Optimization Over Banach Spaces: A Unified View on Supervised Learning and Inverse Problems

Shayan Aziznejad

Thèse N° 8943 (juin 2022)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Dimitri Van De Ville, *président*
Prof. Michael Unser, *directeur de thèse*
Prof. Gabriel Peyré, *rapporteur*
Prof. Anders Hansen, *rapporteur*
Prof. Victor Panaretos, *rapporteur*

École polytechnique fédérale de Lausanne—2022

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with LaTeX

Available at **http://bigwww.epfl.ch/**

Abstract

In this thesis, we reveal that supervised learning and inverse problems share similar mathematical foundations. Consequently, we are able to present a unified variational view of these tasks that we formulate as optimization problems posed over infinite-dimensional Banach spaces. Throughout the thesis, we study this class of optimization problems from a mathematical perspective. We start by specifying adequate search spaces and loss functionals that are derived from applications. Next, we identify conditions that guarantee the existence of a solution and we provide a (finite) parametric form for the optimal solution. Finally, we utilize these theoretical characterizations to derive numerical solvers.

The thesis is divided into five parts. The first part is devoted to the theory of splines, a large class of continuous-domain models that are optimal in many of the studied frameworks. Our contributions in this part include the introduction of the notion of multi-splines, their theoretical properties, and shortest-support generators. In the second part, we study a broad class of optimization problems over Banach spaces and we prove a general representer theorem that characterizes their solution sets. The third and fourth parts of the thesis invoke the applicability of our general framework to supervised learning and inverse problems, respectively. Specifically, we derive various learning schemes from our variational framework that inherit a certain notion of "sparsity" and we establish the connection between our theory and deep neural networks, which are state-of-the-art in supervised learning. Moreover, we deploy our general theory to study continuous-domain inverse problems with multicomponent models, which can be applied to various signal and image processing tasks, in particular, curve fitting. Finally, we revisit the notions of splines and sparsity in the last part of the thesis, this time, from a stochastic perspective.

Keywords: Supervised learning, inverse problems, nonparametric regression, representer theorem, regularization theory, splines, sparsity, total variation, deep neural networks.

Résumé

Dans cette thèse, nous révélons que l'apprentissage supervisé et les problèmes inverses partagent des fondements mathématiques similaires. Par conséquent, nous présentons une vue variationnelle unifiée de ces tâches que nous formulons comme des problèmes d'optimisation posés sur des espaces de Banach de dimension infinie. Tout au long de la thèse, nous étudions cette classe de problèmes d'optimisation d'un point de vue mathématique. Nous commençons par spécifier des espaces de recherche adéquats et des fonctions coût qui sont utilisées en pratique. Ensuite, nous identifions les conditions qui garantissent l'existence d'une solution et nous donnons une forme paramétrique (finie) de la solution optimale. Enfin, nous utilisons ces caractérisations théoriques pour en déduire des algorithmes de résolution numérique.

La thèse est divisée en cinq parties. La première partie est consacrée à la théorie des splines, une grande classe de modèles dans le domaine continu qui sont optimaux dans de nombreux cadres étudiés. Nos contributions dans cette partie incluent l'introduction de la notion de multi-splines, de leurs propriétés théoriques et des générateurs de support le plus restreint. Dans la deuxième partie, nous étudions une large classe de problèmes d'optimisation sur des espaces de Banach et nous prouvons un théorème général de représentation qui caractérise leurs ensembles de solutions. Les troisième et quatrième parties de la thèse évoquent l'applicabilité de notre cadre général à l'apprentissage supervisé et aux problèmes inverses, respectivement. Plus précisément, nous dérivons divers schémas d'apprentissage de notre cadre variationnel qui héritent d'une certaine notion de « parcimonie » et nous établissons le lien entre notre théorie et les réseaux de neurones profonds, qui sont à la pointe de l'apprentissage supervisé. De plus, nous déployons notre théorie générale pour étudier des problèmes inverses dans le domaine continu avec des modèles à composantes multiples, qui peuvent être appliqués à diverses tâches de traitement du signal et d’image, en particulier l’ajustement de courbe. Enfin, nous revisitons les notions de splines et de parcimonie dans la dernière partie de la thèse d’une perspective stochastique cette fois-ci.

Mots clefs : Apprentissage supervisé, problèmes inverses, régression non paramétrique, théorème de représentation, théorie de la régularisation, splines, parcimonie, variation totale, réseaux de neurones profonds.

If everything is imperfect in this imperfect world,
love is most perfect in its perfect imperfection.
— The Seventh Seal (1957)

To Joey, Moji, and Shabi.

Abbreviations

ADMM	alternating direction method of multipliers
CDF	cumulative distribution function
CNN	convolutional neural network
CPWL	continuous and piecewise-linear
DCT	discrete cosine transform
DFT	discrete Fourier transform
DNN	deep neural network
FIR	finite impulse response
FISTA	fast iterative shrinkage-thresholding
gTV	generalized total-variation
HTV	Hessian-Schatten total variation
i.i.d	independent and identically distributed
KLT	Karhunen-Loève transform
KS	Kolmogorov-Smirnov
MMSE	minimum mean-squared error
MKL	multiple-kernel learning
MSE	mean-squared error
PDF	probability density function
LASSO	least absolute shrinkage and selection operator
LSI	linear and shift-invariant
PReLU	parametric rectified linear unit
QFE	quadratic fitting error
RBF	radial-basis function
ReLU	rectified linear unit
RI-TV	rotation-invariant total variation
RKBS	reproducing kernel Banach space
RKHS	reproducing kernel Hilbert space
SDE	stochastic differential equation
SNR	signal-to-noise ratio
SVD	singular-value decomposition
SVM	support-vector machine
TV	total-variation

List of Figures

List of Tables

Contents

Introduction

In this thesis, we develop a mathematical framework for the study of optimization problems posed over Banach spaces. We have undertaken this work with the goal of providing a unified theoretical view of two fundamental areas of modern data science; namely, inverse problems and supervised learning. In this chapter, we briefly describe both domains to establish the general context of our work. We then highlight the main contributions of our work and provide a roadmap that outlines the organization of this thesis.

