et bienveillants pendant ce doctorat. Que ce soit en réunion, autour d'un café, pendant le déjeuner ou en pleine course à pied, j'ai passé mon temps à profiter de leurs curiosités scientifiques et de leurs expertises. J'ai beaucoup appris grâce à eux et je souhaite vivement continuer à partager leur passion pour les sciences. Je leur dois donc beaucoup, autant sur le plan humain que scientifique. Merci à eux de m'avoir permis, dans de si bonnes conditions, de réaliser cette thèse de doctorat.

Pendant cette thèse, Nicolas Durrande m'a proposé de le rejoindre dans un projet scientifique Franco-Colombien. Ce fut le départ d'un travail commun avec

Mauricio A. Alvarez, m'ayant conduit à visiter l'université de Pereira en Colombie plusieurs fois. En plus d'avoir l'opportunité d'enseigner dans un contexte exceptionnel, cet échange a donné naissance à l'application présente dans ce document. Pour toutes ces raisons, je dois beaucoup à Nicolas et Mauricio. Bien entendu, je remercie aussi toutes les personnes avec qui je partage maintenant des souvenirs inoubliables, notamment Julian David Echeverry, Julian Gil Gonzalez et Alvaro Angel Orozco.

Je voudrais maintenant remercier mes autres collègues stéphanois, dont j'ai eu le plaisir de partager le quotidien au fil des ans: Rodolphe Leriche, Olivier Roustant, Isabelle Chantrel, Christine Exbrayat, Xavier Delorme, Frédéric Grimaud, Audrey Derrien, Paolo Gianessi, Didier Rullière, David Gaudrie, Marie-Liesse Cauwet, Adrien Spagnol, Andres Felipe Lopez-Lopera, Mohammed-Amine Abdous, Madhi El-Alaoui-El-Abdellaoui, Adrien Spagnol, Serena Finco, Oussama Masmoudi, Marie-Line Barneoud, Bruno Léger, Nicolas Garland, Esperan Padonou, Thierry Garaix, Niloufare Sadr, Jean-François Tchebanoff, Gabrielle Bruyas et Alain Mounier. Nombre d'entre eux sont devenus des amis proches.

Ma famille m'a toujours accompagné, partageant chaque moment de joie et me soutenant dans les autres. J'ai énormément de chance de les avoir tous et même s'ils le savent déjà, je compte le redire ici aussi : je suis fier et infiniment reconnaissant de les avoir dans ma vie.

A Saint-Etienne, il y a deux personnes à qui je dois beaucoup car elles ont tout simplement été une source quotidienne de bonheur. J'ai trouvé chez elles toutes les qualités que l'on recherche chez des proches, allant de la générosité et la gentillesse qui sont nécessaires à toute bonne relation jusqu'à la ténacité et le courage qui forgent les amis. Leurs noms sont Afafe Zehrouni et Romain Noël et je leur dis à tous les deux merci d'avoir rendu ces trois années aussi belles.

Je finirai par remercier mes proches, Clément Pallez, Etienne & Marie Spormeyeur, Nicolas & Pauline Valance, Thomas & Fanny Calmés, Jean-Baptiste Dufour, Mathilde Armant, Claire Fousse, Joël Pfaff, Hugo Bernard-Brunel, Fulvio Mounier, Anne-Kim Lê, Chloé Dasse, Amélie Beyssat, Sylvain Housseman, Alban Derrien et Pauline Julou pour tous les moments vécus et à venir.

Contents

Abstract iii

Résumé v

Remerciements vii

Table of contents x

List of figures xi

Introduction 1

I Approximation of Gaussian priors in Banach spaces 5

1 Introduction to Gaussian random elements 7

1.1 Definition 7

1.2 Integrability 10

1.3 Covariance and Hilbert subspace(s) 12

1.4 Series representation 20

1.5 Additional results in the Hilbert case 26

1.6 Gaussian random elements in $C(\mathcal{K}, \mathbb{R})$ 30

1.7 Conclusion 42

2 Karhunen-Loève decomposition in Banach spaces 43

2.1 Extending the Karhunen-Loève decomposition 43

2.2 The special case of $C(\mathcal{K}, \mathbb{R})$ 54

2.3 Examples in $\mathcal{C}(\mathcal{K}, \mathbb{R})$ 57

2.4 A study of the standard Wiener process 60

2.5 Conclusion 70

II Applications to Bayesian inverse problems 73

3 Introduction to inverse problems 75

3.1 Forward, inverse and ill-posed problems 75

3.2 Examples 77
3.3 Tikhonov-Philips regularization for operators 79
3.4 Other regularizations methods 84

4 Theory of Bayesian regularization in Banach spaces 87
4.1 Kriging: a motivation for Bayesian inversion 87
4.2 Bayesian regularization 92
4.3 Maximum a posteriori 100
4.4 Markov chain Monte-Carlo method 102
4.5 Approximation 108
4.6 Conclusion 111

5 Karhunen-Loève approximation of Bayesian inverse problems 113
5.1 Approximation using projections 113
5.2 Application to advection diffusion model 115
5.3 Numerical results 128
5.4 Conclusion 132

Conclusion and future works 137

Glossary of mathematical notations 139

Bibliography 141

List of Figures

2.1 8 first normalized basis functions (squared exponential kernel). . . . 61
2.2 Evolution of 8 first basis functions variances (squared exponential kernel). . . . 61
2.3 8 first normalized basis functions (Matérn $\nu = \frac{3}{2}$ kernel). . . . 62
2.4 Evolution of 8 first basis functions variances (Matérn $\nu = \frac{3}{2}$ kernel). 62
2.5 8 first normalized basis functions (Fbm $H = \frac{1}{4}$ kernel). . . . 63
2.6 Evolution of 8 first basis functions variances (Fbm $H = \frac{1}{4}$ kernel). . 63
2.7 8 first normalized basis functions (Fbm $H = \frac{3}{4}$ kernel). . . . 64
2.8 Evolution of 8 first basis functions variances (Fbm $H = \frac{3}{4}$ kernel). . 64
2.9 Numerical comparison of l-errors of admissible sequences for the standard Wiener process. . . . 71
2.10 Comparison of a-errors of admissible sequences for the standard Wiener process. . . . 71

3.1 Forward and inverse problems. . . . 76

4.1 Ω-minimizing and Tikhonov-Philips solutions in $C([0,1],\mathbb{R})$ with Matern $\frac{3}{2}$ covariance kernel. . . . 90
4.2 Mean function m with 95% confidence interval and 10 sample functions from $u|u(\mathbf{x}) = \mathbf{y}$ with Matérn $\frac{3}{2}$ covariance kernel. . . . 91
4.3 Bayesian inverse problem. . . . 92

5.1 Trace plots of $\Phi(u;y)$, λ, D, ξ_0, ξ_1 and ξ_2 (the first 6 000 iterations are burned). . . . 133
5.2 Autocorrelation of $\Phi(u;y)$, λ, D, ξ_0, ξ_1 and ξ_2, excluding burning sample. . . . 134
5.3 Marginal posterior distributions for $\lambda|y$ and $D|y$. . . . 134
5.4 MAP estimator from McMC sampling. Left: Estimated source term f_{MAP}, right: estimated solution $z(u_{MAP})$ with absolute error at data locations. . . . 135
5.5 Posterior point-wise variance of source (left) and solution (right). . . 136

Introduction

In experimental sciences and engineering, most of the mathematical models are designed to represent real-world phenomenon with various levels of fidelity. They describe how a well-known initial state evolves under the modelled assumptions in a final outcome. Their quality is then assessed by comparison with a physical experiment for instance. Since this represents a logical link from causes to consequences, this is called a direct approach and most models are designed to be well-posed in the sense of [Had02], meaning that for any set of initial conditions, one and only one solution exists and it is continuous in the causes.

However, the same model can serve different purposes and in particular, it can be reversed. Indeed, since the model gives a tangible link between causes and consequences, information can go both ways. The reverse approach then consists in inferring what would have been the causes leading to a particular outcome. Despite an obvious symmetry of these two concepts, they are very different in nature.

The well-posedness of the direct problem implies the opposite for the inverse approach, especially in infinite dimensional spaces, where an important loss of information (also called *smoothing*) happens. Stronger it is, hardier the reversion will be.

One common methodology to overcome this difficulty is regularization. In essence, it consists in looking for less general causes for the observed output, adding new and model-external constraints. If the parameter is a function, this could be done by imposing a certain regularity. The notion of solution is thus slightly modified, including a compromise between actually explaining observations and respecting these additional constraints. This provides a new problem, which in particular cases can be well-posed (in the previous sense). The probably most known regularization method is Tikhonov-Philips [Tik63, Phi62].

In parallel, it is now well established that errors of different natures may appear in the practical use of mathematical models. For instance, real-world measurements are always limited to a level of precision, which can sometimes be taken into account in a probabilistic framework. In the context of inverse problems, it results in randomness of the associated causes, even with regularization. The study of this

effect is called *quantification of uncertainty* and it is the objective of Bayesian methods to intrinsically provide it as well as regularization [Stu10, DS15]. It consists in choosing an initial probability measure on the causes and use some observations (even poorly related) of the consequences, and then use Bayes theorem to obtain a new distribution reflecting how data are informative. Such inverse problem is said well-posed when the probability a posteriori satisfies Hadamard's definition.

The motivation in this thesis is to investigate Bayesian inverse problems, with a particular focus on their discretization [CDS10]. Indeed, when the mathematical model involves partial derivative equations, the solution is often an infinite dimensional probability distribution. As it is characterized by a density, it is usually necessary to rely on simulation or variational optimization to extract information. These two methods need a discretization and this is the subject of interest here.

When the prior distribution is Gaussian [Bog98], such as continuous Gaussian random fields [RW04], it is well-known that it can be represented as a random series. The approximation then consists in keeping a finite number of terms in a truncated representation. Since it defines a different prior, it results in a distinct posterior and such discretization is consistent when both tend to be equal when the number of terms increases [Hos17a]. It appears that this convergence, measured in the Hellinger metric, can be controlled by the quality of the prior approximation. This naturally lead to the question of optimal series representation.

The Karhunen-Loève decomposition gives such representation in Hilbert spaces. However, it has several limitations for a practical use. First, it is the solution of an eigenvalue problem involving the covariance operator which is rarely known analytically. This restricts heavily the possible choices for the covariance kernel for instance. Secondly, it is based on Hilbert geometry, meanwhile the natural framework for inverse problems is Banach spaces. Both of these questions will be addressed, through a direct generalization of this decomposition to Banach spaces [BC17]. Besides the theoretical result, it appears that this new representation is easy to implement in the space of continuous functions. The optimality properties are discussed in this context, even if they are provided only in the special case of the Wiener process over $[0, 1]$.

The document is written in two distinct parts of respectively two and three chapters. Chapter 1 presents a short introduction to the theory of Gaussian random elements, with particular focus on series representation and covariance factorization. In chapter 2, the new decomposition is given, both in general Banach spaces and in the case of continuous functions over a compact metric set. It also includes both analytical and numerical examples involving Gaussian random processes. Finally, a particular analysis is given concerning the Wiener process, where two distinct optimality cri-

teria are studied. The second part turns to inverse problems. Specifically, chapter 3 gives a short reminder of Tikhonov-Philips regularization. Chapter 4 provides a detailed presentation of the Bayesian methodology for inverse problems, as it can be found in the literature. It includes a unified treatment with the sets of hypotheses given in [Hos17a], adapted to prior measures with exponential tails. Finally, chapter 5 gives an example how the new Karhunen-Loève decomposition can be used in a real-world, non-linear inverse problem. In particular, the complete analysis of the problem is provided, as well as experimental hyper-parameter tuning.

Part I

Approximation of Gaussian priors in Banach spaces

Chapter 1

Introduction to Gaussian random elements

The objective of this very first chapter is to introduce the theory of Gaussian random elements in Banach spaces. It is a direct generalization of Gaussian random vectors to infinite dimensional spaces, which in particular, will be useful as prior distributions in a Bayesian context. The presentation given here only includes results already well-known in the literature, and one may find inspiring treatments in [Bog98, Lif12] for locally convex spaces, [VTC87, Hai09, VV08, Kuo75, QL04, vN07, DZ14] for Banach spaces and [DZ04a] in Hilbert context.

1.1 Definition

Random elements In finite dimensional spaces, the notion of random variable is well-known (see textbooks [Str10, Mél10]) and defined as a Borel measurable application from any probability space. To extend this notion to (general) Banach spaces, measurability must be specified for different reasons. First, in infinite dimensional spaces, cylindrical (smallest such that every bounded linear functional is measurable) and Borel (smallest containing all open sets) σ-algebras are distinct and both leading to different definitions in general. Secondly, a Borel measurable map may not be approximated by simple functions, making the usual construction of Lebesgue's integral intractable. In order to circumvent both these difficulties, we stick to the notion of strong (or separably-valued) random elements, which offers a sufficiently general setup while avoiding unnecessary technical complications. From now on, we will always consider $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\mathcal{U}, Bor(\mathcal{U}))$ as prototypes of probability space and measurable Banach space (equipped with its Borel σ-algebra).

Definition 1 (Random element)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{U}$ a Banach space, a random element in $\mathcal{U}$ is a map $X : \Omega \to \mathcal{U}$ such that*

- *X is (strongly) measurable: $\forall A \in Bor(\mathcal{U}),\ X^{-1}(A) \in \mathcal{F}$,*

- *X is almost-surely separably valued: $\exists \mathcal{U}_0 \subset \mathcal{U}$, a separable Banach space such that $X \in \mathcal{U}_0$ almost-surely.*

With this definition, it is equivalent to consider the Banach space $\mathcal{U}$ separable, since one could restrict himself to $\mathcal{U}_0$. In this case, both cylindrical and Borel σ-algebras are the same. The space of all (equivalence classes of) random elements from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{U}, Bor(\mathcal{U}))$ will be noted $L^0(\mathcal{U})$ instead of $L^0(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{U}, Bor(\mathcal{U}))$ since we always assume the same probability space and σ-algebra. An alternative treatment could be made with adequate notions of probability measures, since random elements as previously defined immediately induce Radon probability measures. However, the random element point of view is used here, since it appears more intuitive in applications.

Proposition 1 (Distribution of a random element)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a random element, then the map*

$$\mu_X := A \in Bor(\mathcal{U}) \to \mathbb{P}(X \in A) \in [0, 1],$$

is a Borel probability measure, called the distribution of X and it is Radon.

Proof. Since X is measurable, μ_X is well-defined as a push-forward measure. The fact that μ_X is a Radon is a classical result from measure theory, see proposition I.2 in [QL04] for instance. □

Two random elements sharing the same distribution will be called identically distributed. One particularly important property of such probability measures is their characterization through all one-dimensional projections.

Definition 2 (Fourier transform)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a random element, the associated Fourier transform is defined as*

$$\widehat{X} := l \in \mathcal{U}^* \to \mathbb{E}\left[\exp\left(i \langle X, l \rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] \in \mathbb{C}.$$

Proposition 2. *Let $\mathcal{U}$ be a Banach space, $X, Y \in L^0(\mathcal{U})$ two random elements, if $\hat{X} = \hat{Y}$ then X and Y are identically distributed.*

Proof. This result lies on a fundamental property, two finite dimensional probability measures with the same Fourier transform are equal. Now, since the set of all cylinders is a Dynkin system and generates the Cylindrical σ-algebra, which in separable spaces coincides with Borel σ-algebra, the result is clear. □

From now on, the notation $\langle u, l \rangle$ will always denote the image of a vector u by a bounded linear functional l, it is the duality pairing with in $\mathcal{U}$. The dual space will be noted $\mathcal{U}^*$. When different spaces will be involved, the complete notation $\langle u, l \rangle_{\mathcal{U},\mathcal{U}^*}$ will be preferred.

Gaussian random elements The role of Gaussian distributions is ubiquitous in probability theory, statistics and most of their applications in science and engineering. In Euclidean spaces, they are usually defined by a density w.r.t. the fundamental Lebesgue's measure. However, when the dimension is infinite, this cannot be done since Lebesgue's measure does not exist any more. However, since all one dimensional projections of random elements totally characterize their underlying distribution, the following definition is interesting.

Definition 3 (Gaussian random element)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a random element, it is Gaussian if*

$$\forall l \in \mathcal{U}^*,\ \langle X, l\rangle \text{ is a Gaussian random variable.}$$

This definition is clearly a generalization of a fundamental fact about Gaussian vectors. Now, in terms of vocabulary, a random element is said centred when all one-dimensional projections $\langle X, l\rangle$ are, and only this case will be considered here for simplicity (**without further mention**). The general theory is obtained through a translation. In terms of distribution, since all projections are Gaussian it imposes a distinct form for the Fourier transform.

Proposition 3. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a random element, it is Gaussian if and only if*

$$\hat{X}(l) = \exp\left(-\frac{1}{2}\varphi_X(l,l)\right),$$

where:

$$\varphi_X : (l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^* \to \mathbb{E}\left[\langle X, l_1\rangle \langle X, l_2\rangle\right] \in \mathbb{R}$$

is a symmetric, non-negative, bilinear form.

Proof. Let $X \in L^0(\mathcal{U})$ be a random element. If it is Gaussian, by definition $\langle X, l\rangle$, $\forall l \in \mathcal{U}^*$ is a Gaussian random variable and its Fourier transform is:

$$\widehat{\langle X, l\rangle} : t \in \mathbb{R} \to \exp\left(-\frac{t^2}{2}\mathbb{E}\left[\langle X, l\rangle^2\right]\right) \in \mathbb{R}.$$

Now, it is enough to see that $\forall l \in \mathcal{U}^*$, $\hat{X}(l) = \widehat{\langle X, l\rangle}(1)$ and take $\varphi_X(l_1, l_2) := \mathbb{E}\left[\langle X, l_1\rangle \langle X, l_2\rangle\right]$ which is clearly bilinear, non-negative and symmetric. Conversely, suppose the Fourier transform of X has the previous form, then for all $l \in \mathcal{U}^*$, $\langle X, l\rangle$ is Gaussian, thus X is a Gaussian random element and the proof is complete. □

The bilinear form φ_X in proposition 3 is called the covariance of X and totally characterizes the distribution of a Gaussian random element (since it describes the Fourier transform). In some cases, it enjoys definiteness as well, and such covariance will be said non-degenerated and provides a pre-Hilbert structure to $\mathcal{U}^*$. The finite dimensional analogue of φ_X is the usual covariance matrix. One easy (and useful) consequence of proposition 3 is the conservation of Gaussian distributions under bounded linear transformations.

Proposition 4. *Let $\mathcal{U}_1, \mathcal{U}_2$ be Banach spaces, $X \in L^0(\mathcal{U}_1)$ a Gaussian random element and $A \in \mathcal{L}(\mathcal{U}_1, \mathcal{U}_2)$ a bounded operator, then $AX \in L^0(\mathcal{U}_2)$ is a Gaussian random element.*

Proof. $A : \mathcal{U}_1 \to \mathcal{U}_2$ is a bounded operator, thus the adjoint $A^* : \mathcal{U}_2^* \to \mathcal{U}_1^*$ exists and $\forall (x, l) \in \mathcal{U}_1 \times \mathcal{U}_2^*,\ \langle Ax, l \rangle_{\mathcal{U}_2, \mathcal{U}_2^*} = \langle x, A^* l \rangle_{\mathcal{U}_1, \mathcal{U}_1^*}$. Now:

$$\forall l \in \mathcal{U}_2^*,\ \hat{AX}(l) = \hat{X}(A^* l) = \exp\left(-\frac{1}{2}\varphi_X(A^* l, A^* l)\right),$$

which is the Fourier transform of a centred Gaussian random element by proposition 3. □

Corollary 1. *Let $\mathcal{U}$ a Banach space, $X = (X_1, X_2)$ a Gaussian random element in $L^0(\mathcal{U} \times \mathcal{U})$, then $X_1 + X_2$ is a Gaussian random element in $\mathcal{U}$.*

Proof. Apply proposition 4 with the bounded operator $A := (x, y) \to x + y$. □

Now that Gaussian random elements are defined, the presentation will now turn to advanced properties, starting with Bochner integrability.

1.2 Integrability

There are multiple notions of vector-valued integration in Banach spaces (Pettis and Bochner for instance), referring to the different notions of measurability mentioned earlier and here, only Bochner's theory will be used. In this area, Fernique's theorem is a very powerful result, stating that all Gaussian random elements have exponential tails. This will have deep consequences, in particular w.r.t. the covariance form. Fernique's theorem only needs a rotation (of angle $\frac{\pi}{4}$) invariance principle, stated in the next lemma.

Lemma 1. *Let $\mathcal{U}$ be a Banach space, $X, Y \in L^0(\mathcal{U})$ independent and identically distributed Gaussian random elements with distribution μ, then $\frac{\sqrt{2}}{2}(X - Y), \frac{\sqrt{2}}{2}(X + Y)$ are both Gaussian random elements with distribution μ and are independent.*

Proof. Let $(l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^*$ then

$$\mathbb{E}\left[\langle X - Y, l_1 \rangle \langle X + Y, l_2 \rangle\right] = \varphi_X(l_1, l_2) - \varphi_Y(l_1, l_2) = 0,$$

which gives independence and

$$\frac{1}{2}\mathbb{E}\left[\langle X - Y, l_1 \rangle^2\right] = \frac{1}{2}(\varphi_X(l_1, l_1) + \varphi_Y(l_1, l_1)) = \varphi_X(l_1, l_1).$$

The computation is similar for the second random element. □

Theorem 1 (Fernique's)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then there exists $\alpha > 0$ such that:*

$$\forall \kappa \in [0, \alpha],\ \mathbb{E}\left[\exp\left(\kappa \|X\|_{\mathcal{U}}^2\right)\right] < \infty.$$

Proof. The proof is based on an estimation of the type:

$$\mathbb{E}\left[\exp(\kappa \|X\|_{\mathcal{U}}^2)\right] \leq \mathbb{P}\left(\|X\|_{\mathcal{U}} \leq t_0\right) \exp\left(\kappa t_0^2\right) + \sum_{n=0}^{\infty} \mathbb{P}\left(t_n \leq \|X\|_{\mathcal{U}} \leq t_{n+1}\right) \exp\left(\kappa t_{n+1}^2\right),$$

where the sequence $(t_n)_{n\in\mathbb{N}}$ is chosen diverging and such that previous series converges.

- Let $X, Y \in L^0(\mathcal{U})$ be independent and identically distributed Gaussian random elements, $(s,t) \in \mathbb{R}^2$ such that $t \geq s > 0$ then it comes from the previous lemma:

$$\begin{aligned}
&\mathbb{P}\left(\|X\|_{\mathcal{U}} \leq s\right) \mathbb{P}\left(\|X\|_{\mathcal{U}} > t\right) \\
&= \mathbb{P}\left(\|X+Y\|_{\mathcal{U}} \leq \sqrt{2}s\right) \mathbb{P}\left(\|X-Y\|_{\mathcal{U}} > \sqrt{2}t\right) \\
&\leq \mathbb{P}\left(|\, \|X\|_{\mathcal{U}} - \|Y\|_{\mathcal{U}} \,| \leq \sqrt{2}s,\ \|X\|_{\mathcal{U}} + \|Y\|_{\mathcal{U}} > \sqrt{2}t\right).
\end{aligned}$$

  Clearly,

$$\begin{aligned}
\|X\|_{\mathcal{U}} + \|Y\|_{\mathcal{U}} - |\, \|X\|_{\mathcal{U}} - \|Y\|_{\mathcal{U}} \,| &> \sqrt{2}t - \sqrt{2}s, \\
\Leftrightarrow 2\min(\|X\|_{\mathcal{U}}, \|Y\|_{\mathcal{U}}) &> \sqrt{2}(t-s),
\end{aligned}$$

  thus

$$\mathbb{P}\left(\|X\|_{\mathcal{U}} \leq s\right) \mathbb{P}\left(\|X\|_{\mathcal{U}} > t\right) \leq \mathbb{P}\left(\|X\|_{\mathcal{U}} > \frac{t-s}{\sqrt{2}}\right)^2.$$

  Now, t and get s will be fixed along a specific sequence $(t_n)_{n\in\mathbb{N}}$.

- Now, let $t_0 > 0$ such that $\mathbb{P}(\|X\|_{\mathcal{U}} \leq t_0) \geq \frac{2}{3}$ and define $t_{n+1} = t_0 + \sqrt{2}t_n,\ \forall n \in \mathbb{N}$, then by induction $t_n = r\frac{\sqrt{2}^{n+1}-1}{\sqrt{2}-1}$ and thus $t_n \leq t_0\sqrt{2}^{n+4}$. Furthermore, let $u_n := \frac{\mathbb{P}(\|X\|>t_n)}{\mathbb{P}(\|X\|\leq t_0)}$ which leads to $u_{n+1} \leq u_n^2$ from the previous equation with $s = t_0$ and $t = t_{n+1}$. Finally, $u_n \leq u_0^{2^n} \leq 2^{-2^n}$ and $\mathbb{P}(\|X\| \geq t_n) = u_n\mathbb{P}(\|X\| \leq t_0) \leq 2^{-2^n}$.

- Injecting these elements in the initial expression leads to

$$\begin{aligned}
\mathbb{E}\left[\exp(\kappa \|X\|_{\mathcal{U}}^2)\right] &\leq \exp\left(\kappa t_0^2\right) + \sum_{n=0}^{\infty} 2^{-2^n} \exp\left(\kappa t_0^2 2^{n+5}\right), \\
&\leq \exp\left(\kappa t_0^2\right) + \sum_{n=0}^{\infty} \exp\left(-2^n\left(\log(2) - \kappa t_0^2 2^5\right)\right)
\end{aligned}$$

  which converges for $\kappa > 0$ and sufficiently small.

□

From this theorem, we know that Gaussian random elements have thin tails, similarly to the finite dimensional case. In particular, it implies the existence of moments of (strong) order p, for all $p \in \mathbb{N}$.

Corollary 2. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then*

$$\forall \beta \in \mathbb{R}^+,\ \mathbb{E}\left[\exp\left(\beta \|X\|_{\mathcal{U}}\right)\right] < +\infty.$$

Proof. Let κ from theorem 1, then $\forall \alpha \in [0, \kappa]$

$$x \in \mathbb{R} \to \exp(\beta x) \exp\left(-\alpha x^2\right) \in \mathbb{R}^+,$$

is bounded for all $\beta \in \mathbb{R}$. □

Corollary 3. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then:*

$$\forall p \in \mathbb{N},\ \mathbb{E}[\|X\|_{\mathcal{U}}^p] < +\infty.$$

Proof. Let κ from theorem 1, then for all $\alpha \in [0, \kappa]$, the function:

$$x \in \mathbb{R}^+ \to x^p e^{-\kappa x^2} \in \mathbb{R}$$

is bounded for all $p \in \mathbb{N}$. □

Corollary 4. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then the covariance bilinear form φ_X is continuous and satisfies the following estimate:*

$$\|\varphi_X\| \leq \mathbb{E}\left[\|X\|_{\mathcal{U}}^2\right].$$

Proof.

$$\forall (l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^*,\ \mathbb{E}\left[\langle X, l_1\rangle \langle X, l_2\rangle\right] \leq \mathbb{E}\left[\|X\|_{\mathcal{U}}^2\right] \|l_1\|_{\mathcal{U}^*} \|l_2\|_{\mathcal{U}^*},$$

taking the supremum on the unit ball of $\mathcal{U}^*$ in both arguments completes the proof. □

Next section will continue on properties of the covariance form, using Hilbert spaces.

1.3 Covariance and Hilbert subspace(s)

In the definition of a Gaussian random element, the covariance is a central notion and encloses valuable information. In particular, it completely characterizes the distribution, and as it will be shown, provides a fundamental Hilbert structure.

Covariance operator In the spirit of finite dimensional spaces, the covariance can be expressed in operator language instead of bilinear forms (the covariance matrix may also be seen as a linear map). This will provide an alternative point of view, particularly adapted to standard functional analysis tools.

Definition 4. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then its covariance operator $\mathcal{C}_X$ is defined as follows:*

$$\mathcal{C}_X := l \in \mathcal{U}^* \to \mathbb{E}\left[\langle X, l\rangle X\right] \in \mathcal{U}.$$

The covariance operator is well-defined and bounded as a consequence of corollary 3.

Proposition 5. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element then its covariance operator $\mathcal{C}_X$ satisfies the following relations:*

- $\forall (l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^*,\ \langle \mathcal{C}_X l_1, l_2\rangle_{\mathcal{U},\mathcal{U}^*} = \varphi_X(l_1, l_2)$,
- $\|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \|\varphi_X\|$

Proof. It is a standard property of Bochner's integral that $\forall (l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^*$:

$$\langle \mathcal{C}_X l_1, l_2\rangle = \langle \mathbb{E}\left[\langle X, l_1\rangle X\right], l_2\rangle = \mathbb{E}\left[\langle X, l_1\rangle \langle X, l_2\rangle\right] = \varphi_X(l_1, l_2).$$

The equality of the norms is a direct application of the definition and Cauchy-Schwarz inequality. Start with

$$\forall (l_1, l_2) \in \mathcal{B}_{\mathcal{U}^*} \times \mathcal{B}_{\mathcal{U}^*},\ \langle \mathcal{C}_X l_1, l_2\rangle \leq \|\varphi_X\|,$$

thus $\|\mathcal{C}_X\| \leq \|\varphi_X\|$. The other inequality is obtained because

$$\langle \mathcal{C}_X l_1, l_2\rangle \leq \|\mathcal{C}_X\| \, \|l_1\|_{\mathcal{U}^*} \, \|l_2\|_{\mathcal{U}^*},$$

which gives $\|\varphi_X\| \leq \|\mathcal{C}_X\|$. □

Since the covariance operator and bilinear forms are linked, the notation $X \sim \mathcal{N}(0, \mathcal{C}_X)$ will be used to define a Gaussian random element in $L^0(\mathcal{U})$ together with its distribution. The covariance operator enjoys more regularity than simply boundedness, it is nuclear in the following sense.

Proposition 6. *Let $\mathcal{U}$ be a Banach space and $X \sim \mathcal{N}(0, \mathcal{C}_X)$, then $\mathcal{C}_X$ is a nuclear operator, that is $\exists (u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ such that*

$$\forall l \in \mathcal{U}^*,\ \mathcal{C}_X l = \sum_{n\geq 0} \langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*} \, u_n,$$

and $\left(\|u_n\|_{\mathcal{U}}\right)_{n\in\mathbb{N}} \in l^2(\mathbb{N})$.

Proof. The proof can be found in [VTC87]. □

The particular properties (symmetry and non-negativity) of the covariance operator are almost those of an inner product, however it is not necessarily non degenerated. This can be exploited using a quotient space construction.

Lemma 2. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, then there exists a Hilbert space H and a bounded operator $A \in \mathcal{L}(\mathcal{U}^*, H)$ such that*

- $\mathcal{C}_X = A^*A$,
- $R(A)$ *is dense in* H,
- $R(A^*) = R(\mathcal{C}_X)$,
- $\|A\|_{\mathcal{L}(\mathcal{U}^*,H)} = \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}$.

Furthermore, the factorization is unique in the sense that if $A_1 \in \mathcal{L}(\mathcal{U}^*, H_1)$ and $A_2 \in \mathcal{L}(\mathcal{U}^*, H_2)$ are two operators such that $\mathcal{C}_X = A_1^*A_1 = A_2^*A_2$ with $R(A_1)$, $R(A_2)$ respectively dense in H_1, H_2 then there exists an isomorphism $u : H_2 \to H_1$ satisfying $A_1 = uA_2$.

Proof. The operator $\mathcal{C}_X$ is bounded, thus $\ker(\mathcal{C}_X)$ is closed in $\mathcal{U}^*$. Let $H_0 = \mathcal{U}^*/\ker(\mathcal{C}_X)$ be the quotient space equipped with the following bilinear application:

$$\langle [l_1], [l_2] \rangle_{H_0} = \langle \mathcal{C}_X l_1, l_2 \rangle ,$$

which is well defined. It is symmetric, bilinear and positive-definite. Let H be the topological completion of H_0 and $A : \mathcal{U}^* \to H$ the natural embedding operator. It is clearly linear and has a dense image (again, by construction). To see that A is bounded, one has:

$$\forall l \in \mathcal{U}^*,\ \|Al\|^2_H = \langle Al, Al \rangle_H = \langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*} \leq \|\mathcal{C}_X\| \, \|l\|^2_{\mathcal{U}^*} ,$$

thus $\|A\| \leq \|\mathcal{C}_X\|^{\frac{1}{2}}$. Conversely, the Cauchy-Schwarz inequality gives

$$\forall (l, g) \in \mathcal{U}^* \times \mathcal{U}^*,\ \langle \mathcal{C}_X l, g \rangle_{\mathcal{U},\mathcal{U}^*} = \langle Al, Ag \rangle_H \leq \|Al\|_H \, \|Ag\|_H ,$$

and $\|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} \leq \|A\|^2_{\mathcal{L}(\mathcal{U}^*,H)}$, thus $\|A\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U}_X)} = \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}$. Now, let $(l_1, l_2) \in \mathcal{U}^* \times \mathcal{U}^*$ then

$$\langle \mathcal{C}_X l_1, l_2 \rangle = \langle Al_1, Al_2 \rangle_H = \langle A^*Al_1, l_2 \rangle ,$$

thus $A^*A = \mathcal{C}_X$. Now, let A_1, H_1 and A_2, H_2 two valid factorizations. Let $l \in \mathcal{U}^*$ such that $A_1 l = 0$, then $\mathcal{C}_X l = 0$ thus $A_2 l = 0$ as well. One can then consider the map

$$u : A_1 l \in H_1 \to A_2 l \in H_2,$$

which is linear and isomorphic between $R(A_1)$ and $R(A_2)$. To see that it is isomorphic let $l \in \mathcal{U}^*$. Then it comes:

$$\|A_1 l\|^2_{H_1} = \langle \mathcal{C}_X l_1, l_1 \rangle = \|A_2 l\|^2_{H_2}$$

and it extends to H_1 and H_2 by density. □

Even if this result is rather formal (by the use of quotient operations), it has two concrete instances, namely the Gaussian and Cameron-Martin spaces that are presented below.

Gaussian space The first method to build the Hilbert subspace (from lemma 2) is to follow the definition of a Gaussian random element, and associate every linear functional to a Gaussian random variable by the following application:

$$l \in \mathcal{U}^* \to \langle X, l\rangle_{\mathcal{U},\mathcal{U}^*} \in L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}).$$

This is completely equivalent to consider the equivalence classes (equality μ_X-a.e.) of all bounded linear functionals:

$$l \in \mathcal{U}^* \to l \in L^2(\mathcal{U}, Bor(\mathcal{U}), \mu_X; \mathbb{R}).$$

This last representation will be kept (noted shortly as $L^2(\mathcal{U};\mathbb{R})$, and the Gaussian space defined as the topological closure of the dual space by this linear map.

Definition 5 (Gaussian space)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, the associated Gaussian space is*

$$\mathcal{H}_X = \overline{\mathcal{U}^*}^{L^2(\mathcal{U};\mathbb{R})}.$$

Here, linear functionals are implicitly identified with their equivalence classes and the notation is unchanged and will always be clear in its context. The denomination is justified since all elements in this space have indeed, as random variables, a Gaussian distribution.

Proposition 7. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element and $z \in \mathcal{H}_X$ then z is a Gaussian random variable.*

Proof. Let $(l_n)_{n\in\mathbb{N}} \subset \mathcal{U}^*$ such that $l_n \to z$ in $L^2(\mathcal{U};\mathbb{R})$, then for all $t \in \mathbb{R}$ it comes that $\widehat{l_n}(t) \to \widehat{z}(t)$ since

$$\left|e^{itl_n(u)} - e^{itz(u)}\right| \leq |t||l_n(u) - z(u)|$$

and

$$\int_{\mathcal{U}} \left|e^{itl_n(u)} - e^{itz(u)}\right| \mu_X(du) \leq \sqrt{\int_{\mathcal{U}} t^2 (l_n(u) - z(u))^2 \mu_X(du)}.$$

Since $\forall n \in \mathbb{N},\ \forall t \in \mathbb{R},\ \widehat{\langle X, l_n\rangle}(t) = \exp\left(-\frac{t^2}{2}\|l_n\|^2_{L^2(\mathcal{U};\mathbb{R})}\right)$, one has:

$$\forall t \in \mathbb{R},\ \hat{z}(t) = \exp\left(-\frac{t^2}{2}\|z\|^2_{L^2}\right),$$

which completes the proof. □

This construction does not require any theoretical completion, since the space of bounded linear functionals is seen as a subset of the larger space $L^2(\mathcal{U};\mathbb{R})$.

Proposition 8. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with Gaussian space $\mathcal{H}_X$, then the following operator:*

$$i_{\mathcal{H}_X} := l \in \mathcal{U}^* \to l \in \mathcal{H}_X$$

is bounded and provides a factorization of the covariance operator $\mathcal{C}_X = i^*_{\mathcal{H}_X} i_{\mathcal{H}_X}$.

Proof. Note $i_{\mathcal{H}_X} := l \in \mathcal{U}^* \to l \in \mathcal{H}_X$, then this operator is linear and bounded because $X \in L^2(\mathcal{U})$ and

$$\|l\|^2_{\mathcal{H}_X} = \int_{\mathcal{U}} l(u)^2 \mu_X(du) \leq \int_{\mathcal{U}} \|u\|^2_{\mathcal{U}} \mu_X(du) \|l\|^2_{\mathcal{U}^*}.$$

The adjoint operator is well-defined as

$$i^*_{\mathcal{H}_X} := z \in \mathcal{H}_X \to \int_{\mathcal{U}} z(u) u \mu_X(du) \in \mathcal{U}.$$

It is clear that $\mathcal{C}_X = i^*_{\mathcal{H}_X} i_{\mathcal{H}_X}$. □

Proposition 7 is the reason why this particular factorization is said *canonical* in [Vak91], as the Hilbert space is made of random variables.

Cameron-Martin space The second possibility is to use the covariance operator and build a *proper* subspace of $\mathcal{U}$, the Cameron-Martin space. Indeed, one can define the following bilinear form on $R(\mathcal{C}_X)$ taking:

$$\begin{aligned} f &= \sum_{i=1}^{n} \alpha_i \mathcal{C}_X l_i, \\ g &= \sum_{j=1}^{m} \beta_j \mathcal{C}_X g_j, \\ \langle f, g \rangle_X &:= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \langle \mathcal{C}_X l_i, g_j \rangle_{\mathcal{U},\mathcal{U}^*}, \end{aligned}$$

where $(l_1, ..., l_n, g_1, ..., g_m) \in (\mathcal{U}^*)^{n+m}$, $(\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_m) \subset \mathbb{R}^{n+m}$. $\langle ., . \rangle_X$ is a well-defined, symmetric positive-definite bilinear form. The associated norm will be noted $\|.\|_X$ and the Cameron-Martin space will be defined as the topological completion of $R(\mathcal{C}_X)$ under this norm in $\mathcal{U}$.

Definition 6 (Cameron-Martin space)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, the associated Cameron-Martin space is*

$$\mathcal{U}_X := \widehat{R(\mathcal{C}_X)}^{\langle .,. \rangle_X}.$$

Unlike the previous construction, there is no ambient space for the completion to take place since it is a theoretical topological operation using Cauchy sequences. However, it can be identified with a subspace of $\mathcal{U}$.

Lemma 3. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ and $\langle .,. \rangle_X$ the inner product on $R(\mathcal{C}_X)$ defined as previously, then*

$$\|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \sup_{l \in \mathcal{B}_{\mathcal{U}^*}} \|\mathcal{C}_X l\|_X .$$

Proof. Let $l \in \mathcal{U}^*$, then

$$\|\mathcal{C}_X l\|_{\mathcal{U}} = \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X l, g \rangle_{\mathcal{U},\mathcal{U}^*} = \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X l, \mathcal{C}_X g \rangle_X \leq \|\mathcal{C}_X l\|_X \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \|\mathcal{C}_X g\|_X ,$$

and one gets $\|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} \leq \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \|\mathcal{C}_X g\|_X$. Furthermore,

$$\|\mathcal{C}_X l\|_X = \sqrt{\langle \mathcal{C}_X l, \mathcal{C}_X l \rangle_X} = \sqrt{\langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*}} \leq \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} \|l\|_{\mathcal{U}^*} ,$$

which gives the converse inequality and the proof is complete. □

Proposition 9. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then there is a natural topological injection $i_{\mathcal{U}_X} : \mathcal{U}_X \to \mathcal{U}$ such that*

$$\forall h \in R(\mathcal{C}_X) , \; i_{\mathcal{U}_X} h = h,$$

and $\|i_{\mathcal{U}_X}\|_{\mathcal{L}(\mathcal{U}_X,\mathcal{U})} = \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}$. Furthermore (reproduction property):

$$\forall h \in \mathcal{U}_X , \; \forall l \in \mathcal{U}^* , \; \langle h, l \rangle_{\mathcal{U},\mathcal{U}^*} = \langle h, \mathcal{C}_X l \rangle_X .$$

Proof. Let $i : h \in R(\mathcal{C}_X) \to h \in \mathcal{U}$. By definition of the inner product on $\mathcal{U}_X$, one gets:

$$\forall h \in R(\mathcal{C}_X) , \; \forall l \in \mathcal{U}^* , \; \langle h, l \rangle_{\mathcal{U},\mathcal{U}^*} = \langle h, \mathcal{C}_X l \rangle_X .$$

To see that i is bounded, it is enough to write:

$$\begin{aligned}
\|i\|_{\mathcal{L}(R(\mathcal{C}_X),\mathcal{U})} &= \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \sup_{\|\mathcal{C}_X l\|_X \leq 1} \langle \mathcal{C}_X l, g \rangle_{\mathcal{U},\mathcal{U}^*} , \\
&= \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \sup_{\|\mathcal{C}_X l\|_X \leq 1} \langle \mathcal{C}_X l, \mathcal{C}_X g \rangle_X , \\
&= \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \left\langle \frac{\mathcal{C}_X g}{\|\mathcal{C}_X g\|_X}, \mathcal{C}_X g \right\rangle_X , \\
&= \sup_{g \in \mathcal{B}_{\mathcal{U}^*}} \|\mathcal{C}_X g\|_X ,
\end{aligned}$$

and by lemma 3:

$$\|i\|_{\mathcal{L}(R(\mathcal{C}_X),\mathcal{U})} = \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} .$$

Now, $R(\mathcal{C}_X)$ is dense in $\mathcal{U}_X$ (by construction) thus i uniquely extends by continuity to $\mathcal{U}_X$, this new map being noted $i_{\mathcal{U}_X}$ and $\|i_{\mathcal{U}_X}\|_{\mathcal{L}(\mathcal{U}_X,\mathcal{U})} = \|\mathcal{C}_X\|^{\frac{1}{2}}_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}$. Let $h \in \mathcal{U}_X$, $l \in \mathcal{U}^*$ and $(\mathcal{C}_X l_n)_{n \in \mathbb{N}}$ a sequence converging to h in $\mathcal{U}_X$, then

$$\lim_{n \to \infty} \langle \mathcal{C}_X l_n, \mathcal{C}_X l \rangle_X = \langle h, \mathcal{C}_X l \rangle_X$$

since strong convergence implies weak convergence. However, $i_{\mathcal{U}_X}$ is bounded, thus $i_{\mathcal{U}_X}(\mathcal{C}_X l_n) = \mathcal{C}_X l_n \to i_{\mathcal{U}_X}(h)$ in $\mathcal{U}$ and implies that $\langle \mathcal{C}_X l_n, l\rangle_{\mathcal{U},\mathcal{U}^*} \to \langle h, l\rangle_{\mathcal{U},\mathcal{U}^*}$. Because $\langle \mathcal{C}_X l_n, \mathcal{C}_X l\rangle_X = \langle \mathcal{C}_X l_n, l\rangle_{\mathcal{U},\mathcal{U}^*}$ and the uniqueness of limits the reproduction property is true. It remains to see that $i_{\mathcal{U}_X}$ is injective. Let $h \in \mathcal{U}_X$ such that $i_{\mathcal{U}_X} h = 0$, then by the reproduction property, one has $\langle i_{\mathcal{U}_X} h, l\rangle_{\mathcal{U}} = \langle h, \mathcal{C}_X l\rangle_X = 0$. Since $R(\mathcal{C}_X)$ is dense in $\mathcal{U}_X$, it comes that $h = 0$, thus $i_{\mathcal{U}_X}$ is an injection. □

From now on, the space $\mathcal{U}_X$ is identified with a subspace of $\mathcal{U}$. As it was announced, the Cameron-Martin space and its injection provide a factorization of the covariance operator.

Corollary 5. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ and Cameron-Martin space $\mathcal{U}_X$, then one has the following factorization:*

$$\mathcal{C}_X = i_{\mathcal{U}_X} i_{\mathcal{U}_X}^*.$$

Loève isometry Since both Cameron-Martin and Gaussian spaces provide factorization of the covariance operator, there exists an isometry between them. In particular, it will be possible to use either point of view to prove results on X.

Proposition 10 (Loève isometry)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then*

$$\mathcal{U}_X \cong \mathcal{H}_X.$$

Proof. Consider the following map:

$$\mathcal{C}_X l \in R(\mathcal{C}_X) \to l \in L^2(\mathcal{U};\mathbb{R}).$$

Since $l = 0$ μ_X-a.e. implies that $\mathcal{C}_X l = 0$, it is well-defined and injective. Moreover, for all $l \in \mathcal{U}^*$, $\|\mathcal{C}_X l\|_X = \|l\|_{L^2(\mathcal{U};\mathbb{R})}$. This isomorphism naturally extends by density, thus Cameron-Martin and Gaussian spaces are isomorphic. □

It states that to every element from the Cameron-Martin space corresponds a unique element in the Gaussian space and vice-versa.

Corollary 6. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element then the associated Cameron-Martin space $\mathcal{U}_X$ is separable.*

Proof. Since $\mathcal{U}$ can be taken separable, $L^2(\mathcal{U};\mathbb{R})$ is separable since $Bor(\mathcal{U})$ is countably generated. By the Loève isometry, it comes that $\mathcal{U}_X$ is separable. □

The following theorem will be critical in the next chapter.

Theorem 2. *Let $\mathcal{U}$ be a Banach space, $X \sim \mathcal{N}(0, \mathcal{C}_X)$ then*

$$\mathcal{B}_{\mathcal{U}_X} = \{h \in \mathcal{U}_X,\ \|h\|_X \leq 1\} \subset \mathcal{U}_X$$

is compact in $\mathcal{U}$.

Proof. The proof can be found in [Bog98]. □

Translation of Gaussian elements So far, the presentation only concerned centred Gaussian random elements for simplicity. However, all previous notions are adapted to the case where the mean is non-zero. For any vector $u \in \mathcal{U}$ and Gaussian element X, $X+u$ is also Gaussian. However, in terms of distribution, one may ask if they are related in some sense. The Cameron-Martin theorem states precisely that both distributions are equivalent for translations in the Cameron-Martin space.

Theorem 3 (Cameron-Martin)**.** *Let X be a Gaussian element and $h \in \mathcal{U}_X$ with corresponding element $h^* \in \mathcal{H}_X$ then the distributions of X and $X+h$ are equivalent with Radon-Nikodym density*

$$\frac{\partial \mu_{X+h}}{\partial \mu_X}(u) = \exp\left(h^*(u) - \frac{\|h\|_X^2}{2}\right).$$

Proof. Let $h \in \mathcal{U}_X$, there is a corresponding element h^* in $L^2(\mathcal{U};\mathbb{R})$ (Loève isometry), which is a Gaussian random variable with variance $\|h\|_X^2$. It follows that $\exp(h^*)$ is $L^1(\mathcal{U},\mathbb{R})$ (as a log-normal random variate) and $f_h : u \in \mathcal{U} \to \exp(h^*(u) - \frac{1}{2}\|h\|_X^2)$ is a strictly positive (Radon-Nikodym) density (it integrates to one). Consider the following application for a fixed $l \in \mathcal{U}^*$

$$\phi := t \in \mathbb{R} \to \exp\left(-\frac{1}{2}\|h\|_X^2\right) \int_{\mathcal{U}} \exp\left(i\,\langle u, l\rangle_{\mathcal{U},\mathcal{U}^*} + ith^*(u)\right) \mu_X(du) \in \mathbb{C}.$$

As we have $l + th^*$ Gaussian (it is an element from $\mathcal{H}_X$), the previous expression becomes

$$\phi(t) = \exp\left(-\frac{1}{2}\left(\varphi_X(l,l) + 2t\,\langle h, l\rangle_{\mathcal{U},\mathcal{U}^*} + (1+t^2)\|h\|_X^2\right)\right)$$

using the reproducing property. Now, ϕ can be extended to the complex plane and in particular Lebesgue's dominated convergence theorem gives

$$\lim_{z\to -i} \phi(z) = \exp\left(i\,\langle h, l\rangle_{\mathcal{U},\mathcal{U}^*} - \frac{1}{2}\varphi_X(l,l)\right).$$

The direct Fourier transform of μ_{X+h} is obtained directly as follows:

$$\begin{aligned}\forall l \in \mathcal{U}^*,\ \widehat{X+h}(l) &= \mathbb{E}\left[\exp\left(i\,\langle X+h, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] \\ &= \exp\left(i\,\langle h, l\rangle_{\mathcal{U},\mathcal{U}^*} - \frac{1}{2}\varphi_X(l,l)\right).\end{aligned}$$

Since both Fourier transforms are the same, the underlying distributions are equal. □

In the literature, the Cameron-Martin theorem states as well that if $h \notin \mathcal{U}_X$, the distributions are mutually singular. This result is omitted here since it will not be used. One interesting consequence of the Cameron-Martin theorem is the possibility to quantify the so-called small-ball probability.

Theorem 4 (Onsager-Machlup functional)**.** *Let $\mathcal{U}$ be a Banach space and $X \in L^0(\mathcal{U})$ a Gaussian random element then $\forall (h_1, h_2) \in \mathcal{U}_X^2$, it comes:*

$$\lim_{r\to 0} \frac{\mu_X\left(\mathcal{B}_{\mathcal{U}}(h_1, r]\right)}{\mu_X\left(\mathcal{B}_{\mathcal{U}}(h_2, r]\right)} = \exp\left(\frac{1}{2}\left\|h_2\right\|_X^2 - \frac{1}{2}\left\|h_1\right\|_X^2\right).$$

Proof. The proof may be found in [Bog98] (Corollary 4.7.8). □

1.4 Series representation

Admissible sequences The concept of series representation for a Gaussian random element X consists in finding particular sequences of vectors $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ such that the random series

$$\sum_{n\geq 0} \xi_n u_n,$$

with $(\xi_n)_{n\in\mathbb{N}}$ independent $\mathcal{N}(0,1)$ random variables converges almost-surely in $\mathcal{U}$ with distribution μ_X (the law of X). Every such family is called an admissible sequence for X.

Definition 7 (Admissible sequence)**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, $(\xi_n)_{n\in\mathbb{N}}$ a family of independent standard Gaussian random variables, then a sequence $(u_n)_{n\in\mathbb{N}}$ is admissible for X if*

$$\sum_{n\geq 0} \xi_n u_n \quad \textit{converges almost-surely in } \mathcal{U},$$

and its distribution is μ_X.

The existence of such admissible sequence has been first shown in [Tsi81] and later studied in [Bog98, VTC87, Vak91, LP09, Lif12] for instance. Whenever the series representation is actually a finite sum, the random element is said of finite rank (equal to the number of terms). It is essentially the same problem as finding (tensor) representations of the covariance operator.

Lemma 4 (Lemma 1 in [LP09])**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, $\mathcal{C}_X$ its covariance operator and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ then the following assertions are equivalent:*

1. *$(u_n)_{n\in\mathbb{N}}$ is admissible for X,*
2. *$\forall l \in \mathcal{U}^*$, $\left(\langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)_{n\in\mathbb{N}}$ is admissible for $\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*}$,*
3. *$\forall l \in \mathcal{U}^*$, $\mathcal{C}_X l = \sum_{n\geq 0} \langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*} u_n$, where the convergence is in $\mathcal{U}$,*
4. *$\forall r > 0$, $\forall l, g \in \mathcal{B}_{\mathcal{U}^*}(0, r]$*

$$\langle \mathcal{C}_X l, g\rangle_{\mathcal{U},\mathcal{U}^*} = \sum_{n\geq 0} \langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*} \langle u_n, g\rangle_{\mathcal{U},\mathcal{U}^*},$$

where the convergence is uniform.

Proof. $1 \Rightarrow 2$ is direct while $2 \Rightarrow 1$ is a consequence of the Itô-Nisio theorem [IN68]. Now let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent Gaussian standard random variables and note $X_n = \sum_{k=0}^{n} \xi_k u_k$. $1 \Rightarrow 4$ By hypothesis, $X_n \to Y$ almost-surely with Y a Gaussian random element identically distributed with X, thus $X_n \to Y$ in $L^2(\mathcal{U})$ and one has for all $r > 0$

$$\begin{aligned}\left|\sum_{k=0}^{n} \langle u_k, l\rangle_{\mathcal{U},\mathcal{U}^*} \langle u_k, g\rangle_{\mathcal{U},\mathcal{U}^*} - \langle \mathcal{C}_X l, g\rangle_{\mathcal{U},\mathcal{U}^*}\right| &= \left|\mathbb{E}\left[\langle X_n - Y, l\rangle_{\mathcal{U},\mathcal{U}^*} \langle X_n - Y, g\rangle_{\mathcal{U},\mathcal{U}^*}\right]\right|, \\ &\leq r^2 \mathbb{E}\left[\|X_n - Y\|_{\mathcal{U}}^2\right],\end{aligned}$$

with $(l, g) \in \mathcal{B}_{\mathcal{U}^*}(0, r]^2$. $4 \Rightarrow 3$ is direct. $3 \Rightarrow 2$ Finally,

$$\forall l \in \mathcal{U}^*,\ \mathbb{E}\left[\exp\left(i \langle X_n, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] = \exp\left(-\frac{1}{2}\sum_{k=0}^{n} \langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*}^2\right)$$

thus $\widehat{X_n}(l) \to \widehat{X}(l)$. □

As pointed out in [LP09], the last item in lemma 4 is a general version of Mercer's theorem [Mer09] and all convergences are unconditional. It appears that admissible sequences are completely characterized using covariance factorizations from lemma 2 [LP09].

