Some Contributions to Reconstruction Problems

Cédric Herzet

Mémoire d'HdR

Some Contributions to Reconstruction Problems

Cédric Herzet
INRIA, Rennes

Cette thèse d'HdR a été soutenue le 24 novembre 2022 à l'Université de Rennes 1 en présence du jury suivant:

- Valérie Monbet, professeur, université de Rennes 1 (présidente)
- Laure Blanc-Féraud, directrice de recherche, CNRS (rapporteur)
- Gabriel Peyré, directeur de recherche, CNRS (rapporteur)
- Laurent Jacques, professeur, UCLouvain, Belgique (rapporteur)
- Frédéric Champagnat, chercheur, ONERA (examinateur)
- Thierry Chonavel, professeur, IMT Atlantique (examinateur)

Préface

Ce document contient le résumé d'une partie de mes travaux de recherche depuis la fin de ma thèse de doctorat. Il constitue le document principal sur lequel un jury d'experts s'appuiera afin de décider si j'ai acquis un niveau d'expertise suffisant pour obtenir une "Habilitation à diriger des recherches". Ce manuscrit est donc construit de manière à donner, le plus succinctement possible, un aperçu des thématiques de recherche sur lesquelles je me suis penché ces dernières années.

Le document s'articule autour de trois chapitres principaux. Le premier chapitre présente les différentes questions scientifiques auxquelles je me suis intéressé. Chaque section est dédiée à une thématique particulière, en détaille la problématique générale et décrit certaines de mes contributions dans ce domaine. Une liste de publications illustratives de mes travaux conclut chaque partie. Les références sont assorties d'un hyperlien vers les documents correspondants afin de permettre aux experts d'accéder facilement aux détails techniques de mes contributions ainsi qu'à une revue plus détaillée de la littérature. Le deuxième chapitre contient une description de mes perspectives de recherche à court et moyen termes. Dans ce chapitre, chaque section (au nombre de trois) décrit un axe de recherche particulier. Finalement, le dernier chapitre est consacré à l'exposition de mon curriculum vitae dans lequel les experts pourront trouver des éléments plus factuels sur ma carrière.

Durant la rédaction de ce manuscrit, la question de l'utilisation du *"je"* (certains passages ne se référant qu'à moi ou à mon interprétation personnelle des choses) ou du *"nous"* (la plupart de mes travaux ayant été effectués en collaboration avec d'autres personnnes) s'est posée à moi. J'ai opté pour une utilisation mixte de ces deux pronoms, l'emploi du *"nous"* devant être comprise comme *"moi"* plus une des nombreuses personnes avec qui j'ai eu la chance de travailler durant ma carrière.[1]

Comme vous le constaterez lorsque vous tournerez cette page, ce manuscrit est écrit en langue anglaise. Ce choix découle de l'observation suivante: la majorité des textes scientifiques étant de nos jours rédigés en anglais, une description précise de mes travaux dans

[1] Je réfrène ici mon envie de les citer par peur d'en oublier quelques uns. Ils se reconnaîtront néanmoins aisément à la lecture des parties qui les concernent.

l'état de l'art actuel se fait plus naturellement dans cette langue. J'espère néanmoins que les amoureux de la langue de Molière, dont je fais partie, sauront excuser cette décision purement pragmatique.

Foreword

This document contains a summary of a part of my research work. It constitutes the material on which a panel of experts will leverage to decide if I have attained a sufficient level of expertise to obtain an "Habilitation à diriger des recherches" (HdR), a French diploma that allows to officiate as a PhD advisor. This manuscript has therefore been organized to give, as succinctly as possible, an overview of my work during the last few years.

The document is structured around three main chapters. The first chapter presents some problems I have addressed. In this chapter, each section is dedicated to one particular theme. For each theme, the central question of interest and some of my contributions to the field are described. At the end of each section, I provide a list of contributions illustrative of my work in the domain with hyperlinks to the corresponding documents. The second chapter contains a description of my short- and medium-term research perspectives. Finally, the last chapter is dedicated to the exposition of my curriculum vitae in which experts will find more factual elements about my career.

During the writing of this manuscript, the question of the use of either *"I"* (some parts of the document only referring to me or my personal interpretations) or *"we"* (most of my work having been done in collaboration with colleagues) occurred to me. I have adopted a mixed strategy and decided to use both pronouns; the use of *"we"* has to be understood as "myself plus one of the many people I have had the opportunity to work with during my career".[2]

[2] I refrain from mentioning them here for fear of forgetting some of them. They will nevertheless easily recognize themselves when reading some parts of this manuscript.

Contents

Notational conventions	
a	scalar
$\mathbf{a}$	vector
$\mathbf{A}$	matrix
$\mathcal{A}$	set
$\mathscr{A}$	operator
A	subspace or scalar
$.^{\mathrm{T}}$	transpose
$.^{\dagger}$	pseudo inverse
$\langle \cdot, \cdot \rangle_{\mathcal{X}}$	inner product defined on $\mathcal{X}$
$\| \cdot \|_{\mathcal{X}}$	norm defined on $\mathcal{X}$
$\mathrm{dist}(\cdot, \cdot)$	distance defined on $\mathcal{X}$
$\mathbf{0}$	zero vector (of proper dimension)
$\mathbf{0}_n$	zero vector of dimension n
$\mathbf{1}_n$	one vector of dimension n
$\mathbf{I}_n$	identity matrix of dimension n
$\mathbf{a}(i)$	ith element of $\mathbf{a}$
$\mathbf{a}_i$	ith column of $\mathbf{A}$
$\mathbf{a}_{\mathcal{S}}$	restriction of $\mathbf{a}$ to its elements indexed by $\mathcal{S}$
$\mathbf{A}_{\mathcal{S}}$	restriction of $\mathbf{A}$ to its columns indexed by $\mathcal{S}$
$\bar{\mathcal{A}}$	complement set of $\mathcal{A}$
$\mathrm{card}(\mathcal{A})$	cardinality of $\mathcal{A}$

Notations for reconstruction problems	
$\mathcal{X}$	target space
$\mathcal{X}_{\mathrm{target}}$	target set
$\mathbf{x}$	target vector
$\mathcal{Y}$	observation space
$\mathbf{y}$	observed vector
$\boldsymbol{\epsilon}$	additive noise
$\mathscr{M}$	observation operator
$\mathscr{D}$	decoder
$\mathcal{D}$	set of decoders

Notations for communication problems	
$\mathbf{b}$	information sequence
k	length of the information sequence
$\mathcal{C}$	codebook
$\mathscr{C}$	coding operator
m	length of the coded sequence
$\mathbf{x}$	vector of synchronization parameters
n	length of $\mathbf{x}$

Notations for sparse representations	
$\mathbf{x}$	sparse vector
n	length of $\mathbf{x}$
$\|\mathbf{x}\|_0$	number of non-zero elements in $\mathbf{x}$
k	$\max_{\mathbf{x}\in\mathcal{X}_{\text{target}}} \|\mathbf{x}\|_0$
m	length of the observed vector
$\mathbf{A}$	dictionary
$\mathbf{a}_i$	ith atom of $\mathbf{A}$

Notations for Bayesian methods	
$\mathbf{s}$	support vector
$\mathbf{w}$	amplitude vector
γ^2	variance of the elements of $\mathbf{w}$
σ^2	variance of the elements of $\boldsymbol{\epsilon}$
$\mathbf{b}$	Boltzmann-machine parameter
$\mathbf{C}$	Boltzmann-machine parameter
$\propto$	equality up to a normalization factor

Notations for safe screening	
$\mathbf{z}$	primal variable
$\mathbf{u}$	dual variable
$\mathbf{x}^\star$	minimizer of the primal problem
$\mathbf{u}^\star$	maximizer of the dual problem
$\mathcal{R}$	safe region ($\mathbf{u}^\star \in \mathcal{R}$)
$\mathcal{A}$	subset of $\mathbb{R}^m$
$\mathcal{B}(\mathbf{c}, r)$	safe ball with center $\mathbf{c}$ and radius r

Notations for theoretical guarantees	
$\operatorname{supp}(\mathbf{x})$	set of indices of the nonzero elements of $\mathbf{x}$
$\operatorname{kruskal}(\mathbf{A})$	Kruskal rank of $\mathbf{A}$
$\mu(\mathbf{A})$	mutual coherence of $\mathbf{A}$, see (1.31)
$\mathcal{H}$	Hilbert space
Θ	set of parameters defining a "continuous" dictionary
$\mathbf{a}(\theta)$	$\mathbf{a}\colon \mathbb{R}^d \to \mathcal{H}$ evaluated at θ
$\mathcal{A}$	$\{\mathbf{a}(\theta)\colon \theta \in \Theta\}$
$\mathcal{S}$	support of a sparse vector
$\mathcal{Q}$	side information on the support $\mathcal{S}$

Notations for reduced-order models	
$\boldsymbol{\theta}$	parameter
Θ	set of parameters
ε	accuracy parameter, see (1.37)
$\mathrm{PDE}(\cdot, \boldsymbol{\theta})$	differential operator parameterized by θ
$\mathrm{PDE}_{\mathrm{relax}}(\cdot, \boldsymbol{\theta})$	surrogate differential operator
$\mathcal{X}_{\mathrm{relax}}$	surrogate target set
V_n	n-dimensional subspace of $\mathcal{X}$
$\hat{\epsilon}_n$	parameter of $\mathcal{X}_{\mathrm{relax}}$, see (1.40) and (1.59)
$\mathscr{D}_{\mathrm{wc}}$	worst-case optimal decoder (1.45)
$\varepsilon_{\mathrm{wc}}$	worst-case optimal reconstruction error (1.46)
σ_n	inf-sup constant (1.49)
κ_n	Kolmogorov n-width of a set, see (1.54)

Chapter 1

Summary of my work

During my thesis, my postdocs and, later on, during my professional life as a researcher at Inria, I have been interested in different facets of the following general problem:

"Can we recover the value of some quantity of interest from observations?"

This seemingly very simple question underpins many fundamental problems that the scientific community has been working on for centuries. The goal of this document is to provide an overview of some of my contributions to this general question.

Before going any further, as in any sound scientific process, it is first necessary to clarify the terms appearing in the question at hand. In particular, the terms "*quantity of interest*", "*observations*" or "*recover*" need to be defined properly to provide meaningful answers to the above problem. To this end, I identify below four fundamental ingredients of reconstruction problems. These ingredients allow to formally characterize most of the reconstruction problems encountered in the literature, and will be used throughout this document to describe precisely the nature of my contributions.

1. **The target set**: the term "quantity of interest" generally refers to some vector, matrix or function which one would like to have access to. We will denote this quantity as $\mathbf{x}$ hereafter. In practice, one is rarely interested in the recovery of one single element $\mathbf{x}$ but rather targets the reconstruction of any element of some set $\mathcal{X}_{\text{target}}$. We will refer to this set as "*target set*" hereafter. Depending on the application, the target set may correspond to images, audio signals, responses of a physical system, etc. It can also have more formal definitions as *e.g.*, "the set of k-sparse vectors of dimension n" or "the set of finite-energy band-limited functions with maximum frequency f_c".

   The target set $\mathcal{X}_{\text{target}}$ is typically a subset of a more general space $\mathcal{X}$ with some desirable mathematical properties (for instance a structure of Hilbert or Banach space).

For example, "the set of k-sparse vectors of dimension n" is embedded in $\mathcal{X} = \mathbb{R}^n$; the set of finite-energy band-limited functions with maximum frequency f_c is contained in the (Hilbert) space of square-integrable functions $\mathcal{X} = L_2(\mathbb{R})$.

2. **The observation model**: an "observation" is some quantity (*e.g.*, a vector, a function, a sequence of real numbers, etc.) that the practitioner has at its disposal to attempt to identify one element $\mathbf{x}$ of the target set $\mathcal{X}_{\text{target}}$. This quantity will be denoted $\mathbf{y}$ hereafter. As the elements of the target set, the observation $\mathbf{y}$ belongs to some mathematical space which will be denoted $\mathcal{Y}$ hereafter. The term "observation" intuitively suggests that $\mathbf{y}$ is somehow related to the element $\mathbf{x} \in \mathcal{X}_{\text{target}}$ we wish to recover. We refer to the "*observation model*" as the mathematical characterization of the relation between a target vector $\mathbf{x} \in \mathcal{X}_{\text{target}}$ and the observation $\mathbf{y}$.

   In its most general formulation, the model relating the observation to the target vector $\mathbf{x} \in \mathcal{X}_{\text{target}}$ is probabilistic and can be fully characterized by the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$. A simplified version of this model that I have considered in several of my works is as follows:

$$\mathbf{y} = \mathscr{M}(\mathbf{x}) + \boldsymbol{\epsilon} \tag{1.1}$$

   where $\mathscr{M} : \mathcal{X}_{\text{target}} \to \mathcal{Y}$ is some function often referred to as *measurement operator* and $\boldsymbol{\epsilon}$ is some additive perturbation. Model (1.1) corresponds to the quite common situation where $\mathbf{y}$ stems from some deterministic measurements of $\mathbf{x}$ and is corrupted by noise $\boldsymbol{\epsilon}$.

3. **The decoder**: We will call a "*decoder*" any function taking an observation $\mathbf{y}$ as input and returning an element of $\mathcal{X}$ as output, that is

$$\mathscr{D} : \mathcal{Y} \to \mathcal{X}. \tag{1.2}$$

   In practice, a "good" decoder must try to attain two (often contradictory) objectives. First, the decoder must meet certain *accuracy* requirements. More specifically, the output of the decoder should be close in some sense to the element $\mathbf{x}$ which generated its input (see item "The accuracy criterion" below).

   Second, the decoder must usually be compliant with some *complexity* constraints. In particular, any practical decoder should exhibit a complexity scaling (at most) as a polynomial function of the dimensions of the problem at stake. This constraint can be accounted for by requiring $\mathscr{D}$ to belong to some prescribed set of decoders $\mathcal{D}$.

4. **The accuracy criterion**: The purpose of the accuracy criterion is to define mathematically what is meant by the fuzzy statement *"the elements of $\mathcal{X}_{\text{target}}$ are correctly recovered by decoder $\mathscr{D}$ from the received observations"*. In other words, the accuracy criterion specifies what is the level precision that should be attained by the output of the decoder to be considered as "correct".

The simplest accuracy criterion encountered in practice is probably *"exact recovery"*. In this case, the decoder $\mathscr{D}$ is assumed to be successful if it identifies unambiguously any element of $\mathcal{X}_{\text{target}}$ from the received observations. In the noiseless setting, this criterion formally writes as

$$\forall \mathbf{x} \in \mathcal{X}_{\text{target}} : \mathscr{D}(\mathscr{M}(\mathbf{x})) = \mathbf{x}. \tag{1.3}$$

Achieving exact recovery is unfortunately rarely possible in practice, for instance when the observations are corrupted by unbounded noise or $\mathcal{X}_{\text{target}}$ is not a low-dimensional manifold of $\mathcal{X}$. In these cases, more refined accuracy criteria allowing for discrepancies between the target vector and the output of the decoder, can be considered.

For example, a standard accuracy criterion used to assess the robustness of a decoder to noise reads:

$$\forall \mathbf{x} \in \mathcal{X}_{\text{target}}, \forall \boldsymbol{\epsilon} \in \mathcal{Y} : \|\mathscr{D}(\mathscr{M}(\mathbf{x}) + \boldsymbol{\epsilon}) - \mathbf{x}\|_{\mathcal{X}} \leq \delta \|\boldsymbol{\epsilon}\|_{\mathcal{Y}}, \tag{1.4}$$

where δ is some constant and $\|\cdot\|_{\mathcal{X}}$, $\|\cdot\|_{\mathcal{Y}}$ are norms on $\mathcal{X}$ and $\mathcal{Y}$ respectively.

The four ingredients described above can be used to give a more precise formulation of our initial (naive) question, *e.g.*,

"Given a target set $\mathcal{X}_{\text{target}}$, a family of decoders $\mathcal{D}$ and some accuracy criterion, is there some $\mathscr{D} \in \mathcal{D}$ such that any element of $\mathcal{X}_{\text{target}}$ can be reconstructed to the prescribed accuracy from $\mathbf{y} = \mathscr{M}(\mathbf{x}) + \boldsymbol{\epsilon}$?"

Depending on the definition of $\mathcal{D}$, this general question may have different scope and meaning. If $\mathcal{D}$ corresponds to the set of all possible decoders (that is, there is no constraints on the decoders), the question addresses the intrinsic possibility of solving the reconstruction problem. If $\mathcal{D}$ is the set of polynomial-time decoders, it raises the question of the existence of some "practical" solving procedures. In the latter case, providing an answer may sometimes be incredibly difficult since it underpins fundamental questions such as "Does P equal NP?".

A more pragmatic point of view is therefore to particularize this question to $\mathcal{D} = \{\mathscr{D}\}$ where $\mathscr{D}$ is some particular polynomial-time procedure. The question of interest then reads as follows:

"Given a target set $\mathcal{X}_{\text{target}}$, a decoder $\mathscr{D}$ and some accuracy criterion, under which conditions on $\mathscr{M}$ and $\boldsymbol{\epsilon}$ does decoder $\mathscr{D}$ reconstruct any element of $\mathcal{X}_{\text{target}}$ to the prescribed accuracy from $\mathbf{y} = \mathscr{M}(\mathbf{x}) + \boldsymbol{\epsilon}$?"

The practitioner is then faced with two subtasks:

1. For a given decoder $\mathscr{D}$, he/she wishes to find the most general conditions (ideally necessary and sufficient) ensuring that the decoder carries out the prescribed reconstruction task.

2. For a given computational budget, she/he wants to find the decoder that performs correctly under the most general operating conditions.

In my work, I have addressed different instantiations of these two subtasks. In the rest of this chapter, I give a brief overview of some of them and describe how they relate to the general framework of reconstruction problems.

1.1 Synchronization of digital receivers

In digital communications, one wishes to send a sequence of bits $\mathbf{b} \in \{0,1\}^k$ through a noisy channel (wireless environment, digital subscriber line, etc). In order to counter the noise induced by the transmission channel, some redundancy is added to the transmitted sequence. This process is known as *"channel coding"*.

In its simplest version, channel coding consists in designing a "codebook" $\mathcal{C} \subset \{-1,1\}^m$ ($m > k$) and a coding operator

$$\mathscr{C}\colon \{0,1\}^k \to \mathcal{C} \tag{1.5}$$

which associates one (different) element of $\mathcal{C}$ to each element of $\{0,1\}^k$. The idea underpinning channel coding is as follows: if the element of $\mathcal{C}$ are sufficiently "far apart" in $\{-1,1\}^m$ (for example in terms of Hamming distance), it may be possible to unambiguously identify a codeword (and therefore the corresponding bitstream $\mathbf{b}$) from its noisy observation.

Codebooks are typically designed to deal with "additive noise" channels where the codeword associated to any $\mathbf{b} \in \{0,1\}^k$ is observed up to some additive noise:

$$\mathbf{y} = \mathscr{C}(\mathbf{b}) + \boldsymbol{\epsilon}. \tag{1.6}$$

The elements of $\boldsymbol{\epsilon}$ are commonly assumed to be independent and identically distributed. During the last decades, many procedures have been proposed to design good (and tractable) codebooks for additive channels [1]. This line of search has culminated in the 1990's with the discovery of the "Turbo" [2, 3] codes and the rediscovery of "LDPC" codes [4, 5].