Inverse Problems

In a broad sense, the term *inverse problem* refers to the task of recovering an unknown signal of interest from a collection of (possibly corrupted) observations. The problem has been extensively studied over the past decades due to its wide range of applications. To illustrate the generality of this concept, we point out that any imaging system at its core solves a specific inverse problem. For example, in computed tomography, the image of a biological tissue is being reconstructed from a set of X-ray projections taken at various angles.

Generally speaking, an inverse problem is defined through the specification of three components, namely

1. a hypothesis space in which the signal of interest is presumed to exist;
2. a forward model that, for any element in the hypothesis space, generates a set of observations; and
3. the observed data that is often stored as an array.

In this thesis, we focus on continuous-domain linear inverse problems with finitely many observations that are corrupted by noise. Precisely, the hypothesis space is assumed to be an infinite-dimensional function space $\mathcal{F}(\mathbb{R}^d)$ (typically a Banach space) that contains the signal of interest $f : \mathbb{R}^d \to \mathbb{R}$. The forward model $\boldsymbol{\nu} = (\nu_m) : \mathcal{F}(\mathbb{R}^d) \to \mathbb{R}^M$ is assumed to be a continuous vector-valued linear functional. Finally, the observed data is stored in a finite-dimensional vector $\mathbf{y} = (y_m) \approx \boldsymbol{\nu}(f) \in \mathbb{R}^M$ and the goal is to recover the unknown function $f \in \mathcal{F}(\mathbb{R}^d)$ from the vector of observations $\mathbf{y}$.

Supervised Learning

Given a sequence of data points $\{(x_m, y_m)\}_{m=1}^M \subseteq \mathcal{X} \times \mathcal{Y}$, the goal of supervised learning is to find a mapping $f : \mathcal{X} \to \mathcal{Y}$ that adequately explains the data, so that $f(x_m) \approx y_m$ for all $m = 1, \dots, M$, while avoiding overfitting. Supervised learning methods are now being vastly exploited in various research areas such as drug discovery, decision making, medical image analysis, and artificial intelligence.

In this thesis, we are interested in nonparametric regression, a branch of supervised learning whose objective is to learn real-valued functions $\mathbb{R}^d \to \mathbb{R}$ from a prescribed infinite-dimensional function space $\mathcal{F}(\mathbb{R}^d)$. Given that the sampling functional $\delta_{\mathbf{x}_0} : \mathcal{F}(\mathbb{R}^d) \to \mathbb{R} : f \mapsto f(\mathbf{x}_0)$ is linear for any $\mathbf{x}_0 \in \mathbb{R}^d$, we highlight that nonparametric regression can be viewed as a particular continuous-domain linear inverse problem with the forward operator

$$\boldsymbol{\nu} : f \mapsto (f(\mathbf{x}_1), \dots, f(\mathbf{x}_M)) \in \mathbb{R}^M. \tag{1}$$

It is this last observation that provides the intellectual foundation of this thesis, where we propose to construct a unified variational formalism that encompasses both frameworks.

Variational Formalism

From a variational perspective, inverse problems can be formulated as a minimization of the form

$$\min_{f\in\mathcal{F}(\mathbb{R}^d)}\left(\sum_{m=1}^{M} E(\nu_m(f),y_m)+\lambda\mathcal{R}(f)\right), \tag{2}$$

where $E:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is a loss function measuring the discrepancy between the acquired data vector $\mathbf{y}$ and the reconstructed signal f, and $\mathcal{R}:\mathcal{F}(\mathbb{R}^d)\to\mathbb{R}$ is a regularization functional that enforces some prior knowledge on the signal of interest (*e.g.* regularity, sparsity) whose significance is adjusted by the free parameter $\lambda>0$. The aim of this thesis is to study optimization problems of the form (2) and to use our theoretical findings in both supervised learning and inverse problems.

Organization of The Thesis

We have divided the main contributions of this thesis into five groups. Each group is represented by an independent chapter (see, Fig 1, for a schematic illustration). We then conclude the thesis in Chapter 6, where we provide a summary of our work (Section 6.1) that is followed by an outlook on the future research directions (Section 6.2). In what follows, we describe the principal chapters of this thesis in more details.

Spline Theory (Chapter 1)

Providing a link between the digital and analog worlds, splines play a key role in our study of optimization problems over Banach spaces. In fact, we shall reveal

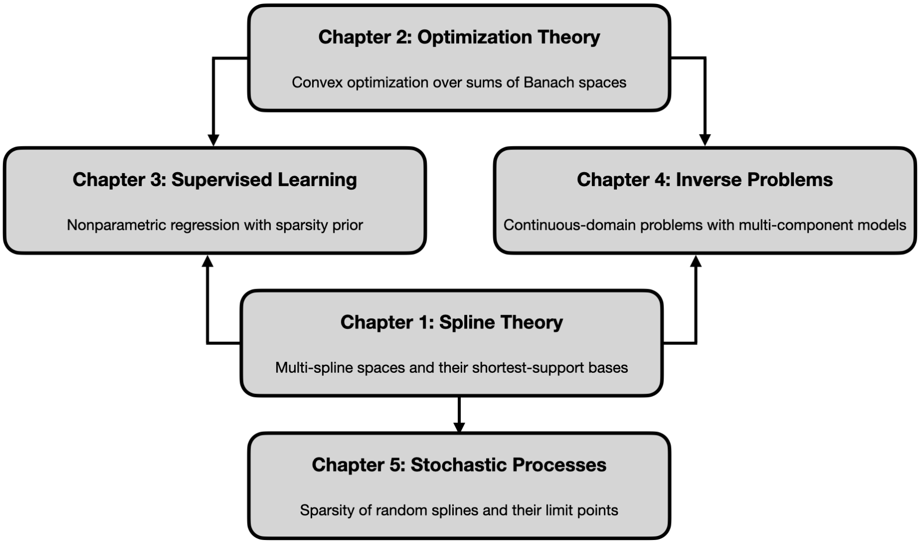

Chapter 2: Optimization Theory
Convex optimization over sums of Banach spaces
Chapter 3: Supervised Learning
Nonparametric regression with sparsity prior
Chapter 4: Inverse Problems
Continuous-domain problems with multi-component models
Chapter 1: Spline Theory
Multi-spline spaces and their shortest-support bases
Chapter 5: Stochastic Processes
Sparsity of random splines and their limit points