Theorem 5. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element and $\mathcal{H}$ a separable Hilbert space with $(e_n)_{n\in\mathbb{N}}$ a basis. A sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is admissible for X if and only if there is an operator $S \in \mathcal{L}(\mathcal{H},\mathcal{U})$ such that $\mathcal{C}_X = SS^*$ and $\forall n \in \mathbb{N},\ Se_n = u_n$.*

Proof. Let $\mathcal{H}$ be a separable Hilbert space, $S \in \mathcal{L}(\mathcal{H},\mathcal{U})$ such that $\mathcal{C}_X = SS^*$ and $(e_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ a Hilbert basis. It comes that

$$\forall l \in \mathcal{U}^*,\ S^*l = \sum_{n\geq 0} \langle S^*l, e_n\rangle_{\mathcal{H}}\, e_n = \sum_{n\geq 0} \langle Se_n, l\rangle_{\mathcal{U},\mathcal{U}^*}\, e_n,$$

and applying S, it comes $\forall l \in \mathcal{U}^*,\ \mathcal{C}_X l = SS^*l = \sum_{n\geq 0} \langle Se_n, l\rangle_{\mathcal{U},\mathcal{U}^*} Se_n$. Using lemma 4, $(Se_n)_{n\in\mathbb{N}}$ is admissible. Conversely, let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ be an admissible sequence for X then $\forall (c_n)_{n\in\mathbb{N}} \in l^2(\mathbb{N})$, one has

$$\begin{aligned}\left\|\sum_{k=n}^{m} c_k u_k\right\|_{\mathcal{U}}^2 &\leq \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \left\langle \sum_{k=n}^{m} c_k u_k, l\right\rangle_{\mathcal{U},\mathcal{U}^*}^2, \\ &\leq \sum_{k=n}^{m} c_k^2 \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \sum_{i\geq 0} \langle u_i, l\rangle_{\mathcal{U},\mathcal{U}^*}^2, \\ &\leq \sum_{k=n}^{m} c_k^2 \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*}, \\ &\leq \|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} \sum_{k=n}^{m} c_k^2,\end{aligned}$$

using lemma 4, thus the sequence is Cauchy in $\mathcal{U}$ and converges. Since $(e_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ is a Hilbert basis, Parseval equality gives

$$\forall h \in \mathcal{H},\ \left(\langle h, e_n\rangle_{\mathcal{H}}\right)_{n\in\mathbb{N}} \in l^2(\mathbb{N}),$$

and $\sum_{n\geq 0} \langle h, e_n\rangle_{\mathcal{H}}\, u_n \in \mathcal{U}$. Define S as the following operator:

$$S := h \in \mathcal{H} \to \sum_{n\geq 0} \langle h, e_n\rangle_{\mathcal{H}}\, u_n,$$

which is linear. To see that it is bounded, write

$$\begin{aligned}
\|Sh\|_{\mathcal{U}}^2 &= \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \langle Sh, l\rangle_{\mathcal{U},\mathcal{U}^*}^2\,, \\
&\leq \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \left(\sum_{n\geq 0} \langle Sh, e_n\rangle_{\mathcal{H}}\, \langle u_n, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)^2\,, \\
&\leq \|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}\, \|h\|_{\mathcal{H}}^2\,,
\end{aligned}$$

and now $\forall n \in \mathbb{N},\ Se_n = u_n$. The proof is complete. □

In practice, one finds admissible sequences using covariance factorization and pre-existing Hilbert bases. In [LP09], the authors emphasize a more general notion of Parseval frames instead of Hilbert bases, allowing redundancy and the possibility to work with wavelets for instance.

Corollary 7. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with Cameron-Martin space $\mathcal{U}_X$, then all Hilbert bases $(h_n)_{n\in\mathbb{N}} \subset \mathcal{U}_X$ are admissible for X.*

Proof. Since the injection $i_{\mathcal{U}_X} : \mathcal{U}_X \to \mathcal{U}$ provides a valid factorization, it is a consequence of theorem 5. □

Among all possible bases in the Cameron-Martin space, there are some satisfying a strong summability condition, linked with the nuclearity of the covariance operator.

Proposition 11. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then there exists an admissible sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ such that*

$$\sum_{n\geq 0} \|u_n\|_{\mathcal{U}}^2 < +\infty.$$

Proof. The proof can be found in [Bog98]. □

An admissible sequence with this last property will be called a nuclear representation of X. The general concept of admissible sequences only provides a Gaussian random element with distribution target distribution μ_X. However, it is possible to find particular sequences such that they lie in the range of $\mathcal{C}_X$ (as it is dense in $\mathcal{U}_X$). In that case, the approximation is much stronger, since it holds almost-surely.

Proposition 12. *Let $\mathcal{U}$ be a Banach space, $X \sim \mathcal{N}(0, \mathcal{C}_X)$, $(l_n)_{n\in\mathbb{N}} \subset \mathcal{U}^*$ such that $(\mathcal{C}_X l_n)_{n\in\mathbb{N}}$ is admissible for X, then*

$$X = \sum_{n\geq 0} \langle X, l_n\rangle_{\mathcal{U},\mathcal{U}^*} \mathcal{C}_X l_n, \ a.s.$$

Proof. For all $n \in \mathbb{N}$, let $X_n = \sum_{k=0}^{n} \langle X, l_k\rangle_{\mathcal{U},\mathcal{U}^*} \mathcal{C}_X l_k$, then for every $l \in \mathcal{U}^*$, $\left(\langle X_n, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)_{n\in\mathbb{N}}$ converges to $\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*}$ in $L^2(\Omega;\mathbb{R})$ by lemma 4, thus in probability and Itô-Nisio theorem gives the conclusion. □

This (stronger) type of representations is known as stochastic bases in [Her81, Oka86].

Definition 8. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ and $(u_n^*)_{n\in\mathbb{N}} \subset \mathcal{U}^*$ then the system $(u_n, u_n^*)_{n\in\mathbb{N}}$ is a stochastic basis if*

- $\forall (n,m) \in \mathbb{N}^2, \ \langle u_n, u_m^*\rangle_{\mathcal{U},\mathcal{U}^*} = \delta_{nm}$,
- $X = \sum_{n\geq 0} \langle X, u_n^*\rangle_{\mathcal{U},\mathcal{U}^*} u_n, \ a.s.$

Optimality of representations Since there are multiple admissible sequences (and stochastic bases) to represent a Gaussian random element, it is natural to study the speed of convergence of such representations, in particular in applied contexts. Indeed, the practitioner will be often interested in finite approximations using an admissible sequence $(u_n)_{n\in\mathbb{N}}$ of the following form:

$$\forall n \in \mathbb{N}^*, \ X_n = \sum_{k=0}^{n} \xi_k u_k.$$

It is then of interest to look for optimal representations, provided a measure of convergence speed. The usual notions of distances (Hellinger, total-variation) between distributions μ_X and μ_{X_n} are not adapted (in fact not defined), since both measures are mutually singular. However, a first criterion could be a *uniform* approximation of the type (see [KL02] for a precise discussion):

$$\forall n \in \mathbb{N}^*, \ \mathbb{E}\left[\left\|\sum_{k\geq n-1} \xi_k u_k\right\|_{\mathcal{U}}^2\right]^{\frac{1}{2}}.$$

This leads to the important notion of l-numbers for a Gaussian random element.

Definition 9. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then its sequence of l-numbers is defined as $\forall n \in \mathbb{N}^*$:*

$$l_n(X) = \inf\left\{\mathbb{E}\left[\left\|\sum_{k\geq n-1} \xi_k u_k\right\|_{\mathcal{U}}^2\right]^{\frac{1}{2}}, \ (u_n)_{n\in\mathbb{N}} \text{ is admissible for } X\right\}.$$

In particular, these l-numbers where shown closely related to the metric entropy of the underlying probability measure [LLW99]. In fact, this notion can be linked to the covariance operator using factorizations.

Proposition 13. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, $\mathcal{H}$ a separable Hilbert space and $A \in \mathcal{L}(\mathcal{H},\mathcal{U})$ such that $\mathcal{C}_X = AA^*$ then $\forall n \in \mathbb{N}^*$:*

$$l_n(X) = \inf \left\{ \mathbb{E}\left[\left\| \sum_{k \geq n-1} \xi_k A e_k \right\|_{\mathcal{U}}^2 \right]^{\frac{1}{2}}, \ (e_n)_{n\in\mathbb{N}} \text{ Hilbert basis in } \mathcal{H} \right\},$$

where $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables.

However, it is not clear that there exists an admissible sequence $(u_n)_{n\in\mathbb{N}}$ such that

$$\forall n \in \mathbb{N}^*, \ l_n(X) = \mathbb{E}\left[\left\| \sum_{k \geq n-1} \xi_k u_k \right\|_{\mathcal{U}}^2 \right]^{\frac{1}{2}}.$$

In other words, the infimum may not be a minimum in the definition of l-numbers. Nevertheless, if one can at least estimate the sequence of l-numbers, there exists at least one admissible sequence achieving this speed of approximation.

Proposition 14. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, $\mathcal{H}$ a Hilbert space, $A \in \mathcal{L}(\mathcal{H},\mathcal{U})$ such that $\mathcal{C}_X = AA^*$, $\alpha > 0$ and $\beta \in \mathbb{R}$ then the following assertions are equivalent:*

1. *$\exists c_1 > 0, \ \forall n \in \mathbb{N}^*, \ l_n(X) \leq c\frac{(1+\log(n))^\beta}{n^\alpha}$,*
2. *there exists $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ an admissible sequence for X and $c_2 > 0$ such that*

$$\forall n \in \mathbb{N}^*, \ \mathbb{E}\left[\left\| \sum_{k \geq n-1} \xi_k u_k \right\|_{\mathcal{U}}^2 \right]^{\frac{1}{2}} \leq c_2 \frac{(1+\log(n))^\beta}{n^\alpha}.$$

3. *there exists a sequence independent Gaussian random elements $(X_k)_{k\in\mathbb{N}} \subset L^0(\mathcal{U})$, $c_3 > 0$ such that*

$$X = \sum_{k\geq 0} X_k,$$
$$\forall k \in \mathbb{N}, \ \operatorname{rank}(X_k) < 2^k,$$
$$\forall k \in \mathbb{N}, \ \mathbb{E}\left[\|X_k\|_{\mathcal{U}}^2 \right]^{\frac{1}{2}} \leq c_3 k^\beta 2^{-\alpha k}.$$

Proof. The proof can be found in [Pis89]. □

In view of proposition 14, we see that if one estimates the l-numbers for a particular factorization of the covariance operator, then there exists an admissible sequence with the same approximation error estimate. However, in some cases, it is also possible to bound the l-numbers below and when both estimates are similar, the decomposition is said asymptotically optimal.

Definition 10. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element then an admissible sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is asymptotically optimal for X if $\exists c_1, c_2 > 0$, $\forall n \in \mathbb{N}^*$,*

$$c_1 l_n(X) \leq \mathbb{E}\left[\left\|\sum_{k\geq n-1} \xi_k u_k\right\|_{\mathcal{U}}^2\right]^{\frac{1}{2}} \leq c_2 l_n(X).$$

The second notion of approximation of Gaussian random element is a *worst-case* type. It is based on the approximation of the covariance operator by finite dimensional operators and it leads to the notion of approximation numbers for Gaussian random elements.

Definition 11. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, the sequence of approximation numbers is defined as:*

$$\forall n \in \mathbb{N}^*,\ a_n(X) = \inf\left\{\|\mathcal{C}_X - S\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})},\ S \in \mathcal{L}(\mathcal{U}^*,\mathcal{U}),\ \operatorname{rank}(S) < n\right\}.$$

A first practical interest of these numbers is in particular their interpretation of linear reconstruction of Gaussian random elements.

Proposition 15 (Proposition 6.2 in [KL02])**.** *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, $n \in \mathbb{N}^*$ and $\epsilon > 0$ then $a_n(X) < \epsilon$ if and only if there exist $(l_0, ..., l_{n-1}) \in (\mathcal{U}^*)^n$ and $(\alpha_0, ..., \alpha_{n-1}) \in \mathbb{R}^n$ such that*

$$\forall l \in \mathcal{U}^*,\ \mathbb{E}\left[\left(\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*} - \sum_{k=0}^{n-1} \alpha_k \langle X, l_k\rangle_{\mathcal{U},\mathcal{U}^*}\right)^2\right] \leq \epsilon \|l\|_{\mathcal{U}^*}.$$

Proof. The proof is in [KL02]. □

The second interest in approximation numbers is the possibility to provide a lower bound to l-numbers.

Proposition 16. *Let $\mathcal{H}$ be a Hilbert space, $\mathcal{U}$ a Banach space and $A \in \mathcal{L}(\mathcal{H},\mathcal{U})$, then there exists $c > 0$,*

$$\forall m, n \geq 1,\ \sqrt{\log(m)}\, a_{n+m-1}(A) \leq c l_n(A).$$

1.5 Additional results in the Hilbert case

In this section, some important results are provided concerning the very specific case where $\mathcal{U}$ is a Hilbert space. Indeed, there are some important strengthenings of previous general concepts in this very special type of Banach spaces. Here, the dual space $\mathcal{U}^*$ will be identified with $\mathcal{U}$ using Riesz representation theorem. In particular, it implies that the covariance operator is an endomorphism and can be diagonalized by the spectral theorem (since it is compact and self-adjoint). Moreover, it is non-negative thus all the eigenvalues $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$ and the nuclearity translates into $(\lambda_n)_{n\in\mathbb{N}} \subset l^1(\mathbb{N})$.

Characterization of Gaussian covariance operators In the section 1.3, the covariance operator has been defined using Bochner's integral and shown to be non-negative, symmetric and nuclear. In Hilbert spaces, the converse is also true, that is every operator sharing these properties is the covariance of a Gaussian random element.

Theorem 6 (Prohorov)**.** *Let $\mathcal{U}$ be a separable Hilbert space and $\mathcal{C} \in \mathcal{L}(\mathcal{U},\mathcal{U})$, it is the covariance of a Gaussian random element in $\mathcal{U}$ if and only if it is symmetric, non-negative and nuclear.*

Proof. Let $\mathcal{U}$ be a separable Hilbert space and $X \sim \mathcal{N}(0, \mathcal{C}_X)$ a Gaussian random element. It has been already shown that its covariance operator is self-adjoint and non-negative. The nuclearity can be directly shown here. Indeed, let $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ be a Hilbert basis, then

$$\begin{aligned}\int_{\mathcal{U}} \|u\|_{\mathcal{U}}^2 \,\mu_X(du) &= \int_{\mathcal{U}} \sum_{n\geq 0} \langle u, u_n\rangle_{\mathcal{U}}^2 \,\mu_X(du) \\ &= \sum_{n\geq 0} \int_{\mathcal{U}} \langle u, u_n\rangle_{\mathcal{U}}^2 \,\mu_X(du) \\ &= \sum_{n\geq 0} \langle \mathcal{C}_X u_n, u_n\rangle_{\mathcal{U}}\end{aligned}$$

from Lebesgue's dominated convergence theorem, which implies a finite trace, thus $\mathcal{C}_X$ is nuclear. Now, let $\mathcal{C}_X \in \mathcal{L}(\mathcal{U},\mathcal{U})$ be a symmetric, non-negative and nuclear operator, the spectral theorem implies the existence of a Hilbert basis $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ and a sequence $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$ (by non-negativity) such that

$$\forall u \in \mathcal{U},\ \mathcal{C}_X u = \sum_{n\geq 0} \lambda_n \langle u, u_n\rangle_{\mathcal{U}} \, u_n,$$

and $(\lambda_n)_{n\in\mathbb{N}} \in l^1(\mathbb{N})$ (by nuclearity). Now, let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables (on the same probability space) and consider the following quantity:

$$\tilde{X} = \sum_{n\geq 0} \sqrt{\lambda_n}\xi_n u_n.$$

It comes that

$$\mathbb{E}\left[\left\|\tilde{X}\right\|_{\mathcal{U}}^2\right] = \mathbb{E}\left[\sum_{n\geq 0}\lambda_n\xi_n^2\right] = \sum_{n\geq 0}\lambda_n < +\infty,$$

using Lebesgue's dominated convergence theorem and thus $\tilde{X} \in L^2(\mathcal{U})$. Now it remains to see that it is a Gaussian random element with covariance $\mathcal{C}_X$. Again, Lebesgue's dominated convergence theorem gives

$$\forall u \in \mathcal{U},\ \mathbb{E}\left[e^{i\langle \tilde{X},u\rangle_{\mathcal{U}}}\right] = \prod_{n\geq 0}\mathbb{E}\left[e^{i\sqrt{\lambda_n}\langle u,u_n\rangle_{\mathcal{U}}\xi_n}\right] = \mathbb{E}\left[-\frac{1}{2}\langle \mathcal{C}_X u,u\rangle_{\mathcal{U}}\right].$$

In conclusion, $\mu_{\tilde{X}} = \mu_X$ by proposition 2. □

In general Banach spaces, the characterization of Gaussian covariance operators is still an open problem (except in l^p spaces, see [KT14] and the reference therein).

Karhunen-Loève representation In separable Hilbert spaces, the covariance operator admits a spectral representation. This eigendecomposition naturally leads to a Hilbert basis, which will be used to give an important admissible sequence, namely the Karhunen-Loève decomposition.

Proposition 17. *Let $\mathcal{U}$ be a separable Hilbert space, $X \in L^2(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ which eigendecomposition is $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$, $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ then*

1. $U_X = \overline{R\left(\mathcal{C}_X^{\frac{1}{2}}\right)}^{\mathcal{U}}$,
2. $(\sqrt{\lambda_n}u_n)_{n\in\mathbb{N}}$ *is a Hilbert basis in* U_X,
3. $\forall (h_1,h_2) \in \mathcal{U}_X,\ \langle h_1,h_2\rangle_X = \sum_{n\geq 0}\lambda_n^{-1}\langle h_1,u_n\rangle_{\mathcal{U}}\langle h_2,u_n\rangle_{\mathcal{U}}$.

Proof. Let $u \in \mathcal{U}$ and note $h = \mathcal{C}_X u$, then

$$\|h\|_X^2 = \langle h,h\rangle_X = \langle \mathcal{C}_X u,u\rangle_{\mathcal{U}} = \left\|\mathcal{C}_X^{\frac{1}{2}}u\right\|_{\mathcal{U}}^2$$

using the reproduction property. This isometry extends by density, giving item 1. Similarly, let $(u_n)_{n\in\mathbb{N}}$ be a spectral basis of $\mathcal{C}_X$, $u \in \mathcal{U}$ and $h = \mathcal{C}_X u$, then

$$\forall n \in \mathbb{N},\ \langle h,u_n\rangle_X = \langle u,u_n\rangle_{\mathcal{U}}.$$

This implies that if $\langle h,u_n\rangle_X = 0$ for all $n \in \mathbb{N}$, $h = 0$ thus $(u_n)_{n\in\mathbb{N}}$ is dense in $\mathcal{U}_X$. Finally,

$$\left\|\sqrt{\lambda_n}u_n\right\|_X^2 = \langle \lambda_n u_n,u_n\rangle_X = \langle u_n,u_n\rangle_{\mathcal{U}} = 1,$$

and similarly for $\lambda_n > 0$, $\left\langle \sqrt{\lambda_n}u_n,\sqrt{\lambda_p}u_p\right\rangle_X = \frac{\sqrt{\lambda_p}}{\sqrt{\lambda_n}}\langle u_n,u_p\rangle_{\mathcal{U}} = \delta_{np}$. The specific form of the inner product is deduced directly. □

Theorem 7. *Let $\mathcal{U}$ be a separable Hilbert space, X a Gaussian random element in $\mathcal{U}$ with covariance operator $\mathcal{C}_X$, then using previous notations, the sequence $(\sqrt{\lambda_n}u_n)_{n\in\mathbb{N}}$ is a Hilbert basis of $\mathcal{U}_X$ and one has*

$$X = \sum_{n\geq 0} \langle X, u_n\rangle_{\mathcal{U}}\, u_n \ a.s.$$

Proof. Since $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is a Hilbert basis, the result is clear. □

Besides of having independent Gaussian components, this decomposition enjoys two fundamental properties: a precise quantification of errors under truncation (low-rank approximation) and optimality in both trace and operator norms.

Theorem 8. *Let $\mathcal{U}$ be a Hilbert space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X \in \mathcal{L}(\mathcal{U},\mathcal{U})$. Let $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ and $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ respectively its eigenvalues (in decreasing order) and eigenvectors, then $\forall n \in \mathbb{N}^*$:*

- $l_n(X) = \left(\sum_{k\geq n} \lambda_k\right)^{\frac{1}{2}}$,
- $a_n(X) = \lambda_{n-1}^{\frac{1}{2}}$.

In other words, the problem of approximation of Gaussian random elements in Hilbert spaces is solved, the optimal basis being essentially unique and given by the (rescaled) eigenvectors. In finite dimensional spaces, this is also known as Eckart-Young-Mirksy theorem for low-rank approximation.

Feldman-Hajek theorem The second important result is that equivalence and singularity properties from Cameron-Martin theorem may be extended to measures with different covariance operators.

Theorem 9 (Feldman-Hajek)**.** *Let $\mathcal{U}$ a Hilbert space, $X_i \sim \mathcal{N}(m_i, \mathcal{C}_i),\ i \in \{1,2\}$ two Gaussian random elements, then:*

1. *The distributions μ_{X_1} and μ_{X_2} are equivalent if and only if:*
   - *(a)* $\mathcal{U}_{X_1} = \mathcal{U}_{X_2}$
   - *(b)* $m_1 - m_2 \in \mathcal{U}_{X_1}$
   - *(c)* $\left\| \left(\mathcal{C}_1^{-\frac{1}{2}}\mathcal{C}_2^{\frac{1}{2}}\right)\left(\mathcal{C}_1^{-\frac{1}{2}}\mathcal{C}_2^{\frac{1}{2}}\right)^* - \mathcal{I} \right\|_{HS} < \infty$
2. *They are mutually singular otherwise.*

Proof. The proof is given in [DZ14], theorem 2.25. □

Here, the Hilbert-Schmidt norm of an operator $A \in \mathcal{L}(\mathcal{U},\mathcal{U})$ is:

$$\|A\|_{HS} = \left(\sum_{n\geq 0} \|Ae_n\|_{\mathcal{U}}^2\right)^{\frac{1}{2}},$$

where $(e_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is a Hilbert basis (this quantity is invariant under change of bases).

Examples

Example 1 (Gaussian vectors)**.** *This first example shows how the previous general theory applies in an Euclidean space $\mathcal{U} = \mathbb{R}^n$ with $n \in \mathbb{N}^*$. Let $X \in L^0(\mathcal{U})$ be a Gaussian random element. In the canonical basis $(e_i)_{i\in[1,n]}$, the covariance operator is the usual matrix*

$$\forall (i,j) \in [1,n]^2,\ (\Sigma_X)_{i,j} = \mathbb{E}\left[\langle X, e_i\rangle_{\mathcal{U}} \langle X, e_j\rangle_{\mathcal{U}}\right].$$

Because this matrix is symmetric, one can find an eigendecomposition and represent it as a diagonal matrix with eigenvalues $(\lambda_1, ..., \lambda_n) \in [0, +\infty[^n$. The Cameron-Martin space is then equal to the span of eigenvectors with strictly positive eigenvalues, while the inner product is defined $\forall (\Sigma x, \Sigma y) \in \Sigma(\mathbb{R}^n),\ \langle \Sigma y, \Sigma x\rangle_X = \langle \Sigma y, x\rangle_{\mathcal{U}}$. In this very special case, the classical terminology is Gaussian random vectors.

Example 2 (Inverse Laplace covariance.)**.** *Let $\Omega =]0,1[$, $f \in L^2(\Omega;\mathbb{R})$ and consider the following equation:*

$$\begin{aligned} -\Delta u &= f \text{ on } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned}$$

Using standard variational analysis (Lax-Milgram theorem), one shows that a (unique) weak solution u exists in the Sobolev space $H_0^1(\Omega)$, defining a solution map $C := f \in L^2(\Omega;\mathbb{R}) \to u \in H_0^1(\Omega;\mathbb{R})$. This map is linear and bounded. Furthermore, since Ω is bounded itself, the Sobolev embedding $H_0^1(\Omega)$ in $L^2(\Omega)$ is compact and C is thus a compact operator. To see that it is self-adjoint, let $(f,g) \in L^2(\Omega)^2$ and $u = Cf$, $v = Cg$. It is clear that

$$\langle Cf, g\rangle_{L^2} = \langle u', v'\rangle_{L^2} = \langle f, Cg\rangle_{L^2},$$

and C is self-adjoint in L^2. By the spectral theorem, it admits a spectral representation and in this particular case, it is given $\forall n \in \mathbb{N}$:

$$\begin{aligned} Cf_n &= \lambda_n f_n, \\ \lambda_n &= \frac{1}{(n+1)^2\pi^2}, \\ f_n : &= x \in]0,1[\to \sqrt{2}\sin((n+1)\pi x) \in \mathbb{R}. \end{aligned}$$

It is thus clearly trace-class, and the following random series

$$X = \sum_{n\geq 0} \sqrt{\lambda_n}\xi_n f_n,$$

with $(\xi_n)_{n\in\mathbb{N}}$ a sequence of independent standard normal variates defines a Gaussian random element in $\mathcal{U} = L^2(\Omega,\mathbb{R})$ with covariance operator $\mathcal{C}$ (known as the Brownian bridge). The Cameron-Martin space can be identified with the subspace $H_0^1(\Omega)$, $\forall f \in L^2(\Omega)$, note $u = Cf$, then

$$\|u\|_X^2 = \langle u, u\rangle_X = \langle u, f\rangle_{L^2} = \langle u', u'\rangle_{L^2} = \|u\|_{H_0^1}^2.$$

1.6 Gaussian random elements in $C(\mathcal{K}, \mathbb{R})$

In this section, the presentation focuses on the particular case of continuous functions over a compact metric set, that is $\mathcal{U} = C(\mathcal{K}, \mathbb{R})$ from now on. First of all, it is a Banach space when the supremum norm is considered and it is separable (as a consequence of Stone-Weierstrass theorem). Moreover, the dual space $\mathcal{U}^*$ is identified with the space of finite, signed measures on $\mathcal{K}$ equipped with the Borel σ-algebra (Riesz-Markov theorem, see [Rud09]). In this context, the theory of continuous Gaussian random fields will provide an alternative viewpoint on the previous material. Indeed, they can be seen as Gaussian random elements of $L^0(\mathcal{U})$.

Continuous Gaussian random fields The theory of random fields is a practical tool to define probability measures over function spaces [AT07]. Indeed, let $\mathcal{K}$ be a set and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, then a family $\{X_s\}_{s\in\mathcal{K}}$ of real random variables, all defined on the same probability space, maps Ω to $\mathbb{R}^{\mathcal{K}}$ (the set of all maps from $\mathcal{K}$ to $\mathbb{R}$).

Definition 12 (Random field)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, $\mathcal{K}$ a set, then a random field on $\mathcal{K}$ is a family of real random variables $(X_s)_{s\in\mathcal{K}}$ defined on the same probability space.*

This formally defines a measurable mapping X between $(\Omega, \mathcal{F})$ and $\mathbb{R}^{\mathcal{K}}$ equipped with the cylindrical σ-algebra $Cyl(\mathcal{K})$ (smallest such that all point-wise evaluations are measurable). Let $\omega \in \Omega$ such that the previous map is well-defined, then $X(\omega) := (X_s(\omega))_{s\in\mathcal{K}}$ is called a sample function. From now on, we will always suppose that a unique probability space is given, and will not mention it any more. Now, given a random field on $(\mathcal{K}, d)$, one can consider an associated probability measure, its distribution.

Definition 13 (Distribution of a random field)**.** *Let $\mathcal{K}$ a set and $X = (X_s)_{s\in\mathcal{K}}$ a random field, its distribution is defined as:*

$$\mu_X := A \in Cyl(\mathcal{K}) \to \mathbb{P}\left(X^{-1}(A)\right) \in [0, 1].$$

The particular nature of the Cylindrical σ-algebra gives the possibility to consider the previous distribution on cylinders, providing a family of finite distributions.

Definition 14 (Finite distribution)**.** *Let $\mathcal{K}$ a set and $X = (X_s)_{s\in\mathcal{K}}$ a random field, then for all $n \in \mathbb{N}$ and all $(s_0, ..., s_n) \in \mathcal{K}^{n+1}$, the following map is a probability measure on $(\mathbb{R}^{n+1}, Bor(\mathbb{R}^{n+1}))$:*

$$\mu_X^{(s_0,...,s_n)} := A \in Bor(\mathbb{R}^{n+1}) \to \mathbb{P}\left((X_{s_0}, ..., X_{s_n}) \in A\right) \in [0, 1].$$

Obviously, the distribution of a random field uniquely defines a family of finite distributions. However, it is possible to build a distribution on $Cyl(\mathcal{K})$ by specifying a family of finite distributions, provided they are consistent (Kolmogorov extension theorem). Now, since a random field is a family of random variables, one can consider important mappings, generalizing the usual notion of moments.

Definition 15. *Let $\mathcal{K}$ a set and $X = (X_s)_{s\in\mathcal{K}}$ a random field on $\mathcal{K}$:*

- *if $\forall s \in \mathcal{K},\ X_s \in L^1(\mathbb{R})$, then the map*

$$m_X := s \in \mathcal{K} \to \mathbb{E}[X_s] \in \mathbb{R}$$

is the mean function of X,

- *if $\forall s \in \mathcal{K},\ X_s \in L^2(\mathbb{R})$, then the map*

$$k_X := (s,t) \in \mathcal{K}^2 \to \mathbb{E}[(X_s - \mathbb{E}[X_s])(X_t - \mathbb{E}[X_t])] \in \mathbb{R}$$

is the covariance kernel of X.

A random field with a mean function is said to be integrable, and if additionally it has a covariance kernel, it is square integrable. A random field with trivial mean function will be called centred, and again, only this case will be considered here such that

$$\forall (s,t) \in \mathcal{K}^2,\ k_X(s,t) = \mathbb{E}[X_s X_t].$$

An important consequence of the definition is that a covariance kernel is symmetric and semi-positive definite.

Proposition 18. *Let $\mathcal{K}$ a set and $X = (X_s)_{s\in\mathcal{K}}$ a square integrable random field on $\mathcal{K}$, then its covariance kernel k_X has the following properties:*

- *Symmetry: $\forall (s,t) \in \mathcal{K}^2,\ k_X(s,t) = k_X(t,s)$,*
- *Semi-positive definiteness:*

$$\forall n \in \mathbb{N},\ \forall (s_1, ...s_n) \in \mathcal{K}^n,\ \forall (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n,\ \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j k_X(s_i, s_j) \geq 0.$$

The terminology of semi-positive definite kernel is usual in the literature of covariance kernels, while non-negativity is used in functional analysis for operators. Now, the definition of Gaussian random field is based on all its finite distributions, as follows.

Definition 16 (Gaussian random field)**.** *Let $\mathcal{K}$ be a set and $(X_s)_{s\in\mathcal{K}}$ a random field on $\mathcal{K}$, it is Gaussian if $\forall n \in \mathbb{N},\ \forall (s_0, ..., s_n) \in \mathcal{K}^{n+1},\ (X_{s_0}, ..., X_{s_n})$ is a Gaussian vector on $(\mathbb{R}^{n+1}, Bor(\mathbb{R}^{n+1}))$.*

One particularity of such field is that they are square integrable, thus the mean function and covariance kernel are well-defined. As a consequence, all finite distributions can be expressed using these two notions. Indeed, let $n \in \mathbb{N}$, $\underline{s} = (s_0, ..., s_n)$ then

$$(X_{s_0}, ..., X_{s_n}) \sim \mathcal{N}(0, k_X(\underline{s}, \underline{s}))$$

where $k_X(\underline{s}, \underline{s})$ is a square matrix of dimension $n+1$, such that $(k_X(\underline{s}, \underline{s}))_{(i,j)} = k_X(s_i, s_j)$. Since all finite distributions are parametrized by the covariance kernel

(the mean function being taken null) and as it bears the necessary consistency, it provides a characterization of the random fields' law. It is also important to mention that given a symmetric, semi-positive definite kernel, one can build a Gaussian process from which it is the covariance.

Example 3. *Let $\mathcal{K} \subset \mathbb{R}^n$, $n \geq 1$, then the following applications are valid covariance kernels:*

- *Squared-exponential kernel:*

$$\forall (s,t) \in \mathcal{K}^2,\ k(s,t) = \exp\left(-d(s,t)^2\right),$$

- *Exponential kernel:*

$$\forall (s,t) \in \mathcal{K}^2,\ k(s,t) = \exp\left(-d(s,t)\right),$$

- *Matérn kernel:*

$$\forall (s,t) \in \mathcal{K}^2,\ k(s,t) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} d(s,t)\right)^{\nu} K_\nu \left(\sqrt{2\nu} d(s,t)\right),$$

  where Γ, K_ν are respectively the gamma and modified Bessel functions and $\nu > 0$.

There exists a large family of well-known kernels in the literature, one could consult [RW04, BTA04] for instance. In the special case where the compact set $\mathcal{K}$ is a Cartesian product, it is possible to define covariance kernels by tensorization.

Proposition 19. *Let $\mathcal{K}_1$, $\mathcal{K}_2$ be two sets, $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$, k_1 and k_2 two symmetric, semi-positive definite kernels on $\mathcal{K}_1$, $\mathcal{K}_2$ respectively, then*

$$\forall (s,t) \in \mathcal{K}^2,\ k(s,t) = k_1(s_1,t_1) k_2(s_2,t_2)$$

is a symmetric, semi-positive definite kernel on $\mathcal{K}$.

Example 4. *Let $\mathcal{K} = [0,1]$ and consider the Wiener process kernel $\forall (s,t) \in [0,1]^2,\ k_W(s,t) = \min(s,t)$, then the Brownian sheet is defined as the centred Gaussian process with the following covariance kernel*

$$\forall (s_1,t_1,s_2,t_2) \in [0,1]^4,\ k\left((s_1,t_1),(s_2,t_2)\right) = k_W(s_1,t_1) k_W(s_2,t_2).$$

Now, concerning the associated sample functions, their continuity (as well as other types of regularity) has been widely studied in the literature and [AT07] provides a detailed introduction. The analysis of general necessary conditions ensuring this continuity is beyond the scope of this work, but there is a practical criterion when $\mathcal{K}$ is a subset of the Euclidean space.

Theorem 10 (Kolmogorov continuity test)**.** *Let $n \in \mathbb{N}^*$, $\mathcal{K} \subset \mathbb{R}^n$, $X = (X_s)_{s\in\mathcal{K}}$ a random field on $\mathcal{K}$, if*

$$\forall (s,t) \in \mathcal{K}^2,\ \mathbb{E}\left[|X_s - X_t|^p\right] \leq C \left\|s-t\right\|_{\mathbb{R}^n}^{p\alpha+n}$$

where $p > 1$, $\alpha > 0$ are constants, then there exists a version $\tilde{X}$ of X with almost-surely β-Hölder continuous sample functions, for all $0 < \beta < \alpha$.

Example 5. *Let $\mathcal{K} = [0,1]$ and consider $W = (W_t)_{t\in[0,1]}$ the standard Wiener process, that is $W_0 = 0$ almost-surely and $k_W(s,t) = \min(s,t)$. One important property is that increments are normally distributed with a proportional variance $\forall (s,t) \in [0,1]^2,\ \mathbb{E}\left[(W_t - W_s)^2\right] = |t-s|$, thus*

$$\forall (s,t) \in [0,1],\ s \neq t,\ \mathbb{E}\left[\left(\frac{W_s - W_t}{\sqrt{|s-t|}}\right)^4\right] = 3,$$

and one can apply theorem 10 with $C = 3$, $p = 4$ and $\alpha = \frac{1}{4}$. In other words, there exists a modification of the standard Wiener process with β-Hölder continuous sample functions, provided $0 < \beta < \frac{1}{4}$.

It appears that continuous Gaussian random fields have necessarily a continuous covariance kernel, but the converse needs not be true.

Proposition 20. *Let $(\mathcal{K}, d)$ be a compact metric space and $X = (X_s)_{s\in\mathcal{K}}$ a Gaussian random field with almost-surely continuous sample functions, then k_X is continuous.*

General Gaussian random fields on a set $\mathcal{K}$ are initially defined as measurable mappings from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^{\mathcal{K}}, Cyl(\mathcal{K}))$. In the case of almost-sure continuity of sample functions, the random field takes its values in $C(\mathcal{K},\mathbb{R})$, where both Borel and Cylindrical σ-algebras are equal when $(\mathcal{K}, d)$ is a compact metric set. In this case, the random field may then be viewed as a measurable map from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{U}, Bor(\mathcal{U}))$, and one can show that it is a Gaussian random element (in the sense of definition 3). Consequently, both concepts of covariance kernel and operator are linked in the following sense.

Proposition 21. *Let $(\mathcal{K}, d)$ be a compact metric space, $X = (X_s)_{s\in\mathcal{K}}$ a continuous Gaussian random field on $\mathcal{K}$ with covariance kernel k_X, then it is a Gaussian random element $\mathcal{U} = C(\mathcal{K},\mathbb{R})$ with covariance operator:*

$$\mathcal{C}_X : \mu \in \mathcal{U}^* \to \left(s \in \mathcal{K} \to \int_{\mathcal{K}} k_X(s,t)\mu(dt)\right) \in \mathcal{U}.$$

Conversely, a Gaussian random element $X \in L^0(\mathcal{U})$ has a covariance kernel defined as follows:

$$\forall (s,t) \in \mathcal{K}^2,\ k_X(s,t) = \left\langle \mathcal{C}_X \delta_s, \delta_t \right\rangle_{\mathcal{U},\mathcal{U}^*}.$$

Proof. Using the definition, one has $\forall(\mu,\nu)\in\mathcal{U}^*\times\mathcal{U}^*$:

$$\begin{aligned}\langle\mathcal{C}_X\mu,\nu\rangle_{\mathcal{U},\mathcal{U}^*} &= \mathbb{E}\left[\int_{\mathcal{K}} X_t\mu(dt)\int_{\mathcal{K}} X_s\nu(ds)\right],\\ &= \mathbb{E}\left[\int_{\mathcal{K}}\int_{\mathcal{K}} X_sX_t\mu(dt)\nu(ds)\right],\\ &= \int_{\mathcal{K}}\int_{\mathcal{K}}\mathbb{E}\left[X_sX_t\right]\mu(dt)\nu(ds),\\ &= \int_{\mathcal{K}}\int_{\mathcal{K}} k_X(s,t)\mu(dt)\nu(ds),\end{aligned}$$

from which the covariance operator is identified. Conversely, one has

$$\forall(s,t)\in\mathcal{K}^2,\ k_X(s,t)=\mathbb{E}\left[X_sX_t\right]=\langle\mathcal{C}_X\delta_s,\delta_t\rangle_{\mathcal{U},\mathcal{U}^*}.$$

□

Reproducing Kernel Hilbert Space (RKHS) Given a symmetric, semi-definite kernel (or equivalently a covariance kernel), one can build an important Hilbert subspace of $\mathbb{R}^{\mathcal{K}}$, construction similar in many senses with the Cameron-Martin space of Gaussian random elements (see [Aro50, BTA04, VV08]).

Proposition 22. *Let $\mathcal{K}$ a set and k a symmetric, semi-positive definite kernel and the following vector space*

$$\mathcal{H}_0 := \operatorname{span}\left\{k(s,.),\ s\in\mathcal{K}\right\}.$$

Let $(f,g)\in\mathcal{H}_0^2$, such that

$$\begin{aligned}f &= \sum_{i=0}^{n}\alpha_i k_X(.,s_i),\\ g &= \sum_{j=0}^{m}\beta_j k_X(.,t_j),\end{aligned}$$

with $n,m\in\mathbb{N}$, $(\alpha_0,...,\alpha_n,\beta_0,...,\beta_m)\in\mathbb{R}^{m+n+2}$ and $(s_0,...,s_n,t_0,...,t_m)\in\mathcal{K}^{n+m+2}$, then the following map:

$$(f,g)\in\mathcal{H}_0^2\to\langle f,g\rangle_k=\sum_{i=0}^{n}\sum_{j=0}^{m}\alpha_i\beta_j k_X(s_i,t_j)\in\mathbb{R},$$

is well-defined and an inner product. Moreover, it satisfies the reproduction property:

$$\forall f\in\mathcal{H}_0,\ \forall s\in\mathcal{K},\ f(s)=\langle f,k(.,t)\rangle_k.$$

Proof. Let $f\in\mathcal{H}_0$ and $s\in\mathcal{K}$, then:

$$\langle f,k(.,s)\rangle_k=\sum_{i=0}^{n}\alpha_i k(s_i,s)=f(s),$$

thus the application is well-defined (it does not depend on the particular representation of vectors), bilinear, symmetric and non-negative. To see that it is definite, it remains to use Cauchy-Schwarz inequality:

$$\forall s \in \mathcal{K},\ |f(s)| = |\langle f, k(.,s)\rangle_k| \leq \|f\|_k \sqrt{k(s,s)},$$

thus $\|f\|_k = 0 \Rightarrow f = 0$. □

Proposition 23 (Reproducing Kernel Hilbert Space)**.** *Let $\mathcal{K}$ a set and k a symmetric and semi-positive definite kernel, then the space $\mathcal{H}_0 = \operatorname{span}\{k(.,s),\ s \in \mathcal{K}\}$ is continuously embedded in $\mathbb{R}^{\mathcal{K}}$. The associated Reproducing Kernel Hilbert Space is its topological completion in $\mathbb{R}^{\mathcal{K}}$:*

$$\mathcal{H}_k = \widehat{\mathcal{H}_0}^{\langle .,. \rangle_k} \subset \mathbb{R}^{\mathcal{K}}.$$

Moreover, it satisfies the reproduction property:

$$\forall f \in \mathcal{H}_k,\ \forall s \in \mathcal{K},\ \langle f, k(s,.)\rangle_k = f(s).$$

The notion of Gaussian space may be defined as well, using the embedding $s \in \mathcal{K} \to X_s \in L^2(\Omega; \mathbb{R})$ and it is isometric (Loève isometry) with the Reproducing Kernel Hilbert space. When the kernel is continuous, the embedding will be in $C(\mathcal{K}, \mathbb{R})$.

Proposition 24. *Let $(\mathcal{K}, d)$ be a compact metric space, k a symmetric, positive-definite and continuous kernel, then $\mathcal{H}_k \subset C(\mathcal{K}, \mathbb{R})$.*

Proof. Let $f \in \mathcal{H}_k$, then for all $(s,t) \in \mathcal{K}^2$, it comes:

$$(f(s) - f(t))^2 = \langle f, k(s,.) - k(t,.)\rangle_k^2 \leq \|f\|_k^2 \|k(s,.) - k(t,.)\|_k^2 .$$

Now, $\|k(s,.) - k(t,.)\|_k^2 = k(s,s) + k(t,t) - 2k(s,t)$ and since k is continuous, $\lim_{s \to t} \|k(s,.) - k(t,.)\|_k^2 = 0$. □

Corollary 8. *Let $(\mathcal{K}, d)$ be a compact metric space, k a symmetric, semi-positive definite and continuous kernel, then $\mathcal{H}_k$ is separable.*

Concerning continuous Gaussian random fields on a compact metric space, both definitions of RKHS and Cameron-Martin space are well-defined and lead to the very same notion (Theorem 2.1 in [VV08]), but it needs not be true in general.

Series representation of continuous Gaussian random fields The previous Reproducing Kernel Hilbert space is useful in the analysis of Gaussian random fields. In particular, when it is separable, one can use Hilbert bases to give series representations of the underlying random field.

Proposition 25. *Let $(\mathcal{K}, d)$ be a compact metric space, $X = (X_s)_{s\in\mathcal{K}}$ a Gaussian random field with continuous covariance kernel k_X and $(h_n)_{n\in\mathbb{N}} \in \mathcal{H}_k$ a Hilbert basis, then*

$$\forall s \in \mathcal{K},\ X_s = \sum_{n\geq 0} \xi_n h_n(s),\ a.s.$$

with $(\xi_n)_{n\in\mathbb{N}}$ a sequence of independent standard Gaussian random variables and the convergence is in $L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$.

Proof. Since k_X is continuous, one has $\mathcal{H}_{k_X}$ separable and since it is isometric to the Gaussian space, the latter is separable as well. Let $s \in \mathcal{K}$ and $(h_n^*)_{n\in\mathbb{N}}$ a Hilbert basis in the Gaussian space, then

$$X_s = \sum_{n\geq 0} \mathbb{E}\left[X_s h_n^*\right] h_n^*,$$

and using the Loève isometry $\mathbb{E}\left[X(s)h_n^*\right] = \langle k_X(s,.), h_n\rangle_{k_X} = h_n(s)$, thus the result follows. □

However, the very nature of continuous Gaussian random fields can be use to significantly increase the convergence, passing from $L^2(\Omega, \mathbb{R})$ to $C(\mathcal{K}, \mathbb{R})$.

Theorem 11. *Let $(\mathcal{K}, d)$ be a compact metric space, $X = (X_s)_{s\in\mathcal{K}}$ a continuous Gaussian random field with covariance kernel k_X and $(h_n)_{n\in\mathbb{N}} \subset \mathcal{H}_k$ a Hilbert basis, then*

$$\forall s \in \mathcal{K},\ X_s = \sum_{n\geq 0} \xi_n h_n(s),\ a.s.$$

with $(\xi_n)_{n\in\mathbb{N}}$ a sequence of i.i.d $\mathcal{N}(0,1)$ random variables, the convergence being in $C(\mathcal{K}, \mathbb{R})$.

Proof. Let $s \in \mathcal{K}$ and $(h_n)_{n\in\mathbb{N}} \subset \mathcal{H}_k$ a Hilbert basis, then one has

$$k_X(s,.) = \sum_{n\geq 0} \langle k_X(s,.), h_n\rangle_{k_X} h_n = \sum_{n\geq 0} h_n(s) h_n,$$

by the reproduction property and it implies

$$\forall (s,t) \in \mathcal{K}^2,\ k_X(s,t) = \sum_{n\geq 0} h_n(s) h_n(t).$$

In particular, taking $t = s$ gives

$$\forall s \in \mathcal{K},\ k_X(s,s) = \sum_{n\geq 0} h_n(s)^2,$$

the convergence being point-wise and monotonous, Dini's theorem establishes the uniform convergence. Consider X as a Gaussian random element and note $X_n =$

$\sum_{k=0}^{n-1}\xi_k h_k$.

$$\begin{aligned}\forall \mu\in\mathcal{U}^*,\ \mathbb{E}\left[|\,\langle X_n,\mu\rangle-\langle X,\mu\rangle\,|\right] &= \mathbb{E}\left[\left|\int X_n(s)-X(s)\mu(ds)\right|\right],\\ &\leq \mathbb{E}\left[\int |X_n(s)-X(s)|\,|\mu|(ds)\right],\\ &\leq \int \mathbb{E}\left[|X_n(s)-X(s)|\right]|\mu|(ds),\\ &\leq \int \mathbb{E}\left[(X_n(s)-X(s))^2\right]^{\frac{1}{2}}|\mu|(ds),\\ &\leq \int \left(\sum_{k\geq n} h_k(s)^2\right)^{\frac{1}{2}}|\mu|(ds).\end{aligned}$$

Since it is the rest of a uniformly convergent series, the limit of the right hand side in last inequality tends to 0. This implies that $\langle X_n,\mu\rangle\to\langle X,\mu\rangle$ in $L^1(\Omega,\mathcal{F},\mathbb{P};\mathbb{R})$ thus in probability. Applying Itô-Nisio theorem, it follows that $X_n\to X$ almost-surely. □

Similarly to Gaussian random elements and their covariance factorization, Hilbert bases in the Cameron-Martin space can be identified by decomposing the covariance kernel.

Proposition 26. *Let $(\mathcal{K},d)$ be a compact metric space, k a continuous, symmetric, positive-definite kernel, $(\lambda_n)\subset\mathbb{R}^+$ and $(f_n)_{n\in\mathbb{N}}\subset C(\mathcal{K},\mathbb{R})$ such that*

$$\forall (s,t)\in\mathcal{K}^2,\ k(s,t)=\sum_{n\geq 0}\lambda_n f_n(s)f_n(t),$$

where the convergence is uniform then $(\sqrt{\lambda_n}f_n)_{n\in\mathbb{N}}$ is a Hilbert basis in $\mathcal{H}_k$.

However, it is a notoriously difficult problem to find such decompositions in general and the following section provides a canonical example of such method. Note that there are numerous recent contributions in this field (see [Git12, Zha17, BCM18] for instance).

Karhunen-Loève decomposition in $L^2(\mathcal{K},\mu)$ The Karhunen-Loève decomposition is a very important result in the theory and applications of random fields, since it gives a *practical* method to obtain bases in the Reproducing Kernel Hilbert space. It is based on a spectral representation of the covariance operator with sample functions in $L^2(\mathcal{K},\mu;\mathbb{R})$ where μ is a finite Borel measure on $\mathcal{K}$.

Proposition 27. *Let $(\mathcal{K},d)$ be a compact metric space, k a continuous, symmetric and semi-positive definite kernel and μ a finite, Borel measure on $\mathcal{K}$ then*

$$C_k : f\in L^2(\mathcal{K},\mu;\mathbb{R})\to\int_{\mathcal{K}} k(.,t)f(t)\mu(dt)\in L^2(\mathcal{K},\mu;\mathbb{R}),$$

is well-defined, self-adjoint and compact.

Proof. Let $f \in L^2(\mathcal{K}, \mu; \mathbb{R})$, then

$$\|C_k f\|_{L^2}^2 = \int_{\mathcal{K}} \left(\int_{\mathcal{K}} k(s,t) f(t) \mu(dt) \right)^2 \mu(ds) \leq \int_{\mathcal{K}} \int_{\mathcal{K}} k(s,t)^2 |\mu|(ds)|\mu|(dt) \, \|f\|_{L^2}^2 .$$

Since k is continuous and $\mathcal{K}$ compact, it is clear that

$$\int_{\mathcal{K}} \int_{\mathcal{K}} k(s,t)^2 |\mu|(ds)|\mu|(dt) < +\infty,$$

thus $C_k f \in L^2$ and the operator C_k is well-defined. The linearity follows from the linearity property of Lebesgue's integral. The operator is self-adjoint because:

$$\begin{aligned} \forall (f,g) \in L^2(\mathcal{K}, \mu; \mathbb{R})^2, \ \langle C_k f, g \rangle_{L^2} &= \int_{\mathcal{K}} \int_{\mathcal{K}} k(s,t) f(t) \mu(dt) g(s) \mu(ds), \\ &= \int_{\mathcal{K}} \int_{\mathcal{K}} k(s,t) g(s) \mu(ds) f(t) \mu(dt), \\ &= \langle f, C_k g \rangle_{L^2} \end{aligned}$$

the exchange of integral being possible thanks to Fubini's theorem. Now, let $f \in L^2(\mu)$, then

$$\begin{aligned} |C_k f(s) - C_k f(t)| &= \left| \int_{\mathcal{K}} (k(s,v) - k(t,v)) f(v) \mu(dv) \right|, \\ &\leq \int_{\mathcal{K}} |(k(s,v) - k(t,v)) f(v)| \, |\mu|(dv), \\ &\leq \|f\|_{L^2} \, \|k(s,.) - k(t,.)\|_{L^2} . \end{aligned}$$

Since k is continuous, $\lim_{s \to t} \|k(s,.) - k(t,.)\|_{L^2} = 0$ and $C_k f$ is continuous as well. Moreover,

$$\forall s \in \mathcal{K}, \ |C_k f(s)| \leq \sup_{t \in \mathcal{K}} |k(s,t)| \, \|f\|_{L^2}$$

thus taking the supremum, we have $C_k(\mathcal{B}_{L^2(\mu)})$ bounded. It remains to see that $C_k(\mathcal{B}_{L^2(\mu)})$ is equicontinuous and conclude. Since $\mathcal{K}$ is compact and k continuous, it is uniformly continuous, thus $\forall \epsilon > 0, \ \exists \delta > 0, \ \forall (s_1, s_2, t_1, t_2) \in \mathcal{K}^4, \ d(s_1, s_2) + d(t_1, t_2) \leq \delta \Rightarrow |k(s_1, t_1) - k(s_2, t_2)| \leq \epsilon$. Let $\epsilon > 0$ and take $d(s,t) \leq \delta$ then

$$|C_k f(s) - C_k f(t)| \leq \int_{\mathcal{K}} \epsilon |f(v)| |\mu|(dv) \leq \epsilon \, \|f\|_{L^2} \leq \epsilon.$$

Now, since $C_k(\mathcal{B}_{L^2})$ is bounded and equicontinuous, it is relatively compact by Arzela-Ascoli theorem and the proof is complete. □

Now, the Karhunen-Loève decomposition uses the spectral theory in Hilbert spaces, as it was presented in section 1.5.

Theorem 12 (Mercer's)**.** *Let $(\mathcal{K}, d)$ be a compact metric space, k a continuous, symmetric and semi-positive definite kernel and μ a finite, Borel measure on $\mathcal{K}$, $(\lambda_n, f_n)_{n \in \mathbb{N}}$ the spectral decomposition of C_k, then:*

- $\lambda_n \geq 0, \ \forall n \in \mathbb{N}$,

- $f_n \in C(\mathcal{K}, \mathbb{R}),\ \forall n \in \mathbb{N}$,
- k *admits the following expansion:*

$$\forall (s,t) \in \mathcal{K}^2,\ k(s,t) = \sum_{n \geq 0} \lambda_n f_n(s) f_n(t),$$

where the convergence is absolute and uniform on $\mathcal{K} \times \mathcal{K}$.