Model (1.6) is unfortunately rarely encountered in practice and more realistic models have therefore to be considered. One refined model that I have considered in my work writes as follows:

$$\mathbf{y} = \mathbf{H}(\mathbf{x})\mathscr{C}(\mathbf{b}) + \boldsymbol{\epsilon} \tag{1.7}$$

where $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{m\times m}$ is a matrix depending on some "external" parameters $\mathbf{x} \in \mathbb{R}^n$. For instance, the elements of $\mathbf{x}$ may correspond to parameters of the electronic devices used to implement the digital transmission (*e.g.*, the frequency and the offset of the carrier, the frequency of the sampling clock, etc).

When $\mathbf{x}$ is known, the presence of $\mathbf{H}(\mathbf{x})$ in model (1.7) can usually be compensated to recover ideal model (1.6) (and thus benefits from the full power of codes designed for additive channels). In practice, $\mathbf{x}$ is unknown and must be estimated. This task is commonly referred to as *"synchronization"* and was the main topic of my PhD. I also published several works in this field at the early stages of my research career.

In the field of synchronization, the set of target signals typically takes the trivial form $\mathcal{X}_{\text{target}} = \mathbb{R}^n$, although some probability distribution $p(\mathbf{x})$ may possibly enforce constraints on the admissible values of $\mathbf{x}$. The observation model is probabilistic and $p(\mathbf{y}|\mathbf{x})$ is defined by (1.7) by using the following typical assumptions: *i)* $\boldsymbol{\epsilon}$ is an i.i.d. Gaussian vector; *ii)* $\mathbf{b}$ is uniformly distributed on $\{0,1\}^k$.

Two relevant questions to address in the field of synchronization are as follows:

1. How to design a *tractable* decoder $\mathscr{D}$ fully exploiting the probabilistic model $p(\mathbf{y}|\mathbf{x})$?[1]
2. Considering the mean-square estimation error

$$\mathrm{E}_{\mathbf{y},\mathbf{x}}[\|\mathscr{D}(\mathbf{y}) - \mathbf{x}\|_2^2] \leq \delta \tag{1.8}$$

   as an accuracy criterion, what is the smallest value of δ attainable within the family of unbiased decoders?

The main difficulty in answering these two questions lies in the mathematical structure of the likelihood function $p(\mathbf{y}|\mathbf{x})$: Using the above assumptions, it can easily be seen that $p(\mathbf{y}|\mathbf{x})$ corresponds to a mixture of 2^k Gaussians with means equal to $\mathbf{H}(\mathbf{x})\mathscr{C}(\mathbf{b})$, $\mathbf{b} \in \{0,1\}^k$. As 2^k becomes rapidly large even for moderate values of k, handling such a huge number of terms may appear numerically intractable at first sight.

In a series of paper, I showed with some co-authors that tractable implementations of maximum a posteriori estimates of $\mathbf{x}$ (or good approximations thereof in some cases) exist by exploiting the specific structure of the channel codebook $\mathcal{C}$, see [6–8]. Similarly, exploiting

[1] In particular, one usually wishes to derive maximum-likelihood or maximum-a-posteriori estimates of $\mathbf{x}$ to benefit from the desirable statistical properties of these estimators.

the particular structure of the probability model at hand, we showed that Bayesian Cramer-Rao bound, lower-bounding the mean-square error achievable by any unbiased decoder (and thus also the parameter δ appearing in (1.8)), can be evaluated with tractable complexity [9, 10].

Some papers representative of my work in this field are:

- C. Herzet *et al.*, "Code-aided Turbo Synchronization", The Proceedings of the IEEE, vol. 95, pp. 1255-1271, June 2007.

  doi.org/10.1109/JPROC.2007.896518

- S. Bay, C. Herzet, J.-M. Brossier, J.-P. Barbot and B. Geller, "Analytic and Asymptotic Analysis of the Bayesian Cramér-Rao Bound for Dynamical Phase Offset Estimation", IEEE Transactions on Signal Processing, vol. 56, pp. 61-70, Jan. 2008.

  doi.org/10.1109/TSP.2007.906760

- C. Herzet, K. Woradit, H. Wymeersch and L. Vandendorpe, "Code-Aided Maximum-Likelihood Ambiguity Resolution Through Free-Energy Minimization", IEEE Transactions on Signal Processing, vol. 58, no. 12, pp. 6238-6250, Dec. 2010.

  doi.org/10.1109/TSP.2010.2068291

1.2 Sparse representations

A second line of research that I have pursued over the last decade is in the field of *"sparse representations"*. In this context, the goal is to identify (to some accuracy) a vector with a few non-zero elements from a set of measurements.

More formally, the set of target vectors $\mathcal{X}_{\text{target}} \subset \mathbb{R}^n$ is here defined as

$$\mathcal{X}_{\text{target}} = \{\mathbf{x} \in \mathbb{R}^n \colon \|\mathbf{x}\|_0 \leq k\} \tag{1.9}$$

for some $k \in \mathbb{N}$, where $\|\cdot\|_0$ denotes the counting function which returns the number of non-zero components in its argument. It is easy to see that any vector of $\mathcal{X}_{\text{target}}$ contains at most k non-zero elements and $\mathcal{X}_{\text{target}}$ is commonly referred to as the set of *"k-sparse"* vectors of $\mathbb{R}^n$. In its standard formulation, the observation model considered in sparse-representation problems is linear:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \tag{1.10}$$

for some matrix $\mathbf{A} = [\mathbf{a}_1 | \dots | \mathbf{a}_n] \in \mathbb{R}^{m \times n}$, $m < n$, and additive noise $\boldsymbol{\epsilon} \in \mathbb{R}^m$. Matrix $\mathbf{A}$ is often referred to as *"dictionary"* and its columns as *"atoms"*.

Despite of its apparent simplicity, addressing a reconstruction problem involving target set (1.9) and observation model (1.10) is often a challenging task. To illustrate our point, let us focus on the simplified case where (1.10) holds with $\boldsymbol{\epsilon} = \mathbf{0}_m$ (noiseless setting) and exact recovery (1.3) is considered as an accuracy criterion. In this setup, it may be shown (see *e.g.*, [11, Theorem 2.13]) that there exist some decoders achieving exact recovery if and only if

$$\text{“Any } 2k \text{ columns of } \mathbf{A} \text{ are linearly independent”.} \tag{EkC}$$

In the sequel, we will refer to this condition as “Exact k-sparse recovery Condition” (EkC). An instance of decoder achieving exact recovery under (EkC) writes as follows:

$$\begin{aligned} \mathscr{D}\colon \mathbb{R}^m &\to \mathbb{R}^n \\ \mathbf{y} &\mapsto \underset{\mathbf{z}\in\mathbb{R}^n}{\arg\min} \|\mathbf{z}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{z}\,. \end{aligned} \tag{P_0}$$

Unfortunately, solving (P_0) turns out to be an NP-hard problem, see *e.g.*, [11, Section 2.3]. Now, in the current state of knowledge no polynomial-time solver has been found for NP-hard problems. In fact, it is conjectured by many mathematicians that no polynomial-time solver exists for NP-hard problems although this result has not been proved yet. If this conjecture is correct, NP-hardness of (P_0) means that there is no polynomial-complexity algorithm solving *any* instance of (P_0) (that is for any choice of $\mathbf{y}$ and $\mathbf{A}$). We note nevertheless that there may possibly exist polynomial-complexity procedures solving *some* instances of (P_0).

We thus face the following question:

“Given $\mathbf{A} \in \mathbb{R}^{m\times n}$, is there some polynomial-complexity algorithm achieving exact recovery of any k-sparse vector $\mathbf{x}$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$? (and if so, under which conditions?)”

This question (extended to more general notions of accuracy such as *e.g.*, stability or robustness of the decoder, see *e.g.*, [12]) has sparked of surge of interest in the scientific community over the two last decades. I personally contributed to this field of research in several of my contributions. My work has addressed methodological, theoretical and application aspects of this question as described below.

1.2.1 Bayesian methods

A first avenue of my research focused on the design of polynomial-complexity decoders achieving good reconstruction performance (at least empirically) for sparse-representation problems. My contributions mainly took place within a Bayesian framework where “sparsity-inducing” probabilistic priors are assumed on $\mathbf{x}$.[2] In this context, a popular prior used in

[2] More formally, $p(\mathbf{x})$ is such that $\int_{\mathbf{x}\in\mathcal{X}_{\text{target}}} p(\mathbf{x})d\mathbf{x} \simeq 1$.

many contributions of the literature (see *e.g.*, [13, 14]) reads as follows:

$$\mathbf{x} = \mathbf{s} \odot \mathbf{w} \tag{1.11}$$

where

$$\mathbf{s}(i) \sim \text{Bernoulli}(p) \tag{1.12}$$

$$\mathbf{w} \sim \text{Gaussian}(\mathbf{0}_n, \gamma^2 \mathbf{I}_n) \tag{1.13}$$

and $\odot$ denotes element-wise product. This model is known as the (multiplicative) "Bernoulli-Gaussian" model in the literature and has been declined in other similar variants, see *e.g.*, [13, 15, 16].

Interestingly, when the distribution of the noise $\boldsymbol{\epsilon}$ corrupting the observations is zero-mean i.i.d Gaussian, *i.e.*,

$$\boldsymbol{\epsilon} \sim \text{Gaussian}(\mathbf{0}_m, \sigma^2 \mathbf{I}_m), \tag{1.14}$$

the MAP estimate of $\mathbf{x}$ is closely related to an ℓ_0-penalized sparse-regression problem. More specifically, when $\gamma^2 \to \infty$ the MAP estimate of $\mathbf{x}$ can also be found by solving the following problem for some $\lambda > 0$:

$$\underset{\mathbf{z} \in \mathbb{R}^n}{\arg\min} \, \tfrac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_0, \tag{$P_{2,0}$}$$

see *e.g.*, [17, Section 7], [14, Section II.B]. This observation has motivated three of my lines of research that I briefly describe hereafter.

Bayesian greedy procedures: In a first series of works [17–19], I investigated (in collaboration with A. Drémeau) how several standard greedy procedures (*e.g.*, MP [20], OMP [21], SP [22], CoSaMP [23]) can be revisited in a Bayesian framework. The main idea of these contributions was to depart from the simple i.i.d. model (1.12) to consider more complex dependencies between the activation probabilities of the atoms.

One general model that we considered in these works writes as

$$p(\mathbf{s}) \propto \exp\big(\mathbf{b}^{\mathrm{T}}\mathbf{s} + \mathbf{s}^{\mathrm{T}}\mathbf{C}\mathbf{s}\big), \tag{1.15}$$

where $\propto$ denotes equality up to a normalization constant, and $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{C} \in \mathbb{R}^{n \times n}$ are some model parameters. This model is often known as "Boltzmann machine" or "Ising model" and plays an important role in statistical physics. We then derived several greedy procedures targeting the following problem:

$$\underset{\mathbf{w} \in \mathbb{R}^n, \mathbf{s} \in \mathbb{R}^n}{\arg\max} \; p(\mathbf{w}, \mathbf{s} | \mathbf{y}) \tag{1.16}$$

where $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$ is implicitly defined by (1.13)-(1.14) and (1.15).

The proposed procedures were similar in spirit to standard strategies of the literature but leveraged the cost function in (1.16) rather than the residual "$\mathbf{y} - \mathbf{A}\mathbf{x}$"[3] to make their greedy decisions. This choice had two desirable consequences:

1. A "backward" step (*i.e.*, the removal of some atom previously selected by the greedy algorithm) naturally appears in the considered Bayesian framework.
2. The parameters of the model (and in particular more general priors on the support $\mathbf{s}$) can straightforwardly be taken into account.

For example, when prior (1.15) is considered, the locally-optimal[4] decision regarding the activation coefficient $\mathbf{s}(i)$ reads as

$$\underset{\mathbf{s}(i)\in\{0,1\}}{\arg\max}\left(\max_{\mathbf{w}(i)\in\mathbb{R}} p(\mathbf{w}, \mathbf{s}|\mathbf{y})\right) = \begin{cases} 1 & (\mathbf{a}_i^{\mathrm{T}}\mathbf{r}_i)^2 > T_i \\ 0 & \text{otherwise} \end{cases} \tag{1.17}$$

where

$$T_i = -2\sigma^2 \frac{\sigma^2+\gamma^2}{\gamma^2}\left(\mathbf{b}(i) + 2\sum_{j\neq i}\mathbf{C}(i,j)\,\mathbf{s}(j)\right),$$
$$\mathbf{r}_i = \mathbf{y} - \sum_{j\neq i}\mathbf{a}_j\mathbf{w}(j).$$

It is clear from (1.17) that removing atom $\mathbf{a}_i$ from the sparse decomposition becomes (locally) more advantageous as soon as $(\mathbf{a}_i^{\mathrm{T}}\mathbf{r}_i)^2 \leq T_i$. Moreover, the value of T_i is an explicit function of the model parameters $\mathbf{b}$ and $\mathbf{C}$. We note that, when $\mathbf{b} = \log(\frac{p}{1-p})\mathbf{1}_n$ and $\mathbf{C} = \mathbf{0}_{n\times n}$, prior model (1.15) reduces to the standard i.i.d Bernoulli model (1.12). In that particular case, the Bayesian framework considered in our work thus naturally leads to a "forward-backward" extension of some existing greedy procedures.

Variational methods for sparse representations: In [24, 25], I explored (in collaboration with A. Drémeau) how variational inference tools can be applied to Bayesian formulations of sparse-representation problems.

In a nutshell, the main idea of these contributions was to benefit from the Bayesian framework to "marginalize out" some of the variables of the problem at hand and (hopefully)

[3] As commonly used in standard greedy methods.

[4] That is, when $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$ is optimized with respect to $(\mathbf{s}(i), \mathbf{w}(i))$ while the other components of $(\mathbf{s}, \mathbf{w})$ are fixed to a given value.

obtain a better estimate of the sparse vector underlying the data. More specifically, we targeted the following problem $\forall i \in \{1, \ldots, n\}$:

$$\arg\max_{\mathbf{s}(i)\in\{0,1\}} p(\mathbf{s}(i)|\mathbf{y}) \tag{1.18}$$

where the cost function is the marginal of the joint probability $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$. As previously, we assumed that $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$ is implicitly defined by (1.13)-(1.14) and (1.15).

Unsurprisingly, the evaluation of the marginal $p(\mathbf{s}(i)|\mathbf{y})$ turns out to be intractable. We thus considered a particular variational approximation of $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$, known as "mean-field approximation", where a surrogate probability $q(\mathbf{w}, \mathbf{s})$ verifying

$$q(\mathbf{w}, \mathbf{s}) = \prod_{i=1}^{n} q(\mathbf{w}(i), \mathbf{s}(i)) \quad \text{with} \quad \sum_{\mathbf{s}(i)\in\{0,1\}} \int_{\mathbb{R}} q(\mathbf{w}(i), \mathbf{s}(i)) d\mathbf{w}(i) = 1 \ \forall i \tag{1.19}$$

is computed by optimizing the Kullback-Leibler divergence between $q(\mathbf{w}, \mathbf{s})$ and $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$, see *e.g.*, [26]. We emphasized that the factors $q(\mathbf{w}(i), \mathbf{s}(i))$ appearing in the decomposition (1.19) have a simple parametric form at optimality and used this feature to efficiently search a local minimum (or a saddle point) of the Kullback-Leibler divergence between $q(\mathbf{w}, \mathbf{s})$ and $p(\mathbf{w}, \mathbf{s}|\mathbf{y})$.

Interestingly, the iterates of the numerical procedure derived in our variational framework can also be seen as a "soft-decision" version of the "Bayesian Matching Pursuit" algorithm proposed in [17–19]: here the mean and the variance of the elements of $\mathbf{s}$ and $\mathbf{w}$ (according to the current value of $q(\mathbf{w}, \mathbf{s})$) appear in the recursion of the optimization procedure rather than hard decisions on these variables, see [25, Section IV.D].

Dictionary learning in a Bayesian framework: In [27], I investigated (in collaboration with C. Elvira and H.-P. Dang) the use of some variants of the Bernoulli-Gaussian model (1.12)-(1.14) for dictionary learning. Our goal was not only to estimate the decomposition dictionary $\mathbf{A} \in \mathbb{R}^{m\times n}$ and the sparse vectors corresponding to a set of observations $\{\mathbf{y}_l\}_{l=1}^{L}$ but also to infer the dimension n (that is the number of atoms) of $\mathbf{A}$.

We followed two lines of thought. In a first approach, we assumed that

$$\begin{aligned} \mathbf{s}(i) &\sim \text{Bernoulli}(p_i) \\ p_i &\sim \text{Beta}, \end{aligned}$$

that is the activation parameter $\mathbf{s}(i)$ of each atom is drawn independently from a Bernoulli distribution with parameter p_i but the latter parameter is also considered as a random variable following a non-informative Beta distribution. This over-parameterization of the problem aimed to trigger "Occam's razor" effects by automatically tuning some activation

parameters p_i to zero in the estimation process.[5] In a second approach, we considered the case where $\mathbf{s}$ is a binary *sequence* whose realizations obey an Indian buffet process [28]. This setup enabled us to handle dictionaries containing (virtually) an infinite (countable) number of atoms.

In both approaches, we assumed that the atoms follow a unit-variance centered Gaussian distribution:

$$\mathbf{a}_i \sim \text{Gaussian}(\mathbf{0}_m, \mathbf{I}_m).$$

We then computed estimates of the quantities of interest by using a "small-variance asymptotics" approximation of the Gibbs sampler associated to our probabilistic model, see [29]. This approach allowed us to obtain a Bayesian method automatically tuning the parameters of our model while exhibiting a reasonable computational complexity (as compared to a standard implementation of a Gibbs sampler).

Some papers representative of my work on Bayesian methods for sparse problems:

- C. Herzet and A. Drémeau, "Bayesian pursuit algorithms", EURASIP European Signal Processing Conference (EUSIPCO), Aalborg, 2010.

  ieeexplore.ieee.org/document/7096474

- A. Drémeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions", IEEE Transactions on Signal Processing, 60(7):3425–3438, July 2012.

  doi.org/10.1109/tsp.2012.2192436

- J. Arbel, H.-P. Dang, C. Elvira, C. Herzet, Z. Naulet and M. Vladimirova, "Bayes in action in deep learning and dictionary learning", ESAIM: Proc. and Surveys, 2022

  esaim.org/arbel2022

[5] Strictly speaking, in this setup the size of the dictionary is therefore fixed in advance and equal to n. However, since in practice only a subset of atoms are activated in the learning process, the effective size of the final dictionary is typically much smaller that n.

1.2.2 Screening acceleration methods

One of the most popular alternatives to decoder (P_0) writes as[6]

$$\begin{aligned}\mathscr{D}\colon \mathbb{R}^m &\to \mathbb{R}^n \\ \mathbf{y} &\mapsto \underset{\mathbf{z}\in\mathbb{R}^n}{\arg\min}\, \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{y}=\mathbf{A}\mathbf{z}\,,\end{aligned} \tag{P_1}$$

or in its "noise-aware" version,

$$\begin{aligned}\mathscr{D}\colon \mathbb{R}^m &\to \mathbb{R}^n \\ \mathbf{y} &\mapsto \underset{\mathbf{z}\in\mathbb{R}^n}{\arg\min}\, \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2+\lambda\|\mathbf{z}\|_1\,,\end{aligned} \tag{$P_{2,1}$}$$

for some $\lambda > 0$.

Problems (P_1) and ($P_{2,1}$) are convex programs and are respectively known as "Basis Pursuit" and "Basis Pursuit Denoising" in the literature, see [30, 31]. These decoders owe their popularity to two important facts:

- First, they enjoy very good reconstruction properties under mild conditions (see *e.g.*, [11, Chapter 4] for an overview of standard recovery results). In particular, they correctly identify the elements of (1.9) from (noiseless) partial measurements under conditions slightly stronger than (EkC).
- Second, they can respectively be reformulated as linear or quadratic programs with linear constraints for which it is known to exist generic polynomial-complexity solving procedures, see *e.g.*, [32, Chapters 13 & 16].