Figure 1: Roadmap of the thesis.

that splines are optimal solutions of many optimization problems of the form (2). We are motivated by this connection to deepen our understanding of splines. To that end, we start this thesis with a brief overview of the classical theory of splines, where we discuss basic notions and highlight main properties (Section 1.1). Next, we study the support property of Hermite splines by proving that they are maximally localized among pairs of functions with the same reproduction property (Section 1.2). Finally, we introduce a new family of splines, called multi-splines, and we detail their role in generalized sampling (Section 1.3).

Relevant Publications:

- J. Fageot, S. Aziznejad, M. Unser, and V. Uhlmann, "Support and approximation properties of Hermite splines," *Journal of Computational and Applied Mathematics*, vol. 368, pp. 1-15, 2020.
- A. Goujon, S. Aziznejad, A. Naderi, and M. Unser, "Shortest-support multi-

spline bases for generalized sampling," *Journal of Computational and Applied Mathematics*, vol. 395, pp. 1–18, 2021.

Optimization Over Banach Spaces (Chapter 2)

In the second part of this thesis, we study generic optimization problems of the form (2) that are posed over Banach spaces. We are particularly interested in cases where the search space has a direct sum/product structure, which will act as the foundation of our forthcoming contributions in the areas of inverse problems and supervised learning. We first identify sufficient conditions that guarantee the existence of a solution. This is the minimum requirement to have a well-posed problem. Next, we characterize the solution set of these problems via a representer theorem which shall be used throughout the thesis. Our machinery relies on existing tools from functional analysis, in particular, duality theory.

Relevant Publication:

- M. Unser and S. Aziznejad, "Convex optimization in sums of Banach spaces," *Applied and Computational Harmonic Analysis*, vol. 56, pp. 1–25, 2022.

Supervised Learning (Chapter 3)

In Chapter 3, we leverage our general representer theorm in the context of supervised learning, where we propose novel learning schemes that inherit a certain notion of sparsity. We start by proposing a novel multi-kernel regression scheme that suggests a sparse kernel expansion with adaptive (data-driven) kernel shapes and positions (Section 3.2). Next, we study the problem of learning univariate functions under joint Lipschitz and sparsity constraints which, subsequently, yields a gridless approach for learning linear splines with the fewest number of knots (Section 3.3). Our theoretical findings in the previous section allow us to derive a new method for learning linear spline activation functions for deep neural networks (Section 3.4). Finally, we introduce a novel Hessian-based seminorm for learning continuous and piecewise-linear functions directly from the data (Section 3.5).

Relevant Publications:

- S. Aziznejad, H. Gupta, J. Campos, and M. Unser, “Deep neural networks with trainable activations and controlled Lipschitz constant,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 4688–4699, 2020.
- S. Aziznejad and M. Unser, “Multikernel regression with sparsity constraint,” *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 201–224, 2021.
- S. Aziznejad and M. Unser, “Duality mapping for Schatten matrix norms,” *Numerical Functional Analysis and Optimization*, vol. 42, no. 6, pp. 679–695, 2021.
- S. Aziznejad, T. Debarre, and M. Unser, “Sparsest univariate learning models under Lipschitz constraint,” *IEEE Open Journal of Signal Processing*, vol. 3, pp. 140-154, 2022.
- S. Aziznejad, J. Campos, and M. Unser, “Measuring complexity of learning schemes using Hessian-Schatten total variation,” *arXiv preprint arXiv:2112.06209*, 2021.
- J. Campos, S. Aziznejad, and M. Unser, “Learning of continuous and piecewise-linear functions with Hessian total-variation regularization,” *IEEE Open Journal of Signal Processing*, vol. 3, pp. 36–48, 2021.
- P. Bohra, J. Campos, H. Gupta, S. Aziznejad, and M. Unser, “Learning activation functions in deep (spline) neural networks,” *IEEE Open Journal of Signal Processing*, vol. 1, pp. 295-309, 2020.

Inverse Problems (Chapter 4)

Another consequence of our general representer theorem is discussed in the context of inverse problems. To illustrate the concept, we focus on the family of one-dimensional inverse problems with multicomponent signal priors. Precisely, we consider the problem of recovering signals that can be decomposed as a sum of sparse (Section

4.2) and/or smooth (Section 4.3) components from a set of linear measurements. By adapting our general representer theorem, we show how to solve these problems numerically in order to obtain a desirable reconstruction. As an application, we use the developed schemes for sparse curve fitting which can be practically used to generate stylized fonts and, more generally, any active contour (Section 4.4).

Relevant Publications:

- T. Debarre, S. Aziznejad, and M. Unser, "Hybrid-spline dictionaries for continuous-domain inverse problems," *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5824–5836, 2019.
- T. Debarre, S. Aziznejad, and M. Unser, "Continuous-domain formulation of inverse problems for composite sparse-plus-smooth signals," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 545–558, 2021.
- I. Lloréns Jover, T. Debarre, S. Aziznejad, and M. Unser, "Coupled splines for sparse curve fitting," IEEE Transactions on Image Processing, vol. 31, pp. 4707-4718, 2022.

Splines and Stochastic Processes(Chapter 5)

In the last part, we revisit splines and sparsity from a stochastic point of view. This provides a probabilistic perspective on the two main themes of this thesis. To that end, we study the family of generalized Lévy processes. We highlight that these are limit points of compound-Poisson processes (a.k.a. random splines). This enables us to develop a spline-based method for generating gridless trajectories of any generalized Lévy process (Section 5.2). We then study the compressibility of generalized Lévy processes in wavelet bases which provides a stochastic meaning for the notion of sparsity (Section 5.3). Finally, we provide an in-depth study of the compressibility of compound-Poisson processes (Section 5.4).