In other words, the problem of finding series representation of Gaussian random fields with values in $L^2(\mathcal{K}, \mu; \mathbb{R})$ is instantiated in an eigendecomposition problem. The above examples are taken directly from [Wan08] in the case where $\mathcal{K} = [0,1]$.

Example 6 (Wiener process)**.** *Let* $\mathcal{K} = [0,1]$ *and consider the standard Wiener process* $W = (W_s)_{s \in [0,1]}$ *with covariance kernel* $\forall (s,t),\ k_W(s,t) = \min(s,t)$ *and* $W_0 = 0$ *almost-surely. Since it is a continuous Gaussian random field on a compact metric space, the sample functions are square integrable w.r.t. Lebesgue's measure. The eigendecomposition of the following covariance operator:*

$$C_W := f \in L^2([0,1]) \to \int_0^1 k_W(s,t) f(t) dt \in L^2([0,1]).$$

is given by $\forall n \in \mathbb{N}$ *solving an ordinary differential equation:*

$$\lambda_n = \frac{1}{\left(n + \frac{1}{2}\right)^2 \pi^2},$$
$$\forall s \in [0,1],\ f_n(s) = \sqrt{2} \sin\left(\left(n + \frac{1}{2}\right) \pi s\right).$$

Example 7 (Brownian bridge)**.** *Let* $\mathcal{K} = [0,1]$ *and consider the previous Wiener process* $W = (W_s)_{s \in [0,1]}$*, the Brownian bridge is defined as the Gaussian process* $B = (B_s)_{s \in [0,1]} = (W_s - W_1 s)_{s \in [0,1]}$*. This process corresponds to the conditional Wiener process given* $W_1 = 0$*. The spectral decomposition of the covariance operator gives the following elements:*

$$\lambda_n = \frac{1}{(n+1)^2 \pi^2},$$
$$\forall s \in [0,1],\ f_n(s) = \sqrt{2} \sin\left((n+1) \pi s\right).$$

Example 8 (Ornstein-Uhlenbeck process)**.** *Consider now the Ornstein-Uhlenbeck process* $Z = (Z_s)_{s \in [0,1]}$ *on the time interval* $[0,1]$ *defined as the stationary solution of the following stochastic differential equation:*

$$dZ_t = -\beta Z_t dt + \sigma dW_t,\ \beta > 0,\ \sigma > 0. \tag{1.1}$$

One can show that the solution is a centred, continuous Gaussian random field with covariance kernel

$$K(t,s) = Cov(Z_t, Z_s) = \frac{\sigma^2}{2\beta} e^{-\beta |t-s|}.$$

This kernel is also known as the exponential covariance kernel or Matérn covariance kernel of order $\nu = \frac{1}{2}$*. The associated eigendecomposition is given as* $\forall n \in \mathbb{N}$*,*

$$\lambda_n = \frac{2\beta}{w_n^2 + \beta^2},$$
$$\forall s \in [0,1],\ f_n(s) = \sqrt{\frac{2w_n^2}{2\beta + w_n^2 + \beta^2}} \cos(w_n s) + \sqrt{\frac{2\beta^2}{2\beta + w_n^2 + \beta^2}} \sin(w_n s),$$

where $\forall n \in \mathbb{N}$*,* w_n *is a solution to the following equation:*

$$\left(w^2 - \beta^2\right) \sin(w) = 2\beta w \cos(w).$$

Remark from section 1.5, that this decomposition is optimal in $L^2(\mathcal{K}, \mu; \mathbb{R})$. However, it also provides a series representation in $C(\mathcal{K}, \mathbb{R})$ when the random field is continuous.

Optimality in $C(\mathcal{K}, \mathbb{R})$ The search for optimal series representation of continuous Gaussian random fields over $\mathcal{K} = [0,1]$ is a very active field of research, in particular for the fractional Brownian motion [AT03, DZ04b, DvZ05, Igl05, Nda16]. The results presented here are directly taken from [KL02, LP09] an give practical methods to estimate the l-numbers associated to a continuous Gaussian random field using both covariance kernel and operator. The first result provides an upper bound for l-numbers for continuous Gaussian random fields in $\mathcal{K} = [0,1]$, using the Faber-Schauder basis.

Proposition 28 (Proposition 7.1 in [KL02])**.** *Let* $X = (X_s)_{s\in[0,1]}$ *be a Gaussian random field, if there exist* $c_1 > 0,\ \gamma > 0,\ \beta \in \mathbb{R}$ *such that*

$$\mathbb{E}\left[(2X_s - X_{s+t} - X_{s-t})^2\right] \leq c_1 t^{2\gamma} \log\left(\frac{1}{t}\right)^{2\beta},$$

for all $0 \leq s - t < s < s + t \leq 1$ *then* $\exists c_2 > 0$ *such that,*

$$\forall n \in \mathbb{N},\ l_n(X) \leq c_2 \frac{(1 + \log(n))^{\beta + \frac{1}{2}}}{n^\gamma}.$$

This is the case when $\exists c_3 > 0$ *such that,*

$$\forall (s,t) \in [0,1]^2,\ s \neq t,\ \mathbb{E}\left[(X_s - X_t)^2\right] \leq c_3 |s-t|^{2\gamma} \log\left(\frac{1}{|s-t|}\right)^{2\beta}.$$

Example 9. *Let* $\mathcal{K} = [0,1]$*,* $W = (W_s)_{s\in[0,1]}$ *be the standard Wiener process (*$W_0 = 0$ *a.s.) then*

$$\mathbb{E}\left[(X_s - X_t)^2\right] = s + t - 2\min(s,t) = |s-t|.$$

Using proposition 28 with $\gamma = \frac{1}{2}$*,* $c_1 = 1$ *and* $\beta = 0$ *gives*

$$\forall n \in \mathbb{N}^*,\ l_n(W) \leq c \left(\frac{1 + \log(n)}{n}\right)^{\frac{1}{2}}.$$

In order to get a lower bound, which will be used to obtain asymptotically optimal representations, the space $C(\mathcal{K},\mathbb{R})$ is embedded in $L^2(\mathcal{K},\mu;\mathbb{R})$, using a finite and Borel measure on $\mathcal{K}$ with full topological support ($\mathrm{supp}(\mu)=\mathcal{K}$). Indeed, consider the following operator:

$$i_\mu : f \in C(\mathcal{K},\mathbb{R}) \to f \in L^2(\mathcal{K},\mu;\mathbb{R}),$$

it is linear, bounded and injective. This implies that if one has an operator $S : L^2(\mathcal{K},\mu;\mathbb{R}) \to C(\mathcal{K},\mathbb{R})$, then $i_\mu \circ S$ is an endomorphism in $L^2(\mathcal{K},\mu;\mathbb{R})$. In particular, the approximation numbers of such operators are equal with eigenvalues. This gives an interesting lower bound for the l-numbers, expressed in terms of eigenvalues of $i_\mu \circ S$.

Proposition 29 (Proposition 4 in [LP09])**.** *Let $d \in \mathbb{N}^*$, $\mathcal{K}=[0,1]^d$, $X=(X_s)_{s\in\mathcal{K}}$ a continuous Gaussian random field with covariance kernel k_X and note $\mathcal{C}_X$ the associated integral operator:*

$$\mathcal{C}_X : L^2(\mathcal{K},ds) \to L^2(\mathcal{K},ds)$$

suppose that $(f_n)_{n\in\mathbb{N}} \subset C(\mathcal{K},\mathbb{R})$ is admissible for X, then if

1. *the sequence $(\lambda_n)_{n\in\mathbb{N}}$ of eigenvalues of $\mathcal{C}_X$ satisfy*

$$\forall n \in \mathbb{N},\ \lambda_n \geq c_1 \frac{\log(n+2)^{2\gamma}}{(n+1)^{2\nu}}$$

   with $\nu > \frac{1}{2}$, $\gamma \geq 0$ and $c_1 > 0$,

2. *$\forall n \in \mathbb{N},\ \|f_n\|_{C(\mathcal{K},\mathbb{R})} \leq c_2 \frac{\log(n+2)^\gamma}{(n+1)^\nu}$ with $c_2 > 0$,*

3. *$(f_n)_{n\in\mathbb{N}}$ is β-Hölder continuous with $\forall n \in \mathbb{N},\ \|f_n\|_\beta \leq c_3(n+1)^b$ with $b \in \mathbb{R}$, $\beta \in]0,1]$ and $c_3 > 0$,*

then

$$l_n(X) \approx \frac{\log(n)^{\gamma+\frac{1}{2}}}{n^{\nu-\frac{1}{2}}}$$

and $(f_n)_{n\in\mathbb{N}}$ is asymptotically optimal for X.

Example 10 (Wiener process)**.** *Let $W=(W_s)_{s\in[0,1]}$ be the standard Wiener process, that is $W_0 = 0$ almost-surely and $\forall (s,t) \in \mathcal{K}^2$, $k_W(s,t) = \min(s,t)$, then the l-numbers are such that*

$$l_n(W) \approx \sqrt{\frac{log(n)}{n}}.$$

Indeed, from example 6, the eigendecomposition of the associated covariance operator (seen as an endomorphism in $L^2([0,1],ds;\mathbb{R})$) is $\forall n \in \mathbb{N}$:

$$\lambda_n = \frac{1}{(n+\frac{1}{2})^2\pi^2},$$
$$f_n(s) = \sqrt{2}\sin\left(\left(n+\frac{1}{2}\right)\pi x\right).$$

It is clear that the sequence of eigenvalues satisfies item 1 *in proposition 29 with* $c_1 = \frac{1}{\pi^2}$, $\nu = 1$ *and* $\gamma = 0$. *Here, the admissible sequence is* $\left(\sqrt{\lambda_n} f_n\right)_{n\in\mathbb{N}}$, *thus:*

$$\forall n \in \mathbb{N},\ \left\|\sqrt{\lambda_n} f_n\right\|_{C([0,1],\mathbb{R})} = \frac{\sqrt{2}}{\pi\left(n+\frac{1}{2}\right)} \leq \frac{2\sqrt{2}}{\pi(n+1)},$$

and item 2 *in proposition* 29 *is satisfied for* $\gamma = 0$, $c_2 = \frac{2\sqrt{2}}{\pi}$ *and* $\nu = 1$. *Finally, the* β*-Hölder norms with* $\beta \in]0,1]$ *of* $\left(\sqrt{\lambda_n} f_n\right)_{n\in\mathbb{N}}$ *are:*

$$\forall n \in \mathbb{N},\ \left\|\sqrt{\lambda_n} f_n\right\|_{\beta} = \sqrt{\lambda_n} \sup_{s\neq t} \frac{|f_n(s) - f_n(t)|}{|s-t|^{\beta}} = \frac{2\sqrt{2}}{\pi\left(n+\frac{1}{2}\right)^2} \leq \frac{8\sqrt{2}}{\pi\left(n+1\right)^2},$$

and satisfies item 3 *with* $c_3 = \frac{8\sqrt{2}}{\pi}$ *and* $b = -2$. *Other asymptotically optimal bases can be found in [LP09] and next chapter.*

1.7 Conclusion

In this introductory chapter on Gaussian random elements, the necessary material has been given for the rest of this first part. In particular, the possibility to represent Gaussian random element as series has been studied, with particular focus on the links with the covariance operator and Cameron-Martin space. In particular, covariance factorizations are key in the practical construction of admissible sequences. However, except in the Hilbert case, this factorization is non-constructive. This will be the objective of next chapter, as first contribution in this thesis, to derive such construction. A second difficulty can arise, concerning continuous Gaussian random fields. The Karhunen-Loève representation lies on an eigendecomposition problem, that can be difficult to solve in practice. In this case, the decomposition proposed in next chapter gives a practical optimization problem, easier to solve numerically.

Chapter 2

Karhunen-Loève decomposition in Banach spaces

Previous chapter was dedicated to a short introduction of Gaussian random elements theory in general Banach spaces and in $C(\mathcal{K}, \mathbb{R})$ with $\mathcal{K}$ a compact metric space. In particular, it has been shown that these elements can be represented as random series, leading to the notion of admissible sequences. Finding such families is usually done by factorization of the covariance operator, using a pre-existing Hilbert basis in the intermediate space. However, this method is not always applicable and it is of interest to find constructive solutions for this problem. It is already done in Hilbert spaces under the name of Karhunen-Loève decomposition, since eigendecomposition of the covariance operator provides such factorization. Again, these problems does not have explicit solutions in all cases. In this second chapter, a constructive approach is given to build stochastic bases of any Gaussian random element, extending the Karhunen-Loève decomposition to general Banach spaces. Besides a theoretical interest, it provides a new methodology to decompose Gaussian random fields.

2.1 Extending the Karhunen-Loève decomposition

In chapter 1, it has been showed that if $\mathcal{U}$ is a Banach space, every Gaussian random element $X \in L^0(\mathcal{U})$ can be represented as a random series, using a factorization of its covariance operator. In Hilbert spaces, the spectral representation of the covariance operator provides such decomposition, in a canonical fashion. In particular, the eigenvectors are obtained sequentially, as unit vectors of maximum variance. This principle will be extended to the Banach setup, using adequate notions of Rayleigh quotients in this context. The role of the eigenvector will be played by a pair $(u, l) \in \mathcal{U} \times \mathcal{U}^*$ of respective unit norms, such that $\mathcal{C}_X l = \lambda u$. In particular, this generalization recovers the usual spectral representation of the covariance operator when $\mathcal{U}$ is Hilbert.

Spectral decomposition in Hilbert spaces Before actually going to the Banach setup, the spectral decomposition of non-negative, self-adjoint and compact operators in Hilbert spaces is revisited, emphasizing the point of view that will be fruitful later on. Indeed, let $\mathcal{U}$ be a separable Hilbert space and $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$. Since it is a non-negative, self-adjoint and nuclear operator (see theorem 6), one has the following characterization of its highest eigenvalue.

Proposition 30. *Let $\mathcal{U}$ be a separable Hilbert space and $A \in \mathcal{L}(\mathcal{U},\mathcal{U})$ a non-negative, self-adjoint and compact operator, then:*

$$\|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} = \sup_{u \in \mathcal{B}_{\mathcal{U}}} \langle Au, u\rangle_{\mathcal{U}} = \langle Au_0, u_0\rangle_{\mathcal{U}} = \lambda_0,$$

where λ_0 is the highest eigenvalue of A and u_0 one associated eigenvector of unit norm: $Au_0 = \lambda_0 u_0$.

Proof. First,

$$\forall (u,v) \in \mathcal{B}_{\mathcal{U}}^2,\ \langle Au, v\rangle_{\mathcal{U}} \leq \|Au\|_{\mathcal{U}} \|v\|_{\mathcal{U}} \leq \|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} \|u\|_{\mathcal{U}} \|v\|_{\mathcal{U}},$$

thus $\sup_{u\in\mathcal{B}_{\mathcal{U}}} \langle Au, u\rangle_{\mathcal{U}} \leq \|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})}$. Conversely, Cauchy-Schwarz inequality gives

$$|\langle Au, v\rangle_{\mathcal{U}}| \leq \sqrt{\langle Au, u\rangle_{\mathcal{U}}}\sqrt{\langle Av, v\rangle_{\mathcal{U}}},$$

thus $\|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} \leq \sup_{u\in\mathcal{B}_{\mathcal{U}}} \langle Au, u\rangle_{\mathcal{U}}$. Since $\langle Au_0, u_0\rangle_{\mathcal{U}} = \lambda_0$, $\lambda_0 \leq \|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})}$. Finally, if $(\lambda_n, u_n)_{n\in\mathbb{N}}$ is the spectral representation of A,

$$\forall u \in \mathcal{U},\ \langle Au, u\rangle_{\mathcal{U}} = \sum_{n\geq 0} \lambda_n \langle u, u_n\rangle_{\mathcal{U}}^2 \leq \lambda_0 \|u\|_{\mathcal{U}}^2,$$

thus $\|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} \leq \lambda_0$. The proof is complete. □

Thus, the first element u_0 in the spectral basis is a unit vector of maximum variance:

$$\|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} = \max_{u\in\mathcal{B}_{\mathcal{U}}} \langle \mathcal{C}_X u, u\rangle_{\mathcal{U}} = \max_{u\in\mathcal{B}_{\mathcal{U}}} \mathbb{V}\left[\langle X, u\rangle_{\mathcal{U}}\right] = \|\mathcal{C}_X u_0\|_X^2 = \lambda_0.$$

Since $u_0 \in \mathcal{U}$ is a vector of unit norm, the Gaussian random element can be split using the associated orthogonal projector of unit rank:

$$\begin{aligned} X &= Y + Z, \\ Y &= \langle X, u_0\rangle_{\mathcal{U}}\, u_0, \\ Z &= X - Y. \end{aligned}$$

Then $Y, Z \in L^0(\mathcal{U})$ are independent Gaussian random elements. The covariance operator is decomposed as well:

$$\begin{aligned} \mathcal{C}_X &= \mathcal{C}_Y + \mathcal{C}_Z, \\ \mathcal{C}_Y &= \lambda_0 \langle ., u_0\rangle_{\mathcal{U}}\, u_0, \\ \mathcal{C}_Z &= \mathcal{C}_X - \mathcal{C}_Y. \end{aligned}$$

Here, Z is identically distributed with the conditional Gaussian random element $X|\langle X, u_0\rangle_{\mathcal{U}} = 0$. The second component is chosen similarly, under the additional constraint to be independent (orthogonality in the Cameron-Martin space) with the first one. It is the exact same thing than looking for the unit vector of maximum variance w.r.t. Z and it appears that:

$$\|\mathcal{C}_Z\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \max_{u\in\mathcal{B}_{\mathcal{U}}} \langle \mathcal{C}_Z u, u\rangle_{\mathcal{U}} = \max_{\substack{u\in\mathcal{B}_{\mathcal{U}} \\ \langle u,u_0\rangle_{\mathcal{U}}=0}} \langle \mathcal{C}_X u, u\rangle_{\mathcal{U}} = \|\mathcal{C}_Z u_1\|_X^2 = \lambda_1.$$

This scheme will be repeated to obtain every eigenvector and this is what will be done in the Banach setup. However, the optimality (see chapter 1) of the spectral basis lies on the following observation (Courant-Fisher min-max theorem):

$$\begin{aligned}\forall n \in \mathbb{N},\ \lambda_n &= \min\left\{ \max_{\|u\|_{\mathcal{U}}=1} \langle \mathcal{C}_X u, u\rangle_{\mathcal{U}},\ u \in \mathcal{V}^{\perp},\ \dim(\mathcal{V}) = n \right\}, \\ &= \max\left\{ \min_{\|u\|_{\mathcal{U}}=1} \langle \mathcal{C}_X u, u\rangle_{\mathcal{U}},\ u \in \mathcal{V},\ \dim(\mathcal{V}) = n \right\}.\end{aligned}$$

It implies that sequentially taking the maximum of variance in the orthogonal subspace of previous vectors leads to the best solution. However, in the Banach setting, this principle needs not be true.

Splitting the space Now, let $\mathcal{U}$ be a general Banach space and $X \in L^0(\mathcal{U})$ a Gaussian random element. The first step is to be able to split the space in such a way that the underlying Gaussian random element is decomposed as the sum of two independent components (one being of unit rank). Indeed, one could always consider a non-trivial linear functional and the associated hyperplane which is of codimension 1:

$$\forall l \in \mathcal{U}^* \setminus \{0\},\ \mathcal{U} = \ker l + \operatorname{span}\{u\},$$

where $u \in \mathcal{U} \setminus \ker l$. However, nothing imposes that the resulting decomposition of the Gaussian random element:

$$\begin{aligned} X &= Y + Z, \\ Y &= \langle X, l\rangle\, u, \\ Z &= X - \langle X, l\rangle\, u, \end{aligned}$$

provides two independent elements. Nevertheless, the next lemma shows that it is possible to impose this independence, choosing an adequate vector $u \in \mathcal{U}$, which will result in an orthogonal direct sum of the two resulting Cameron-Martin spaces. Moreover, both spaces inherit the inner product from $\mathcal{U}_X$.

Lemma 5. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with non-trivial covariance operator $\mathcal{C}_X$, $l \in \mathcal{U}^* \setminus \ker(\mathcal{C}_X)$, $h = \|\mathcal{C}_X l\|_X^{-1}\, \mathcal{C}_X l$ and $h^* = \|\mathcal{C}_X l\|_X^{-1}\, l$ then*

$$X = Y + Z,$$

where:

$$Y = \langle X, h^* \rangle_{\mathcal{U},\mathcal{U}^*} h,$$

and $Z = X - Y$ are two independent Gaussian random elements with respectively

$$\mathcal{C}_Y := g \in \mathcal{U}^* \to \langle h, g \rangle_{\mathcal{U},\mathcal{U}^*} h \in \mathcal{U},$$

and $\mathcal{C}_Z = \mathcal{C}_X - \mathcal{C}_Y$ as covariance operators. Concerning the respective Cameron-Martin spaces, it comes that $\mathcal{U}_Y$ is isometric to $\operatorname{span}\{h\} \subset \mathcal{U}_X$ *and $\mathcal{U}_Z$ to* $\operatorname{span}\{h\}^\perp \subset \mathcal{U}_X$*, leading to*

$$\mathcal{U}_X = \mathcal{U}_Y \oplus \mathcal{U}_Z.$$

Moreover $\mathbb{V}[\langle X, h^* \rangle] = \|h\|_X^2 = 1$.

Proof. Since Y and Z are both images of X under bounded operators, they are Gaussian random elements themselves (proposition 4). Concerning Y, the covariance operator may be deduced directly from its Fourier transform:

$$\forall g \in \mathcal{U}^*,\ \widehat{Y}(g) = \mathbb{E}\left[\exp\left(i \langle X, h^* \rangle_{\mathcal{U},\mathcal{U}^*} \langle h, g \rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] = \exp\left(-\frac{1}{2} \langle h, g \rangle_{\mathcal{U},\mathcal{U}^*}^2\right),$$

from which $\mathcal{C}_Y$ is easily identified. For the independence property, let $(l_0, l_1) \in \mathcal{U}^* \times \mathcal{U}^*$, then

$$\begin{aligned}
\mathbb{E}\left[\langle Y, l_0 \rangle \langle Z, l_1 \rangle\right] &= \mathbb{E}\left[\langle Y, l_0 \rangle \langle X - Y, l_1 \rangle\right], \\
&= \mathbb{E}\left[\langle Y, l_0 \rangle \langle X, l_1 \rangle\right] - \mathbb{E}\left[\langle Y, l_0 \rangle \langle Y, l_1 \rangle\right], \\
&= \mathbb{E}\left[\langle X, h^* \rangle \langle h, l_0 \rangle \langle X, l_1 \rangle\right] - \mathbb{E}\left[\langle X, h^* \rangle \langle h, l_0 \rangle \langle X, h^* \rangle \langle h, l_1 \rangle\right], \\
&= \langle h, l_0 \rangle \langle \mathcal{C}_X h^*, l_1 \rangle - \langle h, l_0 \rangle \langle h, l_1 \rangle \langle \mathcal{C}_X h^*, h^* \rangle, \\
&= 0,
\end{aligned}$$

since $\mathcal{C}_X h^* = h$, which is sufficient. Now, because $X = Y + Z$ and Y, Z are independent,

$$\begin{aligned}
\forall l_0, l_1 \in \mathcal{U}^*,\ \langle \mathcal{C}_X l_0, l_1 \rangle &= \mathbb{E}\left[\langle X, l_0 \rangle \langle X, l_1 \rangle\right], \\
&= \mathbb{E}\left[\langle Y + Z, l_0 \rangle \langle Y + Z, l_1 \rangle\right], \\
&= \mathbb{E}\left[\langle Y, l_0 \rangle \langle Y, l_1 \rangle\right] + \mathbb{E}\left[\langle Z, l_0 \rangle \langle Z, l_1 \rangle\right], \\
&= \langle \mathcal{C}_Y l_0, l_1 \rangle + \langle \mathcal{C}_Z l_0, l_1 \rangle,
\end{aligned}$$

and $\mathcal{C}_X = \mathcal{C}_Y + \mathcal{C}_Z$. Let P be the orthogonal projector from $\mathcal{U}_X$ to $\operatorname{span}\{h\}$, as $\|h\|_X = 1$ then one has $\forall g \in \mathcal{U}^*$,

$$\mathcal{C}_Y g = \langle h, g \rangle_{\mathcal{U},\mathcal{U}^*} h = \langle \mathcal{C}_X g, h \rangle_X h,$$

so $\mathcal{C}_Y g = P(\mathcal{C}_X g)$ and similarly $\mathcal{C}_Z g = (I - P)(\mathcal{C}_X g)$ thus

$$\forall g \in \mathcal{U}^*,\ \langle \mathcal{C}_Z g, \mathcal{C}_Y g \rangle_X = 0.$$

It remains to see that on $R(\mathcal{C}_Y)$ and $R(\mathcal{C}_Z)$, both respective Cameron-Martin norms are inherited from $\mathcal{U}_X$. Let $g \in \mathcal{U}^*$, then $\forall i \in \{1,2\}$ one has:

$$\begin{aligned}\|\mathcal{C}_{X_i} g\|_{X_i}^2 &= \langle \mathcal{C}_{X_i} g, \mathcal{C}_{X_i} g \rangle_{X_i}, \\ &= \langle \mathcal{C}_{X_i} g, g \rangle_{\mathcal{U},\mathcal{U}^*}, \\ &= \langle \mathcal{C}_{X_i} g, \mathcal{C}_X g \rangle_X, \\ &= \langle \mathcal{C}_{X_i} g, \mathcal{C}_{X_i} g \rangle_X = \|\mathcal{C}_{X_i} g\|_X^2.\end{aligned}$$

Now, clearly $\mathcal{U}_Y = R(\mathcal{C}_Y) = \operatorname{span}\{h\}$ and $\mathcal{U}_Z = \overline{R(\mathcal{C}_Z)}^{\|.\|_X} = \operatorname{span}\{h\}^\perp$. Finally,

$$\mathbb{V}[\langle X, h^* \rangle] = \langle \mathcal{C}_X h^*, h^* \rangle = \frac{\langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*}}{\langle \mathcal{C}_X l, \mathcal{C}_X l \rangle_X} = \frac{\|\mathcal{C}_X l\|_X^2}{\|\mathcal{C}_X l\|_X^2} = 1.$$

□

One very important interpretation in previous lemma is that Z has the same distribution as $X|\langle X, l \rangle = 0$ and it will be called the residual Gaussian random element. If Z has a non-trivial covariance operator, it can be used to iterate and obtain the following result.

Proposition 31. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ and Cameron-Martin space $\mathcal{U}_X$. If $\mathcal{U}_X$ is infinite dimensional, there exists $(h_n)_{n\in\mathbb{N}} \subset \mathcal{U}_X$ orthonormal and $(h_n^*)_{n\in\mathbb{N}} \subset \mathcal{U}^*$ such that*

$$X = \sum_{n \geq 0} Y_n + X_\infty,$$

with $\forall n \in \mathbb{N}$, $Y_n = \langle X, h_n^* \rangle h_n$ and X_∞ mutually independent Gaussian random elements in $L^0(\mathcal{U})$. For all $n \in \mathbb{N}$, $\mathbb{V}[\langle X, h_n^* \rangle] = \|h_n\|_X = 1$, $\mathcal{U}_{Y_n} = \operatorname{span}\{h_n\}$ (with $\langle ., . \rangle_{Y_n} = \langle ., . \rangle_X$). Furthermore,

$$\mathcal{U}_X = \overline{\operatorname{span}\{h_n,\ n \in \mathbb{N}\}}^X \oplus \mathcal{U}_{X_\infty},$$

where $\mathcal{U}_{X_\infty} = \operatorname{span}\{h_n,\ n \in \mathbb{N}\}^\perp$ and $\langle ., . \rangle_{X_\infty} = \langle ., . \rangle_X$.

Proof. The proof is done by induction and the initialization ($n = 0$) is the result of lemma 5 applied to X, with notations $l = l_0$, $h_0 = \|\mathcal{C}_X l_0\|_X^{-1} \mathcal{C}_X l_0$, $h_0^* = \|\mathcal{C}_X l_0\|_X^{-1} l_0$ and:

$$X = Y_0 + X_1,$$

where $Y_0 = \langle X, h_0^* \rangle_{\mathcal{U},\mathcal{U}^*} h_0$ and $X_1 = X - Y_0$ are independent Gaussian random element in $L^0(\mathcal{U})$ with orthogonal Cameron-Martin spaces $\mathcal{U}_{Y_0} = \operatorname{span}\{h_0\}$ and $\mathcal{U}_{X_1} = \mathcal{U}_{Y_0}^\perp$. The associated covariance operators are:

$$\begin{aligned}\mathcal{C}_Y &= \langle ., h_0 \rangle_{\mathcal{U},\mathcal{U}^*} h_0, \\ \mathcal{C}_{X_1} &= \mathcal{C}_X - \mathcal{C}_Y.\end{aligned}$$

For the induction step, let $n \in \mathbb{N}$ and suppose the proposition true at this stage, meaning that

$$X = \sum_{k=0}^{n} Y_k + X_{n+1},$$

with $\forall k \in [0,n],\ Y_k = \langle X, h_k^* \rangle h_k$ and X_{n+1}, $(n+1)$ mutually independent Gaussian random elements in $L^0(\mathcal{U})$ with respective Cameron-Martin spaces $\mathcal{U}_{Y_k} = \operatorname{span}\{h_k\}$, $\forall k \in [0,n]$ and $\mathcal{U}_{X_{n+1}} = \operatorname{span}\{h_k,\ k \in [0,n]\}^{\perp}$. Since $\mathcal{U}_X$ is infinite dimensional, there exists $l_{n+1} \in \operatorname{span}\{h_k,\ k \in [0,n]\}^{\perp} \setminus \{0\}$ and:

$$h_{n+1} = \left\| \mathcal{C}_{X_{n+1}} l_{n+1} \right\|_X^{-1} \mathcal{C}_{X_{n+1}} l_{n+1},$$

is well-defined. Now it comes:

$$\begin{aligned}
X_{n+1} &= \left\langle X_n, \frac{l_{n+1}}{\left\| \mathcal{C}_{X_{n+1}} l_{n+1} \right\|_X} \right\rangle h_{n+1} + X_{n+2}, \\
&= \left\langle X - \sum_{k=0}^{n} \langle X, h_k^* \rangle h_k, \frac{l_{n+1}}{\left\| \mathcal{C}_{X_{n+1}} l_{n+1} \right\|_X} \right\rangle h_{n+1} + X_{n+2}, \\
&= \left\langle X, \frac{l_{n+1} - \sum_{k=0}^{n} \langle h_k, l_{n+1} \rangle h_k^*}{\left\| \mathcal{C}_{X_{n+1}} l_{n+1} \right\|_X} \right\rangle h_{n+1} + X_{n+2}, \\
&= \left\langle X, h_{n+1}^* \right\rangle h_{n+1} + X_{n+2},
\end{aligned}$$

with obvious notations. Since $h_{n+1} \in \mathcal{U}_{X_{n+1}} = \operatorname{span}\{h_0, ..., h_n\}^{\perp}$ and has unit norm, the family $(h_0, ..., h_{n+1}) \subset \mathcal{U}_X$ is orthonormal and one has $\mathcal{U}_{X_{n+2}} = \operatorname{span}\{h_0, ..., h_{n+1}\}^{\perp}$.The independence property is obtained directly and the proposition is true at step $(n+1)$. Using the induction principle, the proof is complete. □

In the degenerated case where the covariance operator has finite rank, which is equivalent with $\mathcal{U}_X$ of finite dimension, it is clear that previous proposition is enough to provide an orthonormal basis of the Cameron-Martin space. However, the case where $\mathcal{U}_X$ is infinite dimensional requires to choose the linear functionals more specifically at each step.

Maximum variance functionals Here, it is shown that one can choose $l \in \mathcal{B}_{\mathcal{U}^*}$ maximizing the one-dimensional projected variance. This principle is directly motivated by analogy with the Hilbert case where eigenvectors (seen as linear functionals) are maximizing the Rayleigh quotient.

Lemma 6. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, then*

$$\exists l \in \mathcal{B}_{\mathcal{U}^*},\ \langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*} = \max_{g \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X g, g \rangle_{\mathcal{U},\mathcal{U}^*}.$$

Moreover, it can be chosen with unit norm and satisfies the following relation:

$$\left\| \mathcal{C}_X l \right\|_X = \sup_{g \in B_{\mathcal{U}^*}} \sup_{h \in \mathcal{B}_{\mathcal{U}_X}} \langle h, g \rangle_{\mathcal{U},\mathcal{U}^*} = \left\| \mathcal{C}_X \right\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}^{\frac{1}{2}} = \left\| \mathcal{C}_X l \right\|_{\mathcal{U}}^{\frac{1}{2}}.$$

Note $\lambda = \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*} = \mathbb{V}\left[\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*}\right]$, $h = \|\mathcal{C}_X l\|_X^{-1} \mathcal{C}_X l$ *and* $u = \|\mathcal{C}_X l\|_{\mathcal{U}}^{-1} \mathcal{C}_X l$ *then one has the following properties:*

1. $\mathcal{C}_X l = \lambda u = \sqrt{\lambda} h$,
2. $\|h\|_X = 1$, $\|h\|_{\mathcal{U}} = \sqrt{\lambda}$,
3. $\|u\|_X = \frac{1}{\sqrt{\lambda}}$, $\|u\|_{\mathcal{U}} = 1$,
4. $\langle u, l\rangle_{\mathcal{U},\mathcal{U}^*} = 1$.

Proof. Let $(l_n)_{n\in\mathbb{N}} \subset \mathcal{B}_{\mathcal{U}^*}$ be such that $\langle \mathcal{C}_X l_n, l_n\rangle_{\mathcal{U},\mathcal{U}^*} \to \sup_{l\in\mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*}$ (a maximizing sequence). Because $\mathcal{B}_{\mathcal{U}^*}$ is compact in the $\sigma(\mathcal{U}^*,\mathcal{U})$-topology (Banach-Alaoglu theorem), there exists $l \in \mathcal{B}_{\mathcal{U}^*}$ such that $l_n \rightharpoonup l$ in the $\sigma(\mathcal{U}^*,\mathcal{U})$ topology. Now,

$$\hat{X}(l_n) = \mathbb{E}\left[\exp\left(i\,\langle X, l_n\rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] \to \mathbb{E}\left[\exp\left(i\,\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*}\right)\right] = \hat{X}(l),$$

by Lebesgue's dominated convergence theorem. Using proposition 3 it implies that $\langle \mathcal{C}_X l_n, l_n\rangle_{\mathcal{U},\mathcal{U}^*} \to \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*} \in \mathbb{R}$. If $\mathcal{C}_X$ is trivial, then one can take whatever element in $\mathcal{B}_{\mathcal{U}^*}$. If $\mathcal{C}_X \neq 0$, then l must be of unit norm. Indeed, suppose $\|l\|_{\mathcal{U}^*} < 1$, then $g = \|l\|_{\mathcal{U}^*}^{-1} l$ would contradict the maximality. Now, let $h \in \mathcal{B}_{\mathcal{U}_X}$, then the reproducing property gives $\forall g \in \mathcal{U}^*$, $\langle h, g\rangle_{\mathcal{U},\mathcal{U}^*} = \langle h, \mathcal{C}_X g\rangle_X$ and taking the supremum in h over the unit ball $\mathcal{B}_{\mathcal{U}_X}$ leads to

$$\sup_{h\in\mathcal{B}_{\mathcal{U}_X}} \langle h, g\rangle_{\mathcal{U},\mathcal{U}^*} = \|\mathcal{C}_X g\|_X = \langle \mathcal{C}_X g, g\rangle_{\mathcal{U},\mathcal{U}^*},$$

which leads to the announced equality. Now, from the definition, it is clear that

$$\|h\|_X = \|u\|_{\mathcal{U}} = 1,$$

and $\langle u, l\rangle_{\mathcal{U},\mathcal{U}^*} = \langle \mathcal{C}_X l, u\rangle_X = \left\langle \sqrt{\lambda} h, \frac{h}{\sqrt{\lambda}}\right\rangle_X = \langle h, h\rangle_X = 1$. Finally,

$$\|\mathcal{C}_X l\|_X = \sqrt{\langle \mathcal{C}_X l, \mathcal{C}_X l\rangle_X} = \sqrt{\langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*}} \leq \|\mathcal{C}_X l\|_{\mathcal{U}}^{\frac{1}{2}} \|l\|_{\mathcal{U}^*}^{\frac{1}{2}} \leq \|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})}^{\frac{1}{2}},$$

and conversely

$$\langle \mathcal{C}_X l, g\rangle_{\mathcal{U},\mathcal{U}^*} = \langle \mathcal{C}_X l, \mathcal{C}_X g\rangle_X \leq \|\mathcal{C}_X l\|_X \|\mathcal{C}_X g\|_X,$$

thus $\|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} \leq \|\mathcal{C}_X l\|_X^2$ and both are equal. □

The similarity with the Hilbert case is emphasized here, where $l \in \mathcal{U}^*$ and $u \in \mathcal{U}$ will play the role of eigenvectors associated to λ, since $\mathcal{C}_X l = \lambda u$. Remark that uniqueness of the maximizing linear functional in previous lemma need not be true in general, the set of solutions even being possibly infinite. However, one can always choose this element among the extremal points of the unit ball of $\mathcal{U}^*$, and this will be particularly important in practice.

Proposition 32. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$, then there exists an extremal point δ of $\mathcal{B}_{\mathcal{U}^*}$ such that $\langle \mathcal{C}_X \delta, \delta \rangle_{\mathcal{U},\mathcal{U}^*} = \sup_{l \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*}$.*

Proof. Since the functional $l \in \mathcal{U}^* \to \langle \mathcal{C}_X l, l \rangle_{\mathcal{U},\mathcal{U}^*}$ is quadratic, this is a general result from convex optimization theory. □

Splitting the space using maximum variance functionals It has been proved previously that whenever the Cameron-Martin space is finite dimensional, one can build a basis only using linear functionals of strictly positive residual variance at each stage (proposition 31). Now, if one chooses these functionals such that they maximize the residual variance at each step, the family $(h_n)_{n \in \mathbb{N}}$ will be a Hilbert basis in $\mathcal{U}_X$. The first step is to give an analogue of proposition 31, taking into account that linear functionals are maximizing the variance.

Lemma 7. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ and suppose $\mathcal{U}_X$ infinite dimensional. Note $X_0 := X$ and define by induction on $n \in \mathbb{N}$:*

$$\begin{aligned} l_n &= \arg\max_{l \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_{X_n} l, l \rangle, \\ \lambda_n &= \langle \mathcal{C}_{X_n} l_n, l_n \rangle_{\mathcal{U},\mathcal{U}^*}, \\ h_n &= \lambda_n^{-\frac{1}{2}} \mathcal{C}_{X_n} l_n, \\ u_n &= \lambda_n^{-1} \mathcal{C}_{X_n} l_n, \\ X_{n+1} &= X_n - \langle X_n, l_n \rangle_{\mathcal{U},\mathcal{U}^*} u_n, \end{aligned}$$

then one has the following properties:

- $\forall n \in \mathbb{N},\ \mathcal{C}_{X_n} l_n = \sqrt{\lambda_n} h_n = \lambda_n u_n,$
- $\forall n \in \mathbb{N},\ \|h_n\|_X = \|u_n\|_{\mathcal{U}} = \langle u_n, l_n \rangle_{\mathcal{U},\mathcal{U}^*} = 1,$
- $\forall (n,m) \in \mathbb{N}^2,\ \langle h_n, h_m \rangle_X = \delta_{nm}.$

Furthermore, $(\lambda_n)_{n \in \mathbb{N}} \subset \mathbb{R}^+$ is non-increasing and $\lim_{n \to \infty} \lambda_n = 0$.

Proof. Since $\mathcal{U}_X$ is infinite dimensional, the construction from proposition 31 is licit. From lemma 6, linear functionals of maximum residual variance (a fortiori strictly positive) exist and the sequences $(u_n)_{n \in \mathbb{N}}, (l_n)_{n \in \mathbb{N}}, (\lambda_n)_{n \in \mathbb{N}}$ are well-defined. Now,

$$\forall n \in \mathbb{N},\ X_n = \langle X_n, l_n \rangle_{\mathcal{U},\mathcal{U}^*} u_n + X_{n+1},$$

where $\langle X_n, l_n \rangle_{\mathcal{U},\mathcal{U}^*} u_n$ and X_{n+1} are independent Gaussian random elements, it follows that:

$$\forall n \in \mathbb{N},\ \forall l \in \mathcal{U}^*,\ \langle \mathcal{C}_{X_n} l, l \rangle_{\mathcal{U},\mathcal{U}^*} \geq \langle \mathcal{C}_{X_{n+1}} l, l \rangle_{\mathcal{U},\mathcal{U}^*},$$

and taking the supremum in the above relation on the unit ball of $\mathcal{U}^*$ shows that $(\lambda_n)_{n \in \mathbb{N}}$ is non-increasing. Moreover, $(h_n)_{n \in \mathbb{N}}$ is an orthonormal system in $\mathcal{U}_X$, hence

$$\forall l \in \mathcal{U}^*,\ \langle h_n, l \rangle_{\mathcal{U},\mathcal{U}^*} = \langle h_n, \mathcal{C}_X l \rangle_X \to 0,$$

as a consequence of Bessel's inequality. In other words, $h_n \rightharpoonup 0$ in the weak topology $\sigma(\mathcal{U},\mathcal{U}^*)$. Now, since the unit ball of $\mathcal{U}_X$ is precompact in $\mathcal{U}$, there exists a convergent subsequence $(h_{n_k})_{k\in\mathbb{N}}$ such that $h_{n_k} \to_k h_\infty$ in the strong topology of $\mathcal{U}$. Using the uniqueness of limits in $\mathcal{U}$ equipped with the weak topology $\sigma(\mathcal{U},\mathcal{U}^*)$, it comes that $h_\infty = 0$ and therefore, $\|h_{n_k}\|_{\mathcal{U}} = \sqrt{\lambda_{n_k}} \to 0$, which implies that $(\lambda_n)_{n\in\mathbb{N}} \to 0$. The relations are obtained directly using the reproducing property at each stage similarly to lemma 6. □

Now, it remains to prove that the orthonormal sequence $(h_n)_{n\in\mathbb{N}} \subset \mathcal{U}_X$ from lemma 7 is actually a Hilbert basis.

Theorem 13. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$. Consider the orthogonal family $(h_n)_{n\in\mathbb{N}} \subset \mathcal{U}_X$ as built in lemma 7, then it is a Hilbert basis in $\mathcal{U}_X$.*

Proof. The family $(h_n)_{n\in\mathbb{N}}$ is orthonormal by lemma 7, it remains to see that $\text{span}\{h_n,\ n\in\mathbb{N}\}$ is dense in $\mathcal{U}_X$ to conclude. Let $h \in \mathcal{U}_X$ such that $\forall n \in \mathbb{N},\ \langle h, h_n\rangle_X = 0$, then clearly, $h \in \mathcal{U}_{X_{n+1}},\ \forall n \in \mathbb{N}$. Using the reproducing property in $\mathcal{U}_{X_{n+1}}$, it comes $\forall n \in \mathbb{N},\ \forall l \in \mathcal{B}_{\mathcal{U}^*}$:

$$\begin{aligned}\langle h, l\rangle_{\mathcal{U},\mathcal{U}^*} &= \left\langle h, \mathcal{C}_{X_{n+1}} l\right\rangle_{X_{n+1}},\\ &\leq \|h\|_{X_{n+1}} \sqrt{\left\langle \mathcal{C}_{X_{n+1}} l, \mathcal{C}_{X_{n+1}} l\right\rangle_{X_{n+1}}},\\ &\leq \|h\|_X \sqrt{\lambda_{n+1}}.\end{aligned}$$

This implies that $\forall l \in \mathcal{B}_{\mathcal{U}^*},\ \langle h, l\rangle_{\mathcal{U},\mathcal{U}^*} = 0$, therefore $h = 0$ and $\text{span}\{h_n,\ n\in\mathbb{N}\}$ is dense in $\mathcal{U}_X$. □

Consequences The previous construction leads to a Hilbert basis of $\mathcal{U}_X$ in $R(\mathcal{C}_X)$ using a constructive approach. In fact, the system $(u_n, h_n^*)_{n\in\mathbb{N}} \subset \mathcal{U}\times\mathcal{U}^*$ is a stochastic basis.

Corollary 9. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element with covariance operator $\mathcal{C}_X$ and consider notations from lemma 7, then*

- $X = \sum_{n\geq 0} \langle X, h_n^*\rangle_{\mathcal{U},\mathcal{U}^*} h_n$ *a.s.,*
- $\forall l \in \mathcal{U}^*,\ \mathcal{C}_X l = \sum_{n\geq 0} \langle h_n, l\rangle h_n$,

moreover one has $\|\mathcal{C}_X\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \lambda_0$ and:

$$\forall n \in \mathbb{N},\ \left\|\mathcal{C}_X - \sum_{k=0}^{n} \langle h_k, .\rangle\, h_k\right\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \lambda_{n+1}.$$

Proof. The sequence $(h_n)_{n\in\mathbb{N}}$ is a Hilbert basis in $\mathcal{U}_X$, thus admissible for X and the two representations are consequences of lemma 4 and corollary 7. Now, the truncation error norm is

$$\left\|\mathcal{C}_X - \sum_{k=0}^{n} \langle h_k, .\rangle_{\mathcal{U},\mathcal{U}^*}\, h_k\right\|_{\mathcal{L}(\mathcal{U}^*,\mathcal{U})} = \left\langle \mathcal{C}_{X_{n+1}} l_{n+1}, l_{n+1}\right\rangle_{\mathcal{U},\mathcal{U}^*} = \lambda_{n+1}.$$

□

However, since it is possible to find non-unique maximizing linear functionals at each steps, the error λ_{n+1} may be dependent on these choices. It implies that it is not necessarily minimal, thus the obtained basis may not be optimal in the sense $\forall n \in \mathbb{N},\ a_n(X) = \sqrt{\lambda_{n+1}}$. For instance, one could start with linear functionals having strictly positive residual variance on a finite number of steps, and then apply the proposed decomposition. However, it is still of important practical interest for the following reason.

Corollary 10. *Let $\mathcal{U}$ be a Banach space, $X \in L^0(\mathcal{U})$ a Gaussian random element, then with the notations from lemma 7,*

$$\sup_{l \in \mathcal{B}_{\mathcal{U}^*}} \mathbb{E}\left[\left\langle X - \sum_{k=0}^{n} \langle X, h_k^* \rangle_{\mathcal{U},\mathcal{U}^*}\, h_k, l \right\rangle_{\mathcal{U},\mathcal{U}^*}^2\right]^{\frac{1}{2}} = \sqrt{\lambda_{n+1}}.$$

Proof. By construction, $\forall n \in \mathbb{N}$,

$$\forall l \in \mathcal{U}^*,\ \mathbb{E}\left[\left\langle X - \sum_{k=0}^{n} \langle X, h_k^* \rangle_{\mathcal{U},\mathcal{U}^*}\, h_k, l \right\rangle_{\mathcal{U},\mathcal{U}^*}^2\right] = \left\langle \mathcal{C}_{X_{n+1}} l, l \right\rangle_{\mathcal{U},\mathcal{U}^*},$$

thus the result is immediate. □

In particular, this gives an estimation of the approximation numbers of the Gaussian random element X:

$$\forall n \in \mathbb{N},\ a_n(X) \leq \sqrt{\lambda_{n+1}}.$$

The question whether one has

$$a_n(X) \approx \sqrt{\lambda_n},$$

in general is still open, but it will be positively answered in the particular study of the Wiener process for instance (see section 2.4 below).

Non-Gaussian case It will now be showed that the Gaussian hypothesis in previous construction is not necessary. Indeed, all the above theory can be applied to a wider class, namely the random elements $X \in L^2(\mathcal{U})$ of strong order 2. First, notions of covariance operator and Cameron-Martin spaces are well-defined and share similar properties.

Proposition 33. *Let $\mathcal{U}$ be a Banach space, $X \in L^2(\mathcal{U})$ a random element of strong order 2, then there exists a unique covariance operator $\mathcal{C}_X : \mathcal{U}^* \to \mathcal{U}$ such that:*

$$\forall l, g \in \mathcal{U}^*,\ \langle \mathcal{C}_X l, g \rangle_{\mathcal{U},\mathcal{U}^*} = \mathbb{E}\left[\langle X, l \rangle_{\mathcal{U},\mathcal{U}^*} \langle X, g \rangle_{\mathcal{U},\mathcal{U}^*}\right].$$

This operator is linear, non-negative, symmetric and nuclear.

Proof. Since $X \in L^2(\mathcal{U})$, the covariance operator is directly defined using Bochner's integral:

$$\mathcal{C}_X := l \in \mathcal{U}^* \to \mathbb{E}\left[\langle X, l\rangle_{\mathcal{U},\mathcal{U}^*} X\right] \in \mathcal{U}.$$

The above relations is a property of this integral, thus symmetry, non-negativity and linearity are immediate. The nuclearity is proved in [VTC87] (chapter 3, theorem 2.3). □

In particular, the factorization lemma and the construction of the Cameron-Martin space are still valid.

Proposition 34. *Let $\mathcal{U}$ be a Banach space, $X \in L^2(\mathcal{U})$ a random element of strong order 2 with covariance operator $\mathcal{C}_X$, then the associated Cameron-Martin space $\mathcal{U}_X$ is well-defined. It is separable and injects compactly in $\mathcal{U}$.*

Proof. The proof is similar with the Gaussian case, the separability being given in [VTC87] (chapter 3, corollary 1). The compacity comes from the nuclearity of $\mathcal{C}_X$. □

Now that the essential ingredients are similar, it remains to see that the maximizing linear functionals are still well-defined, which is the object of next proposition.

Proposition 35. *Let $\mathcal{U}$ be a Banach space, $X \in L^2(\mathcal{U})$ a random element of strong order 2 with covariance operator $\mathcal{C}_X$, then it comes that*

$$l_n \rightharpoonup l \Rightarrow \langle \mathcal{C}_X l_n, l_n\rangle_{\mathcal{U},\mathcal{U}^*} \to \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*}.$$

As a consequence, $\exists l \in \mathcal{B}_{\mathcal{U}^*}$ such that $\langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*} = \sup_{g\in\mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X g, g\rangle_{\mathcal{U},\mathcal{U}^*}$.

Proof. Let $(l_n)_{n\in\mathbb{N}} \subset \mathcal{U}^*$, $l \in \mathcal{U}^*$ such that $l_n \rightharpoonup l$ in $\sigma(\mathcal{U}^*,\mathcal{U})$. In particular, this implies that $\langle X, l_n\rangle^2_{\mathcal{U},\mathcal{U}^*} \to \langle X, l\rangle^2_{\mathcal{U},\mathcal{U}^*}$. Moreover, one has

$$\langle X, l_n\rangle^2_{\mathcal{U},\mathcal{U}^*} \leq \|X\|^2_{\mathcal{U}} \|l_n\|^2_{\mathcal{U}^*},$$

and since $(\|l_n\|_{\mathcal{U}^*})$ is bounded, Lebesgue's dominated convergence theorem gives:

$$\langle \mathcal{C}_X l_n, l_n\rangle_{\mathcal{U},\mathcal{U}^*} = \mathbb{E}\left[\langle X, l_n\rangle^2_{\mathcal{U},\mathcal{U}^*}\right] \to \mathbb{E}\left[\langle X, l\rangle^2_{\mathcal{U},\mathcal{U}^*}\right] = \langle \mathcal{C}_X l, l\rangle_{\mathcal{U},\mathcal{U}^*}.$$

Now, the rest of the proof is similar with the Gaussian case. □

With these 3 propositions at hand, the previous decomposition can be derived, in the exact same manner. It leads to the following result:

$$X = \sum_{n\geq 0} \langle X, h_n^*\rangle_{\mathcal{U},\mathcal{U}^*} h_n,$$

with $\left(\langle X, h_n^*\rangle_{\mathcal{U},\mathcal{U}^*}\right)_{n\in\mathbb{N}}$ a sequence of square integrable random variables with unit variance and $(h_n)_{n\in\mathbb{N}}$ a basis in the Cameron-Martin space. The main difference with the Gaussian case are:

1. $\forall n \in \mathbb{N}$, $\langle X, h_n^*\rangle_{\mathcal{U},\mathcal{U}^*}$ need not be Gaussian,
2. $\left(\langle X, h_n^*\rangle_{\mathcal{U},\mathcal{U}^*}\right)$ are mutually non-correlated but possibly dependent variables.

2.2 The special case of $C(\mathcal{K}, \mathbb{R})$

In this section, the previous theory is applied to the Banach space $\mathcal{U} = C(\mathcal{K}, \mathbb{R})$, where $(\mathcal{K}, d)$ is a compact metric space. First, the decomposition is directly given by application of previous general theorems. In a second time, a direct proof is given, using standard analysis in Reproducing Kernel Hilbert spaces.

Application of previous theory As stated in chapter 1, any Gaussian random field on $\mathcal{K}$ may be considered as a random element, the covariance operator being obtained as follows (see chapter 1):

$$\mathcal{C}_X : \mu \in \mathcal{U}^* \to \int_{\mathcal{K}} k_X(., t)\mu(dt).$$

One particularly useful result from previous section is the possibility to choose among extremal points of the unit ball in $\mathcal{U}^*$ at each stage. In this context, these points are (signed) Dirac delta measures and lead to a classical optimization problem.