Despite of the existence of generic polynomial-complexity algorithms, many dedicated methods have been proposed in the literature to accelerate the resolution of problems (P_1)-($P_{2,1}$), see *e.g.*, [33–36]. These methods leverage the fact that the solutions of (P_1)-($P_{2,1}$) contain (under mild conditions) at most m non-zero coefficients, see *e.g.*, [11, Theorem 3.1]. This property can be exploited to decrease the storage/computational burden of solving procedures. Some of my works fall within this line of search.

More specifically, I have been interested in a family of acceleration procedures called *"safe screening"* and first proposed by El Ghaoui *et al.* in [37]. The idea of safe screening consists in identifying some of the zeros of the minimizers of ($P_{2,1}$) with a low computational burden.[7] These elements can then be withdrawn from the optimization variables, resulting in a problem of smaller dimension.

[6] To simplify our discussion and notations, we suppose hereafter that the minimizers of these problems are unique.

[7] We focus hereafter on problem ($P_{2,1}$) since standard screening procedures are specifically tailored to this problem.

Standard screening tests leverage the following "complementary slackness" necessary optimality condition:

$$|\mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star| < \lambda \implies \mathbf{x}^\star(i) = 0, \tag{1.20}$$

where $\mathbf{x}^\star \in \mathbb{R}^n$ is a minimizer of ($P_{2,1}$) and $\mathbf{u}^\star$ the maximizer of the following "dual" problem:

$$\mathbf{u}^\star = \underset{\mathbf{u}\in\mathbb{R}^m:\|\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_\infty\leq\lambda}{\arg\max} \tfrac{1}{2}\|\mathbf{y}\|_2^2 - \tfrac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2. \tag{1.21}$$

The inequality in the left-hand side of (1.20) is clearly a sufficient condition to identify if the ith element of $\mathbf{x}^\star$ is equal to zero. Unfortunately, the implementation of (1.20) is conditioned to the knowledge of $\mathbf{u}^\star$ and solving (1.21) turns out to be as difficult as solving ($P_{2,1}$). This prevents from using (1.20) as a simple test to identify the zeros of $\mathbf{x}^\star$.

This difficulty can be nevertheless circumvented by introducing the concept of "safe region", as suggested in [37]. A safe region is a subset $\mathcal{R} \subset \mathbb{R}^m$ verifying $\mathbf{u}^\star \in \mathcal{R}$. If a safe region $\mathcal{R}$ is available to the practitioner, (1.20) can be relaxed to

$$\max_{\mathbf{u}\in\mathcal{R}} |\mathbf{a}_i^{\mathrm{T}}\mathbf{u}| < \lambda \implies \mathbf{x}^\star(i) = 0. \tag{1.22}$$

For proper choices of $\mathcal{R}$, the maximum in the left-hand side of (1.22) is easy to evaluate and the corresponding inequality may become a tractable test to identify (some) zeros of $\mathbf{x}^\star$. For example, if $\mathcal{R}$ is a sphere, that is

$$\mathcal{R} = \mathcal{B}(\mathbf{c}, r) \triangleq \{\mathbf{u} \in \mathbb{R}^m \colon \|\mathbf{u}-\mathbf{c}\|_2 \leq r\}, \tag{1.23}$$

(1.22) takes a particularly simple form:

$$|\mathbf{a}_i^{\mathrm{T}}\mathbf{c}| < \lambda - r\|\mathbf{a}_i\|_2 \implies \mathbf{x}^\star(i) = 0. \tag{1.24}$$

In this case, provided that $\mathbf{c}$ and r are known, the screening test can be implemented at the expense of evaluating one inner product between $\mathbf{c}$ and $\mathbf{a}_i$. We note that, in many setups, this inner product is already computed by the numerical procedure addressing ($P_{2,1}$) and screening can then be implemented at virtually no cost.

Following El Ghaoui's seminal work, screening has become a central research theme for many researchers and has proved to allow dramatic computational savings in many setups. I describe below some of my personal contributions to this field.

Design of new safe regions: As mentioned above, the standard formulation of safe screening is grounded on the concept of "safe region", that is a subset of the dual space provably containing the dual optimal solution $\mathbf{u}^\star$. The choice of a "good" safe region reveals to be crucial to the final effectiveness and efficiency of the screening tests. On the one hand, loosely speaking, "smaller" regions lead to more effective tests, see [38, Lemma 1].[8] On the

[8] In particular, if $\mathcal{R} = \{\mathbf{u}^\star\}$, (1.22) reduces to standard optimality condition (1.20).

other hand, the complexity of the tests is closely related to the geometry of the safe regions (see *e.g.*, (1.23)-(1.24) for the case of a sphere). As a consequence, several geometries have been considered (ball [39], ellipsoid [40], dome [38], truncated dome [38]) and different strategies to tune the parameters of the corresponding regions have been proposed ([39, 41, 42]) in the literature.

In [43], I participated (in collaboration with A. Malti) to this research effort by introducing a strategy to reconcile two of the most effective screening procedures of the literature, the so-called "FNE" [42] and "GAP" [39] balls. More recently, in the thesis of Thu-Le Tran, we initiated a work (still ongoing) to show that these two balls can in fact be seen as particular cases of a more general framework. A first output of our line of thought has been published in [44]. In this contribution, we introduced a new safe dome (called "Hölder dome") whose construction requires the same computational cost as GAP regions [39] but which is always included (strictly under very mild conditions) in the latter, see [44, Theorems 1 and 2].

Safe screening for SLOPE/OWL: Following El Ghaoui's seminal work, it was soon recognized that screening can also be applied to problems involving sparsity-inducing regularization terms different from standard ℓ_1 norm, see *e.g.*, [45, 46]. In [47], I addressed (in collaboration with C. Elvira) the derivation of screening rules for one particular instance of this type of problems:

$$\text{find } \mathbf{x}^\star \in \underset{\mathbf{z}\in\mathbb{R}^n}{\arg\min}\ \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 + \lambda \sum_{i=1}^{n} \gamma_i |\mathbf{z}|_{[i]} \qquad (P_{2,\text{slope}})$$

where $|\mathbf{z}|_{[i]}$ denotes the ith largest element of $\mathbf{z}$ in absolute value, that is

$$\forall \mathbf{z} \in \mathbb{R}^n :\ |\mathbf{z}|_{[1]} \geq |\mathbf{z}|_{[2]} \geq \ldots \geq |\mathbf{z}|_{[n]},$$

and

$$\gamma_1 > 0, \qquad \gamma_1 \geq \cdots \geq \gamma_n \geq 0.$$

Problem ($P_{2,\text{slope}}$) is known as *"Sorted L-One Penalized Estimation" (SLOPE)* or *"Ordered Weighted L-One Linear Regression" (OWL)* in the literature [48, 49] and has been shown to be of interest in many application domains during the last decades, see *e.g.*, [50–55].

One particular feature of SLOPE/OWL is that, in contrast to previous works of the literature, the regularization term appearing in the cost function is not "separable" (that is it cannot be written as a sum over a partition of the elements of $\mathbf{z}$). We showed that a tractable implementation of screening tests is nevertheless possible although the latter take the form of a *set* of n inequalities to verify for each atom,[9] see [47, Theorem 4.3]. We

[9] Rather than a single inequality as in (1.22).

moreover introduced an efficient algorithm to evaluate the proposed screening tests for *all* the atoms of the dictionary with a complexity scaling as $\mathcal{O}(n \log n + LT)$ where L is the number of elements passing the test and T ($\leq n$) is some problem-dependent constant, see [47, Section 4.3].

Scalable screening for large dictionaries: In [56, 57], I addressed (in collaboration with A. Drémeau and C. Dorffer) the problem of implementing screening when the decomposition dictionary contains a huge number of atoms. We proposed a strategy to identify *several* zeros[10] of the solution of ($P_{2,1}$) by performing *one single* test, see [56, 57]. These contributions leverage the following simple implication:

$$\max_{\mathbf{a}\in\mathcal{A}} \max_{\mathbf{u}\in\mathcal{R}} |\mathbf{a}^{\mathrm{T}}\mathbf{u}| < \lambda \implies \forall \mathbf{a}_i \in \mathcal{A} : \ \mathbf{x}^\star(i) = 0. \tag{1.25}$$

Here, $\mathcal{A}$ represents a region of $\mathbb{R}^m$ containing a subset of atoms of the dictionary. Implication (1.25) simply states that if the inequality in the left-hand side is verified, then all the atoms contained in region $\mathcal{A}$ can be safely discarded from the dictionary. In [56, 57], we applied this methodology to dictionaries containing a large (but finite) number of atoms and emphasized that the maximization in (1.25) remains simple for proper choices of $\mathcal{R}$ and $\mathcal{A}$. We are currently investigating the use of this method with Thu-Le Tran to deal with "continuous" setups where the dictionary contains an infinite uncountable number of atoms [58, 59].

Beyond screening for sparse problems: In [60, 61], I emphasized (in collaboration with C. Elvira) that screening can be extended to different optimization problems, not necessarily involving sparse solutions. More specifically, we focused on a variant of ($P_{2,1}$) enforcing an ℓ_∞ penalization (rather than ℓ_1):

$$\text{find } \mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_\infty. \tag{$P_{2,\infty}$}$$

The solutions of this problem are known to be "flat", that is many elements of the minimizers have the same absolute magnitude. We proposed a procedure (similar to screening and dubbed "squeezing") able to identify these elements and we exploited these detections to accelerate the optimization algorithms addressing this problem. Similarly to screening, our test leverages the problem's optimality conditions and takes the following simple form:

$$|\mathbf{a}_i^{\mathrm{T}}\mathbf{c}| < \lambda - r\|\mathbf{a}_i\|_2 \implies \mathbf{x}^\star(i) = \operatorname{sign}(\mathbf{a}_i^{\mathrm{T}}\mathbf{c})\|\mathbf{x}^\star\|_\infty,$$

where $\mathbf{c}$ and r respectively denotes the center and the radius of a safe sphere for the dual problem of ($P_{2,\infty}$), see *e.g.*, [60, Theorem 2].

[10] We note that standard screening tests, as stated in (1.22), require to verify one inequality per element of the primal vector $\mathbf{x}^\star$.

Beyond screening for convex problems: In [62], I extended (in collaboration with T. Guyard and C. Elvira) the concept of screening to non-convex setups. More specifically, we targeted the following problem:

$$\text{find } \mathbf{x}^\star \in \arg\min_{\mathbf{z}\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_0 \quad \text{subject to } \|\mathbf{z}\|_\infty \leq M, \tag{1.26}$$

for some $M > 0$. We note that(1.26) reduces that standard "ℓ_0-penalized sparse-regression problem" ($P_{2,0}$) for sufficiently-large values of M.

In our work, we considered a "branch-and-bound" strategy [63] to search for the minimizers of (1.26). In a nutshell, the crux of branch-and-bound procedures consists in identifying subsets of supports which *cannot* attain the optimal value of (1.26); this knowledge can then be exploited to decrease the dimension of the combinatorial problem at stake. The most common approach of the literature to identify such irrelevant subsets is based on the resolution of a convex relaxation of (1.26) where some constraints on the support of $\mathbf{z}$ are enforced. More specifically, if

$$p_{\text{up}} < p_{\text{low}}(\mathcal{S}_0, \mathcal{S}_1) \tag{1.27}$$

where $\mathcal{S}_0, \mathcal{S}_1$ are two disjoint subsets of $\{1,\ldots,n\}$, p_{up} is an upper bound on the optimal value of (1.26) and

$$p_{\text{low}}(\mathcal{S}_0, \mathcal{S}_1) = \min_{\mathbf{z}\in\mathbb{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 + \tfrac{\lambda}{M}\|\mathbf{z}_{\overline{\mathcal{S}_0\cup\mathcal{S}_1}}\|_1 + \lambda\text{card}(\mathcal{S}_1) \tag{1.28}$$
$$\text{subject to } \begin{cases} \|\mathbf{z}\|_\infty \leq M \\ \mathbf{z}_{\mathcal{S}_0} = \mathbf{0}_{\text{card}(\mathcal{S}_0)}, \end{cases}$$

then there exists no minimizer of (1.26) verifying

$$\begin{cases} \mathbf{x}^\star(i) = 0 & \forall i \in \mathcal{S}_0 \\ \mathbf{x}^\star(i) \neq 0 & \forall i \in \mathcal{S}_1. \end{cases}$$

Unfortunately, when the number of subsets to test increases (*i.e.*, one has to solve many instances of (1.28)), this task may become computationally costly. Our screening method in [62] can be seen as a fast and low-complexity procedure to test many subsets of supports. Our methodology is grounded on the following observations: *i)* condition (1.27) can be weakened by considering the dual function of (1.28) evaluated at any dual feasible point; *ii)* this dual function only differs by one single (easily-computable) term when the couple $(\mathcal{S}_0, \mathcal{S}_1)$ is changed by one element, see [62, Corollary 1]; *iii)* the value of the dual function obeys some nesting property for nested couples $(\mathcal{S}_0, \mathcal{S}_1)$, see [62, Corollary 2].

Some papers representative of my work in screening:

- C. Herzet, C. Dorffer, and A. Drémeau, "Gather and conquer: Region-based strategies to accelerate safe screening tests", IEEE Transactions on Signal Processing, 67(12):3300–3315, 2019.

  doi.org/10.1109/TSP.2019.2914885

- C. Elvira and C. Herzet, "Safe squeezing for antisparse coding", IEEE Transactions on Signal Processing, 68:3252–3265, 2020.

  doi.org/10.1109/TSP.2020.2995192

- T. Guyard, C. Herzet, and C. Elvira, "Node-screening tests for ℓ_0-penalized least-squares problem", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapur, 2022.

  doi.org/10.1109/ICASSP43922.2022.9747563

- T. L. Tran, C. Elvira, H.-P. Dang, and C. Herzet, "Beyond GAP screening for Lasso by exploiting new dual cutting half-spaces", EURASIP European Signal Processing Conference, EUSIPCO'22, 2022.

  preprint/tran:eusipco22

1.2.3 Theoretical guarantees

Another line of research that I have pursued in recent years is the study of theoretical conditions ensuring the "success" of some sparse representation algorithms. This part of my work has been done in collaboration with A. Drémeau, C. Elvira, R. Gribonval, J. Idier and C. Soussen.

In my contributions, I mainly focused on the exact recovery of the *support* of sparse vectors, that is the correct identification of the positions of their non-zero entries:

$$\text{supp}(\mathbf{x}) \triangleq \{i \colon \mathbf{x}(i) \neq 0\}.$$

More precisely, I considered the following general question: If $\mathcal{X}_{\text{target}}$ is equal to (some subset of) (1.9), under which conditions does some decoder $\mathscr{D}$ verify

$$\forall \mathbf{x} \in \mathcal{X}_{\text{target}}, \forall \|\boldsymbol{\epsilon}\|_2 \leq \delta : \ \text{supp}(\mathscr{D}(\mathbf{A}\mathbf{x} + \boldsymbol{\epsilon})) = \text{supp}(\mathbf{x})? \tag{1.29}$$

I described below some variations of this question that I have addressed during the last decade.

Support recovery for OMP/OLS: In [64], I tackled the support-recovery problem when $\mathscr{D}$ corresponds to the well-known "Orthogonal Matching Pursuit" (OMP) [21] or "Orthogonal Least Squares" (OLS) [65] algorithms. Letting $\mathcal{S} \subset \{1, \dots, n\}$ be some subset of k indices, my co-authors and I first studied conditions ensuring that OLS (with $\mathbf{y} = \mathbf{A}\mathbf{x}$ as input) identifies in at most k steps the support of any vector $\mathbf{x}$ in

$$\mathcal{X}_{\text{target}} = \big\{\mathbf{x}' \in \mathbb{R}^n \colon \operatorname{supp}\big(\mathbf{x}'\big) \subseteq \mathcal{S}\big\}. \tag{1.30}$$

This notion of success is often referred to as "k-step recovery" in the literature [66]. We came up to the conclusion that the well-known "Exact Recovery Condition"[11], *i.e.*,

$$\mathbf{A}_{\mathcal{S}} \text{ is full column-rank and } \max_{i \notin \mathcal{S}} \|\mathbf{A}_{\mathcal{S}}^{\dagger}\mathbf{a}_i\|_1 < 1, \tag{ERC}$$

first derived by J. Tropp for OMP in [67, Theorems 3.1 and 3.10], is still necessary and sufficient for OLS to achieve "k-step recovery" of any element of $\mathcal{X}_{\text{target}}$, see [64, Theorems 3 and 4]. We further refined our analysis to identify operating regimes where OLS has provably better reconstruction performance than OMP: On the one hand, we showed that there exist matrices $\mathbf{A}$ such that OMP fails to reconstruct all the elements of (1.30) in at most k steps, see [64, Theorem 7]; on the other hand, when $k \leq \operatorname{kruskal}(\mathbf{A}) - 1$ we emphasized that there always exists a non-empty subset of (1.30) (made up of vectors with "sufficiently decaying" non-zero coefficients) for which OLS achieves k-step recovery of $\mathcal{S}$, see [64, Lemma 3].

Recovery condition for decaying vectors: The favorable behavior of OLS for sparse vectors with decaying non-zero coefficients led me to the following question: Can we relax standard recovery conditions for OMP/OLS when $\mathcal{X}_{\text{target}}$ is restricted to k-sparse vectors with decaying non-zero coefficients?

In [68], we provided some elements of answer to this question by deriving recovery guarantees for OMP/OLS based on mutual coherence:

$$\mu(\mathbf{A}) \triangleq \max_{i \neq j} |\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_j|. \tag{1.31}$$

We note that, although this type of conditions are known to be pessimistic, they can be easily evaluated in practice and lead to interesting insights into the operating regimes favoring the success of reconstruction algorithms.

In the noiseless setting, if $\mathcal{X}_{\text{target}}$ corresponds to the set of k-sparse vectors (1.9), it is a well-established fact that

$$\mu(\mathbf{A}) < \frac{1}{2k-1} \tag{1.32}$$

[11] Here, $\mathbf{A}_{\mathcal{S}}$ denotes the restriction of $\mathbf{A}$ to its columns in $\mathcal{S}$ and $^+$ is the pseudo-inverse operator.

is a sufficient condition for OMP/OLS to achieve "k-step recovery" of any $\mathbf{x} \in \mathcal{X}_{\text{target}}$, see *e.g.*, [64, 67]. This condition is moreover known to be tight in the following sense: There exist a k-sparse vector $\mathbf{x}$ and a dictionary $\mathbf{A}$ with $\mu(\mathbf{A}) = \frac{1}{2k-1}$ such that OMP/OLS selects a wrong atom at the first iteration, see [69, Theorem 3.1]. This shows that (1.32) cannot be weakened for the recovery of arbitrary k-sparse vectors.

In [68, Theorem 1], we emphasized that more favorable conditions can nevertheless be obtained when considering the set of k-sparse vectors with decaying non-zero coefficients: We showed that if $\mathbf{x}$ is a k-sparse vector with non-zero amplitudes *not all equal*, there exists some $\mu^\star > \frac{1}{2k-1}$ such that OMP/OLS identifies the support of $\mathbf{x}$ in at most k steps for any dictionary $\mathbf{A}$ with mutual coherence $\mu(\mathbf{A}) < \mu^\star$. The value of $\mu^\star$ was made precise as a function of the decay of the non-zero coefficients ([68, Theorem 2]) and different variations of the reconstruction problem were considered (partial recovery ([68, Theorem 3]), robust and stable recovery [68, Theorems 4 and 5])).