Relevant Publications:

- S. Aziznejad and J. Fageot, "Wavelet analysis of the Besov regularity of Lévy white noise," *Electronic Journal of Probability*, vol. 25, pp. 1–38, 2020.

- L. Dadi, S. Aziznejad, and M. Unser, "Generating sparse stochastic processes using matched splines," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4397–4406, 2020.

- S. Aziznejad and J. Fageot, "Wavelet compressibility of compound Poisson processes," *IEEE Transactions on Information Theory*, vol. 68, no. 4, pp. 2752-2766, 2022.

Chapter 1

Splines: The Old Ones and The New

This chapter[1] is devoted to the theory of splines. We start with an overview of splines in Section 1.1, where we recall the main notions and highlight the key properties. We then focus on the support property of Hermite splines in Section 1.2. Finally, we introduce the notion of multi-splines and study their shortest-support bases in Section 1.3.

1.1 An Overview on Splines

Introduced by Schoenberg in his seminal paper [4], splines have been appearing in a variety of research areas, such as interpolation [5], signal processing [6], approximation theory [7], and machine learning [8], to name a few. In this section, we provide

[1] This chapter is based on our published works [1, 2]. Additionally, we use contents from [3] in Section 1.1.

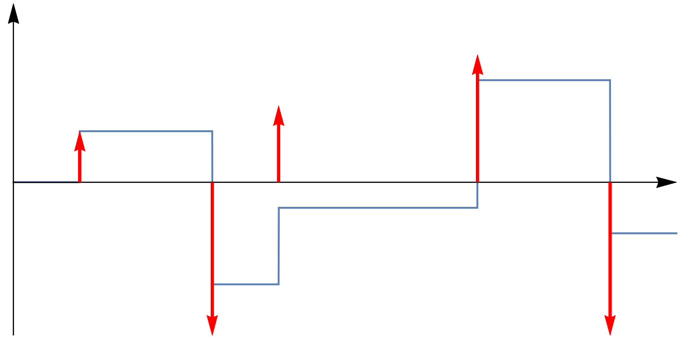


Figure 1.1: A spline of degree zero (blue) and its corresponding innovation model (red).

a very coarse overview of spline theory. Our intent is to prepare the reader for the remainder of this thesis.

1.1.1 Basic Definition

A polynomial spline of degree $n \in \mathbb{N}$ is a piecewise polynomial function $s : \mathbb{R} \to \mathbb{R}$ of the same degree with $(n-1)$ continuous derivatives[2]. Equivalently, one should have that

$$\mathrm{D}^{n+1}\{s\} = w = \sum_{k \in \mathbb{Z}} a_k \delta(\cdot - z_k), \tag{1.1}$$

where D^{n+1} denotes the (weak) derivative operator of order $(n+1)$ and w is the innovation model of s which is specified by the spline coefficients $a_k \in \mathbb{R}\backslash\{0\}$ and distinct knots $z_k \in \mathbb{R}$. An example of a spline of degree 0 with its innovation model is depicted in Figure 1.1.

[2] For degrees $n = 1, 2, 3$, we often use the terms linear, quadratic, and cubic splines, respectively.

1.1.2 Green's Function

The simplest splines of degree n are the polynomials of degree up to n whose associated innovation models are zero. For the simplest nontrivial (with nonzero innovation model) spline of degree n, one requires to solve the differential equation

$$\mathrm{D}^{n+1}\{s\} = \delta. \tag{1.2}$$

Together with an adequate set of boundary conditions, the differential equation (1.2) has a unique solution that is called the Green's function of the $(n+1)$-th order derivative operator. For example, the causal Green's function of D^{n+1} is the one-sided power function $\rho_{\mathrm{D}^{n+1}} : \mathbb{R} \to \mathbb{R}$,

$$\rho_{\mathrm{D}^{n+1}}(x) = (x)_+^n \triangleq \begin{cases} \frac{x^n}{n!}, & x > 0 \\ 0, & x \leq 0. \end{cases} \tag{1.3}$$

Using this notion, one can rewrite (1.1) and obtain a generic expansion for any spline of degree n as

$$s = p + \sum_{k \in \mathbb{Z}} a_k \rho_{\mathrm{D}^{n+1}}(\cdot - z_k), \tag{1.4}$$

where p is a polynomial of degree up to n.

1.1.3 B-Splines

The most popular splines in practice are cardinal splines whose knots are uniformly spaced, so that

$$s = p + \sum_{k \in \mathbb{Z}} a[k] \rho_{\mathrm{D}^{n+1}}(\cdot - kh) \tag{1.5}$$

for some grid size $h > 0$. Following (1.5), we observe that cardinal splines are continuous-domain objects that admit a discrete and linear parametrization, which makes them appealing in practice.

The foundation of cardinal splines was laid by Schoenberg in [5, 9], where he showed that (1.5) can be rewritten as

$$s = \sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{D}^{n+1}}\left(\frac{\cdot - kh}{h}\right) \tag{1.6}$$

for some sequence of coefficients $c[k], k \in \mathbb{Z}$ and some compactly supported function $\beta_{\mathrm{D}^{n+1}} : \mathbb{R} \to \mathbb{R}$ that is called the B-spline of degree n. B-splines allow an efficient and practical implementation, which is exploited in many fields [7, 10, 11, 12].