Proposition 36. *Let $(\mathcal{K}, d)$ be a compact metric space, X a continuous Gaussian random field on $\mathcal{K}$, k_X and $\mathcal{C}_X$ its respective covariance kernel and operator, then*

$$\sup_{\mu \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X \mu, \mu \rangle_{\mathcal{U},\mathcal{U}^*} = \sup_{\mu \in \mathcal{B}_{\mathcal{U}^*}} \int_{\mathcal{K}} \int_{\mathcal{K}} k_X(s,t)\mu(dt)\mu(ds) = \sup_{s \in \mathcal{K}} k_X(s,s).$$

Proof. The functional $\mu \in \mathcal{U}^* \to \langle \mathcal{C}_X \mu, \mu \rangle_{\mathcal{U},\mathcal{U}^*}$ is quadratic. From proposition 32, there exists an extremal point $\delta \in \mathcal{B}_{\mathcal{U}^*}$ such that $\langle \mathcal{C}_X \delta, \delta \rangle_{\mathcal{U},\mathcal{U}^*} = \sup_{\mu \in \mathcal{B}_{\mathcal{U}^*}} \langle \mathcal{C}_X \mu, \mu \rangle_{\mathcal{U},\mathcal{U}^*}$. Now, extremal points of the unit ball in $\mathcal{U}^*$ are Dirac delta measures, thus $\delta = \delta_{s^*}$ or $\delta = -\delta_{s^*}$ with $s^* \in \mathcal{K}$. In this case, it is clear that

$$\int_{\mathcal{K}} \int_{\mathcal{K}} k_X(s,t)\delta_{s^*}(dt)\delta_{s^*}(ds) = k_X(s^*, s^*),$$

and the result follows. □

This principle, jointly with previous decomposition, gives the following result.

Corollary 11. *Let $(\mathcal{K}, d)$ be a compact metric space, $X = (X_s)_{s \in \mathcal{K}}$ a continuous Gaussian random field with covariance kernel k_X. Take $k_0 := k$ and define by induction:*

$$\begin{aligned} s_n &= \arg\max_{s \in \mathcal{K}} k_n(s,s), \\ \lambda_n &= k_n(s_n, s_n), \\ h_n(s) &:= \frac{k_n(s, s_n)}{\sqrt{k_n(s_n, s_n)}}, \\ k_{n+1}(s,t) &:= k_n(s,t) - h_n(s)h_n(t). \end{aligned}$$

The sequence $(h_n)_{n \in \mathbb{N}} \subset \mathcal{H}_k \subset C(\mathcal{K}, \mathbb{R})$ is admissible for X.

Furthermore, the precised quantification of uncertainty from corollary 10 is translated as follows.

Corollary 12. *Let $(\mathcal{K}, d)$ be a compact metric space, $X = (X_s)_{s\in\mathcal{K}}$ a continuous Gaussian random field with covariance kernel k_X. With the notations from corollary 11, note $\forall n \in \mathbb{N}$, X^n the continuous Gaussian process with kernel k_n then:*

$$\sup_{\mu\in\mathcal{U}^*} \mathbb{V}\left[\int_{\mathcal{K}} X_s^n \mu(ds)\right] = \sup_{s\in\mathcal{K}} \mathbb{V}\left[X_s^n\right] = \lambda_n.$$

A direct proof for covariance kernels As it was showed in chapter 1, the series representation of a continuous Gaussian process is the same problem as finding tensor representations of its covariance kernel. The extended Karhunen-Loève decomposition (described for general Banach spaces) is given here directly in the language of Reproducing Kernel Hilbert spaces, the probabilistic aspect being already investigated. It is always assumed that the covariance kernel k is associated to a continuous Gaussian random field $X = (X_s)_{s\in\mathcal{K}}$ with $(\mathcal{K}, d)$ a compact metric set.

Lemma 8. *Let k a continuous, symmetric, semi-positive definite kernel and $\tilde{s} \in \mathcal{K}$ such that $k(\tilde{s}, \tilde{s}) > 0$ then*

$$k = k_1 + k_2,$$

where $\forall (s,t) \in \mathcal{K}^2$,

$$k_1(s,t) = \frac{k(s,\tilde{s})k(t,\tilde{s})}{k(\tilde{s},\tilde{s})},$$
$$k_2(s,t) = k(s,t) - k_1(s,t),$$

are continuous, symmetric, semi-positive definite kernels such that

$$\mathcal{H}_k = \mathcal{H}_{k_1} \oplus \mathcal{H}_{k_2},$$

and $\forall i \in \{1,2\}$, $\|.\|_{k_i} = \|.\|_k$ on $\mathcal{H}_{k_i}$.

Proof. Let $\tilde{s}$ such that $k(\tilde{s}, \tilde{s}) > 0$, then $\forall s \in \mathcal{K}$:

$$X_s = X_s^1 + X_s^2,$$
$$X_s^1 = X_{\tilde{s}} \frac{k(.,s)}{k(\tilde{s},\tilde{s})},$$
$$X_s^2 = X_s - X_s^1.$$

Y^1, Y^2 are two independent Gaussian random fields, the rest of the proof is similar with lemma 5. □

Proposition 37. *Let k a continuous, symmetric, semi-positive definite kernel and $s = \arg\max_{t\in\mathcal{K}} k(t,t)$ then*

$$\sup_{h\in\mathcal{B}_{\mathcal{H}_k}} \|h\|_{C(\mathcal{K},\mathbb{R})} = \sqrt{k(s,s)}.$$

Proof. Using the reproducing property, it comes:

$$\sup_{h\in\mathcal{B}_{\mathcal{H}_k}} \|h\|_{C(\mathcal{K},\mathbb{R})} = \sup_{h\in\mathcal{B}_{\mathcal{H}_k}} \sup_{t\in\mathcal{K}} |h(t)| = \sup_{h\in\mathcal{B}_{\mathcal{H}_k}} \sup_{t\in\mathcal{K}} |\langle h, k(t,.)\rangle_k|.$$

Now, Cauchy-Schwarz inequality implies

$$|\langle h, k(t,.)\rangle_k| \leq \|h\|_k \sqrt{k(t,t)},$$

thus $\sup_{h\in\mathcal{B}_{\mathcal{H}_k}} \|h\|_{C(\mathcal{K},\mathbb{R})} \leq \sup_{t\in\mathcal{K}} \sqrt{k(t,t)}$. However, $f = \frac{k(s,.)}{\sqrt{k(s,s)}}$ is of unit norm in $\mathcal{H}_k$ and $f(s) = \sqrt{k(s,s)}$ thus $\|f\|_{C(\mathcal{K},\mathbb{R})} \geq \sqrt{k(s,s)}$ and the proof is complete. □

Lemma 9. *Let k a symmetric, semi-positive definite and continuous kernel, if $(h_n)_{n\in\mathbb{N}}$ is orthonormal in $\mathcal{H}_k$ then $\|h_n\|_{C(\mathcal{K},\mathbb{R})} \to 0$.*

Proof. The unit ball $\mathcal{B}_{\mathcal{H}_k}$ is compact, thus $\exists h_\infty \in \mathcal{H}_k$ and $(h_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} h_{n_k} = h_\infty$ in $C(\mathcal{K},\mathbb{R})$. Moreover, $\forall s \in \mathcal{K}$, $h_n(s) \to 0$ since (Bessel's inequality):

$$\sum_{n\geq 0} h_n(s)^2 = \sum_{n\geq 0} \langle h_n, k(.,s)\rangle_k^2 \leq \|k(.,s)\|_k^2,$$

$(h_n(s))_{n\in\mathbb{N}} \in l^2(\mathbb{N})$. It follows that $h_{n_k} \to 0$ and $\|h_n\|_{C(\mathcal{K},\mathbb{R})}$ tends to 0. □

Theorem 14. *Let k a symmetric, semi-positive definite and continuous kernel such that $\mathcal{H}_k$ is infinite dimensional. Let $k_0 := k$ and define by induction $\forall n \in \mathbb{N}$:*

$$\begin{aligned} s_n &= \arg\max_{s\in\mathcal{K}} k_n(s,s), \\ \lambda_n &= k_n(s_n,s_n), \\ h_n(s) &:= \frac{k_n(s_n,s)}{\sqrt{\lambda_n}}, \\ k_{n+1}(s,t) &:= k_n(s,t) - h_n(s)h_n(t), \end{aligned}$$

then the family $(h_n)_{n\in\mathbb{N}}$ is a Hilbert basis in $\mathcal{H}_k$.

Proof. Since $\mathcal{H}_k$ is infinite dimensional, $k = k_0$ is non-trivial and one has $k_0(s_0,s_0) > 0$ thus k_1 is well-defined. From lemma 8, both k_1 and $h_0 \otimes h_0$ are symmetric, continuous, semi-positive definite kernels and

$$\mathcal{H}_k = \mathcal{H}_0 \oplus \mathcal{H}_{k_1}$$

with $\mathcal{H}_{k_1}$ infinite dimensional. Now, let $n \in \mathbb{N}$ and suppose that k_n is a well-defined, symmetric, semi-positive definite and continuous kernel with $\mathcal{H}_{k_n}$ an infinite dimensional RKHS. Again, the application of lemma 8 establishes the same properties for k_{n+1}. By induction, it is true $\forall n \in \mathbb{N}$. Now, by construction the family is orthonormal in $\mathcal{H}_k$, it remains to see that it is a basis. Let $h \in \mathcal{H}_k$ such that $\forall n \in \mathbb{N}$, $\langle h, h_n\rangle_k = 0$, it comes that $h \in \mathcal{H}_{k_n}$, $\forall n \in \mathbb{N}$, thus:

$$\forall n \in \mathbb{N},\ \forall s \in \mathcal{K},\ |h(s)| = |\langle h, k_n(.,s)\rangle_k| \leq \|h\|_k \sqrt{k_n(s,s)} \leq \|h\|_k \|h_n\|_{C(\mathcal{K},\mathbb{R})}.$$

From lemma 9, $\|h_n\|_{C(\mathcal{K},\mathbb{R})} \to 0$, thus $h = 0$ and the proof is complete. □

This strategy is distinct from [Git12], as the sequence of points $(s_n)_{n\in\mathbb{N}}$ is not necessary dense, nor given a priori. However, the process of building orthogonal functions is similar to the Gram-Schmidt orthogonalisation.

2.3 Examples in $\mathcal{C}(\mathcal{K},\mathbb{R})$

In this section, different examples of continuous Gaussian random fields are given. In a first part, they are provided analytically and later on by a numerical algorithm. As a reminder of notations, the decomposition is as follows:

$$\forall s\in\mathcal{K},\ X_s=\sum_{n\geq 0}\sqrt{\lambda_n}\xi_n u_n(s)=\sum_{n\geq 0}\xi_n h_n(s),$$

with $\forall n\in\mathbb{N},\ \|h_n\|_k=\|u_n\|_{C(\mathcal{K},\mathbb{R})}=1$. Bases functions will refer to the sequence $(h_n)_{n\in\mathbb{N}}$ while their *normalized* version is $(u_n)_{n\in\mathbb{N}}$.

Analytical examples in $C([0,1],\mathbb{R})$ It appears that the same examples of analytical Karhunen-Loève decomposition in $L^2([0,1],ds;\mathbb{R})$ are available here.

Example 11 (Wiener process)**.** *Consider again the standard Wiener process* $W=(W_s)_{s\in\mathcal{K}}$ *on* $\mathcal{K}=[0,1]$*. Note* $k_0:=k$*, the first step in the decomposition is to find* $s_0\in[0,1]$ *such that*

$$s_0=\arg\max_{s\in[0,1]}k_0(s,s).$$

Here, $\forall s\in[0,1],\ k_0(s,s)=s$*, thus the maximum is obtained at* $s_0=1$*,* $\lambda_0=k_0(s_0,s_0)=1$ *and* $h_0(s)=s$*. Now, the residual process corresponds to the Brownian bridge* $(B_s)_{s\in[0,1]}=(W_s-W_1 s)_{s\in[0,1]}$ *with covariance kernel*

$$k_1:(t,s)\in[0,1]^2\to\mathbb{E}\left[B_t B_s\right]=\min(s,t)-ts.$$

In this case, the maximum of variance is obtained at $s_1=\frac{1}{2}$ *since* $k_1(s,s)=s(1-s)$ *and it comes* $\lambda_1=\frac{1}{4}$ *and* $h_1(s)=\frac{1}{2}\left(\min(s,\frac{1}{2})-\frac{s}{2}\right)$*. The residual process is given by the conditional Wiener process* $W|W_{\frac{1}{2}}=W_1=0$*. Since it is Markovian, the study of the process can be considered independently on* $[0,\frac{1}{2}]$ *and* $[\frac{1}{2},1]$*. On both parts, it is a Brownian bridge. By induction, one has directly the following:*

$$\forall n\in\mathbb{N},\ \lambda_n=\frac{1}{2^{p+2}}\ \text{ for } n=2^p+k,\ k=0,...,2^p-1\ \text{ and } p\geq 0.$$

Furthermore, the Hilbert basis $(h_n)_{n\in\mathbb{N}}\subset\mathcal{H}_k$ *is given by* $h_0(t)=t$ *and* $h_n(t)=\int_0^t h_n'(s)ds,\ n\geq 0$*, where*

$$h_n'(s)=\begin{cases}\sqrt{2^p}\ \text{for}\ \frac{2k}{2^{p+1}}\leq s\leq\frac{2k+1}{2^{p+1}}\\ -\sqrt{2^p}\ \text{for}\ \frac{2k+1}{2^{p+1}}<s\leq\frac{2k+2}{2^{p+1}}\\ 0\ \text{otherwise}\end{cases},$$

if $n=2^p+k,\ k=0,...,2^p-1$ *and* $p\geq 0$*. The family* $(h_n')_{n\in\mathbb{N}}$ *is the usual Haar basis of* $L^2([0,1],\mathbb{R})$*. The normalized functions* $u_n=\frac{h_n}{\sqrt{\lambda_n}}$ *are Schauder's functions*

$$u_n(s)=\sqrt{2^{p+2}}h_n(s)$$

corresponding to hat functions of height 1 *and lying above the intervals* $\left[\frac{k}{2^p}, \frac{k+1}{2^p}\right]$ $(n = 2^p + k)$. *The resulting decomposition* $\sum_{n\geq 0} \xi_n(\omega) h_n$ *is the famous Lévy-Ciesielski construction of Wiener process on the interval* $[0,1]$ *(see [JK69]). Remark that*

$$\sum_{n\geq 0} \|h_n\|_{\mathcal{U}}^2 = \sum_{n\geq 0} \lambda_n = 1 + \frac{1}{4} + 2 \times \frac{1}{8} + 4 \times \frac{1}{16} + ... = +\infty,$$

due to the "multiplicity" and this representation is not nuclear.

The decomposition of the Brownian bridge is obtained in example 11 since it appears as the residual process after the first iteration.

Example 12 (Ornstein-Uhlenbeck process)**.** *We already know that the Ornstein-Uhlenbeck process is Gaussian with covariance kernel given as:*

$$k(s,t) = \frac{\sigma}{2\beta} e^{-\beta|t-s|},$$

and σ *will be supposed equal to* 2β *without loss of generality. Here, the process is stationary and any value* $s \in [0,1]$ *is maximum and valid for the first iteration. However, we choose* $s_0 = 0$ *as it provides an initial condition of equation 1.1. It immediately follows that the conditional process has the following covariance kernel:*

$$k_1 : (t,s) \in [0,1]^2 \to e^{-\beta|t-s|} - e^{-\beta t} e^{-\beta s}.$$

The maximum of $t \to k_1(t,t)$ *is obtained with* $s_1 = 1$ *and it comes*

$$\begin{aligned} \lambda_1 &= 1 - e^{-2\beta}, \\ u_1(t) &= \frac{1}{\lambda_1}\left(e^{-\beta(1-t)} - e^{-\beta} e^{-\beta t}\right). \end{aligned}$$

The next iteration will be the last needed to obtain the full decomposition of the process. The conditional Gaussian process has covariance kernel

$$k_2(t,s) = k_1(t,s) - \lambda_1^{-1} k_1(t,s_1) k_1(s,s_1).$$

A straightforward computation shows that the variance function $t \to k_2(t,t)$ *is maximum at* $s_2 = \frac{1}{2}$ *that leads to*

$$\begin{aligned} \lambda_2 &= 1 - 2\frac{e^{-\beta}}{1+e^{-\beta}}, \\ u_2(t) &= \frac{1}{\lambda_2}\left(e^{-\beta(t-\frac{1}{2})} - e^{-\frac{\beta}{2}} \frac{e^{-\beta t} + e^{-\beta(1-t)}}{1+e^{-\beta}}\right). \end{aligned}$$

From this and since the Ornstein-Ulhenbeck process is Markovian, we deduce the shape of all other functions. Indeed, all further steps will be given on dyadic intervals of the form $[\frac{k-1}{2^p}, \frac{k}{2^p}]$ *with* $k \in [1, ..., 2^p]$, *the process being independent on these intervals. Here is the general form of the basis for* $n = 2^p + k$ *with* $p \geq 0$ *and* $k \in [1, ..., 2^p]$*:*

$$\begin{aligned} \lambda_n &= 1 - 2\frac{e^{-\frac{\beta}{2^p}}}{1+e^{-\frac{\beta}{2^p}}}, \\ u_n(t) &= \frac{1}{\lambda_n}\left[e^{-\beta\left(t-\frac{k-\frac{1}{2}}{2^p}\right)} - e^{-\frac{\beta}{2^{p+1}}} \frac{e^{-\beta\left(t-\frac{k-1}{2^p}\right)} + e^{-\beta\left(\frac{k}{2^p}-t\right)}}{1+e^{-\frac{\beta}{2^p}}}\right]. \end{aligned}$$

Remark that these analytical examples rely heavily on the Markovian nature of the considered processes, as it provides a sort of self-similarity argument.

Numerical examples Since all one dimensional continuous Gaussian random fields are not Markovian, the generalized Karhunen-Loève decomposition is not always available analytically. However, the particular nature of the decomposition provides a practical method to numerically build it. Indeed, the search for maximum variance functionals translates into maximizing a continuous function on a compact space. Algorithm 1 provides a naïve method to build such basis. All the following numerical examples are obtained on a regular desktop machine (late 2015), using Python. The chosen optimization library is Pyswarm (`https://pythonhosted.org/pyswarm/`), as swarm-type algorithms provide robustness to local maximums.

Data: A compact metric set $(\mathcal{K}, d)$;
a continuous, symmetric, semi-positive definite kernel k;
an integer $n \in \mathbb{N}$.
Result: A vector of points $(s_0, ..., s_{n-1})$;
a vector of maximal residual variances $(\lambda_0, ..., \lambda_{n-1})$;
a list of basis functions $(h_0, ..., h_{n-1})$.
Set $k_0 = k$ and $i = 0$;
while $i < n$ **do**
  find $s_i := \arg\max_{t \in \mathcal{K}} k_i(t, t)$;
  put $\lambda_i := k_i(s_i, s_i)$;
  put $h_i := \frac{k_i(.,s_i)}{\sqrt{k_i(s_i,s_i)}}$;
  if $\lambda_i > 0$ **then**
    Set $k_{i+1} := (s, t) \in \mathcal{K}^2 \to k_i(s, t) - h_i(s)h_i(t)$;
    set $i := i + 1$;
  else
    stop;
  end
end

Algorithm 1: Build the n first basis functions.

Example 13 (Squared exponential kernel)**.** *Let $X^1 = (X^1_s)_{s\in[0,1]}$ be the Gaussian process with squared exponential kernel:*

$$k_{X_1}(s, t) = \exp\left(-(t - s)^2\right).$$

A basis is built considering $s_0 = 0$ and using algorithm 1. The 8 first normalized basis functions $(u_0, ..., u_7)$ are presented in figure 2.1 as well as their respective variances $(\lambda_0, ..., \lambda_7)$ in figure 2.2. It appears that most of the variability of the process is

captured with 8 basis functions. Indeed, the maximum of residual variance is less than 10^{-8} of the initial variance.

Example 14 (Matérn $\nu = \frac{3}{2}$ kernel)**.** *Let $X^2 = (X_s^2)_{s\in[0,1]}$ be the Gaussian process with Matérn $\frac{3}{2}$ kernel:*

$$k_{X_2}(s,t) = \left(1 + \sqrt{3}|s-t|\right) \exp\left(-\sqrt{3}|s-t|\right).$$

A basis is built considering $s_0 = 0$ and using algorithm 1. The normalized functions are presented in figure 2.3 as well as the associated variances in figure 2.4.

Example 15 (Fractional Brownian motion $H = \frac{1}{4}$)**.** *Let $X^3 = (X_s^3)_{s\in[0,1]}$ be the Gaussian process with fractional Brownian motion kernel:*

$$k_{X_3}(s,t) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right).$$

where H is the index. A basis is built using algorithm 1 for both $H = \frac{1}{4}$ and $H = \frac{3}{4}$. The normalized functions are presented in figures 2.5 and 2.7 as well as the associated variances in figures 2.6 and 2.8.

2.4 A study of the standard Wiener process

This last section will be dedicated to a comparison of different admissible sequences for the standard Wiener process in $C([0,1],\mathbb{R})$. In particular, if $(h_n)_{n\in\mathbb{N}}$ is an admissible sequence, then

$$\forall s \in [0,1],\ W_s = \sum_{n\geq 0} \xi_n h_n(s),\ a.s.$$

with a uniform convergence. Two distinct measures will be considered for every admissible sequence $(h_n)_{n\in\mathbb{N}}$, namely:

1. l-error:

$$\forall n \in \mathbb{N},\ l_n((h_n)_{n\in\mathbb{N}}) = \mathbb{E}\left[\left(\sup_{s\in[0,1]} \sum_{k\geq n} \xi_k h_k(s)\right)^2\right]^{\frac{1}{2}},$$

2. a-error:

$$\forall n \in \mathbb{N},\ a_n((h_n)_{n\in\mathbb{N}}) = \left(\sup_{s\in[0,1]} \sum_{k\geq n} h_k(s)^2\right)^{\frac{1}{2}}.$$

In this special case, both $l_n(W)$ and $a_n(W)$ convergence rates are known asymptotically and will be compared with $l_n((h_n)_{n\in\mathbb{N}})$ and $a_n((h_n)_{n\in\mathbb{N}})$.

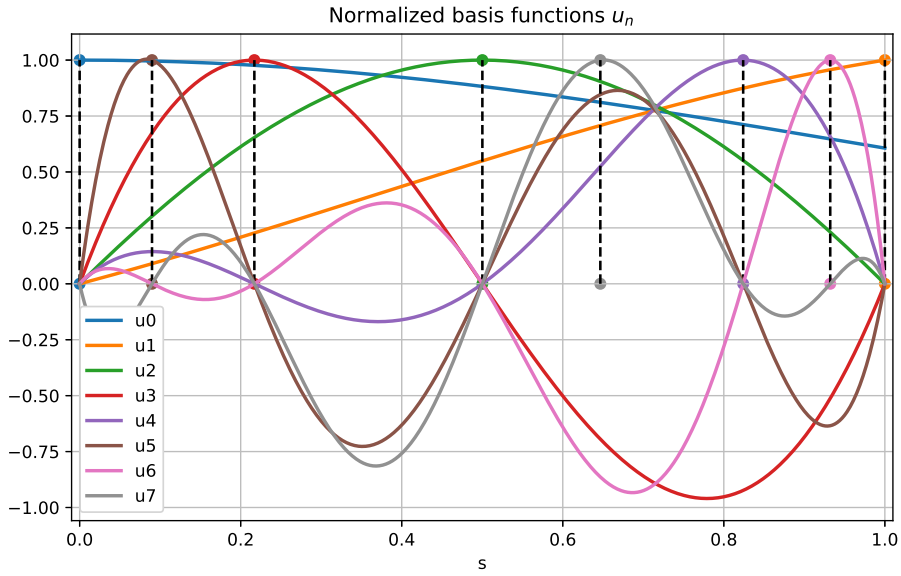

Figure 2.1: 8 first normalized basis functions $(u_0, ..., u_7)$ (squared exponential kernel). The dots represent positions $(s_0, ..., s_7)$ of associated Dirac functionals of maximum variance.

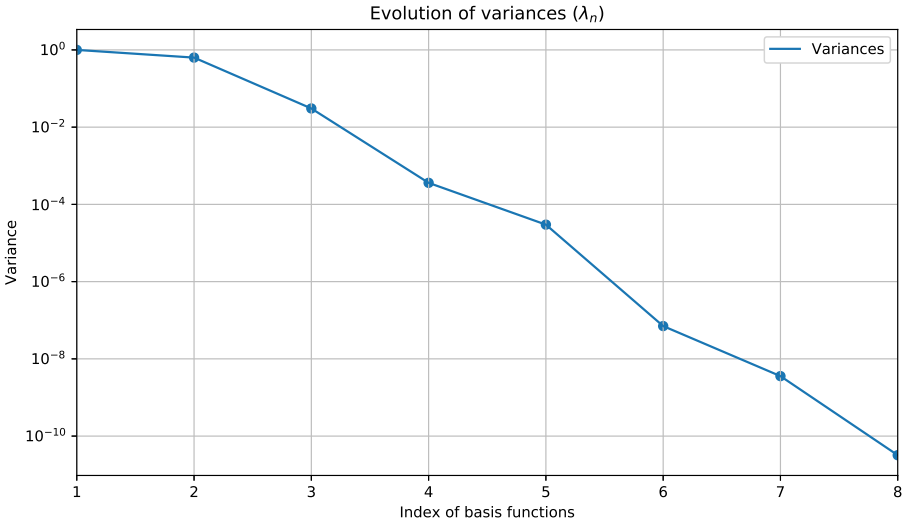

Figure 2.2: Evolution of 8 first basis functions variances $(\lambda_0, ..., \lambda_7)$ (squared exponential kernel).

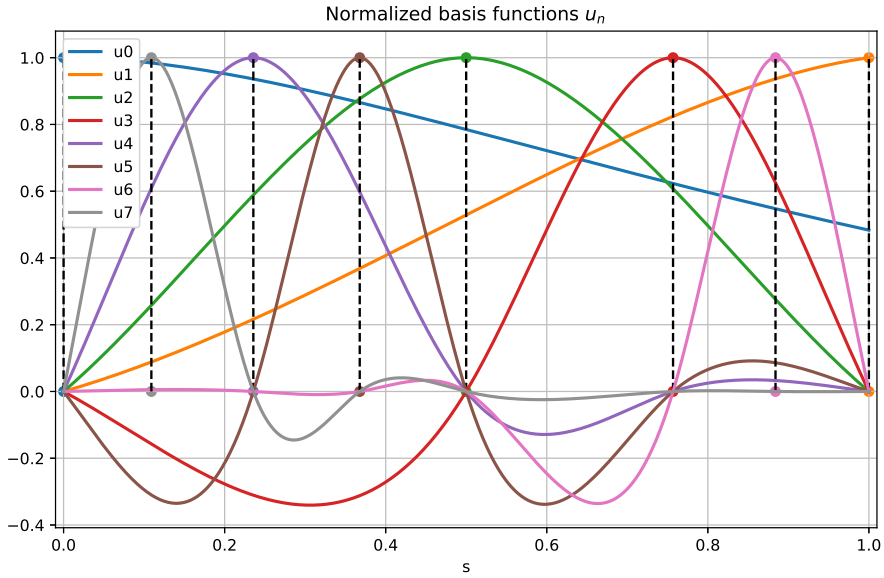

Figure 2.3: 8 first normalized basis functions $(u_0, ..., u_7)$ (Matérn $\nu = \frac{3}{2}$ kernel). The dots represent positions $(s_0, ..., s_7)$ of associated Dirac functionals of maximum variance.

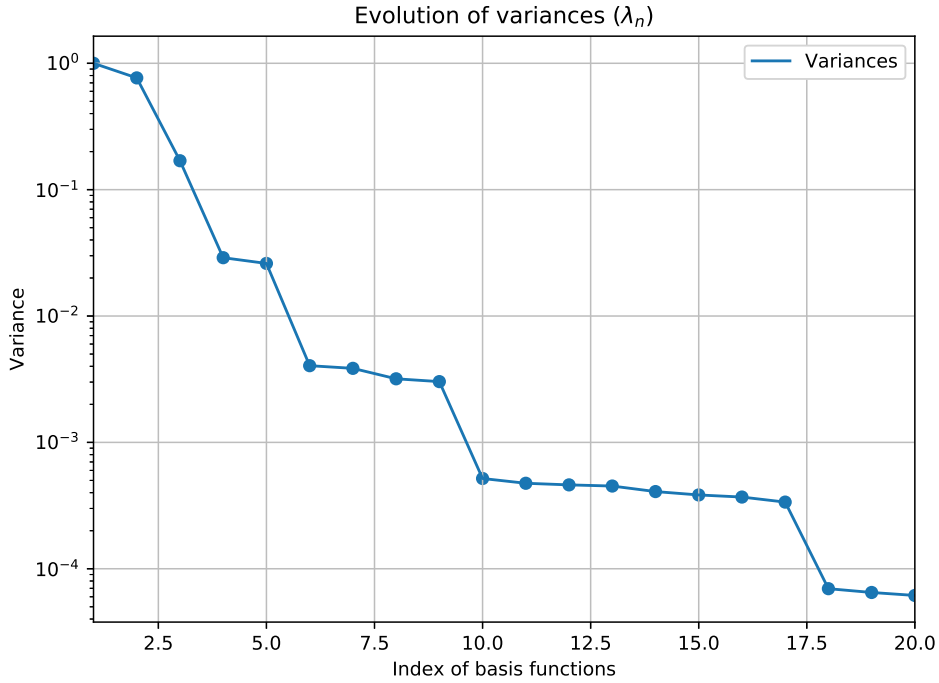

Figure 2.4: Evolution of 8 first basis functions variances $(\lambda_0, ..., \lambda_7)$ (Matérn $\nu = \frac{3}{2}$ kernel).

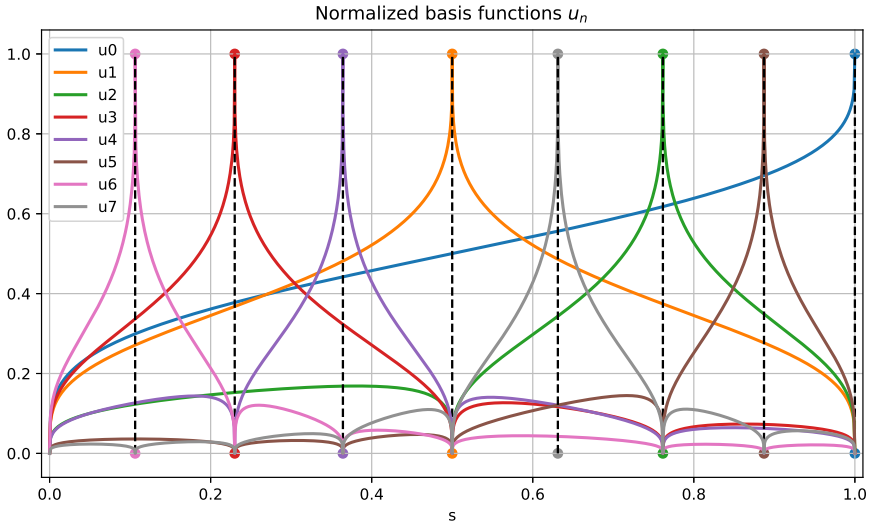

Figure 2.5: 8 first normalized basis functions $(u_0, ..., u_7)$ (Fbm $H = \frac{1}{4}$ kernel). The dots represent positions $(s_0, ..., s_7)$ of associated Dirac functionals of maximum variance.

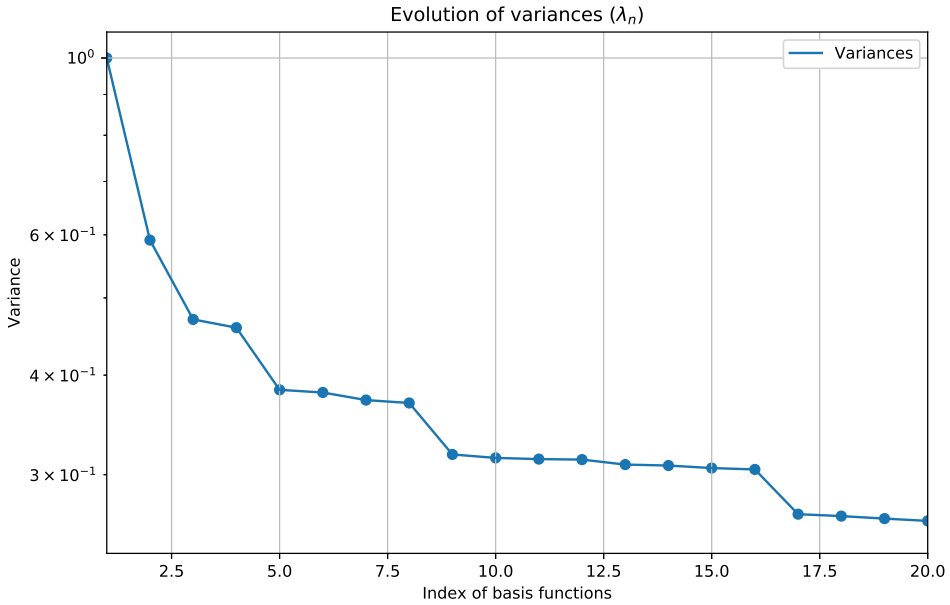

Figure 2.6: Evolution of 8 first basis functions variances (Fbm $H = \frac{1}{4}$ kernel).

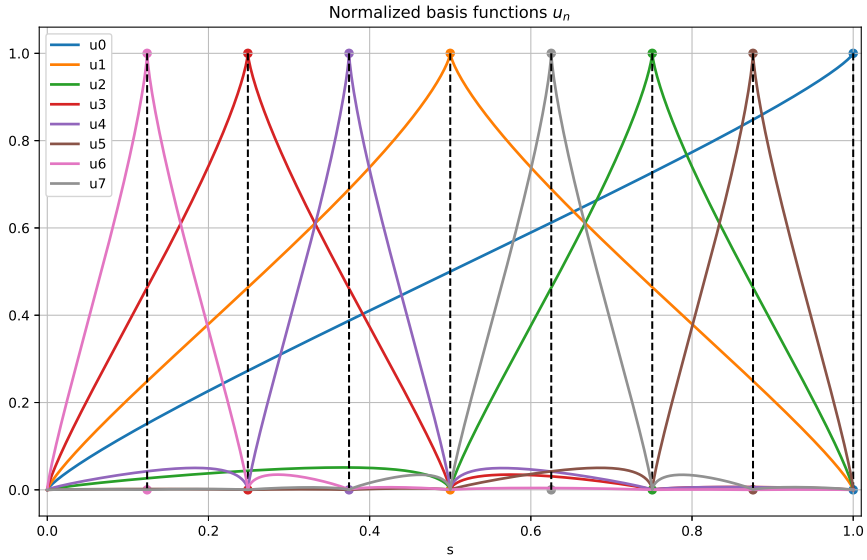

Figure 2.7: 8 first normalized basis functions $(u_0, ..., u_7)$ (Fbm $H = \frac{3}{4}$ kernel). The dots represent positions $(s_0, ..., s_7)$ of associated Dirac functionals of maximum variance.

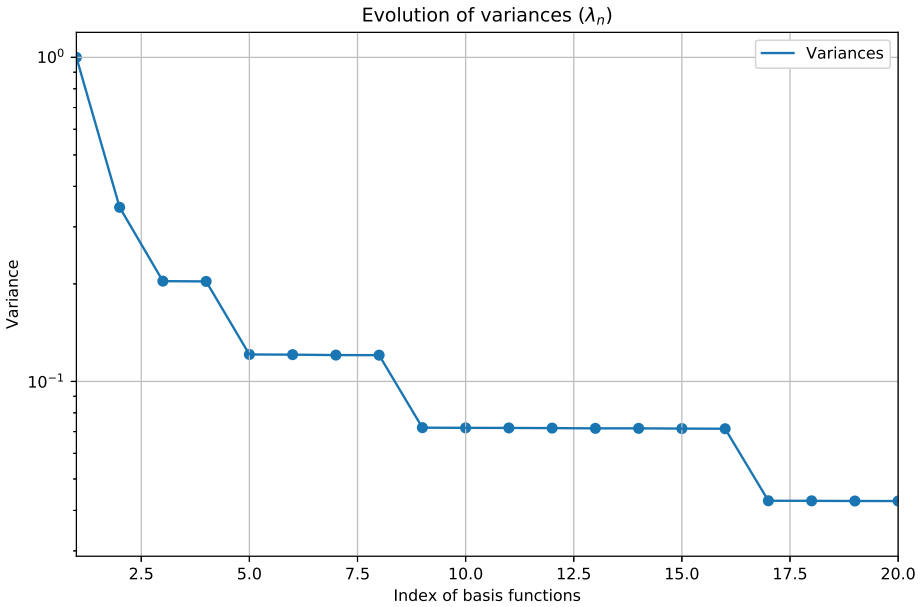

Figure 2.8: Evolution of 8 first basis functions variances (Fbm $H = \frac{3}{4}$ kernel).

A few admissible sequences In this thesis, all admissible sequences presented can be derived using the following covariance factorization.

Proposition 38. *Let $\mathcal{K} = [0,1]$ and consider the standard Wiener process with covariance kernel k_W and operator $\mathcal{C}_W$, then the following operator:*

$$A_W : f \in L^2([0,1], ds) \to \left(s \to \int_0^s f(t)dt \right) \in \mathcal{U} = C\left([0,1], \mathbb{R}\right),$$

provides a factorization of $\mathcal{C}_W$ in $H = L^2([0,1], ds; \mathbb{R})$.

Proof. First, the Cauchy-Schwarz inequality gives

$$\forall f \in L^2([0,1], ds),\ \forall (s,t) \in [0,1]^2,\ |A_W f(s) - A_W f(t) \leq \sqrt{|s-t|}\, \|f\|_{L^2}$$

thus functions $A_W f$ are continuous, the operator is thus well-defined and linear. Furthermore, it is clearly bounded since $f \in L^2([0,1], ds; \mathbb{R})$ and it has an adjoint. It appears that $\forall (\mu, f) \in \mathcal{U}^* \times L^2([0,1], ds; \mathbb{R})$:

$$\begin{aligned}
\langle A_W f, \mu \rangle_{\mathcal{U},\mathcal{U}^*} &= \int_0^1 A_W f(s) \mu(ds), \\
&= \int_0^1 \int_0^1 \chi_{[0,s]}(t) f(t) dt \mu(ds), \\
&= \int_0^1 \int_0^1 \chi_{[0,s]}(t) \mu(ds) f(t) dt, \\
&= \int_0^1 \mu([t,1]) f(t) dt, \\
&= \langle A_W^* \mu, f \rangle_{L^2([0,1],ds)},
\end{aligned}$$

from which $A_W^* := \mu \to (s \to \mu([s,1]))$ is identified. Finally, one has:

$$\begin{aligned}
\forall s \in [0,1],\ A_W A_W^* \mu(s) &= \int_0^s \mu([t,1]) dt, \\
&= \int_0^1 \int_0^1 \chi_{[0,s]}(t) \chi_{[t,1]}(z) \mu(dz) dt, \\
&= \int_0^1 \int_0^1 \chi_{[0,s]}(t) \chi_{[t,1]}(z) dt \mu(dz), \\
&= \int_0^1 k_W(s,z) \mu(dz), \\
&= \mathcal{C}_W \mu(s).
\end{aligned}$$

□

Using theorem 5, the image through A_W of any basis from $L^2([0,1], ds; \mathbb{R})$ provides an admissible sequence for W. This is what will be done in the next examples.

Example 16 (Karhunen-Loève in L^2)**.** *Consider the following Hilbert basis in $L^2([0,1], ds; \mathbb{R})$:*

$$\forall n \in \mathbb{N},\ e_n(s) = \sqrt{2} \cos\left(\pi \left(n + \frac{1}{2} \right) s \right)$$

then the associated admissible sequence using A_W is the usual $L^2([0,1],ds;\mathbb{R})$ Karhunen-Loève basis:

$$\forall n \in \mathbb{N},\ A_W e_n(s) = \frac{\sqrt{2}}{\pi(n+\frac{1}{2})} \sin\left(\pi\left(n+\frac{1}{2}\right)s\right).$$

Example 17 (Lévy-Cielsieski)**.** *Consider the following (Haar) Hilbert basis in $L^2([0,1],ds;\mathbb{R})$:*

$$\forall n \in \mathbb{N},\ n = 2^p + k,\ k \in [0, 2^p - 1],\ p \geq 0,\ e_n(s) = \begin{cases} \sqrt{2^p} \text{ for } \frac{2k}{2^{p+1}} \leq s \leq \frac{2k+1}{2^{p+1}}, \\ -\sqrt{2^p} \text{ for } \frac{2k+1}{2^{p+1}} < s \leq \frac{2k+2}{2^{p+1}}, \\ 0 \text{ otherwise}, \end{cases}$$

then the admissible sequence obtained using A_W is the generalized Karhunen-Loève basis in $C([0,1],\mathbb{R})$ (Lévy-Cielsieski).

Example 18 (Paley-Wiener basis)**.** *Consider the following (Fourier) Hilbert basis in $L^2([0,1],ds;\mathbb{R})$:*

$$\begin{aligned} e_0(s) &= 1, \\ \forall n \in \mathbb{N},\ e_{2n+1}(s) &= \sqrt{2}\cos\left(2\pi(n+1)s\right), \\ \forall n \in \mathbb{N},\ e_{2n+2}(s) &= \sqrt{2}\sin\left(2\pi(n+1)s\right), \end{aligned}$$

then the admissible sequence obtained using A_W is the Paley-Wiener basis:

$$\begin{aligned} A_W e_0(s) &= s, \\ A_W e_{2n+1}(s) &= \frac{1}{\sqrt{2}\pi(n+1)}\left(1 - \cos\left(2\pi(n+1)s\right)\right), \\ A_W e_{2n+2}(s) &= \frac{1}{\sqrt{2}\pi(n+1)}\sin\left(2\pi(n+1)s\right). \end{aligned}$$

Example 19 (Sinus)**.** *Consider the following Hilbert basis in $L^2([0,1],ds;\mathbb{R})$:*

$$\forall n \in \mathbb{N},\ e_n(s) = \sqrt{2}\sin\left(\pi\left(n+1\right)s\right)$$

then the admissible sequence obtained using A_W is the following basis:

$$\forall n \in \mathbb{N},\ A_W e_n(s) = \frac{\sqrt{2}}{\pi(n+1)}\left(1 - \cos\left(\pi\left(n+1\right)s\right)\right).$$

The previous factorization can be used to construct other admissible sequences, using the more general notion of Hilbert frames in $L^2([0,1],ds;\mathbb{R})$ instead of Hilbert basis [LP09] (Wavelets for instance). It is also possible to build bases by specifying a dense sequence $(t_n)_{n\in\mathbb{N}}$ in $[0,1]$, along with a Gram-Schmidt construction [Git12]. The last admissible sequence that will be given here for the standard Wiener process is a transformation of the $C([0,1],\mathbb{R})$ Karhunen-Loève basis (Lévy-Ciesielski).

Example 20 (Rotated Lévy-Ciesielski)**.** *Let $\mathcal{K} = [0,1]$ and $W = (W_t)_{t\in[0,1]}$ the standard Wiener process and the admissible sequence $(h_n)_{n\in\mathbb{N}}$ given in example 11*

(Lévy-Ciesielski basis). The specific structure of this basis will now be exploited to define a new basis. Note $\tilde{h}_0 = h_0$, $\tilde{h}_1 = h_1$ *and consider* $\tilde{h}_2, \tilde{h}_3$, *then*

$$\tilde{h}_2 = \frac{1}{\sqrt{2}}(h_2 + h_3),$$
$$\tilde{h}_3 = \frac{1}{\sqrt{2}}(h_2 - h_3),$$

are both orthogonal with unit norm in $\mathcal{U}_X$ *and* $\operatorname{span}\{h_2, h_3\} = \operatorname{span}\left\{\tilde{h}_2, \tilde{h}_3\right\}$. *Repeating a similar transformation at for* $(h_4, ..., h_7)$, $(h_8, ..., h_{15})$ *and so on and so forth gives a new basis, called rotated Lévy-Ciesielski.*

l-optimality The comparison starts with the question of asymptotic optimality (related to l-numbers). It has been shown in the literature that the optimal rate is the following for the Wiener standard process:

$$l_n(W) \approx \sqrt{\frac{\log(n)}{n}}.$$

Using proposition 29 which gives a sufficient set of conditions, it is immediate to show that the following bases are indeed asymptotically optimal and figure 2.9 provides a Monte-Carlo illustration for the previous admissible sequences (10^4 simulations on a fine grid for each n between 0 and 31).

Proposition 39. *Let* $\mathcal{K} = [0, 1]$, $W = (W_s)_{s\in[0,1]}$ *be the standard Wiener process, then the following admissible sequences:*

- *Karhunen-Loève in* $L^2([0, 1], ds; \mathbb{R})$,
- *Paley-Wiener,*
- *Sinus,*
- *Lévy-Ciesielski (Karhunen-Loève in* $C([0, 1], \mathbb{R})$*),*
- *Rotated Lévy-Ciesielski,*

are all asymptotically optimal for W.

Proof. The proof relies on an application of proposition 29 to the different admissible sequences. First, note that the eigenvalues of the covariance operator $\mathcal{C}_W$ (as an endomorphism in $L^2([0, 1], ds; \mathbb{R})$) are known:

$$\forall n \in \mathbb{N},\ \lambda_n = \frac{1}{\pi^2(n + \frac{1}{2})^2},$$

thus item 1 in proposition 29 is satisfied for $\nu = 1$. It remains to check item 2 and 3 relative to infinite and β-Hölder norms.

- Karhunen-Loève in $L^2([0, 1], ds; \mathbb{R})$. It has already be done in example 10.

- Paley-Wiener. The basis functions are built upon cosines and sinuses, it is clear that

$$\begin{aligned} \|h_0\|_{\mathcal{U}} &= 1, \\ \forall n \in \mathbb{N},\ \|h_{2n+1}\|_{\mathcal{U}} &= \frac{\sqrt{2}}{\pi(n+1)}, \\ \forall n \in \mathbb{N},\ \|h_{2n+2}\|_{\mathcal{U}} &= \frac{1}{\sqrt{2}\pi(n+1)}, \end{aligned}$$

  thus $\|h_n\|_{\mathcal{U}} \leq C(n+1)^{-1}$ with C sufficiently big. Concerning the β-Hölder norms, one has:

$$\begin{aligned} \|h_0\|_{\beta} &= 1, \\ \forall n \in \mathbb{N},\ \|h_{2n+1}\|_{\beta} &= 2^{\beta+1}(n+1)^{\beta}, \\ \forall n \in \mathbb{N},\ \|h_{2n+2}\|_{\beta} &= 2^{\beta+1}(n+1)^{\beta}, \end{aligned}$$

  and the proposition applies.

- The analysis is the same as Karhunen-Loève $L^2([0,1], ds; \mathbb{R})$.

- Rotated Lévy-Ciesielski. The transformation used reduces the infinite norm as follows:

$$\left\|\tilde{h}_2\right\|_{\mathcal{U}} = \left\|\frac{h_2 + h_3}{\sqrt{2}}\right\|_{\mathcal{U}} = \frac{1}{\sqrt{2}} \|h_2\|_{\mathcal{U}} = \sqrt{\frac{\lambda_2}{2}},$$

  since h_2 and h_3 have distinct supports. The result is similar at each step, thus $\forall n \geq 1,\ n = 2^p + k,\ k \in [0, 2^p - 1],\ p \geq 1$:

$$\left\|\tilde{h}_n\right\|_{\mathcal{U}} = \frac{1}{\sqrt{2^p}} \|h_n\|_{\mathcal{U}}.$$

  In other words, $(\tilde{h}_n)_{n\in\mathbb{N}}$ satisfies:

$$\forall n \in \mathbb{N},\ n = 2^p + k,\ k \in [0, 2^p - 1],\ p \geq 0,\ \left\|\tilde{h}_n\right\|_{\mathcal{U}} = \frac{1}{2^{p+1}}.$$

  This implies that:

$$\forall n \in \mathbb{N}^*,\ \left\|\tilde{h}_n\right\|_{\mathcal{U}} \leq \frac{1}{n+1},$$

  thus $(\tilde{h}_n)_{n\in\mathbb{N}}$ satisfies item 2 in proposition 29. Concerning the β-Hölder norms, it comes

$$\left\|\tilde{h}_n\right\|_{\beta} = \sup_{s\neq t} \frac{|\tilde{h}_n(t) - \tilde{h}_n(s)|}{|t-s|^{\beta}} = \frac{\frac{2}{\sqrt{2^{p+1}}}}{\frac{1}{2^{p\beta}}} = 2^{p(\beta-\frac{1}{2})+\frac{1}{2}} = 2^{(p+1)(\beta-\frac{1}{2})-1},$$

  and taking either $n = 2^p$ or $n = 2^{p+1} - 1$ gives a valid rate for item 3.

- Lévy-Cielsieski. The necessary conditions from proposition 29 are not satisfied in this case. Indeed, one only has

$$\|h_n\|_{\mathcal{U}} \leq \frac{C}{\sqrt{n+1}},\ C > 0.$$

However, since the rotated Lévy-Ciesielski admissible sequence is optimal, and by construction one has

$$\sum_{k=2^p}^{2^{p+1}} \xi_k h_k =^{Law} \sum_{k=2^p}^{2^{p+1}} \xi_k \tilde{h}_k$$

with $(\xi_n)_{n\in\mathbb{N}}$ a sequence of i.i.d. $\mathcal{N}(0,1)$, it comes:

$$\forall n \in \mathbb{N},\ \mathbb{E}\left[\left\|\sum_{k=0}^{2^n} \xi_k h_k\right\|_{\mathcal{U}}^2\right] = \mathbb{E}\left[\left\|\sum_{k=0}^{2^n} \xi_k \tilde{h}_k\right\|_{\mathcal{U}}^2\right],$$

from which optimality is deduced.

□

a-optimality The second feature that is interesting to observe is the evolution of a-errors, as they measure the linear reconstruction rate. These numbers have been shown to asymptotically satisfy (corollary 7.6 in [KL02]):

$$a_n(W) \approx \frac{1}{\sqrt{n}}.$$

It appears that all previous admissible sequences achieve this convergence rate.

Proposition 40. *Let $\mathcal{K} = [0,1]$, $W = (W_s)_{s\in[0,1]}$ be the standard Wiener process on $[0,1]$ with covariance kernel $\forall (s,t) \in [0,1]^2,\ k_W(s,t) = \min(s,t)$ and $W_0 = 0$ almost-surely. Note $(h_n)_{n\in\mathbb{N}}$ the admissible sequence obtained by Karhunen-Loève decomposition in $C([0,1],\mathbb{R})$:*

$$a_n((h_n)_{n\in\mathbb{N}}) = \sqrt{\lambda_n} \approx a_n(W).$$

Proof. In this construction, one has:

$$\lambda_n = \frac{1}{2^{p+2}},\ n = 2^p + k,\ k \in [0, 2^p - 1],\ p \geq 0,$$

and it comes that

$$\forall n \in \mathbb{N}^*,\ \frac{1}{4n} \leq \lambda_n \leq \frac{1}{2(n+1)} \leq \frac{1}{2n},$$

thus the proof is complete. □

Proposition 41. *Let $\mathcal{K} = [0,1]$, $W = (W_s)_{s\in[0,1]}$ be the standard Wiener process on $[0,1]$. Note $(h_n)_{n\in\mathbb{N}}$ the admissible sequence obtained by Karhunen-Loève decomposition in $L^2([0,1], ds; \mathbb{R})$ then*

$$a_n((h_n)_{n\in\mathbb{N}}) = \sqrt{\sum_{k\geq n} \lambda_n}.$$

Proof. Clearly, one has

$$\forall n \in \mathbb{N},\ \forall s \in [0,1],\ h_n(s)^2 \leq \lambda_n,$$

thus $\sup_{s\in[0,1]} \sum_{k\geq n} h_k(s)^2 \leq \sum_{k\geq n} \lambda_k$. Conversely,

$$\forall n \in \mathbb{N},\ h_n(1)^2 = \lambda_n,$$

thus $\sup_{s\in[0,1]} \sum_{k\geq n} h_n(s)^2 \geq \sum_{k\geq n} \lambda_k$. □

Figure 2.10 represents the maximum of residual kernels at each steps for the different admissible sequences at each step.

2.5 Conclusion

In this chapter, the important concept of Karhunen-Loève decomposition has been generalized from Hilbert to Banach spaces. It mimics the Hilbert case, where each eigenvector is maximizing the associated Rayleigh quotient. In particular, it provides a precise quantification of uncertainty, since the variance of all linear functionals of the residual Gaussian random elements are dominated. In the case of continuous Gaussian random fields, this decomposition translates in a recursive optimization algorithm, easy to implement. However, it is not clear yet if this new construction leads to l and a optimal admissible sequences. After a short analysis of the standard Wiener process, it appears it is interesting lead. Here are some open questions:

1. Does the Karhunen-Loève decomposition of a Gaussian random element X always lead to rate a-optimal and/or l-optimal bases ?

2. Does the rotation provided for the Lévy-Ciesielski basis for the standard Wiener process generalizes to other Gaussian random elements ?

3. In view of Courant-Fisher minimax principle in Hilbert spaces, is there a dual construction of the presented decomposition ?

4. In the literature of optimal design of experiments, the concept of sequential search for maximum variance points is generalized in *Stepwise Uncertainty Reduction* [BBG16]. Is this possible to adapt alternative criterion (trace, etc.) to derive new admissible sequences for continuous Gaussian random fields ?