Towards "k-step recovery" in continuous dictionaries: In [70], we considered the recent paradigm of sparse representations in *"continuous"* dictionaries. In this context, the dictionary used in the sparse decomposition is assumed to contain an infinite uncountable number of atoms obeying some continuity property, see *e.g.*, [58, 59, 71, 72].

In our work, we supposed that the set of atoms forming the dictionary can be written as

$$\mathcal{A} \triangleq \{\mathbf{a}(\theta) \colon \theta \in \Theta\}$$

where $\mathbf{a} \colon \mathbb{R}^d \to \mathcal{H}$ is a continuous function, $\mathcal{H}$ a Hilbert space and Θ an interval of $\mathbb{R}^d$. We addressed the following question: If

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{a}(\theta_i) x_i$$

for some distinct parameters $\{\theta_i \in \Theta\}_{i=1}^k$ and non-zero scalars $\{x_i \in \mathbb{R}^*\}_{i=1}^k$, are there situations where OMP can identify the set $\{\theta_i \in \Theta\}_{i=1}^k$ in exactly k steps for any choice of x_i's?

Although the existence of a positive answer to this question may be surprising at first sight, we showed that k-step recovery of $\{\theta_i \in \Theta\}_{i=1}^k$ from $\mathbf{y}$ is indeed possible for some particular instances of dictionaries $\mathcal{A}$. In particular, we derived generic conditions on $\mathcal{A}$ and $\{\theta_i \in \Theta\}_{i=1}^k$ ensuring that this type of recovery occurs (see [70, Theorem 1]) and showed that these requirements can be met by some supports $\{\theta_i \in \Theta\}_{i=1}^k$ and some general families of dictionaries (see [70, Theorems 2-6]).

Interestingly, in the particular case where $d = 1$, our results revealed that "universal k-step recovery" is possible in some dictionaries $\mathcal{A}$. More precisely, we found instances of dictionaries for which OMP correctly identifies any subset of parameters $\{\theta_i \in \Theta\}_{i=1}^k$ in k steps for any $k \geq 1$ and any choice of the weighted coefficients x_i's. To our knowledge, this is the first occurrence of a "universal" recovery result for *signed* combinations of atoms.

Support recovery with side information: In [73], we studied a variation of the support-recovery problem where some "side information" about $\mathcal{S}$ is available to the decoder. In our work, we assumed that side information takes the form of a set of indices $\mathcal{Q} \subseteq \{1, \dots n\}$, available to the practitioner, which may contain a subset of the true support $\mathcal{S}$ but also possibly include wrong indices. In this context, we derived conditions ensuring that some decoders exploiting this (possibly imperfect) knowledge can successfully identify the support of some sparse vectors.

The "side-information-aware" decoders considered in our paper were constructed as simple modifications of standard procedures of the literature, namely OMP, OLS and ℓ_0/ℓ_1 decoders (P_0)-(P_1). For example, the "informed" versions of OMP and OLS simply consist in initializing the greedy procedures with $\mathcal{Q}$ rather than the empty support. As for the ℓ_0 and ℓ_1 decoders, the following obvious modification of the original decoders were considered:

$$\begin{aligned} \mathscr{D}_{\ell_p} : \mathbb{R}^m &\to \mathbb{R}^n \\ \mathbf{y} &\mapsto \hat{\mathbf{x}} \in \arg\min_{\mathbf{z}} \|\mathbf{z}_{\bar{\mathcal{Q}}}\|_p \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{z}\,, \end{aligned}$$

where $\mathbf{z}_{\bar{\mathcal{Q}}}$ denotes the restriction of $\mathbf{z}$ to its elements *not* in $\mathcal{Q}$.

The recovery conditions obtained in our work can be regarded as generalizations in an informed setup of standard results of the literature. In particular, our results reduce to standard conditions when $\mathcal{Q} = \emptyset$ (that is no side information is provided to the decoder). A general conclusion of our work is as follows: Informed decoders have more favorable recovery conditions than their standard counterpart as soon as $\mathrm{card}(\mathcal{S} \cap \mathcal{Q}) > \mathrm{card}(\mathcal{Q} \backslash \mathcal{S})$, that is the prior support estimate $\mathcal{Q}$ identifies more correct than incorrect atoms.

This conclusion takes a particularly simple form when recovery conditions are expressed in terms of the mutual coherence of $\mathbf{A}$, see (1.31). Letting $k = \mathrm{card}(\mathcal{S})$, $g \triangleq \mathrm{card}(\mathcal{S} \cap \mathcal{Q})$ and $b \triangleq \mathrm{card}(\mathcal{Q} \backslash \mathcal{S})$, we showed that any k-sparse vector $\mathbf{x}$ can be correctly identified by the above "informed" decoders with $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\mathcal{Q}$ as inputs provided that

$$\mu(\mathbf{A}) < \frac{1}{2k - g + b - 1},$$

see [73, Theorems 3 and 5]. This condition has to be compared to (1.32). We see that a more favorable recovery condition is obtained as soon as $g > b$.

Some papers representative of my work on theoretical guarantees for sparse problems:

- C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares", IEEE Transactions on Information Theory, 59(5):3158–3174, May 2013.

  doi.org/10.1109/TIT.2013.2238606

- C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information", IEEE Transactions on Information Theory, 59(11):7509–7524, Nov. 2013.

  doi.org/10.1109/TIT.2013.2278179

- C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay", IEEE Transactions on Information Theory, 62(1):459–470, Jan. 2016.

  doi.org/10.1109/TIT.2015.2490660

- C. Elvira, R. Gribonval, C. Soussen, and C. Herzet, "When does OMP achieve exact recovery with continuous dictionaries?", Applied and Computational Harmonic Analysis, 51:374–413, 2021.

  doi.org/10.1016/j.acha.2020.12.002

1.3 Model-order reduction

In its most widespread formulation, the *"model-order reduction"* (MOR) problem aims to find a low-complexity approximation to the following problem. Given some operator $\mathrm{PDE}\colon \mathcal{X} \times \Theta \to \mathcal{Y}$ and some parameters $\boldsymbol{\theta} \in \Theta$,

$$\text{find } \mathbf{x} \in \mathcal{X} \text{ verifying } \mathrm{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}. \tag{1.33}$$

In the context of MOR, PDE often denotes a differential operator on $\mathcal{X}$ which depends on some parameter $\boldsymbol{\theta}$, see [74, Chapter 2] for many practical examples. In this case, (1.33) amounts to solving a parametric partial differential equation. In many applications, calculating the solution of (1.33) proves to be too expensive in terms of computing resources. The general question addressed by MOR is therefore as follows: is it possible to transform (1.33) into another problem, easier to solve, while keeping some guarantees of accuracy on the solution? The name "model-order reduction" stems from the fact that standard procedures of the literature achieve this goal by reducing the original model (1.33) into a new system with fewer equations and unknowns.

MOR as a reconstruction problem. MOR can be seen as a particular reconstruction problem and be rephrased in the terminology introduced at the beginning of this chapter. The formulation given hereafter is nevertheless rather personal and is not standard in the literature dedicated to MOR.

The set of target signals is here represented by the set of solutions of (1.33) for all values of $\boldsymbol{\theta} \in \Theta$, that is

$$\mathcal{X}_{\text{target}} = \{\mathbf{x} \colon \text{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}, \boldsymbol{\theta} \in \Theta\}. \tag{1.34}$$

It is usually assumed that (1.33) is well-posed, that is there exists a unique solution to (1.33) for each $\boldsymbol{\theta} \in \Theta$. Hereafter, we denote $\mathscr{S} \colon \Theta \to \mathcal{X}_{\text{target}}$ the mapping which associates the solution of (1.33) to each element of Θ and, assuming it exists, by $\mathscr{S}^{-1} \colon \mathcal{X}_{\text{target}} \to \Theta$ its inverse.[12]

Regarding the observation model, the practitioner has two sources of information to reconstruct some $\mathbf{x} \in \mathcal{X}_{\text{target}}$. First, the value of $\boldsymbol{\theta}$ corresponding to $\mathbf{x}$ (that is $\boldsymbol{\theta} = \mathscr{S}^{-1}(\mathbf{x})$) is usually assumed to be an input of the MOR problem. Second, given some $\boldsymbol{\theta} = \mathscr{S}^{-1}(\mathbf{x})$, the target vector $\mathbf{x}$ must obviously verify "$\text{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$". Gathering these two elements leads to the following observation model:

$$\mathbf{y} = \mathscr{M}(\mathbf{x}) \tag{1.35}$$

where

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix}, \quad \mathscr{M}(\mathbf{x}) = \begin{bmatrix} \mathscr{S}^{-1}(\mathbf{x}) \\ \text{PDE}(\mathbf{x}, \mathscr{S}^{-1}(\mathbf{x})) \end{bmatrix}. \tag{1.36}$$

We note that the construction of the observation model can here appear slightly artificial; nevertheless it obeys the definition given at the beginning of this chapter: it provides a mathematical characterization between what is wanted (*i.e.*, some $\mathbf{x} \in \mathcal{X}_{\text{target}}$) and what is available to the practitioner. We will see hereafter that such a formulation is enlightening to recast MOR into the general framework of reconstruction problems.

The target set and the observation model being clearly defined, we will now emphasize that the main challenge of MOR lies in the design of low-complexity decoders with controlled reconstruction performance. Let us first note that, by definition of the problem at stake, given the target set (1.34) and the observation model (1.35)-(1.36), there always exists a decoder, say $\mathscr{D}_{\text{exact}}$, achieving exact reconstruction (simply choose $\mathscr{D}_{\text{exact}}(\mathbf{y}) = \mathbf{x}$ where $\mathbf{x}$ is the unique solution of $\text{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$). However, as mentioned previously, implementing such a decoder may often be computationally too-demanding. Hence, the idea of

[12] If the inverse does not exist (that is some $\mathbf{x} \in \mathcal{X}_{\text{target}}$ is the solution of (1.33) for several $\boldsymbol{\theta} \in \Theta$), the discussion remains valid as long as $\mathscr{S}^{-1} \colon \mathcal{X}_{\text{target}} \to \Theta$ is chosen so that $\mathscr{S}^{-1}(\mathbf{x})$ returns *some* $\boldsymbol{\theta} \in \Theta$ such that $\mathbf{x}$ is the solution of (1.33).

MOR consists in trading some accuracy against computational savings. More specifically, we wish to design some decoder $\mathscr{D}$ with a complexity significantly smaller than $\mathscr{D}_{\text{exact}}$, while verifying[13]

$$\sup_{\mathbf{x}\in\mathcal{X}_{\text{target}}} \operatorname{dist}(\mathbf{x}, \mathscr{D}(\mathscr{M}(\mathbf{x}))) \leq \varepsilon \tag{1.37}$$

for some pre-specified accuracy $\varepsilon \geq 0$ and metric $\operatorname{dist}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. In the construction of reduced-order models, the complexity of the decoder is usually controlled by imposing $\mathscr{D}$ to belong to a prescribed family of functions $\mathcal{D} = \{\mathscr{D}_{\boldsymbol{\zeta}}\colon \mathcal{Y} \to \mathcal{X} : \boldsymbol{\zeta} \in \mathcal{P}\}$.[14] Our MOR problem can then be expressed as the search of some $\mathscr{D} \in \mathcal{D}$ fitting an accuracy criterion, *e.g.*, (1.37).

Some specificities of MOR when viewed as a learning problem. At this stage, it becomes very tempting to draw some connections between MOR and standard "learning" problems. Indeed, letting $\{(\mathbf{x}_l, \mathbf{y}_l)\colon \mathbf{x}_l \in \mathcal{X}_{\text{target}}, \mathbf{y}_l = \mathscr{M}(\mathbf{x}_l)\}_{l=1}^L$ be a set of samples of the target set together with the corresponding observations, the MOR problem can be seen as the search of some $\boldsymbol{\zeta} \in \mathcal{P}$ optimizing a fitting criterion, *e.g.*,

$$\text{find } \boldsymbol{\zeta}^\star \in \arg\min_{\boldsymbol{\zeta}\in\mathcal{P}} \left(\sup_l \operatorname{dist}(\mathbf{x}_l, \mathscr{D}_{\boldsymbol{\zeta}}(\mathbf{y}_l)) \right). \tag{1.38}$$

This problem closely matches the standard formulation of conventional (supervised) learning problems where one wants to fit the parameters of a function relating a *training set* of "input-output" couples $\{(\mathbf{y}_l, \mathbf{x}_l)\}_{l=1}^L$, see *e.g.*, [75]. As a matter of fact, this observation has led an increasing number of researchers (including myself [76, 77]) to draw their inspiration from the machine-learning community to build new MOR procedures in the last few years, see *e.g.*, [78, 79].

The MOR problem has however certain specific features that distinguish it from classical learning problems:

- First, a closed-form (although implicit) definition (1.34) of the target set $\mathcal{X}_{\text{target}}$ is available. This is in contrast to many learning problems where the model underlying the training samples is usually unknown.
- Second, forming large training sets may often be computationally very demanding since accessing one element of $\mathcal{X}_{\text{target}}$ requires to solve problem (1.33).

These peculiarities have given rise to techniques for the design of "good" decoders specifically tuned to MOR problems.

[13] We restrict here our discussion to this "ℓ_∞" accuracy criterion because it is the most widespread in the MOR community. Our discussion can nevertheless be extended to other accuracy measures.

[14] Here, $\boldsymbol{\zeta}$ must be seen as a set of parameters fully characterizing the decoder.

On the one hand, the availability of a closed-form expression of the target set has two main consequences. First, the knowledge of "$\text{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$" is commonly explicitly exploited in the construction of the decoder. For example, a well-known approach (partially) exploiting this information is the so-called *"Petrov-Galerkin projection"* that I shall describe below. Second, the knowledge of a generative model for $\mathcal{X}_{\text{target}}$ is conducive to the derivation of accuracy guarantees on the decoder's performance. As a matter of fact, contrarily to many contributions in machine learning, the derivation of such guarantees has become quasi-systematic in most MOR papers, and plays an important role in the validation of proposed reduction techniques. I shall briefly describe below one of my contributions ([80]) in this field (see item "**Beyond standard Petrov-Galerkin projection: the multi-slice model**").

On the other hand, the problem of forming large data set has given birth to different strategies to circumvent this issue. I elaborate on this topic in the paragraph "**Building good approximation subspaces from partial observations**" where I shall briefly describe two of my contributions ([81, 82]) dealing with the identification of good approximation subspaces for $\mathcal{X}_{\text{target}}$ from partial observations.

Petrov-Galerkin MOR. The idea behind most MOR techniques of the literature is to replace the original reconstruction problem (involving the target set $\mathcal{X}_{\text{target}}$ and observation model $\mathscr{M}$) with a simpler problem, that allows for a decoder with low complexity and "good" reconstruction performance. The most widespread approach to achieve this goal is known as "Petrov-Galerkin projection" and is based on the two ingredients described hereafter.

Target set relaxation: In order to simplify our original reconstruction problem, the target set $\mathcal{X}_{\text{target}}$ can be relaxed into another set $\mathcal{X}_{\text{relax}}$ such that

$$\mathcal{X}_{\text{target}} \subseteq \mathcal{X}_{\text{relax}}. \tag{1.39}$$

In the conception of the reduced-order model, the relaxed set $\mathcal{X}_{\text{relax}}$ then plays the role of a new (surrogate) target set. The idea is that it can be much easier to design a decoder and study its performance with $\mathcal{X}_{\text{relax}}$ than with the original set $\mathcal{X}_{\text{target}}$. A typical choice for $\mathcal{X}_{\text{relax}}$, encountered in many contributions of the literature, is as follows:

$$\mathcal{X}_{\text{relax}} = \{\mathbf{x} \in \mathcal{X} : \text{dist}(\mathbf{x}, V_n) \leq \hat{\epsilon}_n\} \tag{1.40}$$

where V_n is an appropriate n-dimensional subspace of $\mathcal{X}$, $\hat{\epsilon}_n \geq 0$ and

$$\text{dist}(\mathbf{x}, V_n) \triangleq \inf_{\mathbf{z} \in V_n} \text{dist}(\mathbf{x}, \mathbf{z}). \tag{1.41}$$

In this case, $\mathcal{X}_{\text{relax}}$ can be interpreted as the set of elements of $\mathcal{X}$ whose distance to some subspace V_n is no greater than $\hat{\epsilon}_n$. In the sequel, we will refer to (1.40) as an "n-dimensional slice of $\mathcal{X}$".

Subsampling of the observation constraints: Another simplification of our original reconstruction problem can be made at the level of the observation model (1.35)-(1.36). In the context of Petrov-Galerkin MOR, the standard approach consists in only exploiting a finite subset of the infinite-dimensional set of constraints "$\mathrm{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$". This subset typically takes the form

$$\langle \mathbf{m}_j, \mathrm{PDE}(\mathbf{x}, \boldsymbol{\theta}) \rangle_{\mathcal{Y}} = 0 \quad \forall j \in \{1, \dots m\} \tag{1.42}$$

where $\langle \mathbf{m}_j, \cdot \rangle_{\mathcal{Y}}$ denotes a bounded linear form on $\mathcal{Y}$. In the sequel, we will refer to the operator returning the value of the left-hand side of (1.42) $\forall j \in \{1, \dots m\}$ as $\mathrm{PDE}_{\mathrm{relax}}(\mathbf{x}, \boldsymbol{\theta})$.

The relaxed version of the observation model considered in MOR then reads as

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathscr{S}^{-1}(\mathbf{x}) \\ \mathrm{PDE}_{\mathrm{relax}}(\mathbf{x}, \mathscr{S}^{-1}(\mathbf{x})) \end{bmatrix}. \tag{1.43}$$

We note that the second part of the observed vector is always equal to $\mathbf{0}$ and does therefore not change with $\mathbf{x} \in \mathcal{X}_{\mathrm{target}}$. This is the reason why we will express the reduced decoder as a function of $\boldsymbol{\theta}$ only (that is $\mathscr{D} \colon \Theta \to \mathcal{X}$) in the sequel.

Performance limits of a Petrov-Galerkin "reduced" decoder. As mentioned in the previous paragraph, the construction of a "reduced" Petrov-Galerkin decoder implies the degradation of some parts of the information available to the practitioner. First, our "prior" knowledge "$\mathbf{x} \in \mathcal{X}_{\mathrm{target}}$" is relaxed to "$\mathbf{x} \in \mathcal{X}_{\mathrm{relax}}$". Second, the infinite-dimensional set of constraints "$\mathrm{PDE}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$" is reduced to the finite set "$\mathrm{PDE}_{\mathrm{relax}}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$". On top of these two relaxations, the part "$\boldsymbol{\theta} = \mathscr{S}^{-1}(\mathbf{x})$" of the observation model (1.35)-(1.36) is usually never taken explicitly into account in the construction of the decoder since the knowledge of $\mathscr{S}$ (or its inverse $\mathscr{S}^{-1}$) is basically equivalent to solving our original problem (1.33).

These degradations imply that the reconstruction performance of any decoder based on these new sources of information is somehow lower-bounded. In particular, no decoder verifying (1.37) may exist for arbitrarily small ε. A relevant question is therefore as follows: what is the smallest value of ε achievable by a decoder only exploiting the relaxed prior information $\mathcal{X}_{\mathrm{relax}}$ and constraints $\mathrm{PDE}_{\mathrm{relax}}$?