Interestingly, the B-splines can be constructed recursively with the relation

$$\beta_{\mathrm{D}^{n+1}} = \beta_{\mathrm{D}^n} * \beta_{\mathrm{D}} = \int_{\mathbb{R}} \beta_{\mathrm{D}^n}(t)\beta_{\mathrm{D}}(\cdot - t)\mathrm{d}t, \tag{1.7}$$

starting from β_{D}, which is the rectangular window over $[0, 1)$

$$\beta_{\mathrm{D}}(x) = \begin{cases} 1, & 0 \le x < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{1.8}$$

B-splines have also compact expressions in the Fourier domain. Indeed, by denoting $\widehat{\beta_{\mathrm{D}^n}}$ as the Fourier transform of β_{D^n} and rewriting (1.7) in the Fourier-domain, we observe that

$$\widehat{\beta_{\mathrm{D}^{n+1}}}(\omega) = \widehat{\beta_{\mathrm{D}^n}}(\omega)\left(\frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right). \tag{1.9}$$

Hence, one readily obtains the closed-form expression

$$\widehat{\beta_{\mathrm{D}^{n+1}}}(\omega) = \left(\frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right)^{n+1}, \quad n \in \mathbb{N}. \tag{1.10}$$

Using (1.10), we can also verify that the B-spline of degree n is indeed a spline of the same degree. The corresponding innovation is

$$\mathrm{D}^{n+1}\{\beta_{\mathrm{D}^{n+1}}\} = \sum_{k\in\mathbb{Z}} d_{\mathrm{D}^{n+1}}[k]\delta(\cdot - k), \tag{1.11}$$

L	$\rho_{\mathrm{L}}(x)$	$\beta_{\mathrm{L}}(x)$	d_{L}
D	$\mathbb{1}_{+}(x)$	$\beta_{\mathrm{D}}(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$	$(1, -1)$
D^2	x_{+}	$\beta_{\mathrm{D}^2}(x) = \begin{cases} x, & 0 \leq x < 1 \\ 2 - x, & 1 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$	$(1, -2, 1)$
D^3	$x_{+}^2/2$	$\beta_{\mathrm{D}^3}(x) = \begin{cases} x^2/2, & 0 \leq x < 1 \\ -x^2 + 3x - 3/2, & 1 \leq x < 2 \\ (3 - x^2)/2, & 2 \leq x < 3 \\ 0, & \text{otherwise} \end{cases}$	$(1, -3, 3, -1)$
D^4	$x_{+}^3/6$	$\beta_{\mathrm{D}^4}(x) = \begin{cases} x^3/6, & 0 \leq x < 1 \\ -x^3/2 + 2x^2 - 2x + 2/3, & 1 \leq x < 2 \\ x^3/2 - 4x^2 + 10x - 22/3, & 2 \leq x < 3 \\ (4 - x)^3/6, & 3 \leq x < 4 \\ 0, & \text{otherwise} \end{cases}$	$(1, -4, 6, -4, 1)$

Table 1.1: Characteristics of splines of degree n for $n = 0, 1, 2, 3$.

where $(d_{\mathrm{D}^{n+1}}[k])_{k \in \mathbb{Z}}$ is called the B-spline innovation filter. It is a discrete finite-impulse-response filter that is supported in $\{0, \dots, n+1\}$ and whose z transform is

$$D_{\mathrm{D}^{n+1}}(z) \triangleq \sum_{k \in \mathbb{Z}} d_{\mathrm{D}^{n+1}}[k] z^{-k} = (1 - z^{-1})^{n+1}. \tag{1.12}$$

Using this sequence, one readily verifies that

$$c = a * d_{\mathrm{D}^{n+1}}, \tag{1.13}$$

where the sequences $a[\cdot]$ and $c[\cdot]$ are defined in (1.5) and (1.6), respectively. We provide a summary of the relevant characteristics of B-splines for small values of n in Table 1.1. As shown in Figure 1.2, higher-order derivatives lead to smoother B-splines.

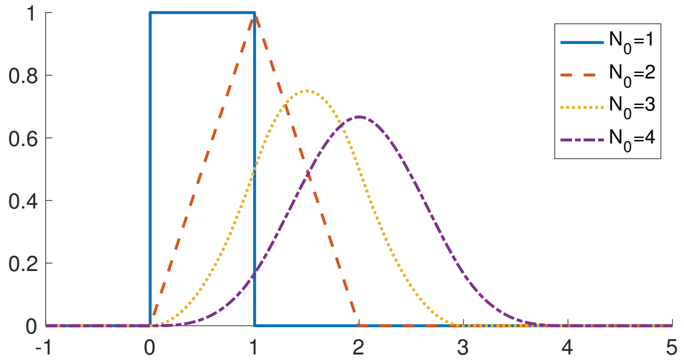


Figure 1.2: The causal B-spline associated to the operator $\mathrm{L} = \mathrm{D}^{N_0}$ for $N_0 = 1, 2, 3, 4$.

1.1.4 Approximation Power of Cardinal Splines

We now briefly discuss the approximation power of cardinal splines. To that end, let us denote by $S_n \subset L_2(\mathbb{R})$, the space of cardinal splines of degree n with integer knots and finite energy. Moreover, the integer-shift-invariant space generated by $\phi \in L_2(\mathbb{R})$ is denoted by $\mathrm{S}(\phi)$ and is defined as [13, 14, 15, 16]

$$\mathrm{S}(\phi) = \overline{\mathrm{span}}\left(\{\phi(\cdot - k)\}_{k\in\mathbb{Z}}\right) \subseteq L_2(\mathbb{R}). \tag{1.14}$$

Remarkably, S_n is an integer-shift-invariant space and is generated by the B-spline of degree n, $S_n = \mathrm{S}(\beta_{\mathrm{D}^{n+1}})$. Moreover, one can express S_n as

$$S_n = \left\{\sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{D}^{n+1}}(\cdot - k) : c[\cdot] \in \ell_2(\mathbb{Z})\right\}. \tag{1.15}$$

To verify this, we note that B-splines generate a Riesz basis (See, Definition 1.2, for more details), *i.e.* there exists $A, B > 0$ such that for any sequence $c \in \ell_2(\mathbb{Z})$, we have that

$$A\,\|c\|_{\ell_2} \leq \left\|\sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{D}^{n+1}}(\cdot - k)\right\|_{L_2} \leq B\,\|c\|_{\ell_2}. \tag{1.16}$$

To study the approximation power of splines, let

$$\mathrm{S}_h(\phi) = \{f(\cdot/h) : f \in \mathrm{S}(\phi)\} \tag{1.17}$$

be the h-dilate of $\mathrm{S}(\phi)$. The space $\mathrm{S}(\phi)$ is said to have an approximation power of order M if any sufficiently smooth and decaying function can be approached by an element of $\mathrm{S}_h(\phi)$ with an error decaying as $O(h^M)$. The so called "Strang-Fix conditions" give sufficient conditions to have a space with an approximation power of order M [1, 15, 17]. In particular, for compactly supported and integrable generating functions, it is sufficient to have the space $\mathrm{S}(\phi)$ reproduce polynomials of degree up to $(M-1)$.