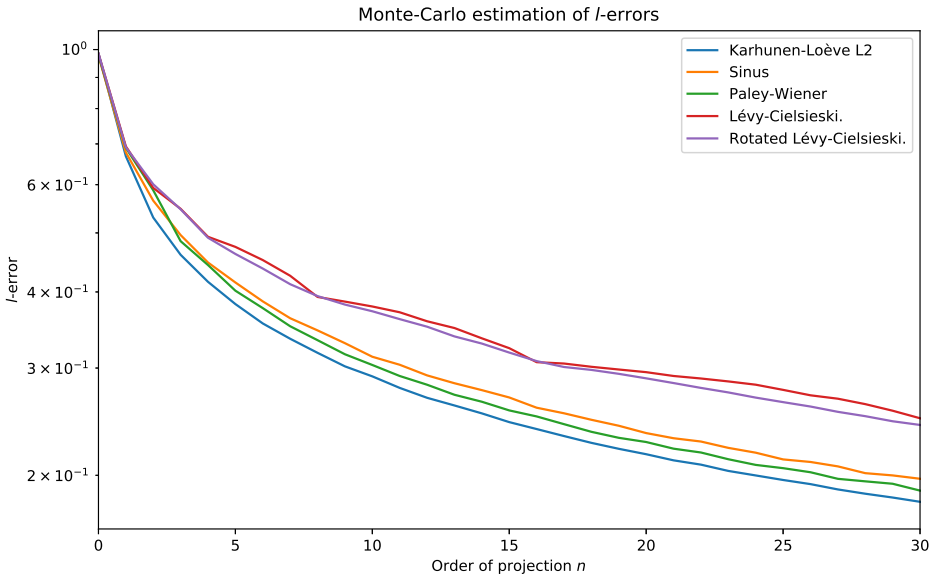

Figure 2.9: Numerical comparison of l-errors of admissible sequences for the standard Wiener process.

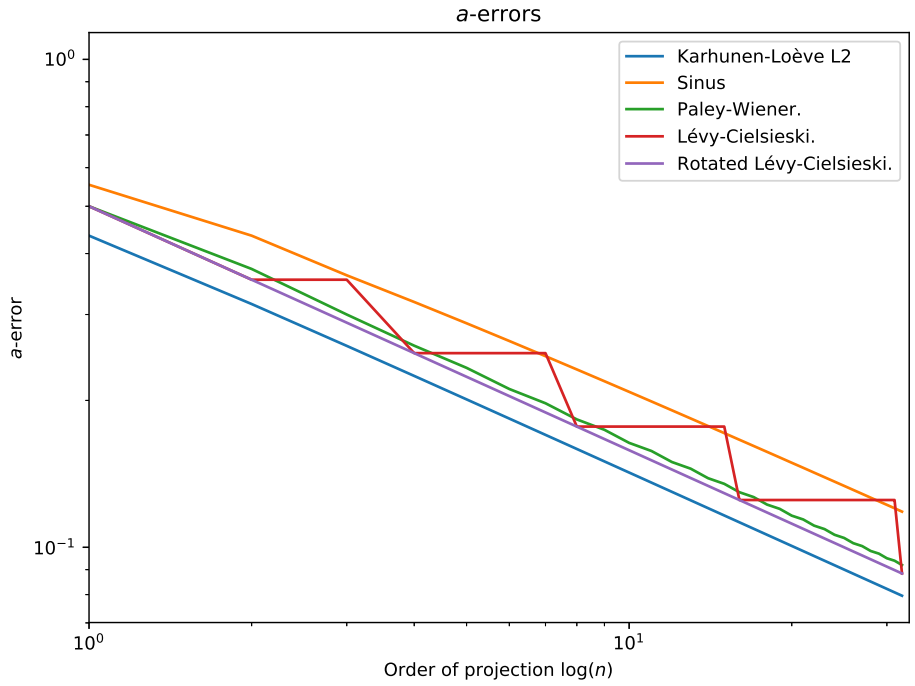

Figure 2.10: Comparison of a-errors of admissible sequences for the standard Wiener process.

Part II

Applications to Bayesian inverse problems

Chapter 3

Introduction to inverse problems

This introductory chapter on inverse problems presents important concepts relative to this area. It starts with notions of forward, inverse and well-posed problems (section 3.1), which are illustrated both on linear and non-linear examples (section 3.2). Then, the well-known Tikhonov-Philips regularization is introduced in the context of reflexive Banach spaces with general convex penalty functionals (section 3.3).

3.1 Forward, inverse and ill-posed problems

Well-posed problem In a large number of fields ranging from engineering to fundamental sciences, mathematics have been applied to represent various phenomena. These models are designed by the formulation of hypotheses ensuring the desired behaviour for their solution. One well accepted set of properties that a model (or problem) should have, has been stated by J. Hadamard [Had02] in his definition of *well-posed problems*.

Definition 17 (Well-posed problem in the sense of Hadamard [Had02])**.** *A problem is well-posed if solutions:*

1. *exist,*
2. *are identical,*
3. *vary continuously with data.*

Let us emphasize the three statements in this definition. Although the absence of solution is an interesting mathematical result on itself, one may generally expect solutions for further analysis, especially in applied contexts. Thus, if condition 1 in the previous definition is not satisfied, the whole model needs adjustments. The second assertion about multiplicity is also important, especially if solutions are

particularly different. Indeed, one needs either to treat them all (which may be impossible) or to choose and more importantly justify his choice. Finally, continuity is also mandatory for theoretical and practical reasons. For instance, any approximate treatment of the problem (numerical discretization, measurement errors, etc.) would have important and undesired consequences. It may be necessary to consider different topologies (weak topology in Banach spaces for instance) to get continuity. Naturally, a none well-posed problem will be called *ill-posed*.

Forward and inverse problems Once a model has been carefully designed and is shown to be well-posed, a typical use (for instance in Physics or Engineering) is to *determine the consequences implied by known causes* through the model, which is called the *forward* or *direct* approach. Since the model is a tangible link between the two, one may want to *determine the causes leading to known consequences*, which is the *inverse* problem. Both approaches are illustrated in figure 3.1.

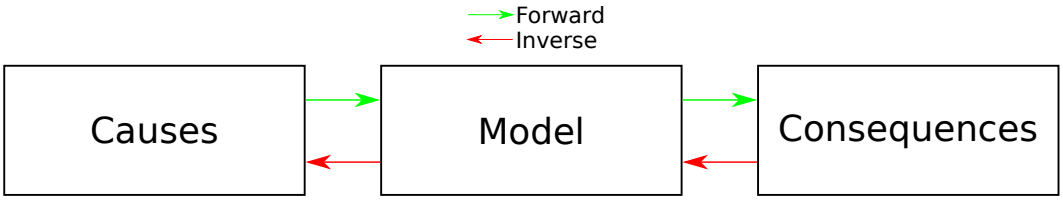

Figure 3.1: Forward and inverse problems.

From a mathematical standpoint, this terminology is conventional as both problems can be considered the inverse of each other. In practice however, forward models are usually studied first to check well-posedness and then their inverse counterparts are considered. Even though they represent the two faces of the same coin, both problems may have very different properties. As previously stated, forward approaches are well-posed almost by definition, while inverses are often ill-posed (see the linear example of compact operators in section 3.2 for instance). This is often due to a *smoothing* effect, meaning that two different causes could lead to similar consequences, or in some sense that the forward model looses information. This phenomenon is particularly clear in the context of linear operators (see section 3.2).

Mathematical formulation In this work, we will use the following mathematical notations. Let $\mathcal{U}$ and $\mathcal{Y}$ be Banach spaces and

$$\mathcal{G} : D(\mathcal{G}) \subset \mathcal{U} \to \mathcal{G}(D(\mathcal{G})) \subset \mathcal{Y}$$

a map. A direct problem will be formally defined as:

$$\text{for } u \in \mathcal{U}, \text{ find } y \in \mathcal{Y} \text{ such that } y = \mathcal{G}(u), \tag{3.1}$$

whereas its inverse:

$$\text{for } y \in \mathcal{Y}, \text{ find } u \in \mathcal{U} \text{ such that } y = \mathcal{G}(u). \tag{3.2}$$

In this setting, the well-posedness of the forward problem 3.1 is equivalent to $\mathcal{U} = D(\mathcal{G})$ and $\mathcal{G}$ continuous. Indeed, in that case $\mathcal{G}(u)$ is defined for all $u \in \mathcal{U}$ and continuous in u. Conversely, the problem 3.2 is well-posed if $\mathcal{G}(D(\mathcal{G})) = \mathcal{Y}$, $\mathcal{G}^{-1}$ exists and is continuous. These particular statements will be detailed in case of a linear map in the next section.

3.2 Examples

Following the previous discussion, examples of forward and inverse problems are provided in this section. The two first examples are given with a linear map $\mathcal{G}$, as it illustrates fundamental phenomena, while the last is taken from a real-world non-linear example from computational Biology.

Closed range operators Let $\mathcal{G} \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$ be a bounded operator such that $R(\mathcal{G})$ is closed. If $\mathcal{U} = D(\mathcal{G})$, the forward model is well-posed. Conversely, if $\mathcal{Y} = R(\mathcal{G})$ there may be multiple solutions, in which case the inverse problem is ill-posed. Indeed, any particular solution $u \in \mathcal{U}$ leads to the full set of solutions $u + \ker(\mathcal{G})$. However, if $\mathcal{G}$ is injective as well, its co-restriction to $R(\mathcal{G})$ is invertible, and is automatically continuous (Open mapping theorem).

Example 21. *Let* $\mathcal{U} = l^2(\mathbb{N})$, $0 < m \leq M < \infty$ $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ *such that* $\lambda_0 = 0$, $\forall n \in \mathbb{N}^*$, $0 < m \leq \lambda_n \leq M < +\infty$ *and*

$$\mathcal{G} : (u_n)_{n\in\mathbb{N}} \in \mathcal{U} \to (\lambda_n u_n)_{n\in\mathbb{N}} \in \mathcal{U}.$$

This operator is non-injective as $\ker(\mathcal{G}) = \operatorname{span}\{e_0\}$ *where* $e_0 = (1, 0, ...)$. *Furthermore, let* $f : (u_n)_{n\in\mathbb{N}} \in \mathcal{U} \to u_0 \in \mathbb{R}$, *then* $R(\mathcal{G}) = \ker(f)$ *is closed. Indeed:*

- $\forall u \in \mathcal{U}$, $f(\mathcal{G}u) = (\mathcal{G}u)_0 = \lambda_0 u_0 = 0$, *thus* $R(\mathcal{G}) \subset \ker(f)$,
- $\forall u \in \ker(f)$, *let* $v := \left(0, \frac{u_1}{\lambda_1}, \frac{u_2}{\lambda_2}, ...\right)$ *then* $v \in \mathcal{U}$ *since* $\|v\|_{\mathcal{U}} \leq \frac{\|u\|_{\mathcal{U}}}{m}$ *and* $u = \mathcal{G}v$ *thus* $\ker(f) \subset R(\mathcal{G})$.

The forward problem is well-posed and its inverse ill-posed because of the multiplicity of solutions. However, this ill-posedness is solved by restriction of $\mathcal{G}$ *to* $\ker(\mathcal{G})^\perp$ *for instance.*

In other words, for closed ranged operators between Banach spaces, the ill-posedness is essentially due to possible multiplicity of solutions. This is the main idea behind the following alternative definition of a well-posed problem.

Definition 18 (Well-posed problem in the sense of Nashed [Nas81])**.** *The inverse problem 3.2 is well-posed if and only if* $R(\mathcal{G})$ *is closed.*

A very important case in practice is given by finite rank operators, because finite dimensional spaces are always topologically closed. In particular, this means that ill-posedness is essentially an *infinite dimensional phenomenon* [EHN96].

Example 22 (Interpolation of continuous functions)**.** *Let $\mathcal{U} = C([0,1],\mathbb{R})$ equipped with the supremum norm, $(x_1,...,x_n) \in [0,1]^n$ and*

$$\mathcal{G} : u \in \mathcal{U} \to (u(x_1),...,u(x_n)) \in \mathbb{R}^n,$$

which is well-defined linear with finite rank. The forward problem of computing $\mathcal{G}(u)$ given $u \in \mathcal{U}$ is well-posed. The inverse problem of finding u given $y \in \mathbb{R}^n$ is ill-posed in the sense of Hadamard (but well-posed in the sense of Nashed).

Compact operators Suppose now that $\mathcal{U}$ and $\mathcal{Y}$ are infinite dimensional Banach spaces and $\mathcal{G} \in \mathcal{K}(\mathcal{U},\mathcal{Y})$ a compact operator, then it is well known that $\mathcal{G}$ can't be homeomorphic.

Proposition 42. *Let $\mathcal{U}$, $\mathcal{Y}$ two infinite dimensional Banach spaces and $\mathcal{G} \in \mathcal{K}(\mathcal{U},\mathcal{Y})$ be an injective compact operator, then $\mathcal{G}^{-1} : R(\mathcal{G}) \to \mathcal{U}$ is not bounded.*

Proof. Let $\mathcal{G} \in \mathcal{K}(\mathcal{U},\mathcal{Y})$ injective and $\mathcal{G}^{-1} \in \mathcal{L}(R(\mathcal{G}),\mathcal{U})$ then $\mathcal{G}^{-1}\mathcal{G} = \mathcal{I}$ is compact, which provides a contradiction with Riesz theorem. □

Example 23 (Poisson's equation)**.** *Let Ω be a smooth domain and consider the following equation:*

$$\begin{aligned} -\Delta u &= f \text{ on } \Omega, \\ u &= 0 \text{ on } \partial\Omega. \end{aligned}$$

The associated weak formulation is

$$\forall v \in H_0^1(\Omega),\ \int_\Omega \nabla u.\nabla v dx = \int_\Omega f v dx,$$

which, according to Lax-Milgram theorem, has a unique solution $u \in H_0^1(\Omega)$ for every $f \in H^{-1}(\Omega)$. The solution map is linear and compact ($H_0^1(\Omega)$ is compactly embedded in $L^2(\Omega)$), thus the forward problem of computing u from f is well-posed. However, the inverse problem is ill-posed according to proposition 42.

Non-linear forward map Here, the following linear evolution equation is considered

$$\begin{aligned} \frac{\partial y}{\partial t}(x,t) + \lambda y(x,t) - D\Delta y(x,t) &= f(x,t),\ \forall (x,t) \in]0,L[\times]0,T] \\ y(x,0) &= 0,\ \forall x \in]0,L[\\ y(0,t) &= 0,\ \forall t \in]0,T], \\ y(L,t) &= 0,\ \forall t \in]0,T]. \end{aligned} \tag{3.3}$$

In other words, the quantity of interest $y(x,t)$ evolves from a null initial state under 3 mechanisms. First, a direct variation, resulting from $f(x,t)$ which depends on time and space. Furthermore, y diffuses at a rate D (strictly positive) and decays

at a rate λ (positive) both being constant in time and space. The direct problem is to compute the solution $y(x,t)$ at all time and space given the model parameters $u = (\lambda, D, f)$. When the source f is the only parameter, this problem is linear and the equation 3.3 is called a linear evolution equation. Here, parameters from the differential operator are considered as well, thus the solution map is non-linear. The inverse problem consists in finding $u = (\lambda, D, f)$ from partial information on y. This example is further studied in chapter 5.

Other well-known examples Different examples might be found in the literature, based on the theory of partial differential equations for instance. These models are usually taken from physics or engineering and tomography, gravitometry or imaging are important examples [Isa17].

3.3 Tikhonov-Philips regularization for operators

A *regularization* method is now presented, which consists in defining alternative notions of solutions to the inverse problem 3.2. These elements will be such that their existence and continuity is proved. It is presented in the case of bounded operators between reflexive Banach spaces, following the lines of [SKHK12], but this approach extends to much more general settings.

Regularized solution Consider a *penalty* functional

$$\Omega : D(\Omega) \subset \mathcal{U} \to \mathbb{R}^+$$

which quantifies a certain property of elements $u \in \mathcal{U}$. A typical example would be a norm which measures the size of all elements in $\mathcal{U}$, but functionals $\Omega(u) := \|Au\|_{\mathcal{X}}$ with $A : \mathcal{U} \to \mathcal{X}$ are also classical (A being a differential operator in a function space for instance). Using this new functional, one can define a new notion of solution to problem 3.2.

Definition 19 (Ω-minimizing solution [SKHK12])**.** *An element $u^\dagger \in \mathcal{U}$ is an Ω-minimizing solution to problem 3.2 if*

$$u^\dagger = \arg\min_{u \in \mathcal{G}^{-1}(y) \cap D(\Omega)} \Omega(u)$$

In other words, we choose among the set of solutions of problem 3.2 those who minimize the penalty functional Ω. The existence and uniqueness of such elements will strongly depends on properties of both spaces $\mathcal{U}$, $\mathcal{Y}$ and the map Ω. Usually, a penalty functional is chosen such that it is:

1. Proper: $D(\Omega) \neq \emptyset$, thus Ω may be used to distinguish some elements in $\mathcal{U}$,
2. Convex: $\forall c \in \mathbb{R}$, the shape of the level set $\{\Omega(u) \leq c\}$ is convex,

3. Lower semi-continuous: $\forall c \in \mathbb{R}$, the level set $\{\Omega(u) \leq c\}$ is closed in the strong topology,

4. Stabilizing: $\forall c \in \mathbb{R}$, $\{\Omega(u) \leq c\}$ is weakly sequentially pre-compact.

These particular hypotheses are exactly what is needed to show the existence of minimizers. Indeed, the combination of items 2 and 3 gives the following general result.

Proposition 43. *Let $\mathcal{U}$ be a Banach space, $A \subset \mathcal{U}$ a convex and (strongly) closed subset, then it is weakly closed as well.*

Proof. See [Bre11], chapter 3. □

Finally, item 4 gives the compacity and the existence of a limit for any minimizing sequence. For instance, the norm of $\mathcal{U}$ is a penalty functional satisfying these hypotheses and this particular choice leads to the concept of minimal-norm solutions. Other penalties may be defined, the following power-type family has been particularly studied (Chapter 5 in [SKHK12] for instance) with $q \in]1, +\infty[$

$$\Omega(u) = \begin{cases} \frac{1}{q}\|u\|_{\mathcal{X}}^q, & \text{if } u \in \mathcal{X} \\ +\infty, & \text{otherwise,} \end{cases}$$

and $\mathcal{X}$ a continuously embedded Banach subspace of $\mathcal{U}$.

Example 24 (Penalty from Reproducing Kernel Hilbert Spaces (RKHS))**.** *For every compact topological space K, the set of real continuous functions on K, equipped with the supremum norm is a Banach space. Choosing a continuous kernel k, the associated (RKHS) injects continuously (see chapter 1 for a proof).*

Now, using the previous properties of penalty functionals and in the context of reflexive Banach spaces, we can show the existence of Ω-minimizing solutions to problem 3.2.

Proposition 44. *Let $\mathcal{U}$ be a reflexive Banach space, $\mathcal{Y}$ be a Banach space and $\mathcal{G} \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$ a bounded operator, then the set of Ω-minimizing solutions of problem 3.2 is non-empty whenever $y \in R(\mathcal{G})$ and $D(\Omega) \cap \mathcal{G}^{-1}(y) \neq \emptyset$.*

Proof. Since $y \in R(\mathcal{G})$ and the set $\mathcal{G}^{-1}(y) \cap D(\Omega)$ is non-empty, it contains a minimizing sequence $(u_n)_{n\in\mathbb{N}}$ such that $\Omega(u_n) \to \inf\{\Omega(u),\ u \in \mathcal{G}^{-1}(y)\} \geq 0$. Because every convergent sequence is bounded, there is $k \in \mathbb{R}$ such that $(\Omega(u_n))_{n\in\mathbb{N}} \in \{u \in \mathcal{U},\ \Omega(u) \leq k\}$. This set being convex and closed, it is weakly closed and even weakly sequentially compact, thus we can extract a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ which is weakly converging to $\tilde{u} \in \mathcal{U}$. Finally, the lower semi-continuity of Ω imposes that $\tilde{u}$ is an Ω-minimizing solution. □

The unicity is more involved, and can be established using geometrical arguments, such as uniform convexity for instance.

Proposition 45. *Let* $\Omega(u) := \frac{1}{q}\|u\|_{\mathcal{U}}^q$ *with* $q \in [1, +\infty[$ *and* $\mathcal{U}$ *uniformly convex, then there exists a unique* Ω*-minimizing solution to problem 3.2.*

Proof. Suppose $q = 1$ without loss of generality. From the Millman-Pettis theorem, $\mathcal{U}$ is reflexive. Moreover, the norm is proper, convex, continuous (thus lower semi-continuous) and stabilizing as any closed ball is weakly compact in a reflexive Banach space, thus from proposition 44 the set of Ω-minimizing solutions is non-empty. Let $u_1^\dagger, u_2^\dagger$ distinct Ω-minimizing solutions (supposed to be of unit norm without loss of generality). Since the set Ω-minimizing solutions is convex, it contains $\tilde{u} := \frac{u_1^\dagger + u_2^\dagger}{2}$ as well. However, the uniform convexity implies $\Omega(\tilde{u}) < \Omega(u_1^\dagger)$ which is a contradiction. □

However, even if the Ω-minimizing solution is a singleton, nor the linearity or continuity of the resulting map is given in proposition 44. In other words, looking directly for Ω-minimizing solutions may be an ill-posed problem. The Tikhonov-Philips regularization method will instead look for approximations of Ω-minimizing solutions.

Remark 1. *If* $\mathcal{U}$ *and* $\mathcal{Y}$ *are Hilbert spaces and* Ω *the norm from* $\mathcal{U}$*, the map* $\mathcal{G}^\dagger : y \in \mathcal{R}(\mathcal{G}) \to u^\dagger \in \mathcal{U}$ *is known as the Moore-Penrose inverse [EHN96]. Its domain of definition is* $R(\mathcal{G}) + R(\mathcal{G})^\perp$*. In particular, if* $R(\mathcal{G})$ *is closed,* $D\left(\mathcal{G}^\dagger\right) = \mathcal{Y}$*.*

Regularized problem The concept of Ω-minimizing solutions is appealing for its clarity but doesn't provide for well-posed inverse problems. Moreover, the data y may not be available, and one can sometimes work with noise $\delta \in \mathcal{Y}$, thus consider $y^\delta = y + \delta$ instead of y. The regularized problem will give a answer to these 2 problems, considering the following Tikhonov-Philips functional $\forall u \in \mathcal{U}$:

$$T_\alpha^\delta(u) := u \in D(\Omega) \cap D(\mathcal{G}) \subset \mathcal{U} \to \frac{1}{p}\left\|\mathcal{G}u - y^\delta\right\|_{\mathcal{Y}}^p + \alpha\Omega(u)$$

where $p \in]1, +\infty[$, $\alpha > 0$. This quantity is composed of two different terms, balancing between proximity to data and penalty as α varies. Taking these 2 constraints in account, a Tikhonov-Philips solution to problem 3.2 can be defined.

Definition 20 (Tikhonov-Philips solution)**.** *An element* $u_\alpha^\delta \in \mathcal{U}$ *is a Tikhonov-Philips solution to problem 3.2 if*

$$T_\alpha^\delta(u_\alpha^\delta) = \inf\{T_\alpha^\delta(u),\ u \in \mathcal{U}\}.$$

Remark that u_α^δ may not be an element from $\mathcal{G}^{-1}(y)$.

Theorem 15. *Let* $\mathcal{U}$*,* $\mathcal{Y}$ *be reflexive Banach spaces,* $\mathcal{G} \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$ *a bounded operator,* $y \in R(\mathcal{G})$ *and* $y^\delta \in \mathcal{Y}$ *then*

1. *(Existence of solutions, proposition 4.1 in [SKHK12]):* $\forall \alpha > 0$, $\forall y^\delta \in \mathcal{Y}$*, the set of Tikhonov-Philips solutions is non-empty.*

2. *(Stability of solutions, proposition 4.2 in [SKHK12]):* $\forall \alpha > 0$, *if* $y^{\delta_n} \to y^\delta$ *then any sequence* $(u_\alpha^{\delta_n})_{n\in\mathbb{N}}$ *of Tikhonov-Philips solutions has a subsequence* $\left(u_\alpha^{\delta_{n_k}}\right)_{k\in\mathbb{N}}$ *such that* $u_\alpha^{\delta_{n_k}} \rightharpoonup u_\alpha^\delta$ *and* u_α^δ *is a Tikhonov-Philips solution.*

3. *(Convergence of solutions, theorem 4.3 in [SKHK12]): Let* $(y^{\delta_n})_{n\in\mathbb{N}}$ *a sequence in* $\mathcal{Y}$ *such that* $y^{\delta_n} \to y$ *and* $\|\delta_n\|_\mathcal{Y}$ *converges monotonically to* 0, $(\alpha_n)_{n\in\mathbb{N}} \in]0, +\infty[$ *and* $\left(u_{\alpha_n}^{\delta_n}\right)_{n\in\mathbb{N}}$ *of Tikhonov-Philips solutions. If*

   - $\limsup_{n\to\infty} \Omega\left(u_{\alpha_n}^{\delta_n}\right) \leq \Omega(u),\ \forall u \in \mathcal{G}^{-1}(y)$,
   - $\left\|\mathcal{G}u_{\alpha_n}^{\delta_n} - y^{\delta_n}\right\|_\mathcal{Y} \to 0$,

   then $\left(u_{\alpha_n}^{\delta_n}\right)_{n\in\mathbb{N}}$ *has a weakly convergent subsequence which limit is an element* $u^\dagger$. *If the set of* Ω*-minimizing solution is a singleton, then* $u_{\alpha_n}^{\delta_n} \rightharpoonup u^\dagger$.

Proof. 1. Let $\alpha > 0$, since Ω is proper, it comes that $D(\Omega) \neq \emptyset$ and it contains a sequence $(u_n)_{n\in\mathbb{N}}$ such that $T_\alpha^\delta(u_n) \to a := \inf\{T_\alpha^\delta(u),\ u \in D(\Omega)\}$. In particular, the sequence $\left(T_\alpha^\delta(u_n)\right)_{n\in\mathbb{N}}$ and a fortiori $(\Omega(u_n))_{n\in\mathbb{N}}$ are bounded in $\mathbb{R}$. Since the sets $\{u \in \mathcal{U},\ \Omega(u) \leq k\}$ are weakly sequentially pre-compact, a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ converges weakly to $\tilde{u} \in \mathcal{U}$, and because $D(\Omega)$ is weakly closed (convex and strongly closed), we have $\tilde{u} \in D(\Omega)$. As a bounded operator for the strong topologies is also weak-to-weak continuous, it comes $\mathcal{G}u_n \rightharpoonup \mathcal{G}\tilde{u}$. Since the norm and Ω are both w.l.s.c., T_α^δ is itself w.l.s.c. and thus:

$$\forall n \in \mathbb{N},\ T_\alpha^\delta(\tilde{u}) \leq T_\alpha^\delta(u_n),$$

which concludes the proof as $\tilde{u}$ is Tikhonov-Philips solution to the problem 3.2.

2. Let $\alpha > 0$, we will first show that the sequence $\left(\Omega\left(u_\alpha^{\delta_n}\right)\right)_{n\in\mathbb{N}}$ is bounded above. By definition of $u_\alpha^{\delta_n}$, it comes that $T_\alpha^{\delta_n}(u_\alpha^{\delta_n}) \leq T_\alpha^{\delta_n}(u),\ \forall u \in \mathcal{U},\ \forall n \in \mathbb{N}$. Using the triangle inequality, monotonicity and convexity it comes:

$$\left\|\mathcal{G}u_\alpha^{\delta_n} - y^\delta\right\|_\mathcal{Y}^p \leq 2^{p-1}\left(\left\|\mathcal{G}u_\alpha^{\delta_n} - y^{\delta_n}\right\|_\mathcal{Y}^p + \left\|y^\delta - y^{\delta_n}\right\|_\mathcal{Y}^p\right).$$

Then

$$\begin{aligned}
\alpha\Omega(u_\alpha^{\delta_n}) &\leq T_\alpha^\delta(u_\alpha^{\delta_n}), \\
&\leq 2^{p-1}T_\alpha^{\delta_n}(u_\alpha^{\delta_n}) + \frac{2^{p-1}}{p}\left\|y^\delta - y^{\delta_n}\right\|_\mathcal{Y}, \\
&\leq 2^{p-1}T_\alpha^{\delta_n}(u) + \frac{2^{p-1}}{p}\left\|y^\delta - y^{\delta_n}\right\|_\mathcal{Y}, \\
&\leq 2^{p-1}T_\alpha^\delta(u) + \frac{2^{p-1}}{p}\left\|\mathcal{G}u - y^\delta\right\|_\mathcal{Y} + \frac{4^{p-1}}{p}\left\|y^\delta - y^{\delta_n}\right\|_\mathcal{Y}.
\end{aligned}$$

Since $\left\|y^\delta - y^{\delta_n}\right\|_\mathcal{Y} \to 0$, it is bounded and $\Omega(u_\alpha^{\delta_n})$ is then bounded above from previous equation. Because Ω is stabilizing, there exists a weakly convergent

subsequence $\left(u_\alpha^{\delta_{n_k}}\right)_{k\in\mathbb{N}}$ which limit is $\tilde{u} \in D(\Omega)$ (as it is weakly closed). Moreover, $y^{\delta_n} \to y^\delta$ and $\mathcal{G}$ is weak-to-weak continuous so $\mathcal{G}u_\alpha^{\delta_n} - y^{\delta_n} \rightharpoonup \mathcal{G}\tilde{u} - y^\delta$. From the w.l.s.c. of the norm and Ω it comes:

$$\begin{aligned} \left\|\mathcal{G}\tilde{u} - y^\delta\right\|_{\mathcal{Y}}^p &\leq \liminf_{k\to\infty} \left\|\mathcal{G}u_\alpha^{\delta_{n_k}} - y^\delta\right\|_{\mathcal{Y}}^p, \\ \Omega(\tilde{u}) &\leq \liminf_{k\to\infty} \Omega\left(u_\alpha^{\delta_{n_k}}\right). \end{aligned}$$

Finally, we have $\forall u \in D(\Omega)$:

$$\begin{aligned} T_\alpha^\delta(\tilde{u}) &\leq \liminf_{k\to\infty} T_\alpha^{\delta_k}\left(u_\alpha^{\delta_{n_k}}\right), \\ &\leq \lim_{k\to\infty} T_\alpha^{\delta_{n_k}}(u), \end{aligned}$$

and since $\forall u \in D(\Omega)$, $\lim_{k\to\infty} T_\alpha^{\delta_{n_k}}(u) = T_\alpha^\delta(u)$ then $\tilde{u}$ is a Tikhonov-Philips solution of problem 3.2.

3. Let $u^\dagger$ be an Ω-minimizing solution, then hypothesis XX gives:

$$\limsup_{n\to\infty} \Omega(u_\alpha^{\delta_n}) \leq \Omega(u^\dagger)$$

In particular, the sequence $\left(\Omega(u_\alpha^{\delta_n})\right)_{n\in\mathbb{N}}$ is bounded above, and because Ω is stabilizing, there exists a weakly convergent subsequence $\left(\Omega\left(u_\alpha^{\delta_{n_k}}\right)\right)_{k\in\mathbb{N}}$ with limit $\tilde{u} \in D(\Omega)$ (as it is weakly closed). Then

$$\left\|\mathcal{G}u_\alpha^{\delta_{n_k}} - y\right\|_{\mathcal{Y}} \leq \left\|\mathcal{G}u_\alpha^{\delta_{n_k}} - y^{\delta_n}\right\|_{\mathcal{Y}} + \left\|y - y^{\delta_n}\right\|_{\mathcal{Y}}.$$

The weak-to-weak continuity of $\mathcal{G}$ and the w.l.s.c. of the norm gives $\mathcal{G}\tilde{u} = y$. Moreover, we have $\forall u \in D(\Omega)$:

$$\Omega(\tilde{u}) \leq \Omega(u^\dagger) \leq \Omega(u).$$

In particular, $\Omega(\tilde{u}) = \Omega(u^\dagger) = \lim_{k\to\infty} \Omega\left(u_\alpha^{\delta_{n_k}}\right)$ which concludes the proof.

□

The convergence of solutions is conditional to two particular hypotheses, which may be attained with specific choices of sequences $(\alpha_n)_{n\in\mathbb{N}}$. We won't go into details about a priori and a posteriori choices here, but it is fully presented in [SKHK12]. Moreover, in particular cases such as power-type penalty functional, strong convergence may be proven using error analysis.

Example 25. *Consider again the problem of interpolation in $\mathcal{U} = C([0,1],\mathbb{R})$. Choosing the following penalty functional:*

$$\Omega(u) = \begin{cases} \frac{1}{2}\|u\|_K^2, & \text{if } u \in K \\ +\infty, & \text{otherwise.} \end{cases}$$

where K is the RKHS with continuous kernel k embedded in $\mathcal{U}$. The solution to this minimization problem is unique:

$$u^\alpha = k(x, X)k(X, X)^{-1}X,$$

due to the Representer theorem. The regularized solution u^α inherits properties from k (as a linear combination). For instance, it is well-known that a Gaussian kernel provides infinitely differentiable solutions whereas for a Brownian one, they are piecewise linear.

Remark 2. *In the Hilbert case, taking $p = q = 2$ in a power-type functional is particularly interesting. Indeed, we have T_α twice differentiable with:*

$$\begin{aligned} T_\alpha(u) &= \frac{1}{2}\left\|\mathcal{G}u - y\right\|_{\mathcal{Y}}^2 + \frac{\alpha}{2}\left\|u\right\|_{\mathcal{U}}^2, \\ \nabla T_\alpha(u) &= \mathcal{G}^*\mathcal{G}u - \mathcal{G}^*y + \alpha u, \\ \nabla^2 T_\alpha(u) &= \mathcal{G}^*\mathcal{G} + \alpha\mathcal{I}. \end{aligned}$$

The Hessian bilinear form being positive definite for all $\alpha > 0$, the associated operator is invertible and we have the normal equations for the regularized solution u^α:

$$(\mathcal{G}^*\mathcal{G} + \alpha\mathcal{I})u^\alpha = \mathcal{G}^*y,$$

thus $u^\alpha = (\mathcal{G}^*\mathcal{G} + \alpha\mathcal{I})^{-1}\mathcal{G}^*y$. In particular, if $\mathcal{U}$, $\mathcal{Y}$ are separable and $\mathcal{G}$ is diagonal $\mathcal{G} : (u_n)_{n\in\mathbb{N}} \to (\lambda_n u_n)_{n\in\mathbb{N}}$ (such as a spectral representation of a compact operator), then:

$$(\mathcal{G}^*\mathcal{G} + \alpha\mathcal{I})^{-1}\mathcal{G}^* : (u_n)_{n\in\mathbb{N}} \in l^2(\mathbb{N}) \to \left(\frac{\lambda_n}{\lambda_n^2 + \alpha}u_n\right)_{n\in\mathbb{N}} \in l^2(\mathbb{N}) \tag{3.4}$$

and this operator is bounded since $\left(\frac{\lambda_n}{\lambda_n^2+\alpha}\right)_{n\in\mathbb{N}}$ is bounded.

3.4 Other regularizations methods

The previous Tikhonov-Philips regularization is among the most well-known methods. However, the field of inverse problems is still under active research, see [BB18], [SKHK12] and the references therein for recent and detailed presentations.

Extension of Tikhonov In section 3.3, the Tikhonov-Philips theory has been presented with particular assumptions such as linearity of $\mathcal{G}$, reflexivity of $\mathcal{U}$ and convexity of Ω. All of these may be circumvented, see [SKHK12].

Iterative methods Iterative methods aim at minimizing the discrepancy $\left\|\mathcal{G}u - y^\delta\right\|_{\mathcal{Y}}$ directly. The regularization consists in choosing a stopping rule $N(\delta, y^\delta)$ and choose $u_{N(\delta,y^\delta)}$ as solution to the inverse problem 3.2. This class of methods have the advantage of requiring much less computations than Tikhonov-Philips counterparts.

Bayesian methods One of the most important drawback of deterministic regularization methods, is that they don't provide for uncertainty quantification. In particular, problems involving measured data are ubiquitous and it is often possible to estimate the uncertainty on these. It is then natural to ask a methodology to take into account this additional information. In the two last decades, this necessity led researchers to consider Bayesian techniques in inverse problems [Stu10, Tar05, CS07] and this methodology will be fully detailed in the next chapter.

Chapter 4

Theory of Bayesian regularization in Banach spaces

This second chapter on regularization methods presents the Bayesian theory for inverse problems as initially developed in [Stu10]. This approach consists in choosing an initial Borel probability measure, which will be updated in a posterior distribution using a likelihood and some available data. This last probability measure will be the solution of the inverse problem and naturally includes a quantification of uncertainty. Hadamard's definition of well-posedness will then refer to existence, uniqueness and continuity of this solution relative to the data in an appropriate distance. Furthermore, a clear link has been established between previous Tikhonov-Philips solutions and posterior modes, providing a useful variational principle. After a short introduction on Kriging, which is a fundamental example of Bayesian regularization, the theory is stated for possibly infinite dimensional Banach spaces. It also includes important aspects regarding Markov chain Monte-Carlo algorithms and consistency of approximation.

4.1 Kriging: a motivation for Bayesian inversion

Before introducing the general Bayesian methodology for inverse problems, we start here with the well-known case of continuous functions interpolation, using both deterministic and stochastic viewpoints for regularization.

Problem statement. Consider a compact metric set $(\mathcal{K}, d)$, an element $\mathbf{x} = (x_1, ..., x_n) \in \mathcal{K}^n$ representing distinct inputs and $\mathbf{y} = (y_1, ..., y_n) \in \mathbb{R}^n$, the corresponding image by a continuous function. The interpolation problem consists in finding a continuous function (an element u in the Banach space $C(\mathcal{K}, \mathbb{R})$), such that $\forall i \in [1, n]$, $u(x_i) = y_i$ (or simply $u(\mathbf{x}) = \mathbf{y}$ in vector notations). According to previous chapter, interpolation is an ill-posed inverse problem in the sense of Hadamard (but not in the sense of Nashed). Two distinct regularizations methods will be presented, an application of previous Tikhonov-Philips theory (see chapter 3), the other

on stochastic modelling with Gaussian random fields (the Kriging approach).

Tikhonov-Philips regularization using RKHS. The first approach to solve the interpolation problem will be of Tikhonov-Philips type, choosing a penalty functional based on the norm of a continuously (and compactly, see chapter 1) embedded subspace of $C(\mathcal{K}, \mathbb{R})$, a Reproducing Kernel Hilbert space H_k (RKHS, see [DM05]) associated to a continuous symmetric, semi-positive kernel k. Here, the considered penalty functional will be:

$$\Omega : u \in C(\mathcal{K}, \mathbb{R}) \to \begin{cases} \frac{1}{2}\|u\|_k^2, \text{ if } u \in H_k, \\ +\infty, \text{ otherwise,} \end{cases} \tag{4.1}$$

as it satisfies the properties presented in previous chapter.

Proposition 46. *Let $(\mathcal{K}, d)$ a compact metric space, k a non-trivial, continuous, symmetric, semi-positive definite kernel and H_k the associated RKHS, then the penalty functional Ω defined in equation 4.1 is convex, lower semi-continuous, proper and stabilizing.*

Proof. Ω is a convex function since $\forall (u_1, u_2) \in H_k^2$, $\forall \lambda \in [0,1]$:

$$\begin{aligned} \Omega(\lambda u_1 + (1-\lambda)u_2) &= \frac{1}{2}\|\lambda u_1 + (1-\lambda)u_2\|_k^2, \\ &\leq \frac{1}{2}\left(\lambda \|u_1\|_k + (1-\lambda)\|u_2\|_k\right)^2, \\ &\leq \frac{\lambda}{2}\|u_1\|_k^2 + \frac{1-\lambda}{2}\|u_2\|_k^2 = \lambda\Omega(u_1) + (1-\lambda)\Omega(u_2). \end{aligned}$$

If $u_1 \notin H_k$ or $u_2 \notin H_k$, the inequality is still correct, thus Ω is convex. Now, let $m \in \mathbb{R}^+$ and $(u_n)_{n\in\mathbb{N}} \subset \{u \in C(\mathcal{K}, \mathbb{R}),\ \Omega(u) \leq m\}$ such that $u_n \to u$ in $C(\mathcal{K}, \mathbb{R})$. Because $(u_n)_{n\in\mathbb{N}}$ is bounded in H_k (by hypothesis), it exists $v \in H_k$ and $(u_{n_k})_{k\in\mathbb{N}} \subset H_k$ such that $u_{n_k} \rightharpoonup v$ in the weak topology of H_k (balls are weakly compact in Hilbert spaces). Now, the reproducing property gives $\forall x \in \mathcal{K}$, $u_{n_k}(x) = \langle u_{n_k}, k(.,x)\rangle_k \to \langle v, k(.,x)\rangle_k = v(x)$ which implies that $v = u$ since $u_{n_k} \to u$ in $C(\mathcal{K}, \mathbb{R})$. It remains to see that $u \in \{\Omega(u) \leq m,\ u \in C(\mathcal{K}, \mathbb{R})\}$, which is a consequence of $\|u\|_{H_k}^2 \leq \liminf \|u_n\|_{H_k}^2 \leq m$, thus level sets associated to Ω are closed in $C(\mathcal{K}, \mathbb{R})$. It is proper because $k \neq 0$ and stabilizing because balls in H_k are compact. □

Following the discussion in chapter 3, this provides both existence and uniqueness of Ω-minimizing and Tikhonov-Philips solutions to the interpolation problem.

Proposition 47. *Let $(\mathcal{K}, d)$ be a compact metric space, k a continuous, symmetric, semi-positive definite kernel and Ω the penalty function defined in equation 4.1, $\mathbf{x} \in \mathcal{K}^n$ such that $k(\mathbf{x}, \mathbf{x})$ is invertible, then the interpolation problem*

1. *has a unique Ω-minimizing solution $u^\dagger$,*

$$u^\dagger : z \in \mathcal{K} \to k(z, \mathbf{x})k(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y} \in \mathbb{R}$$

2. *$\forall \alpha > 0$, has a unique Tikhonov-Philips solution for the following Tikhonov-Philips functional:*

$$T_\alpha(u) = \frac{1}{2} \|\mathcal{G}u - y\|^2_{\mathbb{R}^n} + \alpha\Omega(u),$$

given by $u_\alpha : z \in \mathcal{K} \to k(z, \mathbf{x}) \left(k(\mathbf{x}, \mathbf{x}) + \alpha\mathcal{I}\right)^{-1} \mathbf{y} \in \mathbb{R}$.

with the notations $k(\mathbf{x}, \mathbf{x})_{i,j} = k(x_i, x_j)$ *and* $k(z, \mathbf{x}) = (k(z, x_1), ..., k(z, x_n))$. *Moreover, both applications* $y \to u_\alpha$ *and* $y \to u^\dagger$ *are continuous.*

Proof. Let $\alpha > 0$ and T_α the following Tikhonov-Philips functional:

$$T_\alpha := u \in C\left(\mathcal{K}, \mathbb{R}\right) \to \frac{1}{2} \|y - \mathcal{G}u\|^2_{\mathbb{R}^n} + \frac{\alpha}{2} \|u\|^2_k \in \mathbb{R}^+.$$

Clearly, $D\left(\Omega\right) = D\left(T_\alpha\right) = H_k$, thus solutions are in H_k.

- Let $F = \text{span}\left\{k(., x_i),\ i \in [1, n]\right\}$ which is closed in H_k, then $H_k = F \oplus F^\perp$ and note P_F the associated (bounded) orthogonal projector. Remark that $\forall u \in H_k$, $P_F(u) = u_F$:

$$\Omega(u) = \frac{1}{2} \left(\|u_F\|^2_{H_k} + \|u - u_F\|^2_{H_k} \right) \geq \Omega(u_F),$$

  and by the reproducing property

$$\forall i \in [1, n],\ (u - u_F)(x_i) = \langle u - u_F, k(., x_i) \rangle_k = 0,$$

  thus $u(\mathbf{x}) = u_F(\mathbf{x})$. In other words, both objectives are simultaneously reduced by projecting on F and both type of solutions must be in F, that is $\forall z \in \mathcal{K},\ u(z) = k(z, \mathbf{x})\beta$ with $\beta \in \mathbb{R}^n$.

- Concerning Ω-minimizing solutions, they must satisfy $u(\mathbf{x}) = \mathbf{y}$, which if $u^\dagger \in F$ imposes $\beta = k(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y}$ and the solution is unique.

- About Tikhonov-Philips solutions, it is enough to minimize T_α on F, which directly gives $\beta = (k(\mathbf{x}, \mathbf{x}) + \alpha\mathcal{I})^{-1}\mathbf{y}$ and again it is unique.

- Let $(\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{R}^{2n}$ and define $u^\dagger_i(z) := k(z, \mathbf{x})k(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y}_i,\ i \in \{1, 2\}$. Then it comes:

$$\forall z \in \mathcal{K},\ |u^\dagger_1(z) - u^\dagger_2(z)| \leq \|k(z, \mathbf{x})\|_{\mathbb{R}^n} \left\|k(\mathbf{x}, \mathbf{x})^{-1}\right\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)} \|\mathbf{y}_1 - \mathbf{y}_2\|_{\mathbb{R}^n},$$

  and because $\mathcal{K}$ is compact and $z \to \|k(z, \mathbf{x})\|_{\mathbb{R}^n}$ continuous, $\left\|u^\dagger_1 - u^\dagger_2\right\|_\infty \leq C \|\mathbf{y}_1 - \mathbf{y}_2\|_{\mathbb{R}^n}$. The proof is similar for u_α.

□

In conclusion, the Tikhonov-Philips method based on a (compactly embedded) RKHS norm provides a regularized inverse problem. Both $u^\dagger$ and u_α are continuous w.r.t. $\mathbf{y}$, which is a consequence of the well-posedness in the sense of Nashed. As

an illustration, figure 4.1 provides an example of solutions with $\mathcal{K} = [0,1]$, an k a Matérn $\frac{3}{2}$ covariance kernel:

$$k(x,y) = \left(1 + \sqrt{3}|x-y|\right) \exp\left(-\sqrt{3}|x-y|\right).$$

However, in both cases, the solution is always a single function and no quantification of uncertainty is given.

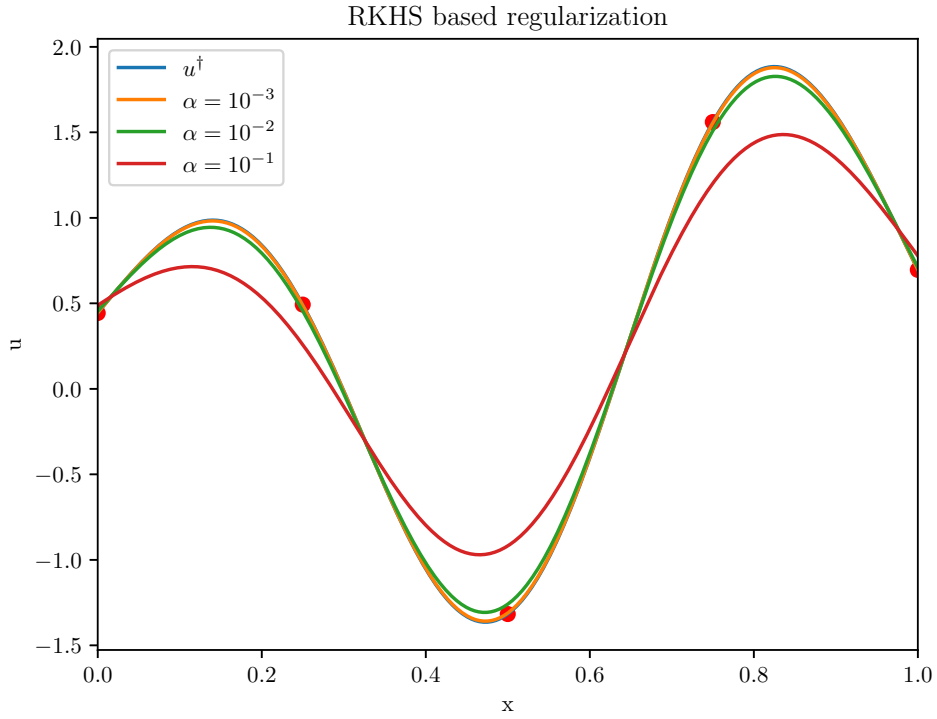

Figure 4.1: Ω-minimizing $u^\dagger$ and Tikhonov-Philips solutions u_α in $C([0,1],\mathbb{R})$ for $\alpha \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ and k the Matern $\frac{3}{2}$ covariance kernel.

Kriging. Kriging has been widely studied to solve the interpolation problem, especially in the field of geosciences [Kri51, Mat62, RW04]. Suppose now that the solution is a sample function from a continuous Gaussian random field with covariance kernel k. Then, for any $x^* = (z, \mathbf{x}) \in \mathcal{K}^{n+1}$, the random vector $u(x^*)$ is Gaussian with zero mean and covariance matrix (in block notations):

$$k(x^*, x^*) := \begin{pmatrix} k(z,z) & k(z,\mathbf{x}) \\ k(\mathbf{x},z) & k(\mathbf{x},\mathbf{x}) \end{pmatrix}.$$

Using conditioning formulas of Gaussian vectors, it comes that $\forall z \in \mathcal{K}$:

$$\begin{aligned} u(z)|u(\mathbf{x}) = \mathbf{y} &\sim \mathcal{N}(m^{\mathbf{y}}(z), v^{\mathbf{y}}(z)), \\ m^{\mathbf{y}}(z) &= k(z,\mathbf{x})k(\mathbf{x},\mathbf{x})^{-1}\mathbf{y}, \\ v^{\mathbf{y}}(z) &= k(z,z) - k(z,\mathbf{x})k(\mathbf{x},\mathbf{x})^{-1}k(\mathbf{x},z). \end{aligned}$$

In other words, the information that $u(\mathbf{x}) = \mathbf{y}$ leads to a new continuous Gaussian random field $u|u(\mathbf{x}) = \mathbf{y}$ with known mean and covariance (the latter being only dependent on $\mathbf{x}$ not $\mathbf{y}$). From this, one can then compute the mean function or sample trajectories, both methods leading to interpolating functions (see figure 4.2). This analysis may also be done under additive Gaussian noise, where the data doesn't directly inform on $u(\mathbf{x})$ but

$$\mathbf{y} = u(\mathbf{x}) + \eta,$$

where $\eta \sim \mathcal{N}(0, \sigma^2\mathcal{I})$. This leads to a different continuous Gaussian random field with parameters:

$$\begin{aligned} u(z)|u(\mathbf{x}) + \eta = \mathbf{y} &\sim \mathcal{N}(m_\sigma^{\mathbf{y}}(z), v_\sigma^{\mathbf{y}}(z)), \\ m_\sigma^{\mathbf{y}}(z) &= k(z,\mathbf{x})\left(k(\mathbf{x},\mathbf{x}) + \sigma^2\mathcal{I}\right)^{-1}\mathbf{y}, \\ v_\sigma^{\mathbf{y}}(z) &= k(z,z) - k(z,\mathbf{x})\left(k(\mathbf{x},\mathbf{x}) + \sigma^2\mathcal{I}\right)^{-1}k(\mathbf{x},z). \end{aligned}$$

Again, it is still possible to extract information from this Gaussian random field by simulation or mean value.

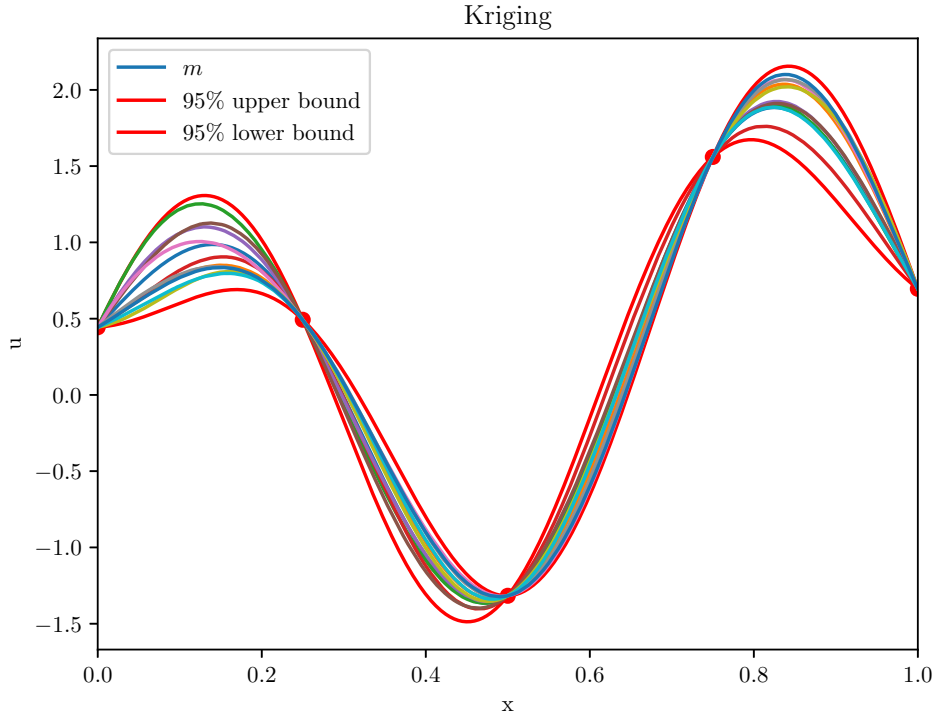

Figure 4.2: Mean function m with 95% confidence interval and 10 sample functions from $u|u(\mathbf{x}) = \mathbf{y}$ with Matérn $\frac{3}{2}$ covariance kernel.

Discussion This simple example shows that, given a continuous, symmetric, semi-positive definite kernel, both Tikhonov-Philips and Kriging approaches lead to similar solutions (if one considers the posterior mean). However, the latter provides also

a probability distribution, thus also quantifies uncertainty. Now, the Kriging approach can be seen through a measure theoretic viewpoint. Indeed, considering the unknown as a sample function from a continuous Gaussian random field is equivalent to choose a prior Borel probability measure on $C(\mathcal{K}, \mathbb{R})$. The conditional random field is also equivalent to a posterior measure, given the data $\mathbf{y}$. The Bayesian theory for inverse problems will generalize this example, involving different priors and likelihoods.