In order to give some insights into this question, let us consider the subset of $\mathcal{X}$ whose elements are compatible with the constraints "$\mathbf{x} \in \mathcal{X}_{\mathrm{relax}}$" and "$\mathrm{PDE}_{\mathrm{relax}}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$", *i.e.*,[15]

$$\mathcal{X}_{\mathrm{post}}(\boldsymbol{\theta}) \triangleq \{\mathbf{z} \in \mathcal{X} \colon \mathrm{PDE}_{\mathrm{relax}}(\mathbf{z}, \boldsymbol{\theta}) = \mathbf{0}\} \cap \mathcal{X}_{\mathrm{relax}}. \tag{1.44}$$

[15] The subscript "post" stands for "posterior" in the sense that $\mathcal{X}_{\mathrm{post}}$ gathers the information provided by the observation model and the prior model $\mathcal{X}_{\mathrm{relax}}$.

Since, from a worst-case perspective, any element of $\mathcal{X}_{\text{post}}(\boldsymbol{\theta})$ can potentially correspond to the target solution, the worst-case optimal decoder $\mathscr{D}_{\text{wc}}\colon \Theta \to \mathcal{X}$ writes as[16]

$$\mathscr{D}_{\text{wc}}(\boldsymbol{\theta}) = \underset{\mathbf{z}' \in \mathcal{X}}{\arg\min} \left(\sup_{\mathbf{z} \in \mathcal{X}_{\text{post}}(\boldsymbol{\theta})} \operatorname{dist}(\mathbf{z}, \mathbf{z}') \right) \tag{1.45}$$

and leads to the following worst-case reconstruction performance:

$$\varepsilon_{\text{wc}}(\boldsymbol{\theta}) = \sup_{\mathbf{z} \in \mathcal{X}_{\text{post}}(\boldsymbol{\theta})} \operatorname{dist}(\mathbf{z}, \mathscr{D}_{\text{wc}}(\boldsymbol{\theta})). \tag{1.46}$$

In view of the definition of $\varepsilon_{\text{wc}}(\boldsymbol{\theta})$, we see that the existence of a decoder verifying (1.37) and only exploiting the constraints "$\mathbf{x} \in \mathcal{X}_{\text{relax}}$"/"$\text{PDE}_{\text{relax}}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}$" is guaranteed provided that

$$\sup_{\boldsymbol{\theta} \in \Theta} \varepsilon_{\text{wc}}(\boldsymbol{\theta}) \leq \varepsilon. \tag{1.47}$$

In particular, if (1.47) holds then decoder (1.45) verifies (1.37).

Affine constraints and n-dimensional slice prior. Although (1.47) gives a precise characterization of the achievable reconstruction performance as a function of $\mathcal{X}_{\text{relax}}$ and $\text{PDE}_{\text{relax}}$, it does not provide any information about the complexity required to attain such a goal.

In practice, tractable (*i.e.*, polynomial-time) complexity can be obtained when $\mathcal{X}_{\text{relax}}$ and $\text{PDE}_{\text{relax}}$ have some desirable mathematical structure. If $\mathcal{X}$ is a Hilbert space and $\operatorname{dist}(\cdot, \cdot) = \|\cdot - \cdot\|_{\mathcal{X}}$, this is for example the case when $\mathcal{X}_{\text{relax}}$ corresponds to an n-dimensional slice of $\mathcal{X}$ and $\text{PDE}_{\text{relax}}(\cdot, \boldsymbol{\theta})$ is an affine operator, that is $\mathcal{X}_{\text{relax}}$ obeys model (1.40) and $\text{PDE}_{\text{relax}}$ is defined as

$$\text{PDE}_{\text{relax}}(\mathbf{x}, \boldsymbol{\theta}) = \begin{bmatrix} \langle \mathbf{w}_1(\boldsymbol{\theta}), \mathbf{x} \rangle_{\mathcal{X}} - b_1(\boldsymbol{\theta}) \\ \vdots \\ \langle \mathbf{w}_m(\boldsymbol{\theta}), \mathbf{x} \rangle_{\mathcal{X}} - b_m(\boldsymbol{\theta}) \end{bmatrix} \tag{1.48}$$

for some $\{ \mathbf{w}_j\colon \Theta \to \mathcal{X} \}_{j=1}^m$, $\{ b_j\colon \Theta \to \mathbb{R} \}_{j=1}^m$. More specifically, when $m = \dim(V_n)$ and the following well-posedness condition is verified[17]

$$\sigma_n \triangleq \inf_{\mathbf{v} \in V_n} \sup_{\mathbf{w} \in W} \frac{\langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{X}}}{\|\mathbf{v}\|_{\mathcal{X}} \|\mathbf{w}\|_{\mathcal{X}}} > 0 \tag{1.49}$$

[16] We again assume that the minimizer exists and is unique to simplify our discussion. We note moreover that the minimization is over $\mathcal{X}$ rather than $\mathcal{X}_{\text{target}}$ because, in the general case, the worst-case optimal estimate does not belong to $\mathcal{X}_{\text{target}}$.

[17] We refer the reader to [83] for more details on the case $m > \dim(V_n)$.

where $W \triangleq \operatorname{span}\Big(\{\mathbf{w}_j(\boldsymbol{\theta})\}_{j=1}^m\Big)$, the worst-case optimal decoder (1.45) takes the following simple form:

$$\mathscr{D}_{\mathrm{wc}}(\boldsymbol{\theta}) = \operatorname*{arg\,min}_{\mathbf{z}\in V_n} \sum_{j=1}^m \Big(\langle \mathbf{w}_j(\boldsymbol{\theta}), \mathbf{z}\rangle_{\mathcal{X}} - b_j(\boldsymbol{\theta})\Big)^2 \tag{1.50}$$

and the corresponding worst-case error reads:

$$\varepsilon_{\mathrm{wc}}(\boldsymbol{\theta}) = \sigma_n^{-1}\hat{\epsilon}_n, \tag{1.51}$$

see [83].

Problem (1.50) is equivalent to solving a system of linear equations (corresponding to the problem's first-order optimality conditions) and can therefore be implemented with a polynomial-time complexity. We also note that (1.50) is equivalent to finding the oblique projection (parallel to the subspace W) onto V_n. It thus justifies (a posteriori) the name "Petrov-Galerkin projection" introduced in the previous paragraph and widely used in the MOR community.

Models (1.40), (1.48) and decoder (1.50) form the core of numerous MOR techniques proposed in the literature during the last decades, see *e.g.*, [74]. In the rest of this section, I shall describe some of my contributions taking place within this context. A first contribution addresses the problem of finding a good approximation subspace V_n when the elements of the target set $\mathcal{X}_{\mathrm{target}}$ are only partially observed, see item "**Building good approximation subspaces from partial observations**". A second contribution deals with the characterization of the worst-case performance of a refined version of decoder (1.50), in which the intersection of several "low-dimensional slices" is exploited as a prior knowledge on the target vector $\mathbf{x}$, see item "**Beyond standard Petrov-Galerkin projection: the multi-slice model**".

On the importance of identifying a good approximation subspace V_n. At this stage, it should become clear that the reconstruction performance achievable by Petrov-Galerkin projection (1.50) depends on the "quality" of the approximation subspace V_n used in the construction of the relaxed set $\mathcal{X}_{\mathrm{relax}}$. In particular, since $\sigma_n \leq 1$ (by Cauchy-Schwarz), we have from (1.51) that attaining a worst-case reconstruction error of ε necessarily requires that

$$\hat{\epsilon}_n \leq \varepsilon. \tag{1.52}$$

Moreover, since $\hat{\epsilon}_n$ can always be chosen as

$$\hat{\epsilon}_n = \sup_{\mathbf{x}\in\mathcal{X}_{\mathrm{target}}} \operatorname{dist}(\mathbf{x}, V_n) \tag{1.53}$$

while verifying the inclusion condition (1.39), (1.52) can be interpreted as follows: the approximation error of any element of $\mathcal{X}_{\text{target}}$ in V_n should be upper bounded by ε. In other words, V_n must be an "ε-good" approximation subspace for $\mathcal{X}_{\text{target}}$.

Many contributions have therefore addressed the problem of finding good approximation subspaces for the elements of $\mathcal{X}_{\text{target}}$. We can for example mention Taylor [84] or Hermite [85] expansions, proper orthogonal decomposition (POD) [86], balanced truncation [87], reduced basis techniques [74], etc.

We note that the smallest worst-case approximation error achievable by some n-dimensional subspace is known as the "Kolmogorov n-width of $\mathcal{X}_{\text{target}}$" and writes:

$$\kappa_n(\mathcal{X}_{\text{target}}) \triangleq \inf_{V:\dim(V)=n} \left(\sup_{\mathbf{x}\in\mathcal{X}_{\text{target}}} \text{dist}(\mathbf{x}, V) \right). \tag{1.54}$$

This metric thus often serves as a gold-standard lower bound on the performance achievable by methods constructing approximation subspaces for $\mathcal{X}_{\text{target}}$. In the next paragraph, I shall emphasize that the Kolmogorov n-width (of a different set) is also of interest to characterize the achievable worst-case performance when only some "partial information" about the target set $\mathcal{X}_{\text{target}}$ is available to the practitioner.

Building good approximation subspaces from partial observations. As mentioned previously, one particular feature of MOR is that the formation of large training sets may be computationally very demanding. This observation has given birth to different strategies to circumvent this issue. A first family of approaches, known as "reduced-basis techniques", constructs (and assesses) the accuracy of a decoder by optimizing some (easily-computable) upper bound on the approximation error, see [74]. Another family of procedures is based on the exploitation of partial observations and/or prior information about the elements of $\mathcal{X}_{\text{target}}$ in the construction of the reduced model, see [88]. Two of my recent contributions focused on the latter approach, see [81, 82].

In [81], we assumed that only the following two sources of information are available to build a proper approximation subspace for $\mathcal{X}_{\text{target}}$:

- *A prior manifold* $\mathcal{X}_{\text{prior}}$, which collects all the knowledge we have "a priori" about $\mathcal{X}_{\text{target}}$. The only constraint we impose on $\mathcal{X}_{\text{prior}}$ is to be such that[18]

$$\mathcal{X}_{\text{target}} \subseteq \mathcal{X}_{\text{prior}}. \tag{1.55}$$

[18] We use the notation $\mathcal{X}_{\text{prior}}$ to avoid any confusion with the set $\mathcal{X}_{\text{relax}}$ used in the construction of the reduced model. Both $\mathcal{X}_{\text{prior}}$ and $\mathcal{X}_{\text{relax}}$ nevertheless play the same role in different contexts and have in common that they must contain the target set $\mathcal{X}_{\text{target}}$, see (1.39) and (1.55)

- *A set of partial observations of the elements of* $\mathcal{X}_{\text{target}}$: we assume that we collect, $\forall \mathbf{x} \in \mathcal{X}_{\text{target}}$, a set of noiseless linear measurements:

$$\{\langle \mathbf{m}_j, \mathbf{x} \rangle_{\mathcal{X}}\}_{j=1}^m \tag{1.56}$$

  where $\{\mathbf{m}_j\}_{j=1}^m$ denotes m orthonormal elements of a Hilbert space $\mathcal{X}$.

On the one hand, the prior information typically derives from some physical considerations and/or constraints we may have about the system under study. An example of construction of $\mathcal{X}_{\text{prior}}$ is given in [81, Section 4]. On the other hand, the nature of the observations available in practice depends on the experimental setup. In our work, we made the assumption that the measurements can be seen as the outputs of some noiseless linear operator to simplify our analysis.

The main question addressed in [81] is then as follows: given (1.55) and (1.56) what is the best worst-case approximation performance achievable when projecting the element of the (unknown) target set $\mathcal{X}_{\text{target}}$ onto an n-dimensional subspace V_n? We showed that the answer to this question is intimately related to the Kolmogorov n-width of the following set

$$\mathcal{X}_{\text{post}} \triangleq \cup_{\mathbf{x} \in \mathcal{X}_{\text{target}}} \left\{ \mathbf{z} \in \mathcal{X} \colon \langle \mathbf{m}_j, \mathbf{z} \rangle_{\mathcal{X}} = \langle \mathbf{m}_j, \mathbf{x} \rangle_{\mathcal{X}} \ \forall j \in \{1, \ldots, m\} \right\} \cap \mathcal{X}_{\text{prior}}, \tag{1.57}$$

and the worst-case optimal n-dimensional approximation subspace for the set of target sets compatible with (1.55) and (1.56) is given by

$$V_n \in \underset{V : \dim(V) = n}{\arg\min} \left(\sup_{\mathbf{z} \in \mathcal{X}_{\text{post}}} \operatorname{dist}(\mathbf{z}, V) \right), \tag{1.58}$$

see [81, Section 3]. This allowed to propose a numerical strategy to try and identify this approximation subspace, see [81, Section 4]. We also derived a theoretical analysis relating, in some simplified setup, an upper bound on $\kappa_n(\mathcal{X}_{\text{post}})$ to the parameters of the prior and observation models (1.55) and (1.56), see [81, Section 6].

In [82], we addressed the same kind of question in a Bayesian setup. In this context, both the prior and observation models (1.55)-(1.56) took a probabilistic form and we derived a methodology based on the sampling of the corresponding posterior probability to construct reduced models combining these two sources of information.

Beyond standard Petrov-Galerkin projection: the multi-slice model. As mentioned previously, "Petrov-Galerkin projection" is the most standard approach to construct reduced-order models. The reasons of its popularity lie both in its simplicity of implementation and in our good understanding of its theoretical performance.

Unfortunately, Petrov-Galerkin projection may sometimes fall short in providing sufficiently accurate reconstruction performance. For example, in the affine/linear case described in

item "**Affine constraints and n-dimensional slice prior**", the worst-case performance attained by Petrov-Galerkin projection is given in (1.51). Now, the factors $\hat{\epsilon}_n$ and σ_n^{-1} appearing in the latter expression have antagonistic behaviors: $\hat{\epsilon}_n$ is a non-increasing function of n whereas σ_n^{-1} is a non-decreasing function of n. As a consequence, even if the choice of the approximation subspace V_n and its dimension n are degrees of freedom to the practitioners, there exist setups where the worst-case error $\sigma_n^{-1}\hat{\epsilon}_n$ may remain large for any choice of n and V_n.

In [80], we studied a possible remedy to this problem by considering a refined definition of the relaxed target set $\mathcal{X}_{\text{relax}}$. Following [83], we considered the following setup. We assumed that $\mathcal{X}$ is a Hilbert space[19] and

$$\mathcal{X}_{\text{relax}} = \cap_{k=0}^{n} \{\mathbf{x} \in \mathcal{X} : \text{dist}(\mathbf{x}, V_k) \leq \hat{\epsilon}_k\} \tag{1.59}$$

where $\hat{\epsilon}_k \geq 0$ and

$$V_0 \subset V_1 \subset \ldots \subset V_n \quad \text{with } \dim(V_k) = k. \tag{1.60}$$

Model (1.59) thus corresponds to the intersection of $n+1$ low-dimensional slices. This is in contrast with the "single-slice" relaxed model (1.40) considered in standard Petrov-Galerkin projection where all the elements of the target set were assumed to be $\hat{\epsilon}_n$-close to some n-dimensional subspace V_n.

We note that model (1.59)-(1.60) is relevant in practice since many procedures of the literature designed to construct "good" n-dimensional approximation subspaces for $\mathcal{X}_{\text{target}}$ proceed sequentially and return a set of approximation subspaces of increasing dimension $\{V_k\}_{k=0}^n$ together with some (possibly approximate) upper bounds $\hat{\epsilon}_k \geq \sup_{\mathbf{x}\in\mathcal{X}_{\text{target}}} \text{dist}(\mathbf{x}, V_k)$, see [74, 86].

Unfortunately, even in the case where the observation operator considered in the construction of the reduced model is affine, the worst-case optimal decoder (1.50) associated to model (1.59) does not have an easily-computable closed-form expression. In [80], we thus considered the following "multi-slice" variation of Petrov-Galerkin projection:

$$\begin{aligned} \mathscr{D}(\boldsymbol{\theta}) = \underset{\mathbf{z}\in V_n}{\arg\min} & \sum_{j=1}^{m} \Big(\langle \mathbf{w}_j(\boldsymbol{\theta}), \mathbf{z} \rangle_{\mathcal{X}} - b_j(\boldsymbol{\theta}) \Big)^2 \\ & \text{subject to } \text{dist}(\mathbf{z}, V_k) \leq \hat{\epsilon}_k \quad \forall k \in \{0, \ldots, n\}. \end{aligned} \tag{1.61}$$

This decoder has the following desirable features: *i)* it exploits all the available prior information (1.59); *ii)* its output is constrained to belong V_n so that only n-dimensional quantities have to be handled in this resolution; *iii)* it can be cast as a quadratic problem with

[19] As previously, the considered metric is then defined as $\text{dist}(\cdot,\cdot) = \|\cdot - \cdot\|_{\mathcal{X}}$.

quadratic (convex) inequality constraints and can therefore be solved to any arbitrary precision with polynomial-time complexity (*e.g.*, with gradient-projected method [89, Chapter 2]).

In [80], we provided a theoretical characterization of the worst-case performance achievable by this decoder. More specifically, we derived a computable upper-bound on the worst-case performance of $\mathscr{D}$ only depending on the problem's main ingredients, see [80, Theorem 2]. Particularizing this result to different setups, we showed that the multi-slice decoder (1.61) can outperform standard Petrov-Galerkin projection by order of magnitude in some cases.

My main contributions in the domain of model-order reduction are:

- C. Herzet, P. Héas, and A. Drémeau, "Model reduction from partial observations", International Journal for Numerical Methods in Engineering, 113(3), 2018.

  doi.org/10.1002/nme.5623

- P. Héas and C. Herzet, "Reduced modeling of unknown trajectories", Archives of Computational Methods in Engineering, 25(1):87–101, 2018.

  doi.org/10.1007/s11831-017-9229-0

- P. Héas and C. Herzet, "Low-rank dynamic mode decomposition: An exact and tractable solution", Journal of Nonlinear Science, 32(1):8, 2021.

  doi.org/10.1007/s00332-021-09770-w

- C. Herzet and M. Diallo, "Performance guarantees for a variational multi-space decoder", Advances in Computational Mathematics, 46(1):10, 2020.

  doi.org/10.1007/s10444-020-09746-6

Chapter 2

Perspectives

In line with of my contributions presented in Chapter 1, my future research will aim to push the frontiers of knowledge of the general reconstruction problem stated at the beginning of my research summary. The three pillars of my previous work, namely *"Models"* - *"Algorithms"* - *"Theoretical Guarantees"*, will remain central to my research themes. I describe below several directions that I wish to pursue in the coming years.

2.1 Low-dimensional invariance-preserving models

One crucial element in the design of efficient and effective reconstruction methods is the identification of "low-dimensional" models accurately describing the signals of interest. From an "effectiveness" point of view, this assertion has for example been mathematically quantified in the field of machine learning (*e.g.*, the celebrated "no-free-lunch theorem" [90]) or in the context of generic reconstruction problems [12]. From an "efficiency" point of view, the choice of an appropriate low-dimensional model generally enables to derive resource-saving reconstruction methods (*e.g.*, computation time/memory space). The "low rank" linear model, ubiquitous in all fields of Sciences, is probably the most basic example of low-dimensional model but other, more refined, options (*e.g.*, kernel methods, sparse models, tensor decompositions, neural networks) have also been successfully considered during the last decades, see [11, 91–93].

In practice, the choice of an appropriate model generally requires answering the following questions:[1]

i) Which family of models should I consider? In particular, what is the type of models (linear, sparse, tensorial, etc.) and the dimensionality best suited to the reconstruc-

[1] Although closely related, these questions are often addressed separately in the literature.

tion task at hand?

ii) In a given family of models, how to select the "best" model according to some accuracy criterion?

The search for relevant answers to these questions has led to the publication of thousands of articles in many scientific fields.