Definition 1.1. *The space $\mathrm{S}(\phi)$ is said to reproduce polynomials of degree up to M if, for all $m = 0, 1, ..., M$, there exists sequences c_m (not necessarily in $(\ell_2(\mathbb{Z}))^N$) such that* [3]

$$\forall x \in \mathbb{R}, \quad x^m = \sum_{k\in\mathbb{Z}} c_m[k]\phi(x-k). \tag{1.18}$$

A straightforward implication is that the space S_n has an approximation power of order $(n+1)$ since

1. it can reproduce polynomials of degree up to n;
2. it can be generated by the compactly supported function $\beta_{\mathrm{D}^{n+1}}$.

To conclude this part, let us highlight the support property of B-splines. Notice that the B-spline of degree n is supported in $[0, n+1]$. Indeed, B-splines are known to be maximally localized, meaning that they are minimally-supported among functions with the same approximation order [18, 19]. This result is classical in approximation theory: Schoenberg showed that B-splines effectively have the adequate approximation order [5]. A complete characterization of the piecewise-polynomial functions of minimal support with a given approximation order can be found in [19, Theorem 1].

[3] for $m = 0$, we use in (1.18) the convention that $x^m = 1$, including for $x = 0$.

1.1.5 Spline Interpolation

To further illustrate the practical convenience of splines, we now detail an interpolation method based on B-splines. Consider the problem of finding a cardinal spline s of degree n that uniformly interpolates a finitely-supported sequence of data points $y[m], m \in \mathbb{Z}$ such that $s(m) = y[m]$ for all $m \in \mathbb{Z}$. Using the expansion (1.6) of s, one can rewrite the latter condition in terms of the B-spline coefficients as

$$y[m] = \sum_{k \in \mathbb{Z}} c[k] \beta_{\mathrm{D}^{n+1}}(m-k), \quad m \in \mathbb{Z}. \tag{1.19}$$

Taking the z-transform from both sides yields

$$Y(z) = C(z)B(z), \quad \forall z \in \mathbb{C} \setminus \{0\}, \tag{1.20}$$

where $Y(z)$, $C(z)$, and $B(z)$ denote the z-transforms of the sequences $y[k]$, $c[k]$, and $b[k] \triangleq \beta_{\mathrm{D}^{n+1}}(k)$, respectively. This means that the sequence $c[k]$ can be obtained via computing the inverse z-transform of $Y(z)/B(z)$. Remarkably, the latter inverse filtering can be implemented efficiently (see, for example, [6]).

1.1.6 Linear Splines: A Special Case

Finally, we review the introduced concepts by focusing on the special case of (first-degree) linear splines. These are, by definition, continuous and piecewise linear (CPWL) functions $\mathbb{R} \to \mathbb{R}$. As we shall see in Chapter 3, the notion of CPWL functions (which can be extended to higher dimensions) plays an important role in nonparametric supervised learning and deep learning theory.

The causal Green's function of degree 1 is the rectified linear unit (ReLU) [20], defined as

$$\mathrm{ReLU}(x) \triangleq x_+ = \max(x, 0). \tag{1.21}$$

This function has brought lots of attention in the past decade due to its usage as the activation functions of deep neural networks (see, Section 3.4). Following (1.7), the B-spline of degree 1 is the triangle function

$$\beta_{\mathrm{D}^2}(\cdot) = \mathrm{ReLU}(\cdot) - 2\mathrm{ReLU}(\cdot - 1) + \mathrm{ReLU}(\cdot - 2) \tag{1.22}$$

that is supported on $[0,2]$ (see, Figure 1.2). This, in particular, implies that the B-spline innovation filter d_{D^2}, introduced in (1.11), can be expressed as

$$d_{\mathrm{D}}^2[k] = \begin{cases} 1, & k = 0 \\ -2, & k = 1 \\ 1, & k = 2 \\ 0, & \text{otherwise.} \end{cases} \tag{1.23}$$

Remarkably, the B-spline of degree 1 is interpolatory. In other words, for any discrete sequence $y[k], k \in \mathbb{Z}$ of data, the cardinal linear spline $s_{\mathrm{int}} : \mathbb{R} \to \mathbb{R}$, defined as

$$s_{\mathrm{int}} = \sum_{k\in\mathbb{Z}} y[k]\beta_{\mathrm{D}^2}(\cdot - k) \tag{1.24}$$

satisfies $s_{\mathrm{int}}(m) = y[m]$ for $m \in \mathbb{Z}$. This is due to the fact that the integer samples of the linear B-spline coincides with the Kronecker delta sequence:

$$b[k] = \beta_{\mathrm{D}^{n+1}}(k) = \delta[k] \triangleq \begin{cases} 1, & k = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1.25}$$

1.1.7 Summary: Think Analog, Act Digital

We have provided a general overview of the theory of polynomial splines. Our main motivation was to allow the general reader to get a sense of what will follow in the rest of this thesis. We took the special case of linear splines as an elaborative example, due to their simple characteristics as well as their particular relevance to this thesis.

Finally, we would like to highlight that the notion of polynomial splines has been extended in various ways. A common thread among all these extensions is the bridge they create between continuous and discrete worlds. These bridges are shown to be optimal in various senses, some of which shall be demonstrated in this thesis.