4.2 Bayesian regularization

Bayesian inversion Keeping the previous example of Kriging in mind, the general theory of Bayesian inversion will now be presented. First, remind the definition of an inverse problem, where one wants to *determine the causes leading to known consequences*. The Bayesian approach consists in specifying an *a priori* on the nature of causes (physical information, constraints, regularity, etc.), choosing a model that links causes and consequences (the likelihood) and deducing *a posteriori* information on the causes given known consequences (figure 4.3). In practice, the a

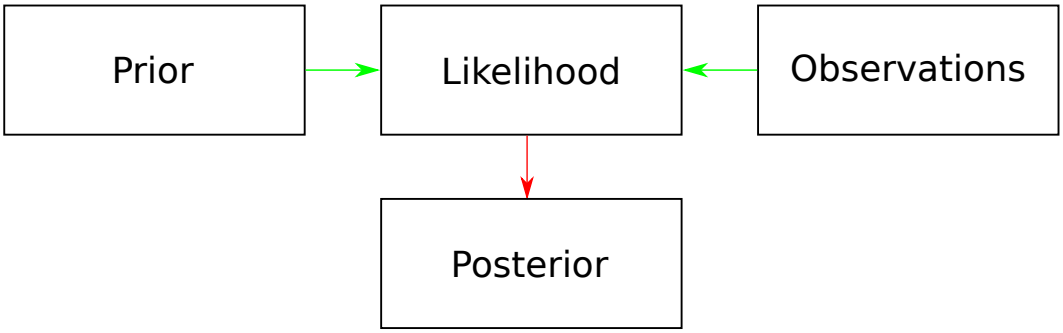

Figure 4.3: Bayesian inverse problem.

priori information is given under the form of a probability distribution μ_0 on the space of causes and the likelihood as a function of both the causes and observations. It generally includes the model but also errors and *deduction* is then given by the celebrated Bayes theorem. This idea of using a probabilistic framework in inverse problems is far from new and has been widely used in finite dimensional contexts [KS05]. However, the theory that will be presented here is infinite dimensional and has been originally developed in the seminal paper [Stu10] (and later as lecture notes [DS15, GHO17]). In a more mathematical formulation, instead of looking for a generalized solution to the inverse problem (Ω-minimizing solutions, etc...), it will taken as a (posterior) probability distribution μ^y on the space of parameters. Similarly to its deterministic counterpart, Bayesian inverse problems may be difficult to solve and Hadamard notion of well-posedness is adapted.

Definition 21 (Well-posed Bayesian inverse problem)**.** *The Bayesian inverse problem is well-posed if the posterior distribution μ^y:*

1. *exists,*

2. *is unique,*

3. *is continuous w.r.t. data y.*

Now, the set of conditions under which a Bayesian inverse problem is well-posed involves both the prior measure μ_0 and the likelihood function, the research of an appropriate setup is still under active development. Indeed, the initial work was done under Gaussian priors [Stu10], later extended to Besov priors [DHS12] and recently to priors with exponential tails [HN17], infinitely-divisible and heavy-Tailed priors [Hos17b] or heavy-tailed stable priors in quasi-Banach spaces [Sul17]. Here, the conditions given in [HN17] are presented as they offer a sufficient level of generality in this thesis.

Mathematical formulation Consider again the notations adopted in chapter 3, that is $\mathcal{U}$, $\mathcal{Y}$ are real Banach spaces, $\mathcal{G} : D(\mathcal{G}) \subset \mathcal{U} \to \mathcal{Y}$ a (possibly non-linear) map and data will be an element $y \in \mathcal{Y}$. The prior μ_0 is a Borel probability distribution on $\mathcal{U}$ and always supposed Radon (which is automatically the case when $\mathcal{U}$ is separable). The likelihood will be given as a positive function, in the form

$$\mathcal{L} : (u, y) \in \mathcal{U} \times \mathcal{Y} \to \exp(-\Phi(u; y)) \in \mathbb{R}^+,$$

and represents how plausible the data y is when the unknown is u (higher values indicate more favourable cases). The function Φ will be called the negative log-likelihood and is chosen according to the problem at hand, for instance an additive model of the form

$$y = \mathcal{G}(u) + \eta$$

where $\eta \sim \mathbb{Q}_0$, a Radon probability measure on $\mathcal{Y}$. In the case where the translated probability distribution $\mathbb{Q}_u = \mathbb{Q}_0(. - \mathcal{G}(u))$ is absolutely continuous w.r.t. $\mathbb{Q}_0$, μ_0-almost surely, the likelihood is a Radon-Nikodym density

$$\frac{d\mathbb{Q}_u}{d\mathbb{Q}_0}(y) = \exp\left(-\Phi(u; y)\right),$$

the evidence, its integral w.r.t. μ_0 such that the posterior is defined by

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{\mathcal{L}(u, y)}{\int_{\mathcal{U}} \mathcal{L}(u, y)\mu_0(du)} = \frac{\exp\left(-\Phi(u; y)\right)}{\int_{\mathcal{U}} \exp\left(-\Phi(u; y)\right) \mu_0(du)},$$

whenever this equation is licit. At this level of generality, the difficulty is to give a set of conditions on both μ_0 and Φ that provides a unique well-defined posterior measure μ^y continuous in the data and this is the objective of next paragraph.

Existence and uniqueness In order to prove the existence and uniqueness of a posterior distribution μ^y, it will be directly defined by the negative log-likelihood Φ such that $\exp(-\Phi)$ is a Radon-Nikodym density w.r.t. μ_0 (up to a constant). In other words, it must be measurable in u (so the Lebesgue integral makes sense), integrable and have a strictly positive integral (the evidence doesn't vanish). Thus, the following set of conditions on the negative log-likelihood is a first step in this direction.

Assumptions 1. *The negative log-likelihood Φ has the following properties:*

1. *Measurability in u: $\forall y \in \mathcal{Y}$, $u \to \Phi(u; y)$ is measurable,*
2. *Lower bound in u, locally in y: $\exists \alpha_1 \geq 0$, $\forall r > 0$, $\exists M(\alpha_1, r) \in \mathbb{R}$, $\forall u \in \mathcal{U}$, $\forall y \in \mathcal{B}_{\mathcal{Y}}(0, r]$,*
$$\Phi(u; y) \geq M(\alpha_1, r) - \alpha_1 \|u\|_{\mathcal{U}}.$$
3. *Boundedness above: $\forall r > 0$, $\exists K(r) > 0$, $\forall u \in \mathcal{B}_{\mathcal{U}}(0, r]$, $\forall y \in \mathcal{B}_{\mathcal{Y}}(0, r]$,*
$$\Phi(u; y) \leq K(r).$$

The particular form of the lower bound in u will be justified by an additional integrability condition on μ_0 (assumption 2), but other alternatives are possible. The measurability condition in u is very often a consequence of continuity.

Assumptions 2 (Prior with exponential tails)**.** *μ_0 is a Borel probability measure on $\mathcal{U}$ such that*

1. *it is Radon,*
2. *$\exists \kappa > 0$,*
$$\int_{\mathcal{U}} \exp(\kappa \|u\|_{\mathcal{U}}) \mu_0(du) < \infty.$$

For instance, convex probability measures automatically satisfy assumption 2 [HN17]. This is in particular the case of Radon Gaussian measures (see theorem 1). Now, under assumptions 1 and 2, the existence and uniqueness of a posterior distribution are established.

Theorem 16 (Theorem 2.1 in [HN17])**.** *Let $\mathcal{U}$ and $\mathcal{Y}$ be Banach spaces, $y \in \mathcal{Y}$, Φ a negative log-likelihood and μ_0 a prior measure respectively satisfying assumption 1 and 2, if $\alpha_1 \leq \kappa$ then μ^y exists, is unique and Radon, characterized by the following Radon-Nikodym density w.r.t. μ_0:*

$$\forall u \in \mathcal{U},\ \frac{\partial \mu^y}{\partial \mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u; y)\right),$$

with $\forall y \in \mathcal{Y},\ Z(y) = \int_{\mathcal{U}} \exp\left(-\Phi(u; y)\right) \mu_0(du) \in \mathbb{R}^+ \setminus \{0\}$.

Proof. Let $y \in \mathcal{Y}$ and consider the evidence $Z(y)$, then

$$\begin{aligned} \forall y \in \mathcal{Y},\ Z(y) &= \int_{\mathcal{U}} \exp\left(-\Phi(u;y)\right) \mu_0(du), \\ &\leq \int_{\mathcal{U}} \exp\left(\alpha_1 \left\|u\right\|_{\mathcal{U}} - M(\alpha_1, r)\right) \mu_0(du), \end{aligned}$$

for all $r \geq \left\|y\right\|_{\mathcal{Y}}$, which is finite if $\alpha_1 \leq \kappa$. Now, $\forall y \in \mathcal{Y}$, $\forall r \geq \left\|y\right\|_{\mathcal{Y}}$,

$$Z(y) \geq \int_{\mathcal{B}_{\mathcal{U}}(0,r]} \exp(-K(r))\mu_0(du) = \mu_0\left(\mathcal{B}_{\mathcal{U}}(0,r]\right) \exp(-K(r)).$$

In particular, this gives constants $c(r), C(r) > 0$ such that $\forall y \in \mathcal{B}_{\mathcal{Y}}(0,r]$, $0 < c(r) \leq Z(y) \leq C(r)$. It remains to see that μ_0 gives a strictly positive measure to $\mathcal{B}_{\mathcal{U}}(0,r]$, which is a consequence of μ_0 being Radon. Since $\mu^y << \mu_0$ then μ^y is Radon itself. □

In the case where $\kappa < \alpha_1$, it is always possible to modify μ_0 using a homothetic transformation (push-forward measure). There are alternative ways to ensure the existence of a posterior distribution, using a Bayes theorem (3.4 in [DS15] for instance) or a different set of conditions on Φ and μ_0. However, assumptions 1 and 2 will show to be general enough and particularly tractable in the rest of this work. The continuity however, requires an adapted notion of distance and additional assumptions on Φ.

Hellinger metric There exists different notions of distance between probability distributions but one shows to be particularly interesting in the context of inverse problems.

Definition 22 (Hellinger distance)**.** *Let $(\Omega, \mathcal{F}, \nu)$ be a probability space, μ_1 and μ_2 two probability measures such that $\mu_i << \nu$ with $i \in \{1,2\}$. The Hellinger distance is defined as*

$$d_{Hell}(\mu_1, \mu_2) = \left(\frac{1}{2} \int_{\Omega} \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 d\nu \right)^{\frac{1}{2}}.$$

Note that it is always possible to find ν satisfying $\mu_1 << \nu$ and $\mu_2 << \nu$, and the value is independent of this choice. One strong motivation for this distance in inverse problems is that it controls the Bochner integral of square integrable functions w.r.t. different measures.

Lemma 10 (7.14 in [DS15])**.** *Let $(\Omega, \mathcal{F}, \nu)$ be a probability space and $\mathcal{U}$ a Banach space, μ_1 and μ_2 two probability measures both absolutely continuous w.r.t. ν, then $\forall f \in L^2(\Omega, \mathcal{F}, \mu_1; \mathcal{U}, Bor(\mathcal{U})) \cap L^2(\Omega, \mathcal{F}, \mu_2; \mathcal{U}, Bor(\mathcal{U}))$, it comes*

$$\left\|\mathbb{E}^{\mu_1}[f(u)] - \mathbb{E}^{\mu_2}[f(u)]\right\|_{\mathcal{U}} \leq 2\sqrt{\mathbb{E}^{\mu_1}\left\|f(u)\right\|_{\mathcal{U}}^2 + \mathbb{E}^{\mu_2}\left\|f(u)\right\|_{\mathcal{U}}^2} d_{Hell}(\mu_1, \mu_2).$$

Proof. Let $f \in L^2(\Omega, \mathcal{F}, \mu_1; \mathcal{U}, Bor(\mathcal{U})) \cap L^2(\Omega, \mathcal{F}, \mu_2; \mathcal{U}, Bor(\mathcal{U}))$ we have

$$\begin{aligned}
& \|\mathbb{E}^{\mu_1}[f] - \mathbb{E}^{\mu_2}[f]\|_{\mathcal{U}} \\
= & \left\| \int_\Omega f(\omega) \left(\frac{d\mu_1}{d\nu} - \frac{d\mu_2}{d\nu} \right) d\nu \right\|_{\mathcal{U}} \\
\leq & \int_\Omega \|f\|_{\mathcal{U}} \left| \frac{d\mu_1}{d\nu} - \frac{d\mu_2}{d\nu} \right| d\nu, \\
= & \int_\Omega \|f\|_{\mathcal{U}} \frac{\left| \sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right|}{\sqrt{2}} \sqrt{2} \left| \sqrt{\frac{d\mu_1}{d\nu}} + \sqrt{\frac{d\mu_2}{d\nu}} \right| d\nu, \\
\leq & \sqrt{\frac{1}{2} \int_\Omega \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 d\nu} \sqrt{2 \int_\Omega \|f\|_{\mathcal{U}}^2 \left(\sqrt{\frac{d\mu_1}{d\nu}} + \sqrt{\frac{d\mu_2}{d\nu}} \right)^2 d\nu}, \\
\leq & d_{Hell}(\mu_1, \mu_2) \sqrt{4 \int_\Omega \|f\|_{\mathcal{U}}^2 \left(\frac{d\mu_1}{d\nu} + \frac{d\mu_2}{d\nu} \right) d\nu},
\end{aligned}$$

which concludes the proof. □

This distance will be used to link how variation in data impacts the associated posterior measures.

Hellinger continuity Now that the existence and uniqueness of the posterior distribution has been established for all $y \in \mathcal{Y}$, a sufficient condition for the continuity of μ^y w.r.t. y in the Hellinger distance will be given here.

Assumptions 3. *The negative log-likelihood is locally Lipschitz in y with μ_0-integrable constant:* $\exists \alpha_2 \geq 0$, $\forall r > 0$, $\exists C(\alpha_2, r) \geq 0$, $\forall u \in \mathcal{U}$*:*

$$\forall y_1, y_2 \in \mathcal{B}_{\mathcal{Y}}(0, r],\ |\Phi(u; y_1) - \Phi(u; y_2)| \leq \exp\left(\alpha_2 \|u\|_{\mathcal{U}} + C(\alpha_2, r)\right) \|y_1 - y_2\|_{\mathcal{U}}.$$

Now, if we combine assumptions 1, 2 and 3, the Bayesian inverse problem is well-posed.

Theorem 17 (Theorem 2.3 in [HN17])**.** *Let $\mathcal{U}, \mathcal{Y}$ be Banach spaces, μ_0 a prior distribution satisfying assumption 2 and Φ a negative log-likelihood such that assumptions 1 and 3 are verified. If $\alpha_1 + 2\alpha_2 \leq \kappa$, the Bayesian inverse problem is well-posed.*

Proof. Let $y_1, y_2 \in \mathcal{Y}$, then $\forall r \geq \max\left(\|y_1\|_{\mathcal{Y}}, \|y_2\|_{\mathcal{Y}}\right)$, $\exists c(r), C(r) > 0$, $c(r) < Z(y_i) < C(r)$, $\forall i \in \{1, 2\}$ (theorem 16). Consider first the following inequality for $a_1, a_2 > 0$ and $(b_1, b_2) \in \mathbb{R}^2$:

$$\begin{aligned}
\left| \frac{e^{-b_1}}{a_1} - \frac{e^{-b_2}}{a_2} \right| &= \left| \frac{e^{-b_1}}{a_1} - \frac{e^{-b_2}}{a_1} + \frac{e^{-b_2}}{a_1} - \frac{e^{-b_2}}{a_2} \right|, \\
&= \left| \frac{e^{-b_1} - e^{-b_2}}{a_1} + e^{-b_2} \left(\frac{1}{a_1} - \frac{1}{a_2} \right) \right|.
\end{aligned}$$

Taking the square of previous equality and using $(a+b)^2 \leq 2(a^2+b^2)$ gives:

$$\left(\frac{e^{-b_1}}{a_1} - \frac{e^{-b_2}}{a_2}\right)^2 \leq 2\left(\frac{e^{-b_1}-e^{-b_2}}{a_1}\right)^2 + 2e^{-2b_2}\left(\frac{1}{a_1}-\frac{1}{a_2}\right)^2.$$

This will now be applied to the Hellinger distance of both posteriors. Let $r > 0$ $(y_1, y_2) \in \mathcal{B}_{\mathcal{Y}}(0,r]$ and apply previous inequality with $b_1 = \frac{\Phi(u;y_1)}{2}$, $b_2 = \frac{\Phi(u;y_2)}{2}$, $a_1 = \sqrt{Z(y_1)}$ and $a_2 = \sqrt{Z(y_2)}$ then

$$d_{Hell}(\mu^{y_1}, \mu^{y_2})^2 \leq 2(I_1 + I_2),$$

with

$$I_1 = \frac{1}{Z(y_1)}\int_{\mathcal{U}}\left(\exp\left(-\frac{\Phi(u;y_1)}{2}\right) - \exp\left(-\frac{\Phi(u;y_2)}{2}\right)\right)^2 \mu_0(du),$$
$$I_2 = \left|\frac{1}{\sqrt{Z(y_1)}} - \frac{1}{\sqrt{Z(y_2)}}\right|^2 Z(y_2).$$

It remains to show that $I_i \leq C_i \left\|y_1 - y_2\right\|_{\mathcal{Y}}^2$, $\forall i \in \{1,2\}$ to complete the proof. Using the lower bound of Φ, its continuity in y and the mean-value theorem on the exponential, it comes $\forall (y_1,y_2) \in \mathcal{B}_{\mathcal{Y}}(0,r]^2$,

$$\begin{aligned}
I_1 &= \frac{1}{Z(y_1)}\int_{\mathcal{U}}\left(\exp\left(-\frac{\Phi(u;y_1)}{2}\right) - \exp\left(-\frac{\Phi(u;y_2)}{2}\right)\right)^2 \mu_0(du),\\
&\leq \frac{1}{Z(y_1)}\int_{\mathcal{U}} \exp\left(\alpha_1 \left\|u\right\|_{\mathcal{U}} - M(\alpha_1,r)\right)\left(\Phi(u;y_1) - \Phi(u;y_2)\right)^2 \mu_0(du),\\
&\leq \frac{1}{Z(y_1)}\int_{\mathcal{U}} \exp\left((\alpha_1+2\alpha_2)\left\|u\right\|_{\mathcal{U}} - M(\alpha_1,r) + 2C(\alpha_2,r)\right)\mu_0(du)\left\|y_1 - y_2\right\|_{\mathcal{U}}^2,\\
&\leq K_1 \left\|y_1 - y_2\right\|_{\mathcal{U}}^2,
\end{aligned}$$

with $0 \leq K_1 < \infty$ if $\kappa \geq \alpha_1 + 2\alpha_2$. Concerning the second term, it comes similarly:

$$\begin{aligned}
&|Z(y_1) - Z(y_2)|\\
&\leq \int_{\mathcal{U}} \exp\left(\alpha_1\left\|u\right\|_{\mathcal{U}} - M(\alpha_1,r)\right)\left|\Phi(u;y_1) - \Phi(u;y_2)\right|\mu_0(du),\\
&\leq \int_{\mathcal{U}} \exp\left[(\alpha_1+\alpha_2)\left\|u\right\|_{\mathcal{U}} - M(\alpha_1,r) + C(\alpha_2,r)\right]\mu_0(du)\left\|y_1 - y_2\right\|_{\mathcal{Y}},\\
&\leq K_2 \left\|y_1 - y_2\right\|_{\mathcal{Y}},
\end{aligned}$$

with $0 \leq K_2 < \infty$ if $\alpha_1 + \alpha_2 \leq \kappa$. Finally,

$$\begin{aligned}
I_2 &\leq \left|\frac{1}{\sqrt{Z(y_1)}} - \frac{1}{\sqrt{Z(y_2)}}\right|^2 Z(y_2),\\
&\leq K_3\left(\Phi(u;y_1) - \Phi(u;y_2)\right)^2,\\
&\leq K_2 K_3 \left\|y_1 - y_2\right\|_{\mathcal{Y}}^2,
\end{aligned}$$

with $K_3 \in \mathbb{R}^+$ by the mean-value theorem and the proof is complete. □

For two distinct data $y_1, y_2 \in \mathcal{Y}$, both posterior distributions are absolutely continuous w.r.t. μ_0 and have strong moments of order 2 (as a consequence of assumption 2). Using lemma 10, one can show continuity of posteriors expectations for instance.

Corollary 13. *Let $\mathcal{U}, \mathcal{Y}$ be Banach spaces, μ_0 a prior distribution satisfying assumption 2 and Φ a negative log-likelihood such that assumptions 1 and 3 are verified and the associated Bayesian inverse problem well-posed, then*

$$\forall r > 0,\ \forall y_1, y_2 \in \mathcal{B}_{\mathcal{Y}}(0, r],\ \left\|\mathbb{E}^{\mu^{y_1}}[u] - \mathbb{E}^{\mu^{y_2}}[u]\right\|_{\mathcal{U}} \leq C(r) \left\|y_1 - y_2\right\|_{\mathcal{Y}}$$

Proof. Since $\mathcal{I}$ is square integrable w.r.t. μ_0 it is integrable w.r.t. μ^y, $\forall y \in \mathcal{Y}$. Now, from lemma 10, it comes $\forall y_1, y_2 \in \mathcal{B}_{\mathcal{Y}}(0, r]$:

$$\left\|\mathbb{E}^{\mu^{y_1}}[u] - \mathbb{E}^{\mu^{y_2}}[u]\right\|_{\mathcal{U}} \leq \sqrt{\mathbb{E}^{\mu^{y_1}}\left[\|u\|_{\mathcal{U}}\right]^2 + \mathbb{E}^{\mu^{y_2}}\left[\|u\|_{\mathcal{U}}\right]^2} d_{Hell}(\mu^{y_1}, \mu^{y_2}).$$

However, the constant can be bounded, indeed

$$\begin{aligned}
\mathbb{E}^{\mu^y}[\|u\|_{\mathcal{U}}^2] &= \int_{\mathcal{U}} \|u\|_{\mathcal{U}}^2 \, \mu^y(du), \\
&= \int_{\mathcal{U}} \|u\|_{\mathcal{U}}^2 \, \frac{d\mu^y}{d\mu_0}(u) \mu_0(du), \\
&\leq \int_{\mathcal{U}} \|u\|_{\mathcal{U}}^2 \, \frac{1}{c(r)} \exp\left(\alpha_1 \|u\|_{\mathcal{U}} - M(\alpha_1, r)\right) \mu_0(du), \\
&\leq C(r),
\end{aligned}$$

thus the proof is complete. □

Gaussian likelihood with finite dimensional data As an example, the special case of Gaussian likelihood with finite dimensional data is studied here. This will be often encountered in practice, when the quantity of interest is measured and corresponds to the following measurement model:

$$y = \mathcal{G}(u) + \eta,$$

where $\eta \sim \mathcal{N}(0, \sigma^2 \mathcal{I}_n)$.

Proposition 48. *Let $\mathcal{U}$ be a Banach space, $\mathcal{Y} = \mathbb{R}^n$ with $n \in \mathbb{N}^*$ and suppose that data is given by previous additive model, then the associated negative log-likelihood can be taken as:*

$$\Phi(u; y) = \frac{1}{2\sigma_\eta^2} \left\|y - \mathcal{G}(u)\right\|_{\mathcal{Y}}^2,$$

whenever $u \in D(\mathcal{G}) \subset \mathcal{U}$.

Proof. Since $u \in D(\mathcal{G})$, one has $\mathcal{G}(u) \in \mathcal{Y}$ and because the covariance of $\mathbb{Q}_0 = \mathcal{N}(0, \sigma^2 \mathcal{I}_n)$ is of full rank, $\mathcal{Y}$ is the Cameron-Martin space associated to η and then

$\mathbb{Q}_u$ and $\mathbb{Q}_0$ are equivalent, the associated Radon-Nikodym density (Cameron-Martin theorem) being:

$$\frac{d\mathbb{Q}_u}{d\mathbb{Q}_0}(u) = \exp\left(\frac{1}{\sigma_\eta^2}\langle \mathcal{G}u, y\rangle_{\mathcal{Y}} - \frac{1}{2\sigma_\eta^2}\|\mathcal{G}u\|_{\mathcal{Y}}^2\right).$$

It remains to see that one can add any constant to the negative log-likelihood without changing the posterior, since it automatically scales the evidence. Here the constant is $-\frac{1}{2\sigma_\eta^2}\|y\|_{\mathcal{Y}}^2$ and the proof is complete. □

Furthermore, when the forward map is a bounded operator in $\mathcal{L}(\mathcal{U}, \mathcal{Y})$, the Gaussian negative log-likelihood automatically satisfies assumption 1 and 3, thus the Bayesian inverse problem is well-posed if and only if μ_0 has exponential tails.

Proposition 49. *Let $\mathcal{U}$, $\mathcal{Y}$ be Banach spaces, $\mathcal{G} \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$ a bounded operator, μ_0 a probability measure satisfying assumption 2 and*

$$\Phi(u; y) = \frac{1}{2\sigma_\eta^2}\|y - \mathcal{G}u\|_{\mathbb{R}^n}^2,$$

then the associated Bayesian inverse problem is well-posed.

Proof. Φ has the following properties:

1. (Lower bound in u). $\forall y \in \mathcal{Y}$, $\Phi(u; y)$ is positive μ_0-a.e. thus one can take $\alpha_1 = 0$ and in particular, $\kappa \geq \alpha_1$.

2. (Boundedness above). For all $(u, y) \in \mathcal{U} \times \mathcal{Y}$:

$$\begin{aligned}\Phi(u; y) &\leq \frac{1}{2\sigma_\eta^2}\left(\|y\|_{\mathcal{Y}} + \|\mathcal{G}u\|_{\mathcal{Y}}\right)^2, \\ &\leq \frac{1}{\sigma_\eta^2}\left(\|y\|_{\mathcal{Y}}^2 + \|\mathcal{G}u\|_{\mathcal{Y}}^2\right), \\ &\leq \frac{1}{\sigma_\eta^2}\left(\|y\|_{\mathcal{Y}}^2 + \|\mathcal{G}\|_{\mathcal{L}(\mathcal{U},\mathcal{Y})}^2\|u\|_{\mathcal{U}}^2\right),\end{aligned}$$

   and in the case where $\max\{\|u\|_{\mathcal{U}}, \|y\|_{\mathcal{Y}}\} \leq r$ one has:

$$K(r) = \frac{r^2\left(1 + \|\mathcal{G}\|_{\mathcal{L}(\mathcal{U},\mathcal{Y})}^2\right)}{\sigma_\eta^2}.$$

3. (Local Lipschitz continuity in y). For all $y_1, y_2 \in \mathcal{B}_{\mathcal{Y}}(0, r]$:

$$\begin{aligned}&|\Phi(u; y_1) - \Phi(u; y_2)| \\ = \;& \frac{1}{2\sigma_\eta^2}\left|\left(\|y_1\|_{\mathcal{Y}} - \|y_2\|_{\mathcal{Y}}\right)\left(\|y_1\|_{\mathcal{Y}} + \|y_2\|_{\mathcal{Y}}\right) - 2\langle y_1 - y_2, \mathcal{G}u\rangle_{\mathcal{Y}}\right|, \\ \leq \;& \frac{1}{\sigma_\eta^2}\left(r + \|\mathcal{G}\|_{\mathcal{L}(\mathcal{U},\mathcal{Y})}\|u\|_{\mathcal{Y}}\right)\|y_1 - y_2\|_{\mathcal{Y}}, \\ \leq \;& \exp\left(\alpha_2\|u\|_{\mathcal{U}} + C(\alpha_2, r)\right)\|y_1 - y_2\|_{\mathcal{Y}},\end{aligned}$$

   with $\alpha_2 > 0$.

□

Finally, if the prior measure is Gaussian, the forward map linear, the posterior distribution is Gaussian itself with known mean and covariance operator.

Proposition 50. *Let $\mathcal{U}$ be a Hilbert space, $\mathcal{Y} = \mathbb{R}^n$ with $n \in \mathbb{N}^*$, $\mathcal{G} \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$, $\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$ and*

$$\Phi(u; y) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}u\|_{\mathcal{Y}}^2,$$

then $\mu^y = \mathcal{N}(m^y, \mathcal{C}^y)$ with:

$$\begin{aligned} m^y &= \mathcal{C}_0\mathcal{G}^* \left(\mathcal{G}\mathcal{C}_0\mathcal{G}^* + \sigma_\eta^2\mathcal{I}\right)^{-1} y, \\ \mathcal{C}^y &= \mathcal{C}_0 - \mathcal{C}_0\mathcal{G}^* \left(\mathcal{G}\mathcal{C}_0\mathcal{G}^* + \sigma_\eta^2\mathcal{I}\right)^{-1} \mathcal{G}\mathcal{C}_0. \end{aligned}$$

4.3 Maximum a posteriori

Once one has proven the Bayesian inverse problem to be well-posed, it is of great interest for the practitioner (and the academician) to extract information from the posterior distribution μ^y. This is usually done by selecting elements from $\mathcal{U}$, in particular these 2 alternatives:

1. The conditional mean (CM). This element is defined as the Bochner integral of the posterior distribution $\mathbb{E}^{\mu^y}[u]$. Its existence is established since the prior is supposed with exponential tails and $\mu^y << \mu_0$ when the inverse problem is well-posed. In particular, it is continuous in the data.

2. The maximum a posteriori (MAP). This element is defined as a mode of the posterior distribution μ^y, which requires a specific treatment at this level of generality. Indeed, the absence of Lebesgue measure will lead to a definition based on the concept of measure differentiation and small balls approach.

The concepts developed here are fully taken from [DHS12, HB15, ABDH17]. The major idea is that MAP estimators are sometimes characterized by a variational problem, linking to the Tikhonov-Philips regularization.

Modes of Radon probability measures In finite dimensional spaces, modes of a probability distribution (having a density w.r.t. Lebesgue's measure), are defined as the maximizers of the probability density function. However, in infinite dimensional spaces, this definition can't be applied and one turns to differentiation of measures (see [Rud09] for instance), or small-ball probabilities.

Definition 23 (Definition 3 in [HB15])**.** *Let $\mathcal{U}$ a Banach space, μ be a Radon probability measure on $\mathcal{U}$ such that supp$(\mu) \neq \emptyset$, then an element u is a mode for μ if*

$$\lim_{\epsilon \to 0} \frac{\mu\left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\sup_{v \in \mathcal{U}} \mu\left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} = 1.$$

The fact that $\mathrm{supp}(\mu) \neq \emptyset$ is essential in previous definition since in the support of a Radon measure, every neighbourhood of $x \in \mathrm{supp}(\mu)$ has strictly positive measure, hence the denominator is non-zero. In the Gaussian case, the small ball probability is known analytically and thus modes are completely characterized by the minimizers of the Onsager-Machlup functional.

Theorem 18 (Mode of Radon Gaussian measures)**.** *Let $\mathcal{U}$ be a Banach space, μ be a Radon Gaussian measure on $\mathcal{U}$ with Cameron-Martin space $\mathcal{U}_\mu$ and such that $\mu(\mathcal{U}) = 1$, then*

$$\lim_{\epsilon \to 0} \frac{\mu\left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu\left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} = \exp\left(\frac{\|v\|_\mu^2 - \|u\|_\mu^2}{2}\right).$$

However, the small ball ratio is unknown most of the time in non-Gaussian cases, thus researchers turn to weak modes, see [HB15, ABDH17].

Maximum a posteriori and Tikhonov-Philips regularization When the measure of interest is the posterior measure given in an inverse problem, modes are called *Maximum a posterior* estimators. The main idea is that if one knows small ball probabilities for the prior distribution, it translates to the posterior using the likelihood function. The Gaussian case is a good illustration.

Theorem 19. *Let $\mathcal{U}$, $\mathcal{Y}$ be Banach spaces, $y \in \mathcal{Y}$, μ_0 a Radon Gaussian measure on $\mathcal{U}$ with Cameron-Martin space $\mathcal{U}_{\mu_0}$, Φ a negative log-likelihood satisfying assumption 1 and: $\forall y \in \mathcal{Y}$, $\forall r > 0$, $\exists L(r) > 0$,*

$$\forall (u, v) \in \mathcal{B}_{\mathcal{U}}(0, r]^2, \ |\Phi(u; y) - \Phi(v; y)| \leq L(r) \|u - v\|_{\mathcal{U}},$$

then the inverse problem is well-posed and the following relation holds:

$$\forall (u, v) \in \mathcal{U}_{\mu_0}^2, \ \lim_{\epsilon \to 0} \frac{\mu^y\left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu^y\left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} = \exp\left(I(v) - I(u)\right),$$

where

$$I(u) = \begin{cases} \Phi(u; y) + \frac{\|u\|_{\mu_0}^2}{2}, & \text{if } u \in \mathcal{U}_{\mu_0}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Proof. Let $(u, v) \in \mathcal{U}_{\mu_0}^2$, then

$$\begin{aligned} \frac{\mu^y\left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu^y\left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} &= \frac{\int_{\mathcal{B}_{\mathcal{U}}(u, \epsilon]} \exp\left(-\Phi(z; y)\right) \mu_0(dz)}{\int_{\mathcal{B}_{\mathcal{U}}(v, \epsilon]} \exp\left(-\Phi(z; y)\right) \mu_0(dz)}, \\ &= \frac{\int_{\mathcal{B}_{\mathcal{U}}(u, \epsilon]} \exp\left(-\Phi(z; y) + \Phi(u; y) - \Phi(u; y)\right) \mu_0(dz)}{\int_{\mathcal{B}_{\mathcal{U}}(v, \epsilon]} \exp\left(-\Phi(z; y) + \Phi(v; y) - \Phi(v; y)\right) \mu_0(dz)}. \end{aligned}$$

Now, take $L_1 = L(\|u\|_{\mathcal{U}} + \epsilon)$ and $L_2 = L(\|v\|_{\mathcal{U}} + \epsilon)$, then

$$\begin{aligned} \frac{\mu^y\left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu^y\left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} &\leq \exp\left(\epsilon(L_1 + L_2)\right) \frac{\int_{\mathcal{B}_{\mathcal{U}}(u, \epsilon]} \exp\left(-\Phi(u; y)\right) \mu_0(dz)}{\int_{\mathcal{B}_{\mathcal{U}}(v, \epsilon]} \exp\left(-\Phi(v; y)\right) \mu_0(dz)}, \\ &\leq \exp\left(\epsilon(L_1 + L_2) + I(v) - I(u)\right). \end{aligned}$$

This clearly provides

$$\limsup_{\epsilon \to 0} \frac{\mu^y \left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu^y \left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} \leq \exp\left(I(v) - I(u)\right).$$

The very same analysis provides the following relation:

$$\liminf_{\epsilon \to 0} \frac{\mu^y \left(\mathcal{B}_{\mathcal{U}}(u, \epsilon]\right)}{\mu^y \left(\mathcal{B}_{\mathcal{U}}(v, \epsilon]\right)} \geq \exp\left(I(v) - I(u)\right),$$

thus the announced equality follows. □

A similar result has been obtained for Besov priors [HB15] and later generalized to a wider class of prior measures [ABDH17].

4.4 Markov chain Monte-Carlo method

Motivation The solution of a well-posed Bayesian inverse problem is a Radon probability measure μ^y, known only through a Radon-Nikodym density w.r.t. the prior distribution. Except in very particular cases, it is difficult to directly extract valuable information just looking at this density. It is often analysed through quantities of interest having the form of integrals w.r.t. μ^y. In a large number of cases, there are no analytical solution and one must consider simulation methods for their approximation. The Monte-Carlo approach is a standard tool, based on relations analogue to the law of large numbers:

$$\int_{\mathcal{U}} f(u)\mu^y(du) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(u_i), \tag{4.2}$$

where f is such that both quantities make sense and $(u_i)_{i \in \mathbb{N}}$ are states drawn from a random process. Markov chain Monte-Carlo methods, and especially the Metropolis-Hastings algorithm, are particularly well-suited for this task and will be presented in this section. After a short presentation of general concepts, the Metropolis-Hastings algorithm is detailed and applied to sample posterior distributions μ^y.

Markov kernels The theory of Markov chains has been developed in general state spaces [RR04] but the presentation is intentionally restricted to Banach spaces, since it is the natural framework for Bayesian inverse problems. Intuitively, a Markov chain is a sequence of random elements, where the conditional distribution of a state only depends on the previous one. In particular, this law is modelled by a Markov kernel.

Definition 24 (Markov kernel)**.** *Let $\mathcal{U}_1$,$\mathcal{U}_2$ two Banach spaces, a Markov kernel P from $\mathcal{U}_1$ to $\mathcal{U}_2$ is a map:*

$$P : \mathcal{U}_1 \times Bor(\mathcal{U}_2) \to [0, 1],$$

such that:

- $\forall u_1 \in \mathcal{U}_1,\ P(u_1, .)$ *is a Borel probability measure on* $\mathcal{U}_2$,
- $\forall A \in Bor(\mathcal{U}_2),\ u \to P(u, A)$ *is measurable.*

Example 26 (Gaussian random walk)**.** *Let* $\mathcal{U} = \mathbb{R}^n,\ n \in \mathbb{N}^*$ *and* Σ *a positive definite matrix of dimension* n*, then* $P(x, dv) = \mathcal{N}(x, \Sigma)$ *is a Markov kernel from* $\mathbb{R}^n$ *to itself. Indeed:*

- $\forall x \in \mathbb{R}^n$, $\mathcal{N}(x, \Sigma)$ *is a probability measure on* $Bor(\mathbb{R}^n)$,
- $\forall A \in Bor(\mathbb{R}^n)$,

$$x \in \mathbb{R}^n \to \int_A \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-x)^T \Sigma^{-1}(y-x)\right) dy$$

is continuous from Lebesgue's dominated convergence theorem, thus measurable.

Considering an initial distribution μ, the modification after a unit of time is obtained in the following manner, defining the Markov operator associated to a kernel.

Definition 25 (Markov operator)**.** *Let* $\mathcal{U}_1$, $\mathcal{U}_2$ *two Banach spaces,* P *a Markov kernel from* $\mathcal{U}_1$ *to* $\mathcal{U}_2$*, then the associated Markov kernel is:*

$$\mathcal{P} := \mu \in \mathcal{M}\left(Bor(\mathcal{U}_1)\right) \to \int_{\mathcal{U}_1} P(u, .)\mu(du) \in \mathcal{M}\left(Bor(\mathcal{U}_2)\right).$$

If there are distributions unchanged under a Markov operator, these are said invariant.

Definition 26 (Invariance)**.** *Let* $\mathcal{U}$ *a Banach space,* P *a Markov kernel from* $\mathcal{U}$ *to itself, then* P *is invariant w.r.t.* $\mu \in \mathcal{M}\left(Bor(\mathcal{U})\right)$ *if* $\mathcal{P}(\mu) = \mu$.

In practice, invariance is often obtained as consequence of a stronger property, the reversibility.

Definition 27 (Reversibility)**.** *Let* $\mathcal{U}$ *be a Banach space,* P *a Markov kernel from* $\mathcal{U}$ *to itself and* $\mu \in \mathcal{M}\left(Bor(\mathcal{U})\right)$ *a probability measure, then* P *is reversible w.r.t.* μ *if*

$$\forall (A, B) \in Bor(\mathcal{U})^2,\ \int_A P(u, B)\mu(du) = \int_B P(v, A)\mu(dv).$$

Lemma 11. *Let* $\mathcal{U}$ *be a Banach space,* $\mu \in \mathcal{M}\left(Bor(\mathcal{U})\right)$ *a Borel probability measure and* P *a Markov kernel from* $\mathcal{U}$ *to itself and reversible w.r.t.* μ *then* P *is also invariant w.r.t* μ.

Proof. Let P a Markov kernel reversible w.r.t. μ, then:

$$\forall A \in \mathcal{F},\ \mathcal{P}\mu(A) = \int_{\mathcal{U}} P(u, A)\mu(du) = \int_A P(u, \mathcal{U})\mu(du) = \mu(A).$$

□

Now that basic concepts have been introduced, the Metropolis-Hastings algorithm will be presented, given a clear methodology on how to build a Markov kernel invariant for a predefined probability measure. In particular, it will be used to build Markov kernel invariant w.r.t. μ^y.

Metropolis-Hastings algorithm The Metropolis-Hastings algorithm [Has70] is a practical way to design Markov kernels with predefined invariant measures in general state spaces [Tie98]. It consists in choosing a proposal Markov kernel Q and an acceptance/rejection step to get a resulting Markov kernel P reversible w.r.t. μ^y. The precise procedure is given in algorithm 2.

$u = u_0$;
$n = 0$;
while $n < N$ **do**
  $v \sim Q(u, dv)$;
  $U \sim \mathcal{U}([0,1])$;
  if $U \leq \alpha(u_n, v)$ **then**
    $u_{n+1} = v$;
  else
    $u_{n+1} = u_n$;
  end
  $n = n + 1$;
end

Algorithm 2: Pseudo-code for the Metropolis-Hastings algorithm.

The resulting Metropolis-Hastings kernel P is given by

$$P(u, dv) = Q(u, dv)\alpha(u, v) + \delta_u(dv) \int_{\mathcal{U}} (1 - \alpha(u, z))Q(u, dz). \tag{4.3}$$

The reversibility of the Markov kernel in equation 4.3 will be obtained by specific choices for Q and α, given the target distribution μ^y. In particular, it is the case when α and Q satisfy the following condition.

Proposition 51. *Let $\mathcal{U}$ be a Banach space, Q a Markov kernel from $\mathcal{U}$ to itself and P be the associated Metropolis-Hastings kernel with acceptance probability α, then P is invariant w.r.t. μ^y if and only if*

$$\nu(du, dv)\alpha(u, v) = \nu^T(du, dv)\alpha(v, u),$$

where $\nu(du, dv) = \mu^y(du)Q(u, dv)$ and $\nu^T(du, dv) = \nu(dv, du)$ are product measures on $Bor(\mathcal{U}) \otimes Bor(\mathcal{U})$.

Proof. Let $(A, B) \in Bor(\mathcal{U})^2$, then

$$\begin{aligned}&\int_A \int_B \int_{\mathcal{U}} (1 - \alpha(u, z))Q(u, dz)\delta_u(dv)\mu(du)\\ &= \int_{A\cap B} \int_{\mathcal{U}} (1 - \alpha(u, z))Q(u, dz)\mu(du),\end{aligned}$$

is symmetric in A and B, thus the equivalence holds. □

This (extended) balance condition translates in conditions on the acceptance probability α, inherited by the target measure μ and the proposal kernel Q. The fundamental idea in the Metropolis-Hastings algorithm is to consider α and Q such that ν and ν^T are equivalent (at least on a sufficient subset of the space) and use the relative Radon-Nikodym density. The existence of such set is given in the following proposition.

Proposition 52 (Proposition 1 in [Tie98])**.** *Let $\mathcal{U}$ be a Banach space, ν and ν^T product measures on $Bor(\mathcal{U}) \otimes Bor(\mathcal{U})$, then there exists a unique, up to null sets under ν and ν^T, symmetric set R in the product σ-algebra such that ν and ν^T are mutually absolutely continuous on R and singular on $R^{\complement}$. Moreover, there exists a version r of the density of ν w.r.t. ν^T (restricted on R) such that $\forall (u, v) \in \mathcal{U}^2$, $0 < r(u, v) < \infty$ and $r(u, v) = r(v, u)^{-1}$.*

Proof. Consider the measure $\kappa = \mu + \mu^T$ which is symmetric and dominates both ν and ν^T. Let h be a Radon-Nikodym density of ν w.r.t. κ then

$$\nu^T(du, dv) = \nu(dv, du) = h(v, u)\kappa(dv, du) = h(v, u)\kappa(du, dv),$$

thus $h(v, u)$ is a Radon-Nikodym density of ν^T w.r.t. κ. Now, let

$$R = \{(u, v) \in \mathcal{U}^2,\ (h(u, v), h(v, u)) \in (\mathbb{R}^+)^2\},$$

then it is symmetric, and ν and ν^T are mutually absolutely continuous. Indeed, one has

$$\forall (u, v) \in R,\ \nu(du, dv) = h(u, v)\kappa(du, dv) = \frac{h(u, v)}{h(v, u)}\nu^T(du, dv).$$

Conversely, on $R^{\complement}$ the measures ν and ν^T are mutually singular, since it is not possible that $h(u, v)$ and $h(v, u)$ are both strictly positive on the same set. Let $r(u, v) = h(u, v)h(v, u)^{-1}$ defined on R, and keep the same notation for its extension by 1 on $\mathcal{U}^2$. Suppose now $\tilde{R}$ a second set such that all previous properties are true, then the measures ν and ν^T must be absolutely continuous and mutually singular on $R \subset \tilde{R}$ and $\tilde{R} \subset R$, which means that $\tilde{R}$ is negligible. □

Now, the extended detailed balance condition can be restated using α and the partition $R, R^{\complement}$.

Theorem 20 (Theorem 2 in [Tie98])**.** *Let P a Metropolis-Hastings kernel as in equation 4.3, $\nu(du, dv) = \mu(du)Q(u, dv)$ and ν^T its symmetric, R the set of mutual absolute continuity between ν and ν^T, then P is reversible w.r.t. μ if and only if:*

1. $\alpha = 0$ *on* $R^{\complement}$, ν*-a.e.,*

2. $\alpha(u,v)\frac{d\nu^T}{d\nu}(u,v) = \alpha(v,u)$ *on* R, ν*-a.e.*

Proof. Let $\eta(du,dv) = \alpha(u,v)\nu(du,dv)$ and $\eta^T(du,dv) = \alpha(v,u)\nu^T(du,dv)$. Suppose that $\alpha = 0$, ν-a.e. on $R^{\complement}$ then $\eta(R^{\complement}) = 0$ and since it is a symmetric set, $\eta^T(R^{\complement}) = 0$ which means detailed balance condition on $R^{\complement}$. Now suppose that $\alpha(u,v)\frac{d\nu^T}{d\nu}(u,v) = \alpha(v,u)$ on R, ν-a.e. then on R:

$$\alpha(v,u)\nu^T(du,dv) = \alpha(u,v)\frac{d\nu}{d\nu^T}(u,v)\nu^T(du,dv) = \alpha(u,v)\nu(du,dv),$$

and finally detailed balance condition holds everywhere. Conversely, if detailed balance condition holds, measures ν and ν^T being singular on $R^{\complement}$ implies that $\alpha = 0$ ν-a.e. on $R^{\complement}$. On R, the measures are equivalent, and detailed balance conditions implies

$$\alpha(u,v)\frac{d\nu^T}{d\nu}(u,v) = \alpha(v,u).$$

□

The Metropolis-Hastings choice of α is:

$$\alpha_{MH}(u,v) = \min\left(1, \frac{d\nu^T}{d\nu}(u,v)\right), \tag{4.4}$$

and it ensures the reversibility w.r.t. μ^y.

Proposition 53. *Let α the Metropolis-Hastings acceptance probability then P is reversible w.r.t. μ.*

Proof. By proposition 52, there exists a set R such that ν and ν^T are mutually absolutely continuous and one can take a version of there density such that it vanishes on $R^{\complement}$, thus $a_{MH} = 0$ on $R^{\complement}$. Now, on R, a version of the Radon-Nikodym density can be taken such that

$$\begin{aligned}
\alpha_{MH}(u,v)\frac{d\nu^T}{d\nu}(u,v) &= \min\left(\frac{d\nu^T}{d\nu}(u,v), \frac{d\nu^T}{d\nu}(u,v)\frac{d\nu}{d\nu^T}(u,v)\right), \\
&= \min\left(\frac{d\nu^T}{d\nu}(u,v), 1\right), \\
&= \alpha_{MH}(v,u)
\end{aligned}$$

and theorem 20 applies. □

In order to use a specific Markov kernel to estimate integrals w.r.t. μ^y as in equation 4.2, it is necessary to choose an initial state $u_0 \in \mathcal{U}$. The convergence then happens under ergodicity conditions, dependent on the Markov kernel at hand. It will be discussed in simple cases later on.

Application to Bayesian inverse problems As previously stated, the target measure in a well-posed Bayesian inverse problem is the posterior distribution μ^y, defined by the following Radon-Nikodym density w.r.t. μ^0:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u;y)\right).$$

Using the Metropolis-Hastings algorithm, the proposal Markov kernel must be chosen such that

$$\nu(du, dv) = Q(u, dv)\mu^y(du)$$

and $\nu^T(du, dv) = \nu(dv, du)$ are both mutually absolutely continuous such that the Metropolis-Hastings acceptance ratio is well-defined:

$$\alpha(u, v)_{MH} = \min\left(1, \frac{d\nu^T}{d\nu}(u, v)\right).$$

Proposition 54. *Let $\mathcal{U}$, $\mathcal{Y}$ be Banach spaces, $y \in \mathcal{Y}$, μ_0 a prior probability distribution and Φ a negative log-likelihood such that the inverse problem is well-posed and the following map:*

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u;y)\right),$$

defines a Radon-Nikodym density w.r.t. μ_0. Let Q be a Markov kernel on $\mathcal{U}$ such that $\nu_0(du, dv) = Q(u, dv)\mu_0(du)$ and $\nu_0^T(du, dv) = Q(v, du)\mu_0(dv)$ are mutually absolutely continuous, then $\nu(du, dv) = Q(u, dv)\mu_0(du)$ and $\nu^T(du, dv) = Q(v, du)\mu_0(dv)$ are absolutely continuous and

$$\frac{d\nu^T}{d\nu}(u, v) = \frac{d\nu_0^T}{d\nu_0}(u, v) \exp\left(\Phi(v;y) - \Phi(u;y)\right).$$

Proof. Given the previous notations, it comes:

$$\begin{aligned}
\nu^T(du, dv) &= Q(v, du)\mu^y(dv), \\
&= \frac{d\mu^y}{d\mu_0}(v)\nu_0^T(du, dv), \\
&= \frac{\frac{d\mu^y}{d\mu_0}(v)}{\frac{d\mu^y}{d\mu_0}(u)} \frac{d\mu^y}{d\mu_0}(u) \frac{d\nu_0^T}{d\nu_0}(u, v)\nu_0(du, dv), \\
&= \exp\left(\Phi(v;y) - \Phi(u;y)\right) \frac{d\nu_0^T}{d\nu_0}(u, v)\nu(du, dv),
\end{aligned}$$

and the proof is complete. □

Corollary 14. *Let $\mathcal{U}$, $\mathcal{Y}$ be Banach spaces, $y \in \mathcal{Y}$, μ_0 a prior probability distribution and Φ a negative log-likelihood such that the inverse problem is well-posed and the following map:*

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u;y)\right),$$

defines a Radon-Nikodym density w.r.t. μ_0. Let Q be a Markov kernel from $\mathcal{U}$ to itself, reversible w.r.t. μ_0, then

$$\forall (u, v) \in \mathcal{U}^2,\ \alpha_{MH}(u, v) = \min\left(1, \exp\left(\Phi(v;y) - \Phi(u;y)\right)\right).$$

The two following examples are fundamental and represent the benchmark in this literature.

Example 27 (Independence sampler)**.** *Let $\mathcal{U}$ and $\mathcal{Y}$ be Banach spaces, $y \in \mathcal{Y}$, μ_0 a Radon probability measure on $\mathcal{U}$, Φ a likelihood such that the Bayesian inverse problem is well-posed and μ^y is well-defined. Consider $Q(u, dv) = \mu_0(dv)$, which is reversible w.r.t. μ_0, then the associated Metropolis-Hastings Markov kernel with*

$$\forall (u,v) \in \mathcal{U}^2,\ \alpha_{MH}(u,v) = \min\left(1, \exp\left(\Phi(v;y) - \Phi(u;y)\right)\right),$$

is reversible w.r.t. μ^y.

Example 28 (preconditioned Crank-Nicholson)**.** *Let $\mathcal{U}$ and $\mathcal{Y}$ be Hilbert spaces, $y \in \mathcal{Y}$, μ_0 a Gaussian probability measure on $\mathcal{U}$ with covariance operator $\mathcal{C}$, Φ a likelihood such that the Bayesian inverse problem is well-posed and μ^y is well-defined. Let $\rho \in [0, 1[$ then $Q(u, dv) = \mathcal{N}(\rho u, (1-\rho^2)\mathcal{C})$ is reversible w.r.t. μ_0. Indeed, it comes that*

$$\begin{aligned}
\nu_0(du, dv) &= \mathcal{N}(0, \mathcal{C}) \otimes \mathcal{N}(\rho u, (1-\rho^2)\mathcal{C}), \\
&= \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathcal{C} & \rho\mathcal{C} \\ \rho\mathcal{C} & \mathcal{C} \end{pmatrix}\right), \\
&= \mathcal{N}(\rho v, (1-\rho^2)\mathcal{C}) \otimes \mathcal{N}(0, \mathcal{C}), \\
&= \nu_0^T(du, dv),
\end{aligned}$$

and Q is reversible w.r.t. μ_0. The associated Metropolis-Hastings Markov kernel with

$$\alpha_{MH}(u,v) = \min\left(1, \exp\left(\Phi(v;y) - \Phi(u;y)\right)\right),$$

is reversible w.r.t. μ^y.

In the context of Gaussian prior measures on Hilbert spaces, a variety of proposals have been developed using infinite dimensional Langevin equations [CRSW13, HSV14, RS18, BS09, Law14]. The case of non-Gaussian priors is less developed, but [Vol13] and [CDPS18] are providing interesting solutions. The necessary ergodicity conditions are given in [HSV14, RS18] using L^2 spectral gaps for pCN and $gpCN$ (generalized pCN)for instance.