It should be noted that the inclusion of any information about the target set of signals in the selection process generally improves the accuracy of both the model and the resulting reconstruction procedure. In particular, many signals of interest verify some invariance or symmetry properties (*i.e.*, some features of the signals remain unchanged under specific transformations). The exploitation of such invariances is for example omnipresent in statistical learning or artificial intelligence where classification methods are often required be robust to translations, rotations or changes of scale. The existence of symmetries and invariances is also a central element in many fields of Sciences (*e.g.*, physics or chemistry) and it can be crucial to exploit them when designing reconstruction methods, see for example [94].

In the line of the work I recently initiated with Frédéric Champagnat (see *e.g.*, [95]), a research direction that I will pursue in the next few years is the exploitation of these invariances (in particular their preservation) in the construction of low-dimensional models. The objects under study are here manifolds (embedded in some Hilbert space) whose elements are continuously indexed by a small number of parameters. We consider the following paradigm: the inner product between two elements of the considered manifold is invariant to some transformations of the parameter space. For example, the set of solutions of some parametric differential equations [74] or some "continuous" dictionaries appearing in sparse representation problems [96–98] are instances of such objects. Our goal is to approach these manifolds by "simple" low-dimensional models while preserving the invariances of the inner product. In our current work and short-term prospects, we will focus on approximations of the manifold in low-dimensional *linear subspaces* preserving the invariance of the inner product with respect to *translation* of the parameter space. In the longer term, our goal is to extend this topic of research to other approximation models and different forms of invariances (rotation, change of scale, etc).

2.2 Generalized screening tests for frugal methods

In the last decades, the increase of both the volume of available data and the size of the problems to be addressed has stressed the need for low-complexity reconstruction methods. The importance of this problem has for example been highlighted in recent years in the field of artificial intelligence, where researchers have been working on the development of so-called "frugal" methods, *i.e.*, procedures requiring less data and computing power [99].

It should be noted that many reconstruction methods of the literature are based on solving optimization problems. A natural way to improve their complexity is therefore to look for numerical procedures achieving the desired solving accuracy at the lowest possible computational cost. Since most standard optimization techniques are iterative, this complexity reduction is generally based on the following two axes:

i) the increase of the *speed of convergence*,

ii) the decrease of the *computational cost per iteration*.

The first axis has been widely explored in the optimization community. Precise bounds on the achievable (worst-case) convergence speeds are known for many families of problems and numerical methods attaining these bounds have been proposed, see *e.g.*, [100]. It should be noted that these convergence rates depend on the regularity of the considered problem, a greater regularity generally leading to better speeds of convergence.

A rather recent technique focusing on the second axis, *i.e.*, the reduction of the complexity per iteration, is the so-called "safe screening" method presented in Chapter 1 of my research summary. In its original version, screening takes place in the context of sparse representations and aims to identify the zero elements in the minimizers of the target optimization problem. Other fields of application were subsequently proposed such as the identification of saturated components in a regularized problem involving an ℓ_∞ penalty [60, 61] or the pruning of the data set in some machine learning problems [101].

It should be noted that, although state-of-the-art techniques as "GAP screening" [46] lead to impressive computational gains, these methods can only be applied to certain families of problems, namely convex problems whose dual obeys some strong-convexity property. In the coming years, another axis of my research will therefore focus on the extension of the principles of screening to more general families of optimization problems. Several directions of research are envisioned here.

First, following the work by Dantas *et al.* [102, 103], a short-term objective will consist in deriving GAP-like[2] screening procedures that are applicable when the strong-convexity constraint on the dual problem is not respected. I recently made (with Clément Elvira and Phuong Dang) a first contribution [104] towards this goal. In this work, we proposed a novel "dual" perspective on GAP procedures which offers a new paradigm for the conception of screening methods. A short-term objective of my future work will therefore consist in pushing forward these developments to derive new GAP-like screening methods applicable to *generic* convex optimization problems.

Another short-term research axis will focus on the extension of existing screening techniques to *"semi-infinite programming"* problems [105], in which there is an infinite uncountable number of constraints. The main focus will be here on the field of sparse representations

[2] That is procedures provably screening all the zeros of the solution in non-degenerated cases.

in continuous dictionaries [96,97] where the dual of the "Beurling Lasso" problem [58,59] is known to be a semi-infinite program.

In the line of [106], one of my medium-term research axes will target the extension of screening principles to *non-convex* optimization problems. The goal will be here twofold: on top of the complexity reduction allowed by screening methods, the proposed method will also be used to certify (from a global-optimality perspective) some of the decisions made by the optimization procedures. In other words, one expected output of this line of search will be the quantification (to some extent) of the uncertainty on the value of the problem's minimizers.

Yet another direction of research aiming at speeding up the resolution of optimization problems is the identification (or the tightening) of some *inactive* constraints. We note that this approach is in contrast with standard screening procedures which rather aim to identify *active* constraints of the problem to reduce the dimensions of the optimization domain. As a motivating example for this "dual" approach, I mention one of my recent contribution [107] with Théo Guyard and Clément Elvira. In this work, we investigate how the identification of inactive constraints can be exploited to improve the "regularity" of the target problem (and therefore the speed of convergence of the solving procedures). More generally, in my future work I intend to investigate how the dynamic[3] refinement of the problem's feasible set can decrease the computational cost needed to attain some prescribed solving accuracy.

Finally, one of my medium-to-longer-term perspective is to extend the principles of screening methods beyond the scope of optimization problems. One target application is the estimation of rare-event probabilities where one wishes to identify the values of some random variable for which some score function exceeds a given threshold. The central question will be here as follows: can we benefit from screening techniques to identify subsets of values provably not exceeding the target threshold and leverage this knowledge to improve the complexity of the sampling procedures used to estimate the rare-event probability?

2.3 Theoretical guarantees for reconstruction problems

A last direction of research that I will pursue in the coming years is the theoretical analysis of the performance achievable by reconstruction methods. This type of studies enable to provide clear answers to the question of the fundamental performance limits attainable in certain application contexts. At a more operational level, they also make it possible to certify the quality of the estimates produced by practical reconstruction methods and characterize their inherent uncertainty.

In line with my recent works in this area [70,80,81], I intend to address the following ques-

[3] For example, through the iterations of a solving procedure.

tions over the next few years. A first question of interest will be the "complexity-accuracy" trade-off that can be attained in some reconstruction problems. While the majority of contributions in the literature address the characterization of the achievable accuracy as a function of the number of observations (see *e.g.*, [11] in the field of sparse representations), several recent contributions focus on the integration of complexity constraints in the characterization of the performance achievable by some reconstruction method, [108–110]. The field of model-order reduction, in which complexity constraints play a predominant role, will be of particular interest to me. The questions addressed here will be at the interface of "standard" reconstruction guarantees (for which the underlying question is of the form "Under which conditions can we recover some quantities of interest up to some precision?") and optimization techniques (the relevant question being here: "How does the solving accuracy of an optimization procedure evolve with the number of iterations performed?" [100]).

A second research theme will be to study how the uncertainty inherent to certain reconstruction methods can impact some other higher-level tasks. An example of interest, in line with the research field of the SIMSMART team, will be the estimation of the probability of rare events involving certain physical, chemical or biological phenomena. In this particular context, the implementation of Monte-Carlo methods generally requires to access the solution of a system of differential equations at each iteration of the process. The evaluation of this solution is often numerically expensive and the implementation of Monte-Carlo methods is then compromised. A family of approaches recently proposed in the literature (see *e.g.*, [111, 112]) consists in approximating the solution of the system of differential equations by the output of a reduced model during certain steps of the process. In the line if these works, part of my research activity will consist in the implementation and the study of reduced models for rare-event simulation. In particular, a special attention will be paid to the following topics: *i)* the design of error bounds on the estimates of the reduced model that can be efficiently computed during the Monte-Carlo process; *ii)* the study of the impact of the error induced by the reduced model on the quality of the final estimate of the probability of the rare event.

Chapter 3

Curriculum vitae

Professional Background

2007- 2022 Researcher at INRIA, Rennes, France.
My research focuses on several aspects of signal processing and statistics including variational inference techniques, optimization techniques, sparse and non-negative representations in redundant dictionaries, model-order reduction.

2016- 2018 Invited Researcher at IMT Atlantique, Brest, France.
During my stay at IMT Atlantique, I focused on the identification, the analysis and the interpretation of oceanic phenomena by using multi-sensors satellite images.

2006-2007 Fulbright Postdoctoral Stay at UC Berkeley, USA.
I focused on the design of efficient message-passing algorithms for joint source-channel decoding.

2006 Research Assistant at ENS Cachan, Paris, France.
I focused on the computation of estimation lower bounds for time-varying parameters.

2002-2006 Research Assistant at UCLouvain, Louvain-la-Neuve, Belgium.
I worked in the Communications and Remote Sensing Laboratory in collaboration with Professor Luc Vandendorpe. I gave lectures in Electromagnetic and Digital Communications to engineering students.

2001 Research Engineer at UCLouvain, Louvain-la-Neuve, Belgium.
I designed VDSL systems in collaboration with Alcatell Bell (Antwerpen Belgium).

Educational Background

2002-2006 PhD. in Applied Sciences, UCLouvain, Louvain-la-Neuve, Belgium.
Thesis supervised by Pr. Luc Vandendorpe: *Turbo Synchronization for Digital Burst Communications.*

2001-2003 DEA in Applied Sciences, UCLouvain, Louvain-la-Neuve, Belgium.
Final grade: *summa cum laude*

1996-2001 Electrical Engineer, UCLouvain, Louvain-la-Neuve, Belgium.
Final grade: *magna cum laude*

Teaching

I describe below the different classes I have taught during my professional career. The number of hours mentioned for each course represents my contribution to the class *per year*.

2021-2022 Signal Processing, Master 1, ENS Rennes (15h).
The goal of this course (given in collaboration with C. Elvira and H. P. Dang) is to introduce the main concepts of signal processing. This course is given to approximately 10 students.

2020-2022 Methodological Project, Master 2, ENSAI (15h).
I follow up methodological projects of ENSAI students in their final year.

2019-2021 Regularization and Model Selection, Master 2, ENSAI (20h).
The goal of this course (given in collaboration with C. Elvira) is to introduce the basic theory of regularization and model selection in inverse problems. This course is given to approximately 50-100 students.

2014-2021 Sparse Representations & Smart Sensing, Master 2, ENSAI (15h).
The goal of this course (given in collaboration with J. Cohen, C. Elvira and A. Roumy) is to introduce the basic theory of sparse representations and compressive sensing. This course is given to approximately 10 students.

2014-2021 Sparse Representations & Smart Sensing, Master 2, INSA (15h).
The goal of this course (given in collaboration with N. Bertin, J. Cohen, C. Elvira, R. Gribonval and A. Roumy) is to introduce the basic theory of sparse representations and compressive sensing. This course is given to approximately 15 students. I am responsible for the organization of the "Sparse Representations and Smart Sensing" module at the INSA since 2019.

2012-2013 Statistics for Data Analysis, Master 1, Université de Rennes (15h).
The goal of this course, coordinated by Valérie Monbet, was to teach the students the basic notions of statistics and their use to analyze high-dimensional data sets. This course was given to approximately 30 students.

2009-2011 Optimization, Master 1, Université de Rennes (15h).
The goal of this course, coordinated by Jean-Jacques Fuchs, was to teach the students the basic notions of optimization: first and second-order optimality conditions, main descent algorithms, etc. This course was given to approximately 20 students.

Miscellaneous

As described below, I have been involved in different scientific projects and events as a participant or a leader. The funding of the projects are indicated when I was the leader.

2021 Steering Committee, iTwist'23.
iTwist is an international conference which takes place every two years and gathers around 100 international researchers and PhD students of the signal processing and optimization communities.

2020-2024 Project Partner, ANR Project MELODY.
Title: *Bridging Geophysics and Machine Learning for the Modeling, Simulation and Reconstruction of Ocean Dynamics.*

2020 Organizing Committee, iTwist'20 (https://itwist20.ls2n.fr/).
iTwist is an international conference which takes place every two years and gathers around 100 international researchers and PhD students of the signal processing and optimization communities.

2016-2020 Workpackage Leader, ANR Project BECOSE.
Title: *Au delà de l'échantillonnage compressé : algorithmes d'approximation parcimonieuse pour les problèmes inverses mal conditionnés.*

2018-2019 Project Leader, action exploratoire du Labex CominLabs.
Title: *Reconciling Sparsity and Parametric Models: Sparse Representations in Continuous Dictionaries.*
Funding : $\sim$ 20 *keuros*

2014-2018 Project Partner, CominLabs Project SEACS.
Title: *Stochastic Model-Data-Coupled Representations for the analysis, simulation and reconstruction of upper ocean dynamics.*

2014-2018 Project Leader, ANR Project JCJC GERONIMO.
Title: *Advanced Geophysical Reduced-Order Model Construction from Image Observations.*
Funding : $\sim$ 140 *keuros*

2013-2014 Project Leader, INSU-LEFE Project.
Title: *Vers de nouvelles méthodes d'estimation de la sous-mésoéchelle océanique.*
Funding : $\sim$ 20 *keuros*

2012 Organizing Committee, Journées thématiques Apprentissage et Parcimonie.

Supervision: Master Theses, PhD's and Postdocs

2020-2021 Master thesis, Théo Guyard, 50%.
Title: *Screening Methods for ℓ_0 Sparse Representations in Continuous Dictionaries*, Co-supervisor: Clément Elvira.

2019-2020 Master thesis, Le Thu Tran, 50%.
Title: *Sparse Representation of Curves*, Co-supervisor: Valérie Monbet.
Le is currently doing a PhD under our supervision.

2016-2017 Master thesis, Said Ouala, 50%.
Title: *Deep learning and data assimilation*, Co-supervisor: Ronan Fablet.
Following his internship, Said completed a PhD under Ronan Fablet's supervision (IMT Atlantique).

2016-2017 Master thesis, Soufiane Ait Tilat, 50%.
Title: *Détection/localisation des sources par méthodes parcimonieuses et dictionnaires continus*, Co-supervisor: Frédéric Champagnat.
Following his internship, Soufiane completed a PhD under our supervision.

2021-2024 PhD. thesis, Théo Guyard, 25%.
Tentative title: *Représentations parcimonieuses dans dictionnaires continus pour le diagnostic automatique de maladies du foie*, HdR supervisor: James Ledoux.

2020-2023 PhD. thesis, Le Tran Thu, 25%.
Tentative title: *Sparse representations in continuous dictionaries. Application to spectrometry data*, co-HdR supervisor: Valérie Monbet.

2018-2021 PhD. thesis, Milan Courcoux-Carro, 25%.
Title: *Conception optimisée d'antenne pour la localisation passive de sources acoustiques*, HdR supervisor: Alexandre Baussard.

2017-2020 PhD. thesis, Soufiane Ait Tilat, 50%.
Title: *Détection et localisation de particules dans des images PIV via des approches parcimonieuses avec grille*, HdR supervisor: Frédéric Champagnat.
Soufiane is currently working as a statistician in a start-up.

2010-2014 PhD. thesis, Ioana Barbu, 75%.
Title: *Estimation tridimensionnelle de vitesse de fluides turbulents*, HdR supervisor: Etienne Mémin.
Ioana is currently "Innovation Project Officer" at Inria.

2016-2017 Postdoctoral fellow, Hassan Maatouk.
Funding: *ANR project GERONIMO.*

2016-2017 Postdoctoral fellow, Mamadou Lamarana Diallo.
Funding: *ANR project GERONIMO.*

2017-2019 Postdoctoral fellow, Clément Dorffer.
Funding: *DGA project.*

2017-2019 Postdoctoral fellow, Clément Elvira.
Funding: *ANR project BECOSE.*

Publications

During my career I have published scientific contributions in national and international journals and conferences. My list of publications is as follows: 1 book chapter, 26 articles in international journals, 77 publications in the proceedings of international conferences, 7

publications in the proceedings of national conferences. The bibliographical references of my work are detailed below.

The full text of most of my contributions is available on my personal webpage at:

http://people.rennes.inria.fr/Cedric.Herzet/

Book Chapter

1. Harold Sneessens, Luc Vandendorpe, Cédric Herzet, Xavier Wautelet, Onur Oguz, "MIMO: from Theory to Implementation", Elsevier, 2011.

International Journals

1. Milan Courcoux-Caro, Charles Vanwynsberghe, Cédric Herzet, Alexandre Baussard, "Sequential sensor selection for the localization of acoustic sources by sparse Bayesian learning", Journal of the Acoustical Society of America, Acoustical Society of America, 2022, 152 (3), pp.1695-1708
2. Patrick Héas and Cédric Herzet, "Low-rank dynamic mode decomposition: An exact and tractable solution", in Journal of Nonlinear Science, 32(1):8, 2021.
3. Clément Elvira, Rémi Gribonval, Charles Soussen, Cédric Herzet, "When does OMP achieve exact recovery with continuous dictionaries?", Applied and Computational Harmonic Analysis, Volume 51, 2021.
4. Clément Elvira and Cédric Herzet, "Safe Squeezing for Antisparse Coding", IEEE Transactions on Signal Processing, vol. 68, pp. 3252-3265, 2020.
5. Cédric Herzet, Mammadou Diallo, "Performance guarantees for a variational "multi-space" decoder", Advances in Computational Mathematics 46, 10, 2020.
6. Cédric Herzet, Clément Dorffer, Angélique Drémeau, "Gather and Conquer: Region-Based Strategies to Accelerate Safe Screening Tests", in IEEE Transactions on Signal Processing, vol. 67, no. 12, pp. 3300-3315, 15 June 15, 2019.
7. Said Ouala, Ronan Fablet, Cédric Herzet et al., "Neural Network Based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-Derived Sea Surface Temperature", Remote Sens., MDPI, 2018.
8. Cédric Herzet, Patrick Héas, Angélique Drémeau, "Model Reduction from Partial Observations", Int. J. Numer. Meth. Engng, 2018.
9. Patrick Héas, Cédric Herzet, "Reduced Modeling of Unknown Trajectories", Archives of Computational Methods in Engineering, pp 1 - 15, 2018.

10. Ioana Barbu, Cédric Herzet, “A New Approach for Volume Reconstruction in TomoPIV with the Alternating Direction Method of Multipliers”, IOP Measurement Science and Technology, vol. 27, nr 19, pp 104002, 2016.

11. Patrick Héas, Angélique Drémeau, Cédric Herzet, “An Efficient Algorithm for Video Superresolution Based on a Sequential Model”, SIAM J. Imaging Sci. 9-2 (2016), pp. 537-572.

12. Cédric Herzet, Angélique Drémeau, Charles Soussens, “Relaxed Recovery Conditions for OMP/OLS by Exploiting Both Coherence and Decay”, IEEE Trans. on Information Theory, vol. 62, pp. 459 - 470, January 2016.

13. Cédric Herzet, Charles Soussens, Jérome Idier, Rémi Gribonval, “Exact Recovery Conditions for Sparse Representations With Partial Support Information”, IEEE Trans. on Information Theory, vol. 59, pp. 7509 - 7524, November 2013.

14. Patrick Héas, Cédric Herzet, Etienne Mémin, Dominique Heitz, Pablo D. Mininni, “Bayesian estimation of turbulent motion”, IEEE Trans. on Pattern Analysis And Machine Learning, vol. 35, pp. 1343-1356, June 2013.

15. Charles Soussens, Rémi Gribonval, Jérome Idier, Cédric Herzet, “Joint K-Step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares”, IEEE Trans. on Information Theory, vol. 59, pp. 3158-3174, May 2013.

16. Pierre Dérian, Patrick Héas, Cédric Herzet and Etienne Mémin, “Wavelets and optic flow motion estimation”, Numerical Mathematics: Theory, Methods and Applications, vol. 6, pp. 116-137, 2013.