1.2 Hermite Splines

Introduced by I.J. Schoenberg [21, 22], the Hermite interpolation problem involves two sequences of discrete samples. They impose constraints not only on the resulting interpolated function but also on its derivatives up to a given order. Hermite splines enjoy excellent approximation power while retaining interpolation properties and closed-form expression, in contrast to existing similar functions. In this section[4], we further demonstrate that Hermite splines are maximally localized, in the sense that the size of their support is minimal among pairs of functions with identical reproduction properties. This sheds a new light on the convenience of Hermite splines in the context of computer graphics and geometrical design.

1.2.1 Background

Schoenberg defines the cardinal cubic-Hermite-interpolation problem as follows [21, 22]: knowing the discrete sequences of numbers $c[k]$ and $d[k]$, $k \in \mathbb{Z}$, we look for a continuously defined function $f_{\mathrm{Her}} : \mathbb{R} \to \mathbb{R}$, that satisfies

$$f_{\mathrm{Her}}(k) = c[k], \qquad f'_{\mathrm{Her}}(k) = d[k] \tag{1.26}$$

for all $k \in \mathbb{Z}$, such that f_{Her} is piecewise polynomial of degree at most 3 and once differentiable with continuous derivative at the integers. The existence and uniqueness of the solution is guaranteed [21, Theorem 1] for any sequences c and d bounded by a polynomial. In [22], it is shown that the Hermite spline f_{Her} associated to the sequences c and d can be expressed as

$$f_{\mathrm{Her}}(t) = \sum_{k \in \mathbb{Z}} \left(c[k]\phi_1(t-k) + d[k]\phi_2(t-k) \right), \tag{1.27}$$

where the functions ϕ_1 and ϕ_2 are given by

$$\phi_1(t) = (2|t|+1)(|t|-1)^2 \mathbb{1}_{0 \le |t| \le 1}, \qquad \phi_2(t) = t(|t|-1)^2 \mathbb{1}_{0 \le |t| \le 1}. \tag{1.28}$$

[4] This section is based on our published work [1].

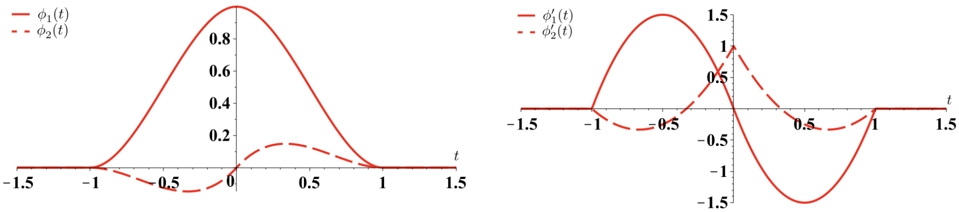


Figure 1.3: Cubic Hermite splines ϕ_1 and ϕ_2. The two functions and their derivatives are vanishing at the integers, with the exception of $\phi_1(0) = 1$ and $\phi_2'(0) = 1$ (interpolation properties). They are supported in $[-1, 1]$.

In addition to their fairly simple analytical expression, the cubic Hermite splines have other important properties. First, they are of finite support in $[-1, 1]$. Moreover, the generating functions ϕ_1, ϕ_2 and their derivatives ϕ_1', ϕ_2' satisfy the joint interpolation conditions

$$\phi_1(k) = \delta[k], \quad \phi_2'(k) = \delta[k], \quad \phi_1'(k) = 0, \quad \phi_2(k) = 0, \tag{1.29}$$

for all $k \in \mathbb{Z}$. The functions and their first derivative are depicted in Figure 1.3, where the interpolation properties can easily be observed. The functions ϕ_1 and ϕ_2 are deeply intertwined as $c[k] = f_{\mathrm{Her}}(k)$ and $d[k] = f'_{\mathrm{Her}}(k)$ in (1.27). The cubic Hermite splines are differentiable with continuous derivatives (C^1-regular) at the integer knots points $t = k$. As a result, functions generated by cubic Hermite splines are C^1-regular piecewise-cubic polynomials with knots at integer locations.

1.2.2 Context

Curves in the plane can be constructed from one-dimensional interpolation schemes by interpolating along each spatial coordinate (see, for example, Section 4.4). The practical value of Hermite splines in this context is to offer tangential control on the interpolated curve. This can be easily understood through their link with Bézier curves [23]. The latter lie at the heart of vector graphics and are popular tools for computer-aided geometrical design and modeling [24, 25, 26]. Because of their small support, Hermite splines are also an interesting option for the design of multiwavelets, which are wavelets with multiple generators [27, 28]. In practice, Hermite splines thus provide a suitable solution to a number of problems, whether with respect to simplicity of construction, efficiency, or convenience. This hands-on intuition can be translated to formal properties of Hermite splines and mathematically characterized. We give as examples the joint interpolation properties of Hermite splines that ensure that, at integer values, the interpolated function exactly matches the sequences of samples and derivative samples that were used to build it; their smoothness properties [29], which guarantee low curvature of the interpolated curve under some mild conditions; and their statistical optimality (in terms of MMSE) for the reconstruction of second-order Brownian motion from direct and first-derivative samples [30].

In that spirit, we investigate the minimal-support property of Hermite splines. The short support of Hermite splines is an important feature that makes them attractive in practice. The size of the support relates to the local extent of modifications on the continuously defined spline curve. More precisely, the modification of a coefficient in the sample sequence only affects the interpolated curve at a distance that equals half of the support size. In the Hermite case, a coefficient k therefore only influences the properties of the curve between coefficients $(k-1)$ and $(k+1)$, which corresponds to the best case scenario. We formally demonstrate that Hermite splines have the minimal support among all basis functions that generate cubic and quadratic splines.