4.5 Approximation

General principle The previous theory of Bayesian inversion leads to a posterior measure on possibly infinite dimensional Banach space. All the valuable information is then encoded in the Radon-Nikodym density, a positive functional possibly difficult to interpret directly. It is then natural to turn to simulation methods and/or variational characterization of point-wise estimators. However, even if these two important steps are well-defined in an infinite dimensional space, their practical

computation requires to cast the Bayesian inverse problem into a finite dimensional setting through discretization. One essential property is *consistency*, meaning that precision increases with the size of the discretization. In order to introduce the main concepts here, a general framework is considered, using the previously defined mathematical objects:

- $\mathcal{U}$, $\mathcal{Y}$ are (real) Banach spaces,
- $y \in \mathcal{Y}$ is the data,
- μ_0 is the prior, a Radon probability measure on $\mathcal{U}$ satisfying assumption 2,
- $\Phi : \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ is a negative log-likelihood satisfying assumptions 1 and 3,
- μ^y is the posterior, a Radon probability measure defined through its Radon-Nikodym density w.r.t. μ_0:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y))$$

where $Z(y) = \int_{\mathcal{U}} \exp(-\Phi(u; y))\mu_0(du)$. Its existence, uniqueness and continuity are established in theorem 17.

In this context, approximation will always refer to the likelihood (through Φ), meaning that one has access to a sequence $(\Phi_n)_{n\in\mathbb{N}}$. If, in particular, $\forall n \in \mathbb{N},\ \Phi_n$ satisfies assumption 1 and 3, a sequence of approximate posterior measures $(\mu_n^y)_{n\in\mathbb{N}}$ is well-defined:

$$\forall n \in \mathbb{N},\ \forall (u, y) \in \mathcal{U} \times \mathcal{Y},\ \frac{d\mu_n^y}{d\mu_0}(u) = \frac{1}{Z_n(y)} \exp(-\Phi_n(u; y)),$$

with $\forall n \in \mathbb{N},\ \forall y \in \mathcal{Y},\ Z_n(y) = \int_{\mathcal{U}} \exp(-\Phi_n(u; y))\mu_0(du)$. This sequence is said *consistent* if it tends to μ^y asymptotically in the Hellinger metric.

Definition 28. *Let $\mathcal{U}, \mathcal{Y}$ be (real) Banach spaces, $y \in \mathcal{Y}$ the data, μ_0 a prior on $\mathcal{U}$ and Φ a negative log-likelihood satisfying respectively assumptions 2, 1 and 3 such that μ^y is well-defined, unique and continuous. If $(\Phi_n)_{n\in\mathbb{N}}$ is a family of negative log-likelihood such that:*

- $\forall n \in \mathbb{N},\ \Phi_n : \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ *satisfies assumption 1 and 3,*
- $\lim_{n\to\infty} d_{Hell}(\mu_n^y, \mu^y) = 0$,

then the sequence $(\mu_n^y)_{n\in\mathbb{N}}$ is a consistent approximation of μ^y.

Since $(\mu_n^y)_{n\in\mathbb{N}}$ is defined by a family of negative log-likelihoods, consistence bears a strong similarity with the analysis of continuity in the data. Indeed, it consists in controlling the Hellinger distance and this remark leads to an analogue of assumption 3.

Assumptions 4. *The sequence of negative log-likelihood is such that* $\forall (u,y) \in \mathcal{U}\times\mathcal{Y}$, *there exist* $\alpha_4 \geq 0$ *and* $C(y) \geq 0$ *such that*

$$\forall n \in \mathbb{N},\ |\Phi(u;y) - \Phi_n(u;y)| \leq \exp\left(\alpha_4 \|u\|_{\mathcal{U}} + C(y)\right) \psi(n)$$

where $\psi(n)$ *is positive and* $\lim_{n\to\infty} \psi(n) = 0$. *Moreover, there exists a lower bound in* u *(uniformly in* n*):*

$$\forall n \in \mathbb{N},\ \Phi_n(u;y) \geq \alpha_5 \|u\|_{\mathcal{U}} - C(\alpha_5),$$
$$\forall u \in \mathcal{B}_{\mathcal{U}}(0,r],\ \forall y \in \mathcal{B}_{\mathcal{Y}}(0,r],\ \forall n \in \mathbb{N},\ \Phi_n(u;y) \leq K(r).$$

Now, it is natural to replicate the methodology from previous chapter to this case.

Theorem 21 (Theorem 2.4 in [CDS10])**.** *Let* $\mathcal{U},\mathcal{Y}$ *be (real) Banach spaces,* $y \in \mathcal{Y}$ *the data,* μ_0 *a prior probability measure on* $\mathcal{U}$ *satisfying assumption 2,* Φ *and* $(\Phi_n)_{n\in\mathbb{N}}$ *all satisfying assumption 1 such that posteriors* μ^y *and* $(\mu_n^y)_{n\in\mathbb{N}}$ *exist. If assumption* 4 *holds then:*

$$\forall n \in \mathbb{N},\ d_{Hell}(\mu^y, \mu_n^y) \leq \psi(n),$$

and the approximation is consistent.

Proof. By the set of hypotheses on Φ and $(\Phi_n)_{n\in\mathbb{N}}$, it comes that $0 < c(r) \leq Z(y) \leq C(r)$ and $\forall n \in \mathbb{N},\ 0 < c'(r) \leq Z_n(y) \leq C'(r)$. Let $n \in \mathbb{N}$, then similarly to theorem 17, it comes:

$$d_{Hell}(\mu^y, \mu_n^y)^2 \leq 2(I_1 + I_2),$$

with

$$\begin{aligned} I_1 &= \frac{1}{Z(y)} \int_{\mathcal{U}} \exp\left(-\Phi(u;y)\right) \left(\Phi(u;y) - \Phi_n(u;y)\right)^2 \mu_0(du), \\ I_2 &= \left| \frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{Z_n(y)}} \right|^2 \int_{\mathcal{U}} \exp\left(-\Phi_n(u;y)\right) \mu_0(du), \end{aligned}$$

It remains to show that $I_i \leq C_i \psi(n), \forall i \in \{1,2\}$ to complete the proof.

$$\begin{aligned} I_1 &\leq \frac{1}{Z(y)} \int_{\mathcal{U}} \exp\left[(\alpha_1 + 2\alpha_4) \|u\|_{\mathcal{U}} - M(\alpha_1, r) + 2C(\alpha_4)\right] \mu_0(du) \psi(n)^2, \\ &\leq K_1 \psi(n)^2. \end{aligned}$$

Concerning the second term, it comes

$$\begin{aligned} |Z(y) - Z_n(y)| &\leq \int_{\mathcal{U}} \exp\left(-\Phi(u;y)\right) |\Phi(u;y) - \Phi_n(u;y)| \, \mu_0(du), \\ &\leq \int_{\mathcal{U}} \exp\left[(\alpha_1 + 2\alpha_4) \|u\|_{\mathcal{U}} - M(\alpha_1, r) + 2C(\alpha_4)\right] \mu_0(du) \psi(n), \\ &\leq K_2 \psi(n). \end{aligned}$$

Finally,

$$\begin{aligned} I_2 &\leq \left| \frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{Z_n(y)}} \right|^2 \int_{\mathcal{U}} \exp\left(\alpha_5 \|u\|_{\mathcal{U}} - C(\alpha_5)\right) \mu_0(du), \\ &\leq K_3 \psi(n)^2, \end{aligned}$$

and the proof is complete. □

An example will be given in chapter 5 using a Schauder basis in the space of parameters as it is done in [Hos17a].

A word on different types of approximations A large number of approaches exists for the approximation of Bayesian inverse problems (Sparse polynomial expansions, [SS12, CS15, Sch13, SS14, SS16], Spectral expansions, [NS16, SSC$^+$15], Goal-oriented [LGW18, SCW$^+$17], Data-driven [CMW14]).

- **Likelihood informed subspace**. This approach is a projection based approximation, developed in the particular case of $\mathcal{U}$ Hilbert and μ_0 Gaussian. The space is decomposed in two distinct parts, one where the posterior differs strongly from the prior (the likelihood-informed subspace) and its complement. In the linear case, this decomposition is done by an analysis of the *update* between the prior and posterior covariances. Since it is often of finite rank, one can. It has been shown to be optimal for different metrics. In case of non-linear forward model, this approach has been extended by the search of a global likelihood-informed subspace, averaging the previous construction by the posterior measure. Complete developments are available in the series of paper [CMW16, CMM$^+$14, SSC$^+$15].
- **Random surrogates** Random surrogates consist in an approximation of Φ by replacing the forward operator $\mathcal{G}$ by a stochastic approximation $\tilde{\mathcal{G}}$. In particular, the error is of random nature, which gives an interesting quantification of uncertainty. In the context of infinite dimensional inverse problems, it has been studied in [LST18, ST16].

4.6 Conclusion

This chapter introduced a general framework to solve Bayesian inverse problems in Banach spaces. In particular, it provides sufficient conditions on the prior distribution μ_0 and the negative log-likelihood for the existence, uniqueness and continuity of the posterior μ^y. The question of approximation is treated in a general case, but next chapter will provide the necessary arguments to use admissible sequences in this purpose.

Chapter 5

Karhunen-Loève approximation of Bayesian inverse problems

This last chapter provides an important application of all the material presented so far in this thesis. It is based on an interesting example of Bayesian inverse problem, where the forward map is non-linear and obtained from a parabolic partial derivative equation. The objective is to recover jointly a positive source term and constant rates of decay and diffusion, from noisy and partial observations of its solution. It is motivated by a real-world Biological problem, which is the subject of [CDA18]. The context will not be detailed here as the presentation focuses on mathematical aspects including: existence and uniqueness of the posterior distribution, its continuity w.r.t. data in Hellinger metric, a variational characterization of the associated posterior modes and finally a consistent approximation by projection. This last item will be based on the extended Karhunen-Loève decomposition from chapter 2, and the speed of convergence is shown to be at least equal to the prior series approximation in Bochner norm (linked to l-numbers). Finally, the problem is illustrated on a real dataset using a numerical solver in Python.

5.1 Approximation using projections

In this chapter, the approximation of Bayesian inverse problems will be specified, using parameters projection on vector bases. This methodology was limited to Banach spaces with pre-existing Schauder bases and linear forward maps in [HN17]. However, it will be shown that when the prior is a Radon Gaussian measure, this restriction is irrelevant.

Schauder bases and projections In vector spaces, finite dimensional projections represent natural approximations. For instance, Hilbert geometry allows to use any basis $(u_n)_{n\in\mathbb{N}}$ to define a sequence of bounded orthogonal projectors with finite rank:

$$\forall n \in \mathbb{N},\ P_n := u \in \mathcal{U} \to \sum_{k=0}^{n} \langle u, u_n \rangle_{\mathcal{U}}\, u_n,$$

such that $\forall u \in \mathcal{U},\ \lim_{n\to\infty} P_n u = u$. Another important property is that these orthogonal projectors are uniformly bounded (with norm lower than one), which can be seen as a consequence of Bessel's inequality:

$$\forall n \in \mathbb{N},\ \forall u \in \mathcal{U},\ \|P_n u\|_{\mathcal{U}} \leq \|u\|_{\mathcal{U}}\,.$$

A similar analysis can be given in Banach settings, using the more general notion of Schauder bases.

Definition 29. *Let $\mathcal{U}$ be a Banach space, a sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is a Schauder basis if*

$$\forall u \in \mathcal{U},\ \exists!(\alpha_n)_{n\in\mathbb{N}} \subset \mathbb{R},\ u = \sum_{n\geq 0} \alpha_n u_n.$$

Moreover, $\forall n \in \mathbb{N},\ u_n^* := u \in \mathcal{U} \to \alpha_n \in \mathbb{R}$ is bounded and linear (coordinate functional).

In particular, existence of a Schauder basis implies separability of the Banach space. However, the converse needs not be true in general [Enf73]. Similarly, a sequence of projectors $(P_n)_{n\in\mathbb{N}} \subset \mathcal{L}(\mathcal{U},\mathcal{U})$ is defined by:

$$P_n := u \in \mathcal{U} \to \sum_{k=0}^{n} \langle u, u_k^* \rangle_{\mathcal{U},\mathcal{U}^*}\, u_k,$$

and again, $\forall u \in \mathcal{U},\ \lim_{n\to\infty} P_n u = u$. Moreover, these operators are uniformly bounded (uniform boundedness principle):

$$\exists K > 0,\ \forall n \in \mathbb{N},\ \forall u \in \mathcal{U},\ \|P_n u\|_{\mathcal{U}} \leq K \|u\|_{\mathcal{U}}\,.$$

Likelihood approximation In the context of Bayesian inverse problems, Schauder bases can be used to define a sequence of approximate negative log-likelihoods by projecting the input as follows:

$$\forall n \in \mathbb{N},\ \forall (u,y) \in \mathcal{U} \times \mathcal{Y},\ \Phi_n(u;y) := \Phi(P_n u; y).$$

The question of the resulting consistency then depends on additional properties of Φ, see for instance [HN17] in case of prior measures with exponential tails or next section for an interesting example.

Links with Karhunen-Loève decomposition As previously announced, existence of Schauder bases in general Banach spaces is a delicate matter. The previous analysis is thus restricted to very few cases, including separable Hilbert spaces. However, when the prior measure is Gaussian, it will be shown that it is applicable to every Banach space, since a natural subspace with a Schauder basis can be constructed.

Theorem 22. *Let $\mathcal{U}$ be a Banach space, μ_0 a Radon Gaussian measure on $\mathcal{U}$ and $(h_n, h_n^*)_{n\in\mathbb{N}}$ a stochastic basis for μ_0, that is*

$$u = \sum_{n\geq 0} \langle u, u_n^* \rangle_{\mathcal{U},\mathcal{U}^*} u_n,\ \mu_0 - a.s.,$$

then there exists a Banach subspace $\mathcal{U}_0 \subset \mathcal{U}$ such that $\mu_0(\mathcal{U}_0) = 1$ and $(h_n, h_n^*)_{n\in\mathbb{N}}$ is a Schauder basis in $\mathcal{U}_0$, that is:

$$\forall u \in \mathcal{U}_0,\ u = \sum_{n\geq 0} \langle u, u_n^* \rangle_{\mathcal{U},\mathcal{U}^*} u_n.$$

The proof of this theorem is the object of [Oka86, Her81]. It is then sufficient to work in $\mathcal{U}_0$ directly, instead of the original Banach space $\mathcal{U}$. In particular, Fernique's theorem remains valid in $\mathcal{U}_0$, since $\mu_0(\mathcal{U}_0) = 1$.

5.2 Application to advection diffusion model

The problem of diffusion, which is ubiquitous in physics, engineering and biology, is usually represented by the following partial differential equation (without transport):

$$\begin{aligned} \frac{\partial z}{\partial t}(x,t) + \lambda(x,t)z(x,t) - D(x,t)\Delta z(x,t) &= f(x,t),\ \forall (x,t) \in \Omega\times]0,T], \\ z(x,t) &= 0,\ \forall (x,t) \in \Omega \times \{t=0\}, \\ z(x,t) &= 0,\ \forall (x,t) \in \partial\Omega\times]0,T]. \end{aligned} \tag{5.1}$$

where the spatial domain is $\Omega \subset \mathbb{R}^n$ ($n \leq 3$) and the final time is $T \in \mathbb{R}^+$. In real world applications, the quantity of interest z (hereafter called the solution of equation 5.1) is often the concentration of some chemical and evolves here from a null initial state under 3 distinct mechanisms: a) direct variation in concentration, given by the source f, b) diffusion at a positive rate D, c) production or depletion at a rate λ. Different hypotheses on these parameters leads to well-posedness of this equation, here we will suppose λ, D as positive constants in time and space and f as continuous, non-negative function. Besides the traditional computation of the solution from the parameters, one can use this model for the determination of an optimal control (e.g. source leading to the optimization of a particular functional) or the identification of parameters from partial knowledge of the solution in an inverse setting. This is the problem that will be of interest in this chapter. The motivation comes from a challenging identification problem in Biology, where the objective is to recover both decay and diffusion positive constant rates as well as a positive source, from finite and noisy observations of the solution z.

Step 1: Forward model analysis Before the application of Bayesian regularization, it is necessary to detail the regularity of z as a map. Using common variational techniques from PDE theory (see [Eva10] or [Bre11] for detailed introductions),

one can show that this equation has a unique weak solution (proposition 55) given $u = (\lambda, D, f)$ in a set $\mathcal{U}$ that will be specified later on. Moreover, this solution evolves smoothly in its parameters.

Proposition 55. *Let $\mathcal{P} = [0, \lambda_M] \times [D_m, D_M] \times L^2([0,T], L^2([0,L]))$ with $\lambda_M \geq 0$ and $0 < D_m \leq D_M$, then for all $u \in \mathcal{P}$, the evolution problem 5.1 has a unique weak solution, such that $\forall u \in \mathcal{P}$:*

1. $z(u) \in L^\infty([0,T], H_0^1[0,L]) \cap L^2([0,T], H^2[0,L])$,
2. $z(u)' \in L^2([0,T], L^2[0,L])$.

Moreover, $u \in \mathcal{P} \to z(u)$ satisfies the following properties:

1. *it is linear in f,*
2. *it satisfies the following energy estimate:*

$$\forall u \in \mathcal{P},\ \|z(u)\|_{L^2([0,T],H^2)} + \|z(u)'\|_{L^2([0,T],L^2[0,L])} \leq C\|f\|_{L^2([0,T],L^2[0,L])}$$

   with $C \geq 0$ a constant independent from u,

3. *it is locally Lipschitz, $\forall u \in \mathcal{P}$, $\forall r > 0$ such that $\mathcal{B}(u,r) \subset \mathcal{P}$, $\exists L(r,u) > 0$ such that $\forall u_1, u_2 \in \mathcal{B}(u,r) \times \mathcal{B}(u,r)$:*

$$\|z(u_1) - z(u_2)\|_{L^2([0,T],H^2)} + \|z(u_1)' - z(u_2)'\|_{L^2([0,T],L^2[0,L])} \leq L(r,u)\|u_1 - u_2\|_{\mathcal{P}}$$

4. *$z := u \in \mathcal{P} \to z(u) \in L^2([0,T], H^2[0,L])$ is twice Fréchet differentiable on $\mathcal{P}$.*

In particular, the solution is continuous in both time and space, with the following estimate:

$$\forall u \in \mathcal{P},\ \|z(u)\|_\infty \leq C\,\|f\|_{L^2([0,T],L^2[0,L])}\,.$$

Proof. Let $\mathcal{P} = \mathbb{R}\times]0,\infty[\times L^2([0,T], L^2[0,L])$ and note $H = L^2(0,L)$, $V = H_0^1(0,L)$ and $V^* = H^{-1}(0,L)$.

- Existence and uniqueness of the weak solution for equation 5.2 is established whenever $D > 0$ and $f \in L^2([0,T], V^*)$ (see theorems 3 and 4, chapter 7 in [Eva10] for a detailed proof using Galerkin approximations), in the sense that:

$$\left\langle z'(t), v(t)\right\rangle_{V,V^*} + \lambda \left\langle z(t), v(t)\right\rangle_H + D\left\langle z(t), v(t)\right\rangle_V = \left\langle f(t), v(t)\right\rangle_H,$$

  for all $v \in L^2([0,T], V)$, for almost-every t in $[0,T]$ and where $z(u) \in L^2([0,T], V)$, $z(u)' \in L^2([0,T], V^*)$ for all $u \in \mathcal{P}$. The problem is linear in f (null boundary and initial conditions), thus the solution map is.

- Furthermore, the source being more regular (here $f \in L^2([0,T], H)$), the solution z satisfies for all $u \in \mathcal{P}$ the following improved regularity :

– $z(u) \in L^2([0,T], H^2[0,L]) \cap L^\infty([0,T], V)$,
– $z(u)' \in L^2([0,T], H)$.

Now, since in $\mathcal{P}$ both λ, D are bounded above and $D \geq D_m > 0$, the following estimate is verified (Theorem 5, chapter 7 in [Eva10]):

$$\forall u \in \mathcal{P},\ \|z(u)\|_{L^2([0,T],H^2[0,L])} + \|z(u)'\|_{L^2([0,T],V)} \leq C\|f\|_{L^2([0,T],H)}$$

where $C \geq 0$ and is independent of u.

- Consider the reduced weak form of equation 5.1, that is

$$\langle F(z,u), v\rangle = 0,\ \forall v \in L^2([0,T], V),$$

with

$$\langle F(z,u), v\rangle = \int_0^T \langle z'(t) - f(t), v(t)\rangle_{V^*,V} + \lambda\langle z(t), v(t)\rangle_H + D\langle z(t), v(t)\rangle_V dt.$$

Let $u = (\lambda, D, f) \in \mathcal{P}$, $h^u = (h^\lambda, h^D, h^f) \in \mathcal{P}$ such that $u + h^u \in \mathcal{P}$, and $h^z \in L^2([0,T], H^2([0,L]))$ with $(h^z)' \in L^2([0,T], H)$ then $\forall v \in L^2([0,T], V)$:

$$\langle F(z + h^z, u + h^u) - F(z,u), v\rangle = \langle F_{z,u}(z,u)[h^z, h^u], v\rangle + c(h^u, h^z),$$

with

$$\begin{aligned} |c(h^u, h^z)| &= \left| h^\lambda \int_0^T \langle h^z, v\rangle_H dt + h^D \int_0^T \langle h^z, v\rangle_V dt \right| \\ &\leq C\|v\|_{L^2([0,T],V)} \left\|(h^\lambda, h^D, h^f)\right\|_{\mathcal{P}} \|h^z\|_{L^2([0,T],V)}, \end{aligned}$$

with C an other constant independent from u, and where:

$$\begin{aligned} \langle F_{z,u}(z,u)[h^z, h^u], v\rangle = &\int_0^T \langle (h^z)', v\rangle_{V^*,V} dt + h^\lambda \int_0^T \langle y, v\rangle_H dt \\ &+ h^D \int_0^T \langle z, v\rangle_V dt + \lambda \int_0^T \langle h^z, v\rangle_H dt \\ &+ D \int_0^T \langle h^z, v\rangle_V dt - \int_0^T \langle f, v\rangle_{V^*,V} dt. \end{aligned}$$

Moreover:

$$|\langle F_{z,u}(z,u)[h^z, h^u], v\rangle| \leq \|(h^u, h^z)\|\|v\|$$

thus $F_{z,u}$ is bounded, which shows that F is Fréchet-differentiable. Consider F_z the partial derivative of F w.r.t. its first variable:

$$\langle F_z(z,u)h^z, v\rangle = \int_0^T \left\langle (h^z)', v\right\rangle_{V,V^*} + \lambda\langle h^z, v\rangle_H + D\langle h^z, v\rangle_V dt,$$

which defines a unique weak solution h whenever $D > 0$ (using same arguments than previously) and F_z^{-1} is bounded. Because F is differentiable and F_z^{-1} exists and is bounded, the implicit function theorem applies and z is differentiable on $\mathcal{P}$. The second order differentiability uses the same arguments.

- Concerning the local Lipschitz continuity, let $u \in \mathcal{P}$, $r > 0$ such that $\mathcal{B}_{\mathcal{P}}(u,r) \subset \mathcal{P}$, $(u_1,u_2) \in \mathcal{B}_{\mathcal{P}}(u,r) \times \mathcal{B}_{\mathcal{P}}(u,r)$ and $z_1 = z(u_1)$, $z_2 = z(u_2)$ the associated weak solutions, then $\forall i \in \{1,2\}$, one has:

$$\langle z_i'(t), v(t)\rangle_{V^*,V} + \lambda_i \langle z_i(t), v(t)\rangle_H + D_i \langle z_i(t), v(t)\rangle_V = \langle f_i(t), v(t)\rangle_{V^*,V},$$

for almost-every t in $[0,T]$. By subtraction, it comes:

$$\begin{aligned}&\langle z_1'(t) - z_2'(t), v(t)\rangle_{V^*,V} + \lambda_1 \langle z_1(t) - z_2(t), v(t)\rangle_H + D_1 \langle z_1(t) - z_2(t), v(t)\rangle_V \\ &\quad = \langle f_1(t) - f_2(t), v(t)\rangle_{V^*,V} + (\lambda_2 - \lambda_1) \langle z_2(t), v(t)\rangle_H + (D_2 - D_1) \langle z_2(t), v(t)\rangle_V .\end{aligned}$$

From previous estimates, z_2 is bounded by $\|f_2\|_{L^2([0,T],H)}$. The very same analysis than previously leads to the following estimate:

$$\|z(u_1) - z(u_2)\|_{L^2([0,T],H^2)} + \|z(u_1)' - z(u_2)'\|_{L^2([0,T],L^2[0,L])} \leq L(r,u)\|u_1 - u_2\|_{\mathcal{P}},$$

where $L(u,r) \geq 0$.

- It is a known result (Theorem 4, section 5.9 in [Eva10]) that

$$\max_{t\in[0,T]} \|z(t)\|_{H^1[0,L]} \leq C \left(\|z\|_{L^2([0,T],H^2)} + \|z'\|_{L^2([0,T],L^2[0,L])} \right), \tag{5.2}$$

for every z such that $z \in L^2([0,T],H^2)$ and $z' \in L^2([0,T],H)$. Since in one dimensional spaces V injects continuously in $\mathcal{C}([0,L],\mathbb{R})$, the ∞-norm is controlled.

The proof is complete. □

Let's emphasize why the properties given in proposition 55 are important for the Bayesian inversion:

1. The energy estimate will be critical to establish the continuity of the posterior w.r.t. data as well as approximation, because it gives sufficient integrability conditions on z w.r.t. u,
2. Continuity (implied by Fréchet differentiability or local Lipschitz behaviour) implies that the solution map is measurable w.r.t. the Borel σ-algebra,
3. Second order Fréchet differentiability will be necessary for geometric methods in optimization (research of modes) and Markov-chain Monte-Carlo sampling (McMC),
4. The local Lipschitz behaviour gives a variational characterization of posterior modes (Maximum a Posteriori).

In the rest of this section, with the notations $\mathcal{U}^\lambda = [0,\lambda_M]$, $\mathcal{U}^D = [D_m, D_M]$ and $\mathcal{U}^f = C([0,L]\times[0,T],\mathbb{R})$, the parameters will be restricted to the subset

$$\mathcal{U} := \mathcal{U}^\lambda \times \mathcal{U}^D \times \mathcal{U}^f,$$

which is implicitly equipped with the norm

$$\|u\|_{\mathcal{U}} = |\lambda| + |D| + \|f\|_{\mathcal{U}^f} .$$

Since $\mathcal{U} \subset \mathcal{P}$ (with continuous injection), the solution map is well defined on this subset and keeps all its properties. Moreover, one can show that a weak solution of equation 5.1 for $u \in \mathcal{U}$ is also a strong solution [Bre11, Eva10], but these regularity results are not needed here. The only, but important result is that all solutions are continuous in both time and space, so point-wise measurements are possible.

Step 2: Choice of prior distribution The second step is to choose a prior probability distribution on $\mathcal{U}$, encoding all knowledge on the physics at hand, while being simple enough to keep analysis tractable. Here are the constraints given by the biological application:

- λ must be non-negative (decay),
- D must be positive (diffusion),
- f must be non-negative and continuous (it is a concentration).

Starting with the decay and diffusion parameters (λ, D), Borel prior distributions μ_0^λ and μ_0^D are chosen with densities w.r.t. Lebesgue's measure on $\mathcal{U}^\lambda$ and $\mathcal{U}^D$ respectively. Now, since f must be positive, the problem is re-parametrized with the new source term:

$$f^* = \exp(f), \tag{5.3}$$

where $f \in \mathcal{U}^f$. By selecting a Radon probability measure μ_0^f on the Banach space $\mathcal{U}^f$, both continuity and positivity of the source will be ensured almost-surely. Here, μ_0^f will be taken as the distribution of a continuous Gaussian random field with covariance operator $\mathcal{C}$. Finally, independence between the three components is imposed, leading to the following product prior distribution:

$$\mu_0(du) := \mu_0^\lambda(d\lambda) \otimes \mu_0^D(dD) \otimes \mu_0^f(df). \tag{5.4}$$

These choices clearly ensure the required constraints on u and $\mu_0(\mathcal{U}) = 1$. The exponential map in equation 5.3 can be replaced with any sufficiently differentiable function from $\mathbb{R}$ to $\mathbb{R}^+$ (to keep the second order Fréchet differentiability of the solution map). Alternative distributions are possible for f (Besov measure from [DLSV13] or more general convex measures from [HN17]) for the regularization. In practice however, this choice is also motivated by the fact that one can find a Gaussian reference measure μ_{ref} in the form

$$\mu_{ref} = \mu_{ref}^\lambda \otimes \mu_{ref}^D \otimes \mu_{ref}^f = \mathcal{N}(\lambda_{ref}, \sigma_\lambda^2) \otimes \mathcal{N}(D_{ref}, \sigma_D^2) \otimes \mathcal{N}(0, \mathcal{C}) = \mathcal{N}(u_{ref}, \mathcal{C}_{ref})$$

where $u_{ref} = (\lambda_{ref}, D_{ref}, 0) \in \mathcal{U}$ and

$$\mathcal{C}_{ref} : (\lambda, D, \mu) \in \mathbb{R}^2 \times \left(\mathcal{U}^f\right)^* \to (\sigma_\lambda^2\lambda, \sigma_D^2 D, \mathcal{C}\mu) \in \mathbb{R}^2 \times \mathcal{U}^f,$$

such that $\mu_0 << \mu_{ref}$. Indeed, choose $(\lambda_{ref}, D_{ref}) \in \mathbb{R}^2$ and $\sigma_\lambda^2, \sigma_D^2 > 0$ then $\mu_0 << \mu_{ref}$ with

$$\frac{d\mu_0}{d\mu_{ref}}(u) = \frac{d\mu_0^\lambda}{d\mu_{ref}^\lambda}(\lambda)\frac{d\mu_0^D}{d\mu_{ref}^D}(D).$$

This reference will be critical for modes analysis and McMC sampling.

Step 3: Well-posedness The third step in Bayesian inversion is to show that previous choices (forward model and prior distribution) lead to a well defined posterior measure. This is the purpose of proposition 56 which is very similar with the theory from chapter 4. Consider a dataset $y = (y_i)_{i\in[1,n]}$ which corresponds to observations at different times and locations $(t_i, x_i)_{i\in[1,n]}$ and assume they are produced from the following model (in vector notation):

$$y = \mathcal{G}(u) + \eta, \tag{5.5}$$

where $\eta \sim \mathcal{N}(0, \sigma_\eta^2 I_n)$ (I_n being identity matrix of dimension n) and $\mathcal{G} : \mathcal{U} \to \mathbb{R}^n$ is the observation operator, mapping directly the PDE parameter u to the value of the associated solution z at measurement locations (composition of solution map z with Dirac measures):

$$\forall u \in \mathcal{U},\ \mathcal{G}(u) = (z(u)[x_i, t_i])_{i\in[1,n]}\,.$$

Here is a simple lemma giving an energy estimate for the observation operator.

Lemma 12. *Let $\mathcal{G}$ be the observation operator from equation 5.5 then*

$$\forall u \in \mathcal{U},\ \|\mathcal{G}(u)\|_{\mathbb{R}^n} \leq C \exp\left(\|f\|_{\mathcal{U}^f}\right),$$

where $C \geq 0$ is constant in u.

Proof. By definition of the observation operator and since there is a Sobolev embedding, it comes

$$\|\mathcal{G}(u)\|_{\mathbb{R}^n} \leq n \|z(u)\|_{C([0,L]\times[0,T],\mathbb{R})} \leq C \|z(u)\|_{W([0,T],L^2,H_0^1)}\,.$$

Now, the energy estimate from proposition 55 jointly with previous equation gives:

$$\|\mathcal{G}(u)\|_{\mathbb{R}^n} \leq C \exp\left(\|f\|_{\mathcal{U}^f}\right),$$

with $C \geq 0$ independent of u and the proof is complete. □

The following proposition establishes the existence of a posterior probability measure μ^y, expressing how observations y changed prior beliefs on the parameter u.

Proposition 56. *Let $\mathcal{G}$ be the observation operator defined in equation 5.5 and μ_0 the Radon probability measure defined in equation 5.4, then there exists a unique*

posterior measure μ^y, characterized by the following Radon-Nikodym density w.r.t. μ_0:

$$\forall y \in \mathcal{Y},\ \forall u \in \mathcal{U},\ \frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\left(-\Phi(u;y)\right),$$

with

$$\forall y \in \mathcal{Y},\ \forall u \in \mathcal{U},\ \Phi(u;y) = \frac{1}{2\sigma_\eta^2}\|y - \mathcal{G}(u)\|^2,$$

and

$$\forall y \in \mathcal{Y},\ Z(y) = \int_{\mathcal{U}} \exp(-\Phi(u;y)\mu_0(du).$$

Proof. Let μ_0 be the Radon probability measure defined in equation 5.4 and consider the following Gaussian negative log-likelihood:

$$\Phi(u;z) = \frac{1}{2\sigma_\eta^2}\left\|y - \mathcal{G}(u)\right\|_{\mathcal{Y}}^2.$$

Since z is continuous in u (proposition 55), Φ is measurable w.r.t. μ_0 for all $y \in \mathcal{Y}$. Now, Φ is non-negative, thus

$$\forall y \in \mathcal{Y},\ Z(y) = \int_{\mathcal{U}} \exp\left(-\Phi(u;y)\right)\mu_0(du) \leq 1.$$

Let $u_0 = (\lambda_0, D_0, f_0) \in \mathcal{U}$ and $r > 0$ such that $\mathcal{B}_{\mathcal{U}}(u_0, r] \subset \mathcal{U}$. Then $\forall u \in \mathcal{B}_{\mathcal{U}}(u_0, r]$ and $y \in \mathcal{B}_{\mathcal{Y}}(0, r]$, lemma 12 gives:

$$\left\|\mathcal{G}(u)\right\|_{\mathcal{Y}} \leq C \exp\left(\|f_0\|_{\mathcal{U}^f} + r\right),$$

and consequently:

$$\begin{aligned}
&\Phi(u;y)\\
&= \frac{1}{2\sigma_\eta^2}\left\|y - \mathcal{G}(u)\right\|_{\mathcal{Y}}^2,\\
&= \frac{1}{2\sigma_\eta^2}\left(\|y\|_{\mathcal{Y}}^2 + \|\mathcal{G}(u)\|_{\mathcal{Y}}^2 - 2\left\langle y, \mathcal{G}(u)\right\rangle_{\mathcal{Y}}\right),\\
&\leq \frac{1}{2\sigma_\eta^2}\left(\|y\|_{\mathcal{Y}}^2 + \|\mathcal{G}(u)\|_{\mathcal{Y}}^2 + 2\left\|y\right\|_{\mathcal{Y}}\left\|\mathcal{G}(u)\right\|_{\mathcal{Y}}\right),\\
&\leq \frac{1}{2\sigma_\eta^2}\left(r^2 + C^2\exp\left(2\left\|f_0\right\|_{\mathcal{U}^f} + 2r\right) + 2rC\exp\left(\|f_0\|_{\mathcal{U}^f} + r\right)\right) = g_1(r, u_0),
\end{aligned}$$

with $C \geq 0$ independent of u and y. It follows that Φ is locally bounded above and

$$\begin{aligned}
\forall y \in \mathcal{Y},\ Z(y) &\geq \int_{\mathcal{B}_{\mathcal{U}}(u_0,r]} \exp\left(-\Phi(u;y)\right)\mu_0(du),\\
&\geq \exp\left(-g_1(r,u_0)\right)\mu_0(\mathcal{B}_{\mathcal{U}}(u_0,r]),\\
&\geq c(u_0,r) > 0.
\end{aligned}$$

Finally, $\forall y \in \mathcal{Y}$, the following application is a Radon-Nikodym density w.r.t. μ_0:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)}\exp(-\Phi(u;y)).$$

□

For the well-posedness, it remains to see that this posterior measure is continuous in y, which is the object of next proposition.

Proposition 57. *Let μ^y the posterior distribution defined by the Radon-Nikodym density from proposition 56, then it is continuous in y w.r.t. Hellinger distance.*

Proof. Using a similar factorization than in the proof of theorem 17, it comes:

$$\forall y_1, y_2 \in \mathcal{Y},\ d_{Hell}(\mu^{y_1}, \mu^{y_2})^2 \leq I_1 + I_2,$$

with

$$I_1 = \frac{1}{Z(y_1)} \int_{\mathcal{U}} \left(\exp\left(-\frac{1}{2}\Phi(u; y_1)\right) - \exp\left(-\frac{1}{2}\Phi(u; y_2)\right)\right)^2 \mu_0(du),$$
$$I_2 = \left| \frac{1}{\sqrt{Z(y_1)}} - \frac{1}{\sqrt{Z(y_2)}} \right|^2 Z(y_2).$$

From proposition 56, it is clear that $\forall r > 0$,

$$\forall y \in \mathcal{B}_{\mathcal{Y}}(0, r],\ Z(y) \geq c(u_0, r) > 0.$$

Now, since Φ is non-negative, it comes that:

$$\begin{aligned} \forall (y_1, y_2) \in \mathcal{Y}^2,\ & |Z(y_1) - Z(y_2)| \\ &= \left| \int_{\mathcal{U}} \exp\left(-\Phi(u; y_1)\right) - \exp\left(-\Phi(u; y_2)\right) \mu_0(du) \right| \\ &\leq \int_{\mathcal{U}} |\Phi(u; y_1) - \Phi(u; y_2)|\, \mu_0(du), \end{aligned}$$

by application of the mean-value theorem to the exponential function. Similarly:

$$\begin{aligned} I_1 &= \frac{1}{Z(y_1)} \int_{\mathcal{U}} \left(\exp\left(-\frac{1}{2}\Phi(u; y_1)\right) - \exp\left(-\frac{1}{2}\Phi(u; y_2)\right)\right)^2 \mu_0(du) \\ &\leq \frac{1}{Z(y_1)} \int_{\mathcal{U}} \frac{1}{2} |\Phi(u; y_1) - \Phi(u; y_2)|^2 \mu_0(du). \end{aligned}$$

Now, $\forall r > 0$ and $\forall y_1, y_2 \in \mathcal{B}_{\mathcal{Y}}(0, r]$

$$\begin{aligned} & |\Phi(u; y_1) - \Phi(u; y_2)| \\ &= \frac{1}{2\sigma_\epsilon^2} \left| \|y_1\|_{\mathcal{Y}}^2 - \|y_2\|_{\mathcal{Y}}^2 + 2\left\langle y_2 - y_1, \mathcal{G}(u)\right\rangle_{\mathcal{Y}} \right| \\ &= \frac{1}{2\sigma_\epsilon^2} \left| (\|y_1\|_{\mathcal{Y}} - \|y_2\|_{\mathcal{Y}})(\|y_1\|_{\mathcal{Y}} + \|y_2\|_{\mathcal{Y}}) + 2\left\langle y_2 - y_1, \mathcal{G}(u)\right\rangle_{\mathcal{Y}} \right| \\ &\leq \frac{1}{\sigma_\eta^2} (r + \|\mathcal{G}(u)\|_{\mathcal{Y}}) \|y_1 - y_2\|_{\mathcal{Y}}, \\ &\leq \frac{1}{\sigma_\eta^2} \left(r + C \exp\left(\|f\|_{\mathcal{U}^f}\right)\right) \|y_1 - y_2\|_{\mathcal{Y}}, \\ &\leq g_2(u, r) \|y_1 - y_2\|_{\mathcal{Y}}, \end{aligned}$$

by lemma 12 with $g_2(u, r) = \frac{1}{\sigma_\eta^2} \left(r + C \exp\left(\|f\|_{\mathcal{U}^f}\right)\right)$ and C independent of u and r. The hypothesis are precisely given such that $g_2 \in L^2(\mu_0)$, thus $I_1 \leq C(r) \|y_1 - y_2\|_{\mathcal{Y}}^2$.

Moreover, $\forall r > 0,\ \forall y \in \mathcal{B}_{\mathcal{Y}}(0,r],\ Z(y) \geq c(u,r) > 0$ thus by the mean-value theorem

$$I_2 \leq K \left|Z(y_1) - Z(y_2)\right|^2 \leq K' \left\|y_1 - y_2\right\|_{\mathcal{Y}}^2,$$

the proof is complete. □

Step 4: Maximum a posteriori In the previous section, the well-posedness of the Bayesian inverse problem has been proved. However, the posterior distribution is only known up to a multiplicative constant, through its density w.r.t. μ_0. In this application, the posterior measure μ^y will be summarized by a posterior mode (MAP estimator), using next proposition.

Proposition 58. *Let μ_0 be the prior probability measure defined in equation 5.4 and μ_{ref} the Gaussian reference measure from equation 5.2. Suppose additionally that*

$$ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right)$$

is locally Lipschitz, then the modes of μ_z are exactly the minimizers of the following Onsager-Machlup functional:

$$I(u) := \Phi(u;y) + \frac{1}{2}\|u - u_{ref}\|_{\mu_{ref}}^2 - ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right),$$

where $\|.\|_{\mu_{ref}}$ and u_{ref} are respectively the norm of the Cameron-Martin space and the mean of μ_{ref}. A minimizer will be noted $u_{MAP} = (\lambda_{MAP}, D_{MAP}, f_{MAP})$.

Proof. The posterior measure μ^y is absolutely continuous w.r.t. μ_{ref} with the following Radon-Nikodym density:

$$\frac{d\mu^y}{d\mu_{ref}}(u) = \frac{d\mu^y}{d\mu_0}(u)\frac{d\mu_0}{d\mu_{ref}}(u) = \frac{1}{Z(y)}\exp(-\tilde{\Phi}(u;y)).$$

where $\tilde{\Phi}(u;y) = \Phi(u;y) - \ln\left(\frac{d\mu_0}{d\mu_{ref}}(u)\right)$. Let us now show that Φ is locally Lipschitz in its first argument:

$$\begin{aligned}
\|\Phi(u_1;y) - \Phi(u_2;y)\|_Y &= \frac{1}{2\sigma_\eta^2}\left|\|y - \mathcal{G}(u_1)\|_{\mathcal{Y}}^2 - \|y - \mathcal{G}(u_2)\|_Y^2\right|, \\
&= \frac{1}{2\sigma_\eta^2}\left|\|\mathcal{G}(u_1)\|_{\mathcal{Y}}^2 - \|\mathcal{G}(u_2)\|_{\mathcal{Y}}^2 + 2\left\langle y, \mathcal{G}(u_2) - \mathcal{G}(u_1)\right\rangle_{\mathcal{Y}}\right|, \\
&\leq \frac{\|\mathcal{G}(u_1)\|_{\mathcal{Y}} + \|\mathcal{G}(u_2)\|_{\mathcal{Y}} + 2\|y\|_Y}{2\sigma_\eta^2}\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_{\mathcal{Y}},
\end{aligned}$$

and since $\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_{\mathcal{Y}} \leq C\left(\|z(u_1) - z(u_2)\|_{L^2([0,T],H^2[0,L])} + \|z(u_1)' - z(u_2)'\|_{L^2([0,T],V)}\right)$, we conclude that Φ and thus $\tilde{\Phi}$ are locally Lipschitz. Now:

$$\begin{aligned}
\frac{J^r(u_1)}{J^r(u_2)} &= \frac{\int_{\mathcal{B}_{\mathcal{U}}(u_1,r]}\exp(-\tilde{\Phi}(u;y))\mu_{ref}(du)}{\int_{\mathcal{B}_{\mathcal{U}}(u_2,r]}\exp(-\tilde{\Phi}(v;y))\mu_{ref}(dv)} \\
&= \frac{\int_{\mathcal{B}_{\mathcal{U}}(u_1,r]}\exp(-\tilde{\Phi}(u;y) + \tilde{\Phi}(u_1;y))\exp(-\tilde{\Phi}(u_1;y))\mu_{ref}(du)}{\int_{\mathcal{B}_{\mathcal{U}}(u_2,r]}\exp(-\tilde{\Phi}(v;y) + \tilde{\Phi}(u_2;y))\exp(-\tilde{\Phi}(u_2;y))\mu_{ref}(dv)}.
\end{aligned}$$

Now,

$$\frac{J^r(u_1)}{J^r(u_2)} \leq \exp\left(rC - \tilde{\Phi}(u_1, y) + \tilde{\Phi}(u_2; y)\right) \frac{\int_{\mathcal{B}_{\mathcal{U}}(u_1,r]} \mu_{ref}(du)}{\int_{\mathcal{B}_{\mathcal{U}}(u_2,r]} \mu_{ref}(dv)}$$

and finally

$$\limsup_{r\to 0} \frac{J^r(u_1)}{J^r(u_2)} \leq \exp\left(-I(u_1) + I(u_2)\right).$$

A similar argument leads to

$$\liminf_{r\to 0} \frac{J^r(u_1)}{J^r(u_2)} \geq \exp\left(-I(u_1) + I(u_2)\right).$$

In conclusion $\lim_{r\to 0} \frac{J^r(u_1)}{J^r(u_2)} = \exp\left(-I(u_1) + I(u_2)\right)$. For a fixed value u_2, this quantity is maximized when u_1 is a minimizer of I, the proof is then complete. □

Step 5: Karhunen-Loève approximation In this paragraph, the approximation of the previous Bayesian inverse problem is considered. Since μ_0^f is a Gaussian measure, the Karhunen-Loève decomposition from chapter 2 can be used in a projection-based approximation. Indeed, the stochastic basis $(h_n^*, h_n)_{n\in\mathbb{N}} \subset \mathcal{U}_f^* \times \mathcal{U}_f$ built from the Karhunen-Loève decomposition, is a Schauder basis of a Banach subspace $\mathcal{U}_0 \subset \mathcal{U}_f$. As a consequence, the following sequence of projectors:

$$P_n : f \in \mathcal{U}_0 \to \sum_{k=0}^{n} \langle f, h_k^* \rangle_{\mathcal{U}_f, \mathcal{U}_f^*} h_k \in \mathcal{U}^0,$$

are uniformly bounded with a constant K ($\forall n \in \mathbb{N},\ \|P_n\|_{\mathcal{L}(\mathcal{U}_0^f, \mathcal{U}_0^f)} \leq K$). Furthermore, the extended projectors are defined as follows to take into account the other parameters:

$$\forall n \in \mathbb{N},\ \forall u \in \mathcal{U},\ Q_n u = (\lambda, D, P_n f).$$

The notation $\mathcal{U}_0 = \mathcal{U}^\lambda \times \mathcal{U}^D \times \mathcal{U}_0^f$ will also be used to restrict the source space from $\mathcal{U}^f$ to $\mathcal{U}_0^f$. In this work, the considered approximations are of projection-type as follows:

$$\forall n \in \mathbb{N},\ \forall (u, y) \in \mathcal{U} \times \mathcal{Y},\ \Phi_n(u; y) := \Phi(Q_n u, y).$$

The study of consistency starts with the following lemma, giving useful energy type inequalities.

Lemma 13. *Let μ_0 be the prior measure defined in equation 5.4, $\mathcal{G}$ the observation operator from equation 5.5 and $(Q_n)_{n\in\mathbb{N}}$ the sequence of operators defined previously, then it comes $\forall n \in \mathbb{N},\ \forall u = (\lambda, D, f) \in \mathcal{U}_0$:*

$$\|\mathcal{G}(u) - \mathcal{G}(Q_n u)\|_{\mathcal{Y}} \leq C_1 \exp\left(\alpha_1 \|f\|_{\mathcal{U}_0^f}\right) \|f - P_n f\|_{\mathcal{U}^f},$$
$$\|\mathcal{G}(Q_n u)\|_{\mathcal{Y}} \leq C_2 \exp\left(\alpha_2 \|f\|_{\mathcal{U}_0^f}\right),$$

with $C_1, C_2, \alpha_1, \alpha_2 \geq 0$ positive constants independent of $u = (\lambda, D, f)$.

Proof. Let $n \in \mathbb{N}$ and $u \in \mathcal{U}_0$, then using the Sobolev embedding of $W([0,T], L^2, H_0^1)$ in $\mathcal{U}^f = C([0,L] \times [0,T], \mathbb{R})$, the energy estimate from proposition 55 and the embedding of $\mathcal{U}_0^f$ in $\mathcal{U}^f$, it comes:

$$\begin{aligned}
\|\mathcal{G}(Q_n u)\|_{\mathcal{Y}} &\leq C \left(\|z(Q_n u)\|_{L^2([0,T],H^2[0,L])} + \left\|z(Q_n u)'\right\|_{L^2([0,T],H)} \right) \\
&\leq C' \exp\left(\|P_n f\|_{\mathcal{U}^f} \right) \\
&\leq C' \exp\left(c \|P_n f\|_{\mathcal{U}_0^f} \right) \\
&\leq C' \exp\left(cK \|f\|_{\mathcal{U}_0^f} \right).
\end{aligned}$$

with $C, C' \geq 0$ constants in u. Concerning the second inequality, using the linearity of the solution map z in f, one has:

$$\|\mathcal{G}(u) - \mathcal{G}(Q_n u)\|_{\mathcal{Y}} = \|G(\lambda, D, f - P_n f)\|_Y ,$$

then using again the energy estimate from proposition 55 and the mean-value theorem applied to the exponential map, it comes:

$$\begin{aligned}
\|\mathcal{G}(u) - \mathcal{G}(Q_n u)\|_{\mathcal{Y}} &\leq C \|\exp(f) - \exp(P_n f)\|_{\mathcal{U}^f} \\
&\leq C \exp\left(\|f\|_{\mathcal{U}^f} + \|P_n f\|_{\mathcal{U}^f} \right) \|f - P_n f\|_{\mathcal{U}^f} \\
&\leq C \exp\left(c \|f\|_{\mathcal{U}_0^f} + c \|P_n f\|_{\mathcal{U}_0^f} \right) \|f - P_n f\|_{\mathcal{U}^f} \\
&\leq C \exp\left(c(1+K) \|f\|_{\mathcal{U}_0^f} \right) \|f - P_n f\|_{\mathcal{U}^f} ,
\end{aligned}$$

and the proof is complete. □

The result of previous lemma can directly be adapted to the Gaussian negative log-likelihood.

Lemma 14. *Consider the negative log-likelihood from proposition 56 and the sequence of approximations $(\Phi_n)_{n \in \mathbb{N}}$, then $\forall n \in \mathbb{N}$:*

$$\begin{aligned}
\forall (u, y) \in \mathcal{U}_0 \times \mathcal{Y},\ |\Phi(u; y) - \Phi_n(u; y)| &\leq C_1 \exp\left(\alpha_1 \|f\|_{\mathcal{U}_0^f} \right) \|f - P_n f\|_{\mathcal{U}_f} \\
\forall (u, y) \in \mathcal{U}_0 \times \mathcal{Y},\ |\Phi_n(u; y)| &\leq C_2 \exp\left(\alpha_2 \|f\|_{\mathcal{U}_0^f} \right),
\end{aligned}$$

with $C_1, C_2, \alpha_1, \alpha_2 \geq 0$ positive constants independent from both u and n.

Proof. Let $y \in \mathcal{Y}$, $n \in \mathbb{N}$ and $u \in \mathcal{U}_0$, then:

$$\begin{aligned}
|\Phi(u; y) - \Phi_n(u; y)| &= \frac{1}{2\sigma_\eta^2} \left| \|\mathcal{G}(u)\|_{\mathcal{Y}}^2 - \|\mathcal{G}(Q_n u)\|_{\mathcal{Y}}^2 + 2 \langle y, \mathcal{G}(Q_n u) - \mathcal{G}(u) \rangle_{\mathcal{Y}} \right| \\
&\leq \frac{1}{2\sigma_\eta^2} \left(2 \|y\|_{\mathcal{Y}} + \|\mathcal{G}(Q_n u)\|_{\mathcal{Y}} + \|\mathcal{G}(u)\|_{\mathcal{Y}} \right) \|\mathcal{G}(Q_n u) - \mathcal{G}(u)\|_{\mathcal{Y}} .
\end{aligned}$$

Thus, using lemma 13, it comes:

$$|\Phi(u; y) - \Phi_n(u; y)| \leq C \exp\left(\alpha \|f\|_{\mathcal{U}_0^f} \right) \|f - P_n f\|_{\mathcal{U}_f} .$$

with $C, \alpha \geq 0$ and independent from u and n. Similarly, one has:

$$\begin{aligned}\Phi_n(u;y) &= \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(Q_n u)\|_{\mathcal{Y}}^2 \\ &\leq \frac{1}{\sigma_\eta^2} \left(\|y\|_{\mathcal{Y}}^2 + \|\mathcal{G}(Q_n u)\|_{\mathcal{Y}}^2 \right) \\ &\leq C' \exp\left(\alpha' \|f\|_{\mathcal{U}_0^f} \right),\end{aligned}$$

with $C', \alpha' \geq 0$ independent from u and n and the proof is complete. □

With these two lemmas at hand, it will now be easy to show the well-posedness of the approximated posteriors, their consistency and give a speed of convergence.

Proposition 59. *Let $(\Phi_n)_{n\in\mathbb{N}}$ the sequence of approximated negative log-likelihoods, $y \in \mathcal{Y}$ a data, then there exists a sequence of posterior measures $(\mu_n^y)_{n\in\mathbb{N}}$, defined by the following Radon-Nikodym densities $\forall n \in \mathbb{N}$:*

$$\forall u \in \mathcal{U}_0,\ \frac{d\mu_n^y}{d\mu_0}(u) = \frac{1}{Z_n(y)} \exp\left(-\Phi_n(u;y)\right),$$

with $Z_n(y) = \int_{\mathcal{U}_,} \exp\left(-\Phi_n(u;y)\right) \mu_0(du)$.