17. Patrick Héas, Cédric Herzet, Etienne Mémin, “Bayesian inference of models and hyperparameters for robust optic-flow estimation”, IEEE Trans. on Image Processing, vol 21, pp. 1437-1451, April 2012.

18. Angélique Drémeau, Cédric Herzet and Laurent Daudet, “Boltzmann machine and mean-field approximation for structured sparse decompositions”, IEEE Trans. on Signal Processing, vol 60, pp. 3425-3438, July 2012.

19. Cédric Herzet, Kampol Woradit, Henk Wymeersch, Luc Vandendorpe, “Low-complexity MAP Hypothesis Testing by Bethe Free Energy Minimization”, IEEE Trans. on Signal Processing, vol. 58, pp. 6238-6250, Dec 2010. (Best paper award of the Newcom++ European network of excellence)

20. Cédric Herzet, Henk Wymeersch, Frederik Simoens, Marc Moeneclaey and Luc Vandendorpe, “MAP-Based Code-Aided Hypothesis Testing" , IEEE Transactions on Wireless Communications, vol. 7, pp. 2856-2860, August 2008.

21. Cédric Herzet, Valéry Ramon, Alexandre Renaux and Luc Vandendorpe, "A Cramer-Rao Bound Characterization of the EM-Algorithm Mean Speed of Convergence", IEEE Transactions on Signal Processing, vol. 56, pp. 2218-2228, June 2008.

22. Stéphanie Bay, Cédric Herzet, Jean-Marc Brossier, Jean-Pierre Barbot and Benoit Geller, "Analytic and Asymptotic Analysis of the Bayesian Cramer-Rao Bound for Dynamical Phase Offset Estimation", IEEE Transactions on Signal Processing, vol. 56, pp. 61-70, Jan. 2008.

23. Cédric Herzet, Henk Wymeersch, Marc Moeneclaey and Luc Vandendorpe, "On Maximum-Likelihood Timing Estimation", IEEE Transactions on Communications, vol. 55, pp. 1116-1119, June 2007.

24. Cédric Herzet, Henk Wymeersch, Vincenzo Lottici, Nele Noels, Marco Luise, Marc Moeneclaey and Luc Vandendorpe, "Code-aided Turbo Synchronization", The Proceedings of the IEEE, vol. 95, pp. 1255-1271, June 2007.

25. Cédric Herzet, Valéry Ramon and Luc Vandendorpe, "A Theoretical Framework for Iterative Synchronization based on the Sum-Product and the Expectation-Maximization Algorithms", IEEE Transactions on Signal Processing, vol. 55, pp. 1644-1658, May 2007.

26. Xavier Wautelet, Cédric Herzet, Antoine Dejonghe, Jérome Louveaux and Luc Vandendorpe, "Comparison of EM-based algorithms for MIMO channel estimation", IEEE Transactions on Communications, vol. 55, pp. 216-226, Jan. 2007.

27. Valéry Ramon, Cédric Herzet and Luc Vandendorpe, "A Semi-analytical Prediction Method of Performance and Convergence Behavior of Multiuser Turbo-equalization / Demapping Scheme", IEEE Transactions on Signal Processing, vol. 55, pp. 1104-1117, March 2007.

International Conferences

1. Cédric Herzet, Clément Elvira, and Hong-Phuong Dang, "Region-free safe screening tests for l1-penalized convex problems", in EURASIP European Signal Processing Conference, EUSIPCO'22, 2022.

2. Thu-Le Tran, Clément Elvira, Hong-Phuong Dang, and Cédric Herzet, "Beyond GAP screening for Lasso by exploiting new dual cutting half-spaces", in EURASIP European Signal Processing Conference, EUSIPCO'22, 2022.

3. Julyan Arbel, Hong-Phuong Dang, Clément Elvira, Cédric Herzet, Zacharie Naulet and Mariia Vladimirova, "Bayes in action in deep learning and dictionary learning", ESAIM: Proc. and Surveys, 2022.

4. Théo Guyard, Cédric Herzet, Clément Elvira, "Node-Screening Tests For The ℓ_0-Penalized Least-Squares Problem", ICASSP'2022, IEEE International Conference on Acoustics, Speech, and Signal Processing, Singapur, May 2022.

5. Théo Guyard, Cédric Herzet, Clément Elvira, "Screen & Relax: Accelerating the Resolution of Elastic-Net by Safe Identification of the Solution Support", ICASSP'2022, IEEE International Conference on Acoustics, Speech, and Signal Processing, Singapur, May 2022.

6. Clément Elvira, Jérémy Cohen, Cédric Herzet, Rémi Gribonval, "Continuous dictionaries meet low-rank tensor approximations", iTwist 2020, International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques, Nantes, France, 2020

7. Frédéric Champagnat, Cédric Herzet, "Interpolating and translation-invariant approximations of parametric dictionaries", EUSIPCO'2020, European Signal Processing Conference, Amsterdam, The Nederlands, August 2020.

8. Milan Courcoux-Caro, Charles Vanwynsberghe, Cédric Herzet, Alexandre Baussard, "Sequential sensor placement using Bayesian compressed sensing for direction of arrival estimation", EUSIPCO'2020, European Signal Processing Conference, Amsterdam, The Nederlands, August 2020.

9. Clément Elvira, Cédric Herzet, "Short and squeezed: accelerating the computation of antisparse representations with safe squeezing", ICASSP'2020, IEEE International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, May 2020.

10. Patrick Héas, Cédric Herzet, Benoit Combès, "Generalized Kernel-Based Dynamic Mode Decomposition", ICASSP'2020, IEEE International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, May 2020.

11. Clément Dorffer, Cédric Herzet, Angélique Drémeau, "Region-based relaxations to accelerate greedy approaches", EUSIPCO'2019, European Signal Processing Conference, Coruna, Spain, September 2019.

12. Said Ouala, Duong Nguyen, Lucas Drumetz, Cédric Herzet, Ananda Pascual, et al, "Learning ocean dynamical priors from noisy data using assimilation-derived neural nets", IGARSS 2019 - International Geoscience and remote Sensing Symposium, Yokohama, Japan, July 2019.

13. Said Ouala, Ronan Fablet, Cedric Herzet, Lucas Drumetz, Bertrand Chapron, et al, "Sea surface dynamics reconstruction using neural networks based on Kalman filter", IGARSS 2019 - International Geoscience and remote Sensing Symposium, Yokohama, Japan, July 2019.

14. Soufiane Ait Tilat, Frédéric Champagnat, Cédric Herzet, "A new sparsity based particle image reconstruction approach for particle detection", International Symposium on Particle Image Velocimetry (ISPIV2019), Munich, Germany, July 2019.

15. Said Ouala, Ronan Fablet, Cédric Herzet, et al, "Learning stochastic representations of geophysical dynamics", ICASSP'2019, IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, UK, May 2019.

16. Frédéric Champagnat, Cédric Herzet, "Atom Selection in Continuous Dictionaries: Reconciling Polar and SVD Approximations", ICASSP'2019, IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, UK, May 2019.

17. Clément Elvira, Rémi Gribonval, Charles Soussen, and Cédric Herzet, "OMP and continuous dictionaries: is k-step recovery possible?", ICASSP'2019, IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, UK, May 2019.

18. Clément Elvira, Rémi Gribonval, Charles Soussen, and Cédric Herzet, "Uniform k-step recovery with CMF dictionaries", SPARS 2019, Signal Processing with Adaptive Sparse Structured Representations, Toulouse, France, 2019

19. Said Ouala, Cédric Herzet, Ronan Fablet, "Bilinear residual Neural Network for the identification and forecasting of dynamical systems" EUSIPCO 2018, European Signal Processing Conference, Rome, Italy, September 2018.

20. Said Ouala, Cédric Herzet, Ronan Fablet, "Sea surface temperature prediction and reconstruction using patch-level neural network representations", IGARSS 2018, International Geoscience and Remote Sensing Symposium, Valencia, Spain, July 2018.

21. Clément Elvira, Rémi Gribonval, Cédric Herzet and Charles Soussen, "A case of exact recovery using OMP with continuous dictionaries", Curves and Surfaces 2018, Arcachon, France, June 2018. (Best poster award)

22. Cédric Herzet and Angélique Drémeau, "Joint Screening Tests for LASSO", ICASSP'2018, IEEE International Conference on Acoustics, Speech, and Signal Processing, Calgary, Canada, April 2018.

23. Cédric Herzet, Mamadou Diallo and Patrick Héas, "Beyond Petrov-Galerkin projection by using "multi-space" priors", Model Reduction for Parametrized Systems (MoRePas), Nantes, France, March 2018.

24. Abed Malti, Cédric Herzet, "Elastic Shape-from-Template with Spatially Sparse Deforming Forces", CVPR'2017, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, July 2017.

25. Patrick Héas, Cédric Herzet, "Optimal Low-rank Dynamic Mode Decomposition",

ICASSP'2017, IEEE International Conference on Acoustics, Speech, and Signal Processing, New Orleans, USA, March 2017.

26. Angélique Drémeau, Cédric Herzet, "DOA Estimation in Structured Phase-Noisy Environments", ICASSP'2017, IEEE International Conference on Acoustics, Speech, and Signal Processing, New Orleans, USA, March 2017.

27. Patrick Héas, Cédric Herzet, "Reduced-Order Modeling Of Hidden Dynamics", ICASSP'2016, IEEE International Conference on Acoustics, Speech, and Signal Processing, Shanghai, China, March 2016.

28. Cédric Herzet, Abed Malti, "Safe screening tests for LASSO based on firmly non-expansiveness", ICASSP'2016, IEEE International Conference on Acoustics, Speech, and Signal Processing, Shanghai, China, March 2016.

29. Ioana Barbu, Cédric Herzet, "A New Approach for Volume Reconstruction in TomoPIV with the Alternating Directions Method of Multipliers", PIV'15, The 11th International Symposium on Particle Image Velocimetry, Santa Barbara, USA, September 2015.

30. Ioana Barbu, Cédric Herzet, "Accelerated, Sparsity-Aware Generalizations of Classical Algorithms for TomoPIV", PIV'15, The 11th International Symposium on Particle Image Velocimetry, Santa Barbara, USA, September 2015.

31. Ioana Barbu, Cédric Herzet, "A Fast and Sparsity-Aware Generalization of SMART for Tomographic Particle Image Velocimetry", SPARS'15, Signal Processing with Adaptive Sparse Structured Representations, Cambridge, England, July 2015.

32. Cédric Herzet, Charles Soussen, "Enhanced Recovery Conditions for OMP/OLS by Exploiting both Coherence and Decay", in Proc. iTWIST'14, international - Traveling Workshop on Interactions between Sparse models and Technology, August 27-29, 2014, Namur, Belgium.

33. Angélique Drémeau, Patrick Héas, Cédric Herzet, "Combining sparsity and dynamics: an efficient way", in Proc. iTWIST'14, international - Traveling Workshop on Interactions between Sparse models and Technology, August 27-29, 2014, Namur, Belgium.

34. Angélique Drémeau, Patrick Héas, Cédric Herzet, "Sparse Representations in Nested Non-Linear Models", ICASSP'2014, IEEE International Conference on Acoustics, Speech, and Signal Processing, Firenze, Italy, May 2014.

35. Mathieu Desoubeaux, Cédric Herzet, William Puech, Gaetan Le Guelvouit, "Enhanced blind decoding of tardos codes with new MAP-based functions", IEEE 15th International Workshop on Multimedia Signal Processing (MMSP), pp. 283-288, September 2013.

36. Charles Soussen, Cédric Herzet, Jérôme Idier, and Rémi Gribonval, "Unrecoverable subsets by OMP and Basis Pursuit", in Proc. SPARS workshop, Signal Processing with Adaptive Sparse Structured Representations, Lausanne, Switzerland, July 2013.

37. Ioana Barbu, Cédric. Herzet, Etienne Mémin, "Joint Estimation of Volume and Velocity in TomoPIV", The 10th International Symposium on Particle Image Velocimetry, Delft, The Nederlands, July 2013.

38. Angélique Drémeau, Cédric Herzet, Laurent Daudet, "Structured Bayesian Orthogonal Matching Pursuit", ICASSP'2012, IEEE International Conference on Acoustics, Speech, and Signal Processing, Kyoto, Japan, March 2012.

39. Ioana Barbu, Cédric Herzet, Etienne Mémin, "Sparse models and pursuit algorithms for PIV Tomography", Forum on recent developments in Volume Reconstruction techniques applied to 3D fluid and solid mechanics (FVR), Poitiers, France, November 2011.

40. Angélique Drémeau, Cédric Herzet, Laurent Daudet, "Structured Bayesian Orthogonal Matching Pursuit", ICASSP'2011, IEEE International Conference on Acoustics, Speech, and Signal Processing, Kyoto, Japan, March 2011.

41. Pierre Dérian, Patrick Héas, Cédric. Herzet, Etienne Mémin, "Wavelet-Based Fluid Motion Estimation", Third International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), Israel, May-June 2011.

42. Patrick Héas, Cédric Herzet, Etienne Mémin, "Robust optic-flow estimation with Bayesian inference of model and hyper-parameters", Third International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), Israel, May-June 2011.

43. Angélique Drémeau, Cédric. Herzet, Laurent Daudet, "Soft Bayesian pursuit algorithm for sparse representations", IEEE Workshop on Statistical Signal Processing, Nice, France, June 2011.

44. Angélique Drémeau , Mehmet Turkan, Cédric. Herzet , Christine Guillemot and Jean-Jacques Fuchs, "Spatial intra-prediction based on mixtures of sparse representations", IEEE International Workshop on Multimedia Signal Processing (MMSP'10), Saint-Malo, France, October, 2010.

45. Onur Oguz, Cédric Herzet, Luc Vandendorpe, "A Novel Adaptive Iterative Detection Technique for Joint Estimation and Detection", PIMRC'2010, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, September 2010.

46. Cédric Herzet and Angélique Drémeau, "Bayesian Pursuit Algorithms", EUSIPCO'2010, European Signal Processing conference, Aalborg, Danemark, August 2010.

47. Cédric Herzet and Angélique Drémeau, “Sparse Representation Algorithms based on Mean-Field Approximations”, ICASSP’2010, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA, 14-19 March 2010.

48. Angélique Drémeau and Cédric Herzet, “An EM-Algorithm Approach for the Design of Orthonormal Bases adapted to Sparse Representations”, ICASSP’2010, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA, 14-19 March 2010.

49. Angélique Drémeau, Cédric Herzet, Christine Guillemot and Jean-Jacques Fuchs, “Sparse Optimization with Directional DCT Bases for Image Compression”, ICASSP’2010, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA, 14-19 March 2010.

50. Onur Oguz, Cedric Herzet, and Luc Vandendorpe, “Low Complexity Iterative Detection in the Presence of Nuisance Parameters”, ICASSP’2010, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA, 14-19 March 2010.

51. Angélique Drémeau, Cédric Herzet, Christine Guillemot and Jean-Jacques Fuchs, “Anisotropic Multi-scale Sparse Learned Bases for Image Compression”, SPIE Electronic Imaging, 2010.

52. Velotiaray Toto-Zarasoa, Aline Roumy, Christine Guillemot, and Cédric Herzet, “Robust non-asymmetric distributed source coding using turbo codes on the syndrome trellis”, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’2009.

53. Cédric Herzet, Velotiaray Toto-Zarasoa, Aline Roumy, “Error resilient for non-asymmetric Slepian-Wolf Coding”, ICC’2009, IEEE International Conference on Communications, June 14-18, 2009, Dresden, Germany.

54. Shawn Yang, Benoit Geller, Cédric Herzet, Jean-Marc Brossier, “Smoothing PLLs for QAM Dynamical Phase Estimation”, ICC’2009, IEEE International Conference on Communications, June 14-18, 2009, Dresden, Germany.

55. Cédric Herzet, “On the Convergence of the Iterative "Pseudo Likelihood" Maximization Algorithm”, ICASSP’ 2008, IEEE International Conference on Acoustics, Speech and Signal Processing, April, 2008, Las Vegas, USA.

56. Harold Sneessens, Xavier Jaspar, Cédric Herzet and Luc Vandendorpe, “Prediction on turbo-decoding performance for adaptive channel coding”, in 41st Asilomar Conference on Signals, Systems and Computers, Asilomar, USA, November 2007.

57. Harold Sneessens, Xavier Jaspar, Cédric Herzet and Luc Vandendorpe, “Calculating turbo-decoding performance characteristics for adaptative channel coding”, SPAWC’07

- IEEE Workshop on Signal Processing Advances in Wireless Communications, June, 2007, Helsinki, Finland.

58. Stéphanie Bay, Cédric Herzet, Jean-Marc Brossier, Jean-Pierre Barbot, Alexandre Renaux and Benoit Geller, "Bayesian Cramér Rao Bound for Dynamical Phase Offset Estimation", SPAWC'07 - IEEE Workshop on Signal Processing Advances in Wireless Communications, June, 2007, Helsinki, Finland.

59. Cédric Herzet and Luc Vandendorpe, "Code-Aided ML Ambiguity Resolution", ICC' 2007 - IEEE International Conference on Communications, June 2007, Glasgow, Scotland.

60. Valéry Ramon, Cédric Herzet, Xavier Wautelet and Luc Vandendorpe, "Soft estimation of time-varying frequency selective channels using Kalman smoothing", ICC' 2007 - IEEE International Conference on Communications, June 2007, Glasgow, Scotland.

61. Cédric Herzet and Luc Vandendorpe, "Prediction of the EM-algorithm Speed of Convergence with Cramer-Rao Bounds", ICASSP' 2007 - IEEE International Conference on Speech, Acoustic and Signal Processing April 2007, Honolulu, Hawaii.

62. Marilena Maiolo, Cédric Herzet and Marco Luise, " Cyclostationarity-Based Blind Symbol Timing Recovery for Continuous Phase Modulations", 15th IST Mobile & Wireless Communication Summit , June 4-8, 2006, Myconos, Greece.

63. Cédric Herzet, Xavier Wautelet, Valéry Ramon and Luc Vandendorpe, "FER-wise Optimal Code-Aided Hypothesis Testing", ICC' 2006 - IEEE International Conference on Communications, June 11-15, 2006, Istanbul, Turkey.

64. Valéry Ramon, Aline Roumy, Cédric Herzet and Luc Vandendorpe, "Calculating the performance degradation of a MMSE/IC turbo-equalization scheme due to SNR estimation errors", ICC' 2006 - IEEE International Conference on Communications, June 11-15, 2006, Istanbul, Turkey.

65. Cédric Herzet, Xavier Wautelet, Valéry Ramon and Luc Vandendorpe, "Iterative Synchronization: EM Algorithm Versus Newton-Raphson Approach", ICASSP'06 - IEEE Intl conf. on Acoustics, Speech and Signal Processing, May 14-19, 2006, Toulouse, France.

66. Valéry Ramon, Cédric Herzet, Xavier Wautelet and Luc Vandendorpe, "Calculating the performance of a soft-information-based best linear unbiased estimator of amplitude and carrier phase offset", ICASSP'06 - IEEE Intl conf. on Acoustics, Speech and Signal Processing, May 14-19, 2006, Toulouse, France.

67. Cédric Herzet, Valéry Ramon and Luc Vandendorpe, "A Factor Graph Approach to Design Close-to-optimal Receivers in the Presence of a Timing Uncertainty", EU-

SIPCO 2005 - European Signal Processing Conference, September 4-8, 2005, Antalya, Turkey (Invited Paper).

68. Cédric Herzet, Valéry Ramon and Luc Vandendorpe, "Turbo synchronization : A Combined Sum-Product and Expectation-Maximization Algorithm Approach", SPAWC'05 - IEEE Workshop on Signal Processing Advances in Wireless Communications, June 5-8, 2005, New-York, USA.