Proof. Let $n \in \mathbb{N}$ and $y \in \mathcal{Y}$, then by composition of continuous maps, Φ_n is measurable in u and

$$Z_n(y) = \int_{\mathcal{U}_,} \exp\left(-\Phi_n(u;y)\right) \mu_0(du) \leq 1,$$

since $\Phi_n(u;y) \geq 0$. Moreover, using lemma 14 and corollary 4, $u \in \mathcal{U}_0$ and $r > 0$ such that $\mathcal{B}_{\mathcal{U}_0}(u, r] \subset \mathcal{U}_0$:

$$\forall u' \in \mathcal{B}_{\mathcal{U}_0}(u, r],\ \Phi_n(u';y) \leq C \exp\left(\alpha(\|f\|_{\mathcal{U}_0} + r)\right) \leq C' \exp\left(\alpha \|f\|_{\mathcal{U}_0}\right),$$

which implies that

$$Z_n(y) \geq \mu_0\left(\mathcal{B}_{\mathcal{U}_0}(u, r]\right) \exp\left(-C' \exp\left(\alpha \|f\|_{\mathcal{U}_0}\right)\right) > 0,$$

as $\mu_0(\mathcal{U}_0) = 1$. Consequently, the following family of applications

$$u \in \mathcal{U}_0 \to \frac{1}{Z_n(y)} \exp\left(-\Phi_n(u;y)\right) \in \mathbb{R}^+,$$

are Radon-Nikodym densities w.r.t. μ_0 and defines a sequence of posterior measures $(\mu_n^y)_{n\in\mathbb{N}}$. □

Corollary 15. *Let $(\mu_n^y)_{n\in\mathbb{N}}$ the sequence of approximated posterior measures from proposition 59, then $\forall n \in \mathbb{N}$, μ_n^y is continuous in y.*

It remains to see that this sequence of approximated posterior measures $(\mu_n^y)_{n\in\mathbb{N}}$ is consistent and give an estimation of the convergence speed.

Proposition 60. *Let $y \in \mathcal{Y}$, μ^y the posterior measure defined in proposition 56, $(\mu_n^y)_{n\in\mathbb{N}}$ the sequence of approximated posterior measures from 59 then*

$$d_{Hell}(\mu^y, \mu_n^y) \leq C\mathbb{E}^{\mu_0^f}\left[\|f - P_n f\|_{\mathcal{U}^f}^2\right]^{\frac{1}{2}},$$

and $(\mu_n^y)_{n\in\mathbb{N}}$ of approximate posterior measures is consistent.

Proof. Let $n \in \mathbb{N}$ and $y \in \mathcal{Y}$, then the Hellinger distance can be controlled as follows, using the very same method than theorem 21 in chapter 4:

$$d_{Hell}(\mu^y, \mu_n^y)^2 \leq 2(I_1 + I_2),$$

with:

$$I_1 = \frac{1}{Z(y)}\int_{\mathcal{U}_0}\left(\exp\left(-\frac{\Phi(u;y)}{2}\right) - \exp\left(-\frac{\Phi_n(u;y)}{2}\right)\right)^2 \mu_0(du),$$
$$I_2 = \left(\frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{Z_n(y)}}\right)^2 Z_n(y).$$

Using the mean-value theorem (exponential function) with $\Phi(u;y) \geq 0$ and lemma 14, it comes:

$$\begin{aligned}
I_1 &\leq \frac{1}{Z(y)}\int_{\mathcal{U}_0}\left(\Phi(u;y) - \Phi_n(u;y)\right)^2 \mu_0(du),\\
&\leq \frac{1}{Z(y)}\int_{\mathcal{U}_0} C_1 \exp\left(2\alpha_1 \|f\|_{\mathcal{U}_0^f}\right)\|f - P_n f\|_{\mathcal{U}_f}^2 \mu_0(du)\\
&\leq \frac{C}{Z(y)}\int_{\mathcal{U}_0^f} \exp\left(2\alpha_1 \|f\|_{\mathcal{U}_0^f}\right)\|f - P_n f\|_{\mathcal{U}_f}^2 \mu_0^f(df)\\
&\leq \frac{C'}{Z(y)}\mathbb{E}\left[\|P_n f - f\|_{\mathcal{U}^f}^4\right]^{\frac{1}{2}}
\end{aligned}$$

Using the Kahane-Kintchine inequality, it follows that

$$I_1 \leq C\mathbb{E}\left[\|f - P_n f\|_{\mathcal{U}^f}^2\right],$$

with $C \geq 0$ independent from n. Concerning I_2, since $\forall n \in \mathbb{N},\ \min(Z_n(y), Z(y)) \geq c > 0$ and $Z_n(y) \leq 1$ (proposition 59), a new application of the mean-value theorem (on the inverse square-root) gives:

$$I_2 \leq C'|Z(y) - Z_n(y)|^2.$$

Again, a new application of the mean-value theorem to the exponential gives:

$$\begin{aligned}
|Z(y) - Z_n(y)| &= \left|\int_{\mathcal{U}_0} \exp\left(-\Phi(u;y)\right) - \exp\left(-\Phi_n(u;y)\right)\mu_0(du)\right|\\
&\leq \int_{\mathcal{U}_0} |\Phi(u;y) - \Phi_n(u;y)|\mu_0(du)\\
&\leq \int_{\mathcal{U}_0} C_1 \exp\left(\alpha_1 \|f\|_{\mathcal{U}_0^f}\right)\|f - P_n f\|_{\mathcal{U}_f}\mu_0(du)\\
&\leq C''\mathbb{E}\left[\|f - P_n f\|_{\mathcal{U}_f}^2\right]^{\frac{1}{2}}
\end{aligned}$$

using lemma 14 and Cauchy-Schwarz inequality. It is now clear that

$$d_{Hell}(\mu^y, \mu_n^y) \leq C\mathbb{E}\left[\|f - P_n f\|_{\mathcal{U}^f}^2\right]^{\frac{1}{2}}.$$

□

The speed of convergence of the approximate posterior is given by the quality of the series representation of the source. Using the material developed in the first part of this thesis (chapter 1 and 2), it is encouraged to use asymptotically optimal decompositions.

5.3 Numerical results

This section is dedicated to the practical implementation of the previous Bayesian inverse problem. First, the prior distribution and the Onsager-Machlup functional are specified, respecting all previous assumptions. Then, quantitative results on a real-world dataset taken from [BBCCS$^+$13] are given. It consists in $n = 508$ different measures which are non-uniformly spread in time and space (precise repartition can be seen as dots in figure 5.4, right side).

Choice of measures Here is a simple choice of prior achieving all previous requirements.

- **Decay and diffusion**. Here $\mu_0^D = \mathcal{U}([D_m, D_M])$ with $0 < D_m \leq D_M$ and $\mu_0^\lambda = \mathcal{U}([0, \lambda_M])$ with $\lambda_M \geq 0$ will be chosen. The relative upper bounds λ_M and D_M are tuned to 2 to be sufficiently large and respect previous estimations from [BBCCS$^+$13]. The lower bound D_m is arbitrarily fixed at 10^{-8}.

- **Source**. As previously stated, the prior measure on the source term μ_0^f will be chosen as a Radon Gaussian measure on $\mathcal{U}_f = C([0, L] \times [0, L], \mathbb{R})$. More specifically it is a continuous Gaussian random field with covariance kernel (tensor Brownian bridge):

  $$k\left([x,t],[x',t']\right) = \sigma^2 \frac{16}{TL}\left(\min(x,x') - \frac{xx'}{L}\right)\left(\min(t,t') - \frac{tt'}{T}\right).$$

  It is chosen because it is almost-surely null on all the boundaries of $[0, T] \times [0, L]$. The constants are used to rescale the variance globally, such that it has maximum value σ^2 (at the centre).

- Under both these choices, the Radon-Nikodym density of the prior distribution w.r.t. the Gaussian reference measure is

  $$\frac{d\mu_0}{d\mu_{ref}}(u)$$
  $$= \frac{2\pi\sigma_\lambda\sigma_D}{\lambda_M(D_M - D_m)} \exp\left(\frac{(\lambda - \lambda_{ref})^2}{2\sigma_\lambda^2} + \frac{(D - D_{ref})^2}{2\sigma_D^2}\right) \chi_{[0,\lambda_M]}(\lambda)\chi_{[D_m,D_M]}(D).$$

The parameters of μ_{ref} are tuned choosing:

$$\begin{aligned}\lambda_{ref} &= \frac{\lambda_M}{2},\\ \sigma_\lambda^2 &= \frac{\lambda_M^2}{12},\\ D_{ref} &= \frac{D_M - D_m}{2},\\ \sigma_D^2 &= \frac{(D_M - D_m)^2}{12},\end{aligned}$$

which corresponds to a moment matching strategy (or a minimization of the Kullback-Leibler divergence). Finally, the associated Onsager-Machlup functional is obtained using proposition 58:

$$I(u) = \frac{1}{2\sigma_\eta^2}\|y - \mathcal{G}(u)\|^2 + \frac{1}{2}\|f\|^2_{\mu_0^f}.$$

Remark that the regularization only acts on f, while the influence of λ and D only appears in the least-square term. This functional will be used for posterior modes computation as well as hyper-parameter tuning.

Numerical solver In this work, the solution map is approximated using finite elements in space (FEniCS library in Python, see [ABH+15] and [LL17]) and finite differences in time (implicit Euler scheme). A hundred $P1$ finite elements are used along thirty time steps on a desktop computer (Intel i7-3770 with 8Gb of RAM memory, late 2015). The domain parameters are $L = 100$ and $T = 100$ (percentages). All quantities related to negative log-likelihood derivatives (Gradient and Gauss-Newton Hessian matrix) are computed using discrete adjoint methods (see [HPUU09] or [Hei08]) to keep scalability.

Karhunen-Loève basis Concerning the prior measure, the discretization only concerns the source term f since λ and D are taken as scalars. Here, a truncated basis is used:

$$\tilde{f} = \sum_{1\leq i_1,i_2\leq N} \sqrt{\lambda_{i_1,i_2}}\xi_{i_1,i_2}\varphi_{i_1,i_2},$$

where $(\xi_{i_1,i_2})_{N\geq i_1,i_2\geq 0}$ are i.i.d. $\mathcal{N}(0,1)$ random variables. Both the functions φ_{i_1,i_2} and scalars λ_{i_1,i_2} are obtained from a tensor product of the one dimensional Schauder basis on dyadic intervals(given in chapter 2 as the Lévy-Ciesielski basis). In this work, $K = 20$ (400 basis functions) thus $\tilde{u} = (\lambda, D, \xi_1, ..., \xi_{N^2})$ is of dimension 402. Alternative bases for this process can be used in a similar way. In particular, it is emphasized that any Gaussian process with two dimensional covariance kernel can be used thanks to chapter 2.

Computation of Maximum a posteriori estimator The variational characterization of the MAP estimators as minimizers of the Onsager-Machlup functional is very useful in practice. Indeed, the MAP can be computed by a direct optimization of the following discretized functional:

$$\tilde{I}(\tilde{u}) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(\tilde{u})\|_{\mathcal{Y}}^2 + \frac{1}{2} \sum_{i=1}^{N^2} \xi_i^2$$

Practical optimization is done using L-BFGS-B algorithm from the Scipy library [BLNZ95] with explicit gradient:

$$\nabla \tilde{I}(\tilde{u}) = \frac{1}{\sigma^2} \left(y - \mathcal{G}(\tilde{u})\right)^T \nabla \mathcal{G}(\tilde{u}) + \xi,$$

the quantity $\nabla \mathcal{G}$ being obtained by discrete adjoint method, which only cost an additional PDE solve (adjoint model).

Tuning of hyper-parameters In common Bayesian methods, such as Gaussian process regression for instance, hyper-parameters are usually calibrated using maximum likelihood estimation (since there is a closed formulae), moment matching or leave-one-out methods. Here however, to the best of the author knowledge, none of these approaches seem to be directly applicable because of either the computational burden or the absence of closed-forms. Instead, the following simple heuristic using the Onsager-Machlup functional will be used.

- The noise variance parameter σ_η^2 is estimated using algorithm 3 with $N = 20$ iterations. This parameter has an important impact on the quality of the MAP estimator. Indeed, a low value results in important instabilities, since the Onsager-Machlup functional is focusing on least-square error. On the contrary, a significant value results in an over-smoothing of the parameter, leading to important differences with the dataset. This delicate problem is also known in deterministic methods, where the regularization parameter must be tuned [SKHK12].

Data: $y \in \mathcal{Y}$
Result: $\hat{\sigma_\eta}$
Initialize $\hat{\sigma}_\eta^1 = 1$;
for i *in* $[1, K]$ **do**
  Compute u_{MAP}^i using $\hat{\sigma}_\eta^i$;
  Set $\hat{\sigma}_\eta^{i+1} = \frac{1}{\sqrt{n-1}} \left\| y - \mathcal{G}(u_{MAP}^i) \right\|_{\mathbb{R}^n}$
end
Return $\left(\hat{\sigma}_\eta^{N+1}\right)^2$.

Algorithm 3: Calibration of noise variance σ_η^2.

- The variance σ has a similar influence on the MAP solution, since it tends to increase or decrease the regularization term influence. However, it is possible to give an interesting value with the following remark. Since the MAP estimator is maximizing the likelihood, it encloses information about the right level of variance. Indeed, consider the discretized MAP estimate:

$$\tilde{u}_{MAP} = (\lambda_{MAP}, D_{MAP}, \xi_1^{MAP}, ..., \xi_{N^2}^{MAP}).$$

  A Maximum Likelihood estimator can be used here, since the vector $(\xi_1^{MAP}, ..., \xi_{N^2}^{MAP})$ can be considered as N^2 realization of a $\mathcal{N}(0, \sigma^2)$ random variable. The parameter σ^2 can thus be tuned by

$$\hat{\sigma}^2 = \frac{1}{N^2 - 1} \sum_{i=1}^{N^2} (\xi_{MAP}^i)^2.$$

Both of these 2 methods are totally empirical but appear to give interesting values, at least in this application.

Robust MCMC algorithm In this application, it has been showed that there is a Gaussian reference measure μ_{ref} such that $\mu_0 << \mu_{ref}$. This will give the opportunity to use already existing robust MCMC algorithms. Indeed, the reference measure is Gaussian, allowing for a vast catalogue of proposal Markov kernels have been developed [CLM16, BS09, CRSW13, BGL$^+$17, Vol13, RS18]. In particular, these sampling methods are well-defined in infinite dimensional spaces, thus are robust to discretization size [CRSW13]. A simplified version of the infinite dimensional manifold Modified Adjusted Langevin Algorithm (inf-mMALA) from [BGL$^+$17] is used here. Indeed, the original algorithm requires an update of the local preconditioner (Gauss-Newton Hessian) at each iteration, which can be very computationally expensive. However, this quantity can be computed once at the MAP estimate (obtained previously from variational technique). This provides an interesting trade-off between computational time and sampling quality. Indeed, the proposal covariance is scaled in the vicinity of the MAP estimator once and for all, while the gradient is computed at each iteration (its cost is negligible in front of the Gauss-Newton Hessian matrix). The algorithm is initiated at the MAP location and ran for $n = 21\,000$ iterations.

Results The result of the Markov chain are displayed in figure 5.1, showing the evolution of $(\lambda, D, \xi_0, \xi_1, \xi_2)$ during the sampling ($n = 21\,000$ iterations with an overall acceptance probability of 60% for $h = 0.3$). From this figure, the decision has been made to discard the first 6 000 as burning sample. On the remaining samples, the autocorrelation gives a self-correlation as shown in figure 5.2. It appears that taking a sample out of two hundreds is reasonable, which leads to a total of 76 posterior samples. From this, the MAP estimate is taken as the minimizer of the

Onsager-Machlup functional among the posterior sample. Precise values of decay, diffusion, negative log-likelihood and Onsager-Machlup functional are given in table 5.2 for the MAP (obtained through variational deterministic optimization and its analogue from McMC sampling) as well as the conditional mean. Additionally to the estimated values, one can also look at the marginal distribution of (λ, D) on figure 5.3. Concerning the MAP estimator (figure 5.4), both pikes in solution z described in [Bec12] and [BBCCS$^+$13] are recovered. The first happens on the anterior part of the embryo in early experiment ($x = 35, t = 35$). The second is much more intense and happens in the posterior part during the second half of the experiment. The estimated source f explains these with an intense and localized increase in concentration. Finally, the quantification of uncertainty (through the point-wise variance) indicates that the dataset informs well the source close to the dataset, leaving a huge unknown before the first measurements (see figure 5.5).

5.4 Conclusion

In this final chapter, a practical inverse problem has been solved, using the developed Bayesian methodology from chapter 4. In particular, the delicate question of discretization has been solved using a Karhunen-Loève basis from chapter 2. The consistency has been shown theoretically and the rate of convergence estimated to at least the speed of prior approximation. This clearly encourages the search of optimal representation of prior measures. This is what has been done by choosing a tensor basis, constructed on Schauder hat functions. This result has an important practical consequence, since it largely extends the choice of available Gaussian priors. Indeed, the series representation with independent coefficients is useful, both in the variational search of MAP estimators and Monte-Carlo sampling. Moreover, a new method of hyper-parameters tuning has been proposed, addressing empirically a delicate question often discarded in the literature.

Parameters	Estimated values
σ_η^2	19.76
σ^2	0.52

Table 5.1: Estimated values of σ_η^2 and σ^2.

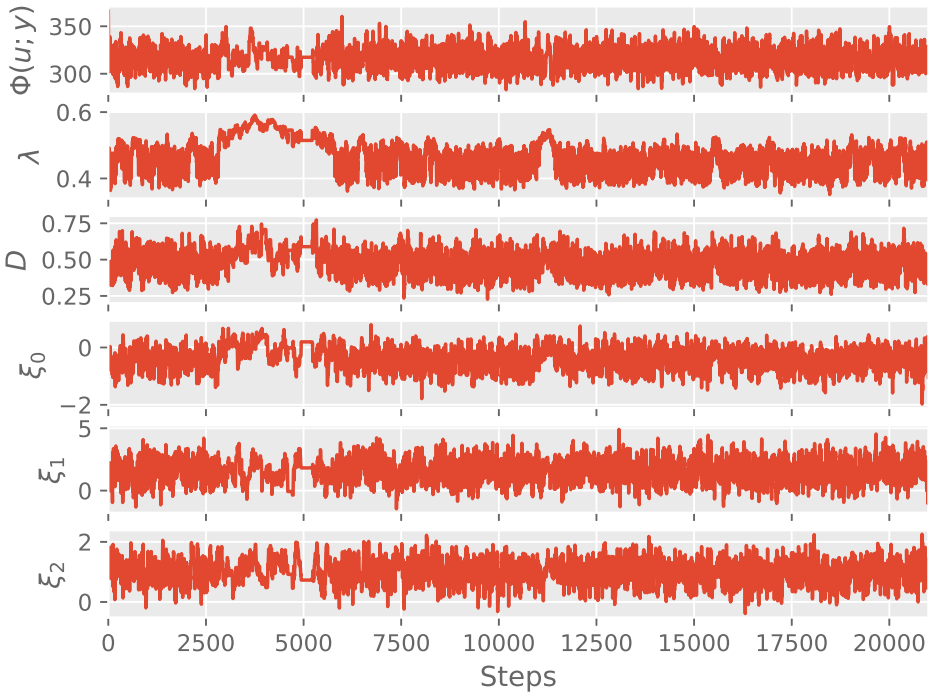

Figure 5.1: Trace plots of $\Phi(u;y)$, λ, D, ξ_0, ξ_1 and ξ_2 (the first 6 000 iterations are burned).

Estimators	λ	D	$\Phi(u;y)$	$I(u)$
MAP (from variational optimization)	0.43	0.34	366.82	566.32
MAP (from McMC sampling)	0.42	0.30	331.26	723.27
Conditional mean (from McMC sampling)	0.45	0.46	303.76	553.40

Table 5.2: Values of decay, diffusion, negative log-likelihood and Onsager-Machlup functional for different estimators.

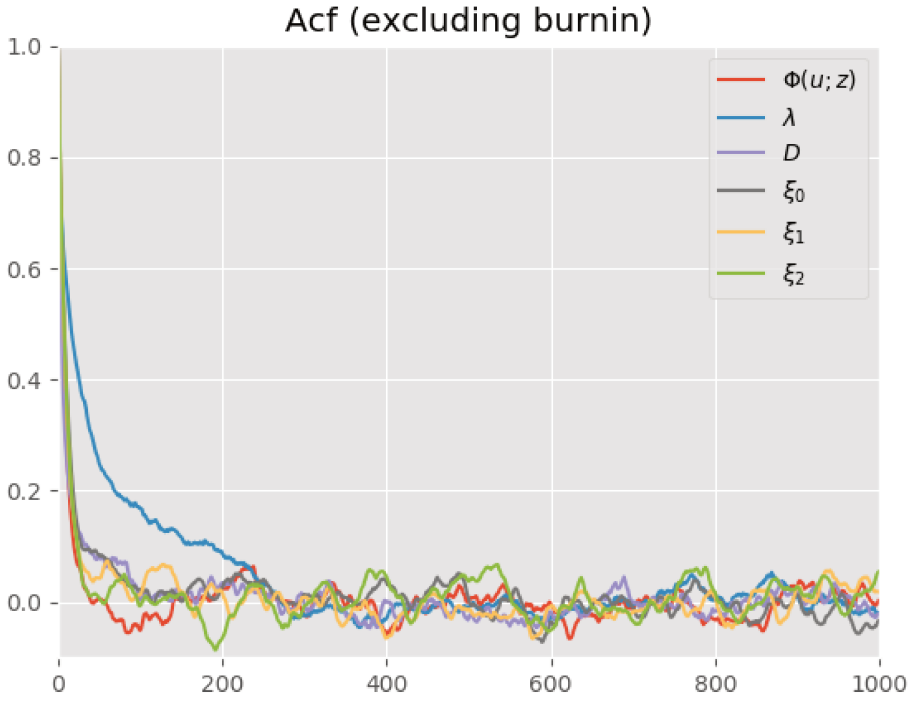

Figure 5.2: Autocorrelation of $\Phi(u;y)$, λ, D, ξ_0, ξ_1 and ξ_2, excluding burning sample.

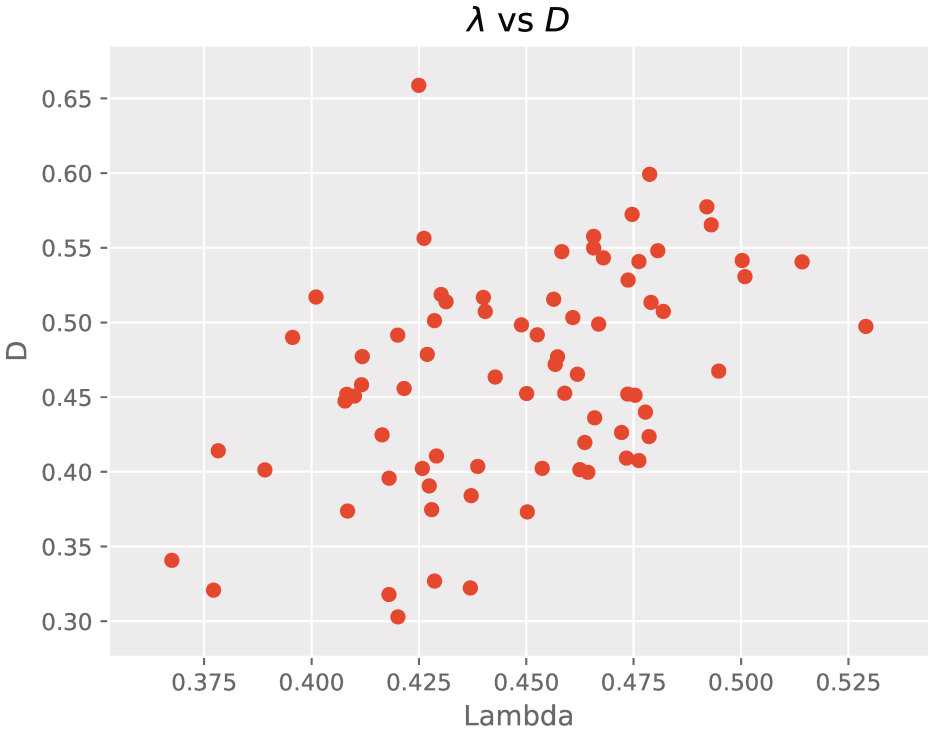

Figure 5.3: Marginal posterior distributions for $\lambda|y$ and $D|y$.

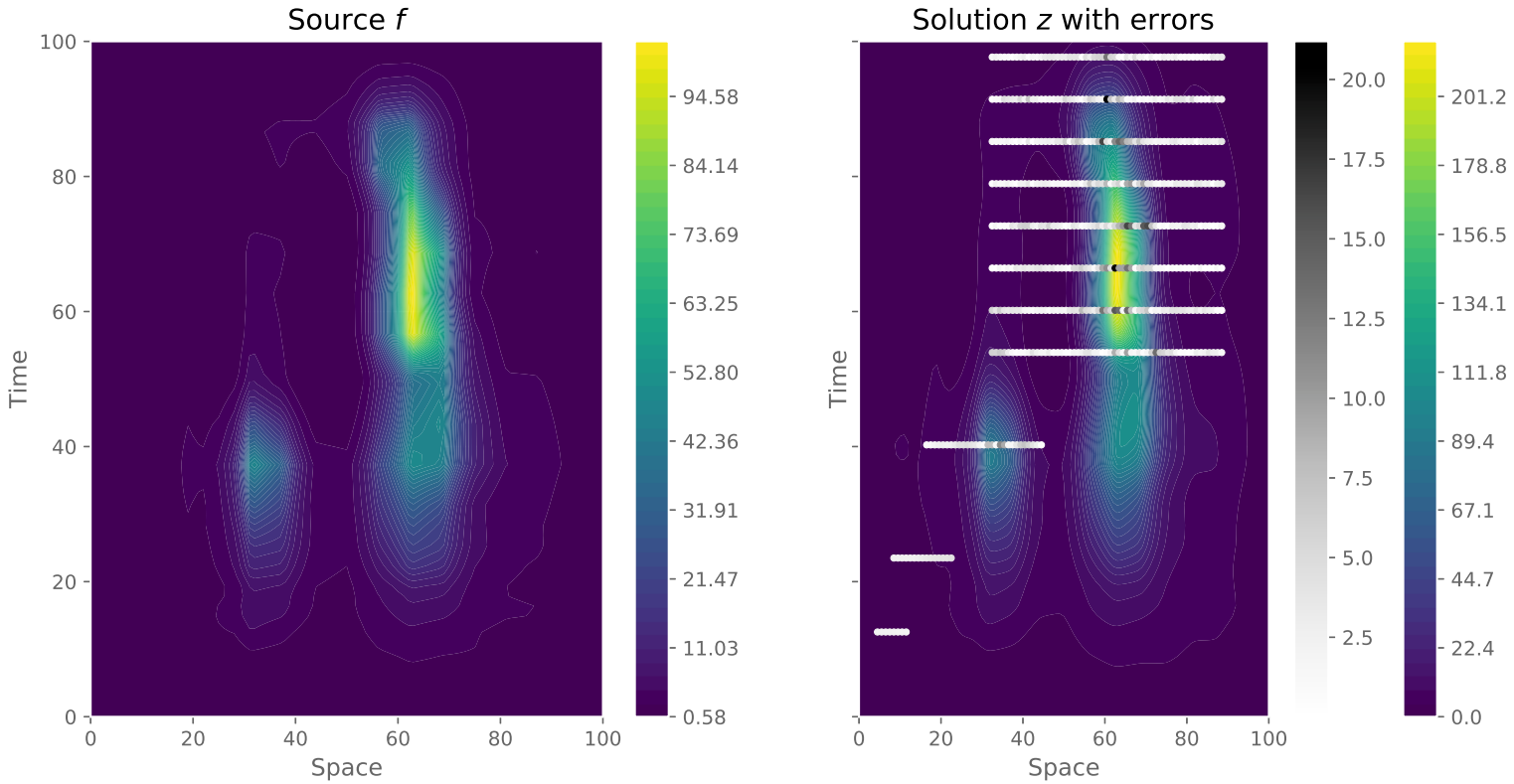

Figure 5.4: MAP estimator from McMC sampling. Left: Estimated source term f_{MAP}, right: estimated solution $z(u_{MAP})$ with absolute error at data locations.

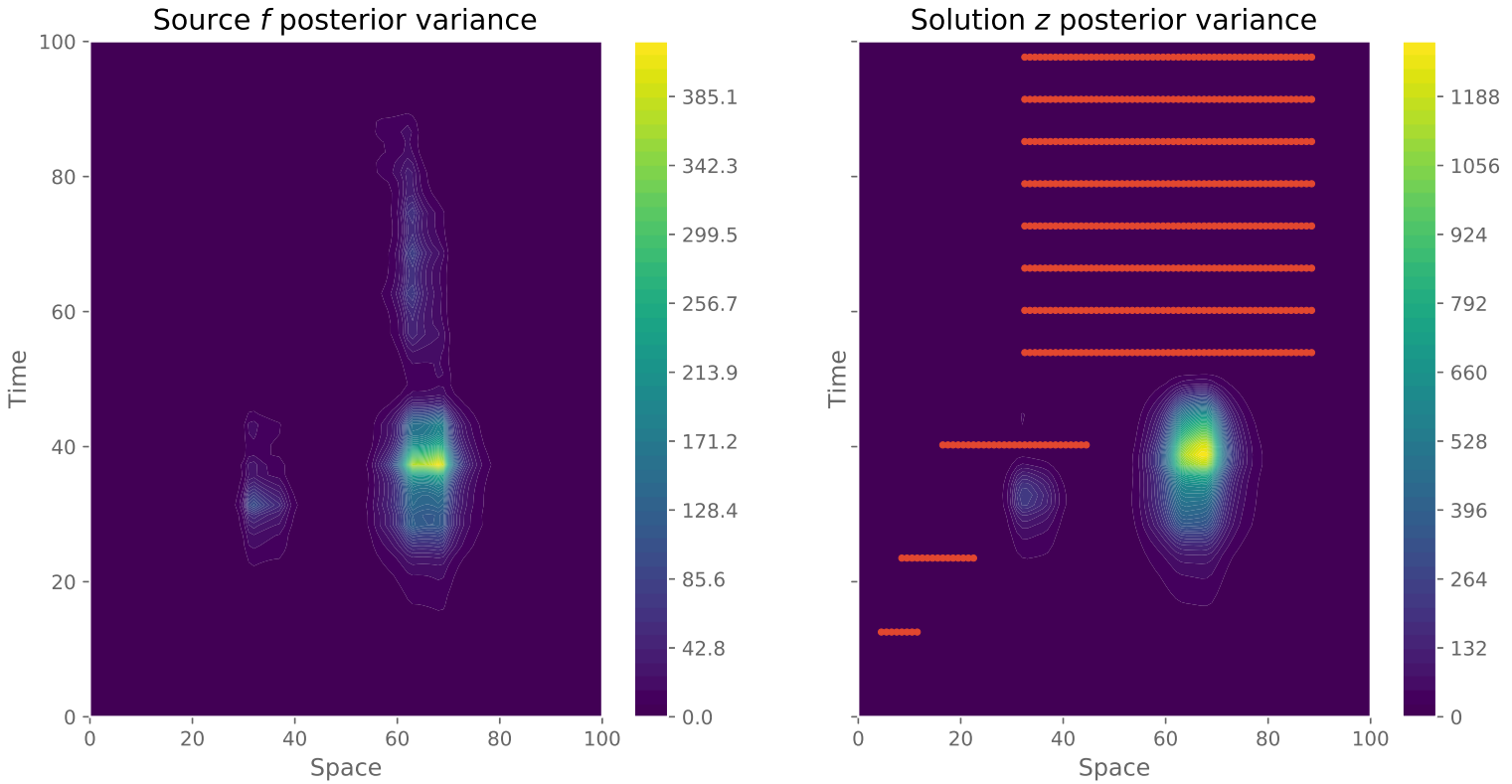

Figure 5.5: Posterior point-wise variance of source (left) and solution (right).

Conclusion and future works

The Bayesian methodology for inverse problems is a very active field of research, generalizing some already well-known and widely-used methods such as Kriging (or Gaussian process regression) to much larger settings. In particular, it is very attractive since it provides simultaneously regularization and quantification of uncertainty. However, it is still in its infancy and a large number of questions, from the measure theoretical foundations to practical implementation, remain open. Hopefully, a large community of researchers have already given satisfying answers, but some can be further studied.

In this thesis, the question that has been investigated is approximation, focusing on a new prior measure discretization method. This is a largely studied aspect with a first theoretical treatment in [CDRS09] and later in [CDS10, Stu10]. The most natural method, at least when the prior is a Gaussian measure, is to use a random series decomposition with independent coordinates. Indeed, it gives a useful representation of the parameters, particularly well-suited to McMC sampling or MAP computation. It is traditionally given by the Karhunen-Loève decomposition when the ambient space has a Hilbert geometry. However, it is limited in practice to a small number of prior distributions, since the useful functions are solutions to possibly difficult eigenproblems. Moreover, when the problem involves functions through (for instance) partial derivative equations, the most natural framework is not necessarily Hilbert.

The generalization of the Karhunen-Loève decomposition to Banach spaces, given in chapter 2 and accepted for publication in [BC17], gives an interesting answer to those limits. Indeed, since it is established in general Banach spaces, it potentially applies to very different settings. However, it also conveys its own difficulties. For instance, instead of looking for eigenvectors associated to a self-adjoint and trace-class operator as it is the case in Hilbert spaces, one must find linear functionals maximizing the projected variance. This is particularly difficult when the dual space has no particular representation. Nevertheless, the space of continuous functions over a compact metric set seems to be exempt and offers a very intuitive interpretation instead. In fact, the previous linear functionals can be limited to Dirac measures in this case. It results in a sequential search for maximum variance locations, a particularly well-

suited problem for numerical optimization (in low dimensional spaces). Moreover, it is possibly applicable to domains where the input set is not a hyper-rectangle, while also preserving the possibility to extend one dimensional decompositions by tensorization. In this way, the series representation of new continuous Gaussian random fields can be considered. Furthermore, this specific setup is particularly interesting, since one can even have both traditional (using a Lebesgue measure for instance) and new Karhunen-Loève decompositions. The main limit remains that optimality of truncated sums, in the Banach norm (the supremum for continuous functions over compact metric sets), has not been established yet and could be possibly wrong. This very interesting question has been answered positively for the standard Wiener process, but needs further research in more general cases.

The second contribution of this thesis is of a practical nature. Indeed, the theory of Bayesian inverse problem may be extensively detailed in the literature, its practical implementation still requires a different type of expertise. The most obvious hole in the methodology would be hyper-parameters tuning. This is a deep and intrinsic problem in Bayesian methodologies. However, when it comes to inverse problems with possibly complex dynamical systems, there are no general method such as leave-one-out or Maximum Likelihood (they are possibly intractable). This is far from being a side dish, since both regularization and quantification of uncertainty are heavily impacted. Concerning posterior modes, it directly influences the trade-off between proximity to data and regularity. In the case of McMC algorithms, it gives the amplitude of proposal jumps, which is critical for a correct exploration of the space. In chapter 5, an attempt for the calibration of the measurement noise and the overall prior variance is given. It is heuristic, but further research could be made to give a theoretical background in this direction. A good start would certainly be an application of these principle in a Kriging context, where one can also use Maximum-Likelihood estimators or Leave-One-Out methods.

Glossary of mathematical notations

$H_0^1(\Omega)$	Sobolev vector space of weakly differentiable functions on the set Ω with null boundary values.
$L^0(\mathcal{U})$	$L^0(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{U}, Bor(\mathcal{U}))$ with $\mathcal{U}$ a Banach space.
$L^k(\Omega, \mathcal{F}, \mu; A, \mathcal{A})$	Vector space of k times integrable equivalence classes of applications from $(\Omega, \mathcal{F})$ to $(A, \mathcal{A})$ w.r.t. the measure μ.
Φ	Negative log-likelihood function.
$\mathcal{B}_{\mathcal{U}}(u, r]$	Closed ball with center u and radius r in the normed space $\mathcal{U}$.
$\mathcal{B}_{\mathcal{U}}(u, r)$	Open ball with center u and radius r in the normed space $\mathcal{U}$.
$\mathcal{B}_{\mathcal{U}}$	Unit ball in the normed space $\mathcal{U}$.
$\mathcal{L}(\mathcal{U}, \mathcal{Y})$	Space of bounded linear operators from $\mathcal{U}$ to $\mathcal{Y}$.
$Bor(\mathcal{U})$	Borel σ-algebra in the Banach space $\mathcal{U}$.
$\widehat{\mathcal{U}}^{\|.\|_Y}$	Topological completion of the set $\mathcal{U}$ w.r.t. to the norm $\|.\|_{\mathcal{Y}}$.
$C(A, B)$	Vector space of continuous functions from A to B.
$\mathcal{C}_X$	Covariance operator of the random element X.
$Cyl(\mathcal{U})$	Cylindrical σ-algebra in the Banach space $\mathcal{U}$.
δ_s	Dirac delta measure at element s.
$D(\mathcal{A})$	Domain of the operator $\mathcal{A}$.
$\langle u, l \rangle_{\mathcal{U}, \mathcal{U}^*}$	Dual pairing between $u \in \mathcal{U}$ and $l \in \mathcal{U}^*$.
$\mathcal{U}^*$	Topological dual space of $\mathcal{U}$.
$\widehat{X}$	Fourier transform of the random element $X \in L^0(\mathcal{U})$.

$\mathcal{H}_X$	Gaussian space of random element $X \in L^0(\mathcal{U})$.
$\mathcal{I}$	Identity operator.
$\langle u, v\rangle_{\mathcal{U}}$	Inner product of $(u, v) \in \mathcal{U}$ in the Hilbert space $\mathcal{U}$.
$\mathcal{K}(\mathcal{U}, \mathcal{Y})$	Space of compact linear operators from $\mathcal{U}$ to $\mathcal{Y}$.
$\mathcal{A}^*$	Adjoint operator of $\mathcal{A}$.
$\mathcal{U}$	A (real) Banach space with norm $\|.\|_{\mathcal{U}}$.
$\mathcal{U}_X$	Cameron-Martin space of random element $X \in L^0(\mathcal{U})$.
μ^y	Posterior probability measure associated to data y.
μ_0	Prior probability measure.
$\mu_1 << \mu_2$	absolute continuity of the measure μ_1 w.r.t. the measure μ_2.
μ_X	Distribution of random element $X \in L^0(\mathcal{U})$.
$\|u\|_{\mathcal{U}}$	Norm of vector u in the Banach space $\mathcal{U}$.
$\mathcal{M}(\mathcal{F})$	Vector space of measures on the σ-algebra $\mathcal{F}$.
$R(\mathcal{A})$	Range of the operator $\mathcal{A}$.
$\mathrm{rank}(\mathcal{A})$	Rank of operator $\mathcal{A}$.
$\sigma(\mathcal{U}^*, \mathcal{U})$	Weak-star topology on $\mathcal{U}^*$.
$\sigma(\mathcal{U}, \mathcal{U}^*)$	Weak topology on the Banach space $\mathcal{U}$.
φ_X	Covariance bilinear form of the random element X.
$f \approx g$	equivalence between functions f and g in the sense that $\exists c, C > 0,\ \forall n \in \mathbb{N},\ cg(n) \leq f(n) \leq Cg(n)$.
CM	Conditional mean.
MAP	Maximum a Posteriori.

Bibliography

[ABDH17] Sergios Agapiou, Martin Burger, Masoumeh Dashti, and Tapio Helin. Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems. pages 1–36, 2017.

[ABH+15] Martin S. Alnaes, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth N. Wells. The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3(100):9–23, 2015.

[Aro50] Nicolas Aronszajn. Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[AT03] Antoine Ayache and Murad S. Taqqu. Rate Optimality of Wavelet Series Approximations of Fractional Brownian Motion. *Journal of Fourier Analysis and Applications*, 9(5):451–471, 2003.

[AT07] Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer, 2007.

[BB18] Martin Benning and Martin Burger. Modern Regularization Methods for Inverse Problems. (1):1–97, 2018.

[BBCCS+13] Kolja Becker, Eva Balsa-Canto, Damjan Cicin-Sain, Astrid Hoermann, Hilde Janssens, Julio R. Banga, and Johannes Jaeger. Reverse-Engineering Post-Transcriptional Regulation of Gap Genes in Drosophilia melanogaster. *PLOS Computational Biology*, 9(10), 2013.

[BBG16] Julien Bect, François Bachoc, and David Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. pages 1–29, 2016.

[BC17] Xavier Bay and Jean-Charles Croix. Karhunen-Loève decomposition of Gaussian measures on Banach spaces. pages 1–14, 2017.

[BCM18] Markus Bachmayr, Albert Cohen, and Giovanni Migliorati. Representations of Gaussian Random Fields and Approximation of Elliptic

PDEs with Lognormal Coefficients. *Journal of Fourier Analysis and Applications*, 24(3):621–649, 2018.

[Bec12] Kolja Becker. *A Quantitative Study of Translational Regulation in Drosophilia Segment Determination*. PhD thesis, University of Mainz, 2012.

[BGL+17] Alexandros Beskos, Mark A Girolami, Shiwei Lan, Patrick E. Farrell, and Andrew M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351, 2017.

[BLNZ95] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[Bog98] Vladimir Igorevich Bogachev. *Gaussian Measures*, volume 62. 1998.

[Bre11] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, volume 2010. Springer, 2011.

[BS09] Alexandros Beskos and Andrew M. Stuart. MCMC methods for sampling function space. page 337, 2009.

[BTA04] Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, 2004.

[CDA18] Jean-Charles Croix, Nicolas Durrande, and Mauricio Alvarez. Bayesian inversion of a diffusion evolution equation with application to Biology. pages 1–23, 2018.

[CDPS18] Victor Chen, Matthew M. Dunlop, Omiros Papaspiliopoulos, and Andrew M. Stuart. Robust MCMC Sampling with Non-Gaussian and Hierarchical Priors in High Dimensions. 2018.

[CDRS09] S. L. Cotter, Masoumeh Dashti, J C Robinson, and Andrew M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25(11):115008, 2009.

[CDS10] S. L. Cotter, Masoumeh Dashti, and Andrew M. Stuart. Approximation of Bayesian inverse probelms. *SIAM Journal on Numerical Analysis*, 48(1):322 – 345, 2010.

[CLM16] Tiangang Cui, Kody J. H. Law, and Youssef M. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.

[CMM+14] Tiangang Cui, James Martin, Youssef M. Marzouk, Antti Solonen, and Alessio Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.

[CMW14] Tiangang Cui, Youssef M. Marzouk, and Karen E. Willcox. Data-Driven Model Reduction for the Bayesian Solution of Inverse Problems. mar 2014.

[CMW16] Tiangang Cui, Youssef M. Marzouk, and Karen E. Willcox. Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction. *Journal of Computational Physics*, 315:363–387, jun 2016.

[CRSW13] S. L. Cotter, Gareth O. Roberts, Andrew M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, aug 2013.

[CS07] Daniela Calvetti and Erkki Somersalo. *Introduction to Bayesian Scientific Computing*. 2007.

[CS15] Peng Chen and Christoph Schwab. Sparse-grid , reduced-basis Bayesian inversion. *Comput. Methods Appl. Mech. Engrg.*, 297:84–115, 2015.

[DHS12] Masoumeh Dashti, Stephen Harris, and Andrew M. Stuart. Besov priors for Bayesian inverse problems. *Inverse Problems and Imaging*, 6(2):183–200, 2012.

[DLSV13] Masoumeh Dashti, Kody J. H. Law, Andrew M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[DM05] Lokenath Debnath and Piotr Mikusinski. *Introduction to Hilbert spaces and applications*. Elsevier a edition, 2005.

[DS15] Masoumeh Dashti and Andrew M. Stuart. The Bayesian Approach to Inverse Problems, 2015.

[DvZ05] Kacha Dzhaparidze and Harry van Zanten. Optimality of an explicit series expansion of the fractional Brownian sheet. *Statistics and Probability Letters*, 71(4):295–301, 2005.

[DZ04a] Guiseppe Da Prato and Jerzy Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces*. 2004.

[DZ04b] Kacha Dzhaparidze and H. Zanten. A series expansion of fractional Brownian motion. *Probability theory and related fields*, 130(1):39–55, 2004.

[DZ14] Guiseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Number 1. Cambrige University Press, 2014.

[EHN96] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. Regularization of inverse problems, 1996.

[Enf73] Per Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Mathematica*, 130(1):309–317, 1973.

[Eva10] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, 2010.

[GHO17] Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of Uncertainty Quantification*. 2017.

[Git12] Claude Jeffrey Gittelson. Representation of Gaussian fields in series with independent coefficients. *IMA Journal of Numerical Analysis*, 32(1):294–319, 2012.

[Had02] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.

[Hai09] Martin Hairer. An Introduction to Stochastic PDEs. 2009.

[Has70] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970.

[HB15] Tapio Helin and Martin Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8), 2015.

[Hei08] Matthias Heinkenschloss. Numerical solution to implicitely constrained optimization problems. Technical report, 2008.

[Her81] Wojciech Herer. Stochastic basis in Fréchet spaces. *Demonstratio Mathematica*, 14(3):719–724, 1981.

[HN17] Bamdad Hosseini and Nilima Nigam. Well-Posed Bayesian Inverse Problems: Priors with Exponential Tails. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):436–465, 2017.

[Hos17a] Bamdad Hosseini. *Finding Beauty in the Dissonance : Analysis and Applications of Bayesian Inverse Problems*. Phd, Simon Fraser University, 2017.

[Hos17b] Bamdad Hosseini. Well-posed Bayesian Inverse Problems with Infinitely-Divisible and Heavy-Tailed Prior Measures. *SIAM Journal on Uncertainty Quantification*, 5:1024–1060, 2017.

[HPUU09] Michael Hinze, René Pinnau, Michael Ulbrich, and Stephan Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.

[HSV14] Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Annals of Applied Probability*, 24(6):2455–2490, 2014.

[Igl05] E. Igloi. A rate-optimal trigonometric series expansion of the fractional brownian motion. *Electronic Journal of Probability*, 10(41):1381–1397, 2005.

[IN68] Kiyosi Itô and Makiko Nisio. On the convergence of sums of independent banach space valued random variables. *Osaka Journal of Mathematics*, 5(1):35–48, 1968.

[Isa17] Victor Isakov. *Inverse Problems for Partial Differential Equations*, volume 127 of *Applied Mathematical Sciences*. Springer International Publishing, Cham, 2017.

[JK69] Naresh C. Jain and G Kallianpur. Norm convergent expansions for Gaussian processes in Banach Spaces. pages 890–895, 1969.

[KL02] Thomas S. Kuhn and Wener Linde. Optimal series representation of fractional Brownian sheets. *Bernoulli*, 8(5):669–696, 2002.

[Kri51] D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.

[KS05] Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[KT14] V. V. Kvaratskhelia and Vazha Tarieladze. Diagonally Canonical and Related Gaussian Random Elements. *Theory of Probability & Its Applications*, 58(2):286–296, 2014.

[Kuo75] Hui-Hsiung Kuo. *Gaussian Measures in Banach Spaces*. Springer, 1975.

[Law14] Kody J. H. Law. Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics*, 262:127–138, 2014.

[LGW18] Harriet Li, Vikram V. Garg, and Karen E. Willcox. Model adaptivity for goal-oriented inference using adjoints. *Computer Methods in Applied Mechanics and Engineering*, 331:1–22, 2018.

[Lif12] Mikhail Lifshits. *Lectures on Gaussian Processes*. Springer, 2012.

[LL17] Hans Petter Langtangen and Anders Logg. *Solving PDEs in Python - The FEniCS Tutorial Volume I.* 2017.

[LLW99] Wenbo V. Li, Werner Linde, and Linde Werner. Approximation, Metric Entropy and Small ball estimates for Gaussian measures. *Annals of Probability*, 27(3):1556–1578, 1999.

[LP09] Harald Luschgy and Gilles Pagès. Expansions For Gaussian Processes And Parseval Frames. *Electronic Journal of Probability*, 14:1198–1221, 2009.

[LST18] H. C. Lie, Tim J Sullivan, and Aretha L. Teckentrup. Random forward models and log-likelihoods in Bayesian inverse problems. pages 1–21, 2018.

[Mat62] Georges Matheron. *Traité de géostatistique appliquée.* Technip edition, 1962.

[Mél10] Sylvie Méléard. *Aléatoire Introduction à la théorie et au calcul des probabilités.* Les éditions de l'Ecole Polytechnique, 2010.

[Mer09] James Mercer. Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philophical Transactions of the Royal Society of London. Series A, Containing Papers of Mathematical or Physical Character*, 209:415–446, 1909.

[Nas81] M. Nashed. Operator-theoretic and computational approaches to Ill-posed problems with applications to antenna theory. *IEEE Transactions on Antennas and Propagation*, 29(2):220–231, 1981.

[Nda16] Mohamed Ndaoud. A New Rate-Optimal Series Expansion of Fractional Brownian Motion. 2016.

[NS16] Joseph Benjamin Nagel and Bruno Sudret. Spectral likelihood expansions for Bayesian inference. *Journal of Computational Physics*, 309:267–294, 2016.

[Oka86] Yoshiaki Okazaki. Stochastic Basis in Fréchet Space. 383:379–383, 1986.

[Phi62] David L. Philips. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery*, 9:84–97, 1962.

[Pis89] Gilles Pisier. *The volume of convex bodies and Banach space geometry.* 1989.

[QL04] Hervé Queffélec and Daniel Li. Introduction à l'étude des espaces de Banach - Analyse et probabilités, 2004.

[RR04] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1(0):20–71, 2004.

[RS18] Daniel Rudolf and Björn Sprungk. On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm. *Foundations of Computational Mathematics*, 18(2):309–343, 2018.

[Rud09] Walter Rudin. *Analyse réelle et complexe*. Dunod, 2009.

[RW04] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.*, volume 14. 2004.

[Sch13] Claudia Schillings. A note on Sparse, Adaptive Smolyak Quadratures for Bayesian Inverse Problems. Technical report, 2013.

[SCW+17] Alessio Spantini, Tiangang Cui, Karen E. Willcox, Luis Tenorio, and Youssef M. Marzouk. Goal-oriented optimal approximations of Bayesian Linear Inverse Problems. 2017.

[SKHK12] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S. Kazimierski. *Regularization Methods in Banach Spaces*. De Gruyter, 2012.

[SS12] Christoph Schwab and Andrew M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003, apr 2012.

[SS14] Claudia Schillings and Christoph Schwab. Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems*, 30(6), 2014.

[SS16] Claudia Schillings and Christoph Schwab. Scaling limits in computational Bayesian inversion. *ESAIM: Mathematical Modelling and Numerical Analysis*, 50(6):1825–1856, 2016.

[SSC+15] Alessio Spantini, Antti Solonen, Tiangang Cui, James Martin, Luis Tenorio, and Youssef M. Marzouk. Optimal low-rank approximations of Bayesian linear inverse problems. *arXiv.org*, page 28, 2015.

[ST16] Andrew M. Stuart and Aretha L. Teckentrup. Posterior Consistency for Gaussian Process Approximations of Bayesian Posterior Distributions. *arXiv*, pages 1–33, 2016.

[Str10] Daniel W Stroock. *Mathematics of Probability*, volume 149. 2010.

[Stu10] Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, may 2010.

[Sul17] Tim J Sullivan. Well-posedness of Bayesian inverse problems in quasi-Banach spaces with stable priors. pages 1–6, 2017.

[Tar05] Albert Tarantola. *Inverse Problem Theory and methods for model parameter estimation*. Siam edition, 2005.

[Tie98] Luke Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9, 1998.

[Tik63] A. N. Tikhonov. Solution of Incorrectly Formulated Problems and the Regularization Method. *Soviet Mathematics Doklady*, 4(4):1035–1038, 1963.

[Tsi81] B. S. Tsirel'son. A natural modification of a random process and its application to stochastic functional series and Gaussian measures. *Journal of Soviet Mathematics*, 16(2):940–956, 1981.

[Vak91] Nicholas Vakhania. Canonical Factorization of Gaussian covariance operators and some of its applications. *Theory of Probability & Its Applications*, 38(3):498–505, 1991.

[vN07] Jan van Neerven. Stochastic Evolution Equations, 2007.

[Vol13] Sebastian J. Vollmer. Dimension-Independent MCMC Sampling for Inverse Problems with Non-Gaussian Priors. feb 2013.

[VTC87] Nicholas Vakhania, Vazha Tarieladze, and S Chobanyan. *Probability Distributions on Banach Spaces*. Springer Netherlands, mathematic edition, 1987.

[VV08] A. W. Van der Vaart and J. H. Van Zanten. Reproducing kernel Hilbert spaces of Gaussian priors. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, 3:200–222, may 2008.

[Wan08] Limin Wang. *Karhunen-Loeve Expansions and their Applications*. PhD thesis, The London School of Economics and Political Sciences, 2008.

[Zha17] Pengchuan Zhang. *Compressing Positive Semidefinite Operators with Sparse / Localized Bases*. Phd thesis, California Institute of Technology, 2017.

Résumé

L'inférence est une activité fondamentale en sciences et en ingénierie: elle permet de confronter et d'ajuster des modèles théoriques aux données issues de l'expérience. Ces mesures étant finies par nature et les paramètres des modèles souvent fonctionnels, il est nécessaire de compenser cette perte d'information par l'ajout de contraintes externes au problème, via les méthodes de régularisation. La solution ainsi associée satisfait alors un compromis entre d'une part sa proximité aux données, et d'autre part une forme de régularité.

Depuis une quinzaine d'années, ces méthodes intègrent un formalisme probabiliste, ce qui permet la prise en compte d'incertitudes. La régularisation consiste alors à choisir une mesure de probabilité sur les paramètres du modèle, expliciter le lien entre données et paramètres et déduire une mise-à-jour de la mesure initiale. Cette probabilité a posteriori, permet alors de déterminer un ensemble de paramètres compatibles avec les données tout en précisant leurs vraisemblances respectives, même en dimension infinie.

Dans le cadre de cette thèse, la question de l'approximation de tels problèmes est abordée. En effet, l'utilisation de lois infini-dimensionnelles, bien que théoriquement attrayante, nécessite souvent une discrétisation pour l'extraction d'information (calcul d'estimateurs, échantillonnage). Lorsque la mesure a priori est Gaussienne, la décomposition de Karhunen-Loève est une réponse à cette question. Le résultat principal de cette thèse est sa généralisation aux espaces de Banach, beaucoup plus naturels et moins restrictifs que les espaces de Hilbert. Les autres travaux développés concernent son utilisation dans des applications avec données réelles.