69. Xavier Wautelet, Cédric Herzet and Luc Vandendorpe, "Cramer-Rao Bounds for Channel Estimation with Symbol a Priori Information", SPAWC'05 - IEEE Workshop on Signal Processing Advances in Wireless Communications, June 5-8, 2005, New-York, USA.

70. Valéry Ramon, Cédric Herzet, Luc Vandendorpe and Marc Moeneclaey, "EM Algorithm-based Estimation of Carrier Phase, Amplitude and Noise", ISSSTA'04 - IEEE International Symposium on Spread Spectrum Techniques and Applications, September 2004 , Sydney, Australia, pp. 550-554.

71. Valéry Ramon, Cédric Herzet, Luc Vandendorpe and Marc Moeneclaey, "EM Algorithm-based Multiuser Synchronization in Turbo Receivers", ICASSP'04 - IEEE Intl conf. on Acoustics, Speech and Signal Processing, May 17-21, 2004, Montreal, Canada, vol. 4, pp. 849-852.

72. Xavier Wautelet, Cédric Herzet, Antoine Dejonghe and Luc Vandendorpe, "MMSE-based and EM Iterative Channel Estimation Methods", SCVT 2003 - IEEE 10th Symposium on Communications and Vehicular Technology in the Benelux, November 13, 2003, Eindhoven, The Netherlands.

73. Luc Vandendorpe, Antoine Dejonghe, Cédric Herzet, Xavier Wautelet and Marc Moeneclaey, "Soft-information-aided Synchronization: Recent Results", SPSC'2003 - 8th Intl Workshop on Signal Processing for Space Communications, September 24-26, 2003, Catania, Italy, pp. 195-202.

74. Cédric Herzet, Valéry Ramon, Luc Vandendorpe and Marc Moeneclaey, "An Iterative Soft-decision-directed Linear Timing Estimator", SPAWC'03 - IEEE Workshop on Signal Processing Advances in Wireless Communications, June 2003, Roma, Italy.

75. Nele Noels, Cédric Herzet, Antoine Dejonghe, Vincenzo Lottici, Luc Vandendorpe and Marc Moeneclaey, "Turbo Synchronization : an EM Algorithm Interpretation", ICC'03 - IEEE Intl Conference on Communications, May 2003, Anchorage, USA, pp. 2933-2937.

76. Cédric Herzet, Valéry Ramon, Luc Vandendorpe and Marc Moeneclaey, "EM algorithm-based Timing Synchronization in Turbo Receivers", ICASSP'03 - ICASSP'03-IEEE International Conference on Acoustics, Speech and Signal Processing, April 2003,

Hong-Kong, China, vol. 4, pp. 612-615.

77. Cédric Herzet, Valéry Ramon and Luc Vandendorpe, "Iterative Soft-decision-directed Timing Estimation for Turbo Receivers", SCVT'02 - IEEE Symposium on Communications and Vehicular Technology in the Benelux, October, 2002, Louvain-la-Neuve, Belgium, pp. 97-101.

National Conferences

1. Patrick Héas, Cédric Herzet, Benoit Combès, "Learning Reduced Models with Kernel Methods", 14-ième colloque national en calcul des structures (CSMA19), Giens, France, May 2019.
2. Clément Elvira, Rémi Gribonval, Charles Soussen, and Cédric Herzet, "Identification de supports en k étapes avec OMP pour les dictionnaires continus", GRETSI 2019, XXVIIème Colloque francophone de traitement du signal et des images, Lille, France, 2019.
3. Soufiane Ait Tilat, Frédéric Champagnat, Cédric Herzet, "Détection et localisation en imagerie de particules par approximation de Taylor de dictionnaires continus", GRETSI 2019, XXVIIème Colloque francophone de traitement du signal et des images, Lille, France, 2019.
4. Ioana Barbu, Cédric Herzet, Etienne Mémin, "Estimation 3D jointe des volumes et des vitesses pour la tomographie PIV", 15ème Congrès Francais de Visualisation et de Traitement d'Images en Mécanique des Fluides (Fluvisu), Orléans, November 2013.
5. Dominique Heitz, Cédric Herzet, Patrick Heas, "Incertitudes de mesures PIV par flot optique", 14eme Congres Francais de Visualisation et de Traitement d'Images en Mécanique des Fluides (Fluvisu), Lille, November 2011.
6. Ioana Barbu, Cédric Herzet, Etienne Mémin, "Sparse models and pursuit algorithms for PIV Tomography", 14eme Congres Francais de Visualisation et de Traitement d'Images en Mécanique des Fluides (Fluvisu), Lille, November 2011.
7. Véronique Souchaud, Cédric Herzet, Etienne Mémin, "Construction d'une base de type POD a partir d'une séquence d'images", colloque GRETSI, Bordeaux, France, September 2011.

Bibliography

[1] S. Lin and D.J. Costello, *Error Control Coding: Fundamentals and Applications*, Pearson-Prentice Hall, 2004.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proceedings of ICC '93 - IEEE International Conference on Communications*, 1993, vol. 2, pp. 1064–1070 vol.2.

[3] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[4] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[5] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[6] V. Ramon, C. Herzet, L. Vandendorpe, and M. Moeneclaey, "EM algorithm-based multiuser synchronization in turbo receivers," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 4, pp. iv–849.

[7] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe, "Code-aided turbo synchronization," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1255–1271, 2007.

[8] C. Herzet, K. Woradit, H. Wymeersch, and L. Vandendorpe, "Code-aided maximum-likelihood ambiguity resolution through free-energy minimization," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6238–6250, 2010.

[9] S. Bay, C. Herzet, J. M. Brossier, J. P. Barbot, A. Renaux, and B. Geller, "Bayesian Cramér-Rao bound for dynamical phase offset estimation," in *2007 IEEE 8th Workshop on Signal Processing Advances in Wireless Communications*, 2007, pp. 1–5.

[10] S. Bay, C. Herzet, J. Brossier, J. Barbot, and B. Geller, "Analytic and asymptotic analysis of bayesian Cramér-Rao bound for dynamical phase offset estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 61–70, 2008.

[11] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing.*, Applied and Numerical Harmonic Analysis. Birkhaüser, 2013.

[12] A. Bourrier, M. E. Davies, T. Peleg, P. Pérez, and R. Gribonval, "Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7928–7946, Dec. 2014.

[13] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Transactions on Information Theory*, vol. 28, pp. 482–488, 1982.

[14] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4572–4584, Oct. 2011.

[15] Hemant Ishwaran and J. Sunil Rao, "Spike and slab variable selection: Frequentist and Bayesian strategies," *The Annals of Statistics*, vol. 33, no. 2, pp. 730–773, 4 2005.

[16] F. Champagnat, Y. Goussard, S. Gautier, and J. Idier, *Deconvolution of Spike Trains*, chapter 5, pp. 117–140, John Wiley & Sons, Ltd., 2008.

[17] C. Herzet and A. Drémeau, "Bayesian pursuit algorithms," in *EURASIP European Signal Processing Conference, EUSIPCO'*, 2010.

[18] C. Herzet and A. Drémeau, "Bayesian pursuit algorithms," Tech. Rep., INRIA-Rennes Bretagne Atlantique, http://hal.inria.fr/hal-00673801, 2012.

[19] A. Drémeau, C. Herzet, and L. Daudet, "Structured bayesian orthogonal matching pursuit," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. Mar. 2012, pp. 3625–3628, IEEE.

[20] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[21] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, 1993.

[22] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[23] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2009.

[24] A. Drémeau and C. Herzet, "Soft Bayesian pursuit algorithm for sparse representations," in *IEEE International Workshop on Statistical Signal Processing 2011 (SSP'11)*, 2011.

[25] A. Drémeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3425–3438, July 2012.

[26] M. Beal and Z. Ghahramani, "The variational bayesian EM algorithm for incomplete data : with application to scoring graphical model structures," *Statistics*, July 2002.

[27] J. Arbel, H.-P. Dang, C. Elvira, C. Herzet, Z. Naulet, and M. Vladimirova, "Bayes in action in deep learning and dictionary learning," in *ESAIM: Proceedings and Surveys*, 2022.

[28] Z. Ghahramani, T.L. Griffiths, and P. Sollich, "Bayesian nonparametric latent feature models," in *Bayesian Statistics*, 2007.

[29] B. Kulis and M. I. Jordan, "Revisiting K-Means: New algorithms via Bayesian nonparametrics," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, Madison, WI, USA, 2012, ICML'12, pp. 1131–1138, Omnipress.

[30] S. Chen and D. L. Donoho, "Basis pursuit," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, 1994, vol. 1, pp. 41–44 vol.1.

[31] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by Basis Pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.

[32] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, second edition, 2006.

[33] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA Journal of Numerical Analysis*, vol. 20, no. 3, pp. 389–403, July 2000.

[34] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[35] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[36] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, June 2010, pp. 1–4.

[37] L. El Ghaoui, V. Viallon, and T. Rabbani, "Safe feature elimination in sparse supervised learning," Tech. Rep. UC/EECS-2010-126, EECS Dept., University of California at Berkeley, Sept. 2010.

[38] Z. J. Xiang, Y. Wang, and P. J. Ramadge, "Screening tests for lasso problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1008–1027, 2017.

[39] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap: safer rules for the Lasso," in *ICML*, 2015.

[40] L. Dai and K. Pelckmans, "An ellipsoid based, two-stage screening test for BPDN," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European.* Aug. 2012, pp. 654–658, IEEE.

[41] J. Liu, Z. Zhao, J. Wang, and J. Ye, "Safe screening with variational inequalities and its application to Lasso," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 289–297.

[42] J. Wang, P. Wonka, and J. Ye, "Lasso screening rules via dual polytope projection," *Journal of Machine Learning Research*, 2015.

[43] C. Herzet and A. Malti, "Safe screening tests for LASSO based on firmly non-expansiveness," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4732–4736.

[44] T. L. Tran, C. Elvira, H.-P. Dang, and C. Herzet, "Beyond GAP screening for Lasso by exploiting new dual cutting half-spaces," in *EURASIP European Signal Processing Conference, EUSIPCO'22*, 2022.

[45] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "Dynamic screening: Accelerating first-order algorithms for the Lasso and Group-Lasso," *Signal Processing, IEEE Transactions on*, vol. 63, no. 19, pp. 5121–5132, Oct. 2015.

[46] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, "Gap safe screening rules for sparsity enforcing penalties," *Journal of Machine Learning Research*, vol. 18, no. 128, pp. 1–33, 2017.

[47] C. Elvira and C. Herzet, "Safe rules for the identification of zeros in the solutions of the SLOPE problem," arXiv 2110.11784, 2021.

[48] M. Bogdan, E. van den Berg, W. Su, and E. Candes, "Statistical estimation and testing via the sorted L1 norm," arXiv1310.1969, 2013.

[49] X. Zeng and Mario A. T. Figueiredo, "Decreasing weighted sorted ℓ_1 regularization," *IEEE Signal Processing Letters*, vol. 21, no. 10, pp. 1240–1244, Oct. 2014.

[50] Xin Xing, Jinjin Hu, and Yaning Yang, "Robust minimum variance portfolio with l-infinity constraints," *Journal of Banking & Finance*, vol. 46, pp. 107–117, 2014.

[51] P. J. Kremer, S. Lee, M. Bogdan, and S. Paterlini, "Sparse portfolio selection via the sorted ℓ_1-norm," *Journal of Banking & Finance*, vol. 110, 2020.

[52] Alexej Gossmann, Shaolong Cao, and Yu-Ping Wang, "Identification of significant genetic variants via slope, and its extension to group slope," in *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, New York, NY, USA, 2015, BCB '15, pp. 232–240, Association for Computing Machinery.

[53] Loubna El Gueddari, Emilie Chouzenoux, Alexandre Vignaud, and Philippe Ciuciu, "Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization," Preprint hal-02292372, Sept. 2019.

[54] U. Oswal and R. Nowak, "Scalable sparse subspace clustering via ordered weighted l1 regression," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018, pp. 305–312.

[55] D. Zhang, H. Wang, M. Figueiredo, and L. Balzano, "Learning to share: simultaneous parameter tying and sparsification in deep learning," in *International Conference on Learning Representations*, 2018.

[56] C. Herzet and A. Drémeau, "Joint screening tests for LASSO," in *ICASSP 2018-IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 1–5.

[57] C. Herzet, C. Dorffer, and A. Drémeau, "Gather and conquer: Region-based strategies to accelerate safe screening tests," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3300–3315, 2019.

[58] Y. de Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336 – 354, 2012.

[59] K. Bredies and H. K. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[60] C. Elvira and C. Herzet, "Safe squeezing for antisparse coding," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3252–3265, 2020.

[61] C. Elvira and C. Herzet, "Short and squeezed: Accelerating the computation of antisparse representations with safe squeezing," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5615–5619.

[62] T. Guyard, C. Herzet, and C. Elvira, "Node-screening tests for the l0-penalized least-squares problem," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5448–5452.

[63] L. A. Wolsey, *Integer Programming*, Wiley, New York, NY, USA, 1998.

[64] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans on Information Theory*, vol. 59, no. 5, 2013.

[65] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[66] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[67] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[68] C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 459–470, Jan. 2016.

[69] T. T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3516–3522, July 2010.

[70] C. Elvira, R. Gribonval, C. Soussen, and C. Herzet, "When does OMP achieve exact recovery with continuous dictionaries?," *Applied and Computational Harmonic Analysis*, vol. 51, pp. 374–413, 2021.

[71] L. Jacques and C. De Vleeschouwer, "A geometrical study of matching pursuit parametrization," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2835–2848, 2008.

[72] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[73] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[74] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations*, vol. 92, Springer International Publishing, 2016.

[75] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.

[76] P. Héas and C. Herzet, "Low-rank dynamic mode decomposition: An exact and tractable solution," *Journal of Nonlinear Science*, vol. 32, no. 1, pp. 8, 2021.

[77] P. Héas and C. Herzet, "Optimal low-rank dynamic mode decomposition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4456–4460.

[78] Kookjin Lee and Kevin T. Carlberg, "Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders," *Journal of Computational Physics*, vol. 404, pp. 108973, 2020.

[79] E. Parish and K. Carlberg, "Time-series machine-learning error models for approximate solutions to parameterized dynamical systems," *Computer Methods in Applied Mechanics and Engineering*, vol. 365, pp. 112990, 2020.

[80] C. Herzet and M. Diallo, "Performance guarantees for a variational "multi-space" decoder," *Advances in Computational Mathematics*, vol. 46, no. 1, pp. 10, 2020.

[81] C. Herzet, P. Héas, and A. Drémeau, "Model reduction from partial observations," *International Journal for Numerical Methods in Engineering*, vol. 113, no. 3, pp. 479–511, 2018.

[82] P. Héas and C. Herzet, "Reduced modeling of unknown trajectories," *Archives of Computational Methods in Engineering*, vol. 25, no. 1, pp. 87–101, 2018.

[83] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, "Data assimilation in reduced modeling," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 5, no. 1, pp. 1–29, 2017.

[84] J. P. Fink and W. C. Rheinboldt, "On the error behavior of the reduced basis technique for nonlinear finite element approximations," *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift fur Angewandte Mathematik und Mechanik*, vol. 63, no. 1, pp. 21–28, 1983.

[85] U. Hetmaniuk, R. Tezaur, and C. Farhat, "An adaptive scheme for a class of interpolatory model reduction methods for frequency response problems," *International Journal for Numerical Methods in Engineering*, vol. 93, no. 10, pp. 1109–1124, 2013.

[86] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, 1996, Cambridge Books Online.

[87] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, Society for Industrial and Applied Mathematics, Jan. 2005.

[88] R. Everson and L. Sirovich, "Karhunen-Loève procedure for gappy data," *J. Opt. Soc. Am. A*, vol. 12, no. 8, pp. 1657–1664, Aug. 1995.

[89] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, USA, 2003.

[90] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, USA, 2nd edition, 2000.

[91] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, illustrated edition, 2004.

[92] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42 of *Springer series in computational mathematics*, Springer, Heidelberg, 2012.

[93] I. Goodfellow, Y. Bengio, and A. Couville, *Deep Learning*, The MIT Press, 2016.

[94] A. Mendible, S. L. Brunton, A. Y. Aravkin, W. Lowrie, and J. N. Kutz, "Dimensionality reduction and reduced-order modeling for traveling wave physics," *Theoretical and Computational Fluid Dynamics*, vol. 34, no. 4, pp. 385–400, 2020.

[95] F. Champagnat and C. Herzet, "Interpolating and translation-invariant approximations of parametric dictionaries," in *28th European Signal Processing Conference (EUSIPCO 2020)*, Amsterdam, Netherlands, Dec. 2021.

[96] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Comm. Pure Appl. Math*, vol. 67, no. 6, pp. 906–956, June 2014.

[97] Q. Denoyelle, V. Duval, G. Peyre, and E. Soubies, "The sliding Frank-Wolfe algorithm and its application to super-resolution microscopy," *Inverse Problems*, 2019.

[98] Bastien Laville, Laure Blanc-Féraud, and Gilles Aubert, "Off-the-grid variational sparse spike recovery: Methods and algorithms," *Journal of Imaging*, vol. 7, no. 12, 2021.

[99] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," 2019, In the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. July 2019.

[100] Y. Nesterov, *Lectures on Convex Optimization*, Springer, second edition, 2018.

[101] W. Zhang, B. Hong, W. Liu, J. Ye, D. Cai, X. He, and J. Wang, "Scaling up sparse support vector machines by simultaneous feature and sample reduction," Sydney,

Australia, 06–11 Aug 2017, vol. 70 of *Proceedings of Machine Learning Research*, pp. 4016–4025, PMLR.

[102] C. F. Dantas, E. Soubies, and C. Févotte, "Safe screening for sparse regression with the Kullback-Leibler divergence," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5544–5548.

[103] C. Dantas, E. Soubies, and C. Févotte, "Expanding boundaries of Gap safe screening," ArXiv 2102.10846, February 2021.

[104] C. Herzet, C. Elvira, and H.-P. Dang, "Region-free safe screening tests for l1-penalized convex problems," in *EURASIP European Signal Processing Conference, EUSIPCO'22*, 2022.

[105] Rembert Reemtsen and Stephan Görner, *Numerical Methods for Semi-Infinite Programming: A Survey*, pp. 195–275, Springer US, Boston, MA, 1998.

[106] A. Atamturk and A. Gomez, "Safe screening rules for l0-regression from perspective relaxations," in *Proceedings of the 37th International Conference on Machine Learning*, Virtual, 13–18 Jul 2020, vol. 119 of *Proceedings of Machine Learning Research*, pp. 421–430.

[107] T. Guyard, C. Herzet, and C. Elvira, "Screen & Relax: Accelerating The Resolution Of Elastic-Net By Safe Identification of The Solution Support," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5443–5447.

[108] V. Chandrasekaran and M. I. Jordan, "Computational and statistical tradeoffs via convex relaxation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. E1181–E1190, 2013.

[109] R. Giryes, Y. C. Eldar, A. M. Bronstein, and G. Sapiro, "Tradeoffs between convergence speed and reconstruction accuracy in inverse problems," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1676–1690, 2018.

[110] V. Roulet, N. Boumal, and A. d'Aspremont, "Computational complexity versus statistical performance on sparse recovery problems," *Information and Inference: A Journal of the IMA*, vol. 9, no. 1, pp. 1–32, 6/29/2021 2020.

[111] B. Peherstorfer, B. Kramer, and K. Willcox, "Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 2, pp. 737–761, 2021/09/10 2018.

[112] P. Héas, "Selecting reduced models in the cross-entropy method," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 8, no. 2, pp. 511–538, 2021/09/07 2020.