1.2.3 Main Results

The Hermite splines ϕ_1 and ϕ_2, given by (1.28), are able to reproduce both β_{D^3} and β_{D^4}, the quadratic and cubic B-splines, respectively [29]. Many other pairs of basis functions, starting with β_{D^3} and β_{D^4} themselves, can also reproduce quadratic and cubic splines. The investigation of the support of a pair of functions that have the same reproduction properties as the Hermite splines follows naturally. We characterize the support of such pairs of functions in Theorem 1.1. This result then allows us to deduce the minimal-support property of Hermite splines in Corollary 1.1.

Theorem 1.1. *Let $\varphi_1, \varphi_2 \in L_2(\mathbb{R})$ be two compactly supported basis functions. We assume that*

$$\beta_{\mathrm{D}^3}(t) = \sum_{k\in\mathbb{Z}} \left(a[k]\varphi_1(t-k) + b[k]\varphi_2(t-k)\right), \tag{1.30}$$

$$\beta_{\mathrm{D}^4}(t) = \sum_{k\in\mathbb{Z}} \left(c[k]\varphi_1(t-k) + d[k]\varphi_2(t-k)\right), \tag{1.31}$$

with reproduction sequences a, b, c, d that satisfy

$$\sum_{k\in\mathbb{Z}} k^3(|a[k]| + |b[k]| + |c[k]| + |d[k]|) < +\infty. \tag{1.32}$$

Then, we have that

$$|\mathrm{supp}(\varphi_1)| + |\mathrm{supp}(\varphi_2)| \geq 4. \tag{1.33}$$

Proof. Firstly, we can restrict ourselves to compactly supported basis functions φ_1 and φ_2 (otherwise, $|\mathrm{supp}(\varphi_1)| + |\mathrm{supp}(\varphi_2)| = \infty$). Moreover, if one of the basis function, for instance φ_2, is identically zero, φ_1 can reproduce β_{D^4} and, subsequently, polynomials up to degree 3. Hence, its support is at least of size four [19, Theorem 1]. Thus, we now assume that φ_1 and φ_2 are compactly supported functions that are not identically 0.

Step 1. We show that the extreme points of the supports of φ_1 and φ_2 are integers. For $x = a, b, c, d$, we set $X(\omega) = \sum_{k\in\mathbb{Z}} x[k]\mathrm{e}^{-\mathrm{j}\omega k}$, the 2π-periodic Fourier transform of the sequence x. Condition (1.32) ensures that X has a periodic continuous third derivative. In the Fourier domain, (1.30) and (1.31) become

$$\widehat{\beta_{\mathrm{D}^3}}(\omega) = \frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}{(\mathrm{j}\omega)^3} = A(\omega)\widehat{\varphi_1}(\omega) + B(\omega)\widehat{\varphi_2}(\omega), \tag{1.34}$$

$$\widehat{\beta_{\mathrm{D}^4}}(\omega) = \frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^4}{(\mathrm{j}\omega)^4} = C(\omega)\widehat{\varphi_1}(\omega) + D(\omega)\widehat{\varphi_2}(\omega). \tag{1.35}$$

We set $\det(\omega) = (A(\omega)D(\omega) - B(\omega)C(\omega))$, which is itself a function with continuous third derivative. From (1.34) and (1.35), we obtain that

$$\det(\omega)\widehat{\varphi}_1(\omega) = D(\omega)\frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}{(\mathrm{j}\omega)^3} - B(\omega)\frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^4}{(\mathrm{j}\omega)^4}, \tag{1.36}$$

$$\det(\omega)\widehat{\varphi}_2(\omega) = -C(\omega)\frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}{(\mathrm{j}\omega)^3} + A(\omega)\frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^4}{(\mathrm{j}\omega)^4}. \tag{1.37}$$

From (1.36), we deduce that, at least when $\det(\omega)$ does not vanish, we have the relation

$$(\mathrm{j}\omega)^4\widehat{\varphi}_1(\omega) = (\mathrm{j}\omega)F(\omega)D(\omega) - (1-\mathrm{e}^{-\mathrm{j}\omega})F(\omega)B(\omega), \tag{1.38}$$

where $F(\omega) = \frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}{\det(\omega)}$. The strategy of the proof is to show that the function F is continuous and periodic, and that (1.38) is therefore valid for any $\omega \in \mathbb{R}$. We study F in two steps: (i) first, we show that $\det(\omega)$ does not vanish for $\omega \notin 2\pi\mathbb{Z}$; and (ii) we then demonstrate that F has a limit at 0 by considering the Taylor expansion of $\det(\omega)$.

(i) Let us start with the first issue. We show that $\det(\omega) \neq 0$ for $\omega \notin 2\pi\mathbb{Z}$ by contradiction. Let us fix $\omega_0 \in (0, 2\pi)$ and assume that $\det(\omega_0) = 0$. We set $\alpha = \frac{1-\mathrm{e}^{-\mathrm{j}\omega_0}}{\mathrm{j}\omega_0}$ and $\gamma = \frac{1-\mathrm{e}^{-\mathrm{j}\omega_0}}{\mathrm{j}(\omega_0+2\pi)}$. Then, $\alpha \neq 0$, $\gamma \neq 0$, and $\alpha \neq \gamma$, while, by periodicity, $\det(\omega_0) = \det(\omega_0 + 2\pi) = 0$. Hence, (1.36) for $\omega = \omega_0$ and $(\omega_0 + 2\pi)$ implies that

$$\begin{pmatrix}\alpha^3 & -\alpha^4\\ \gamma^3 & -\gamma^4\end{pmatrix}\begin{pmatrix}D(\omega_0)\\ B(\omega_0)\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \tag{1.39}$$

The matrix being invertible (with determinant $\alpha^3\gamma^3(\alpha-\gamma) \neq 0$), we deduce that $D(\omega_0) = B(\omega_0) = 0$. Similary, (1.37) with $\omega = \omega_0$ and $(\omega_0 + 2\pi)$ implies that $A(\omega_0) = C(\omega_0) = 0$. Injecting this in (1.34) with $\omega = \omega_0$, we deduce that $\widehat{\beta_{\mathrm{D}^3}}(\omega_0) = \alpha^3 = 0$, which contradicts our initial assumption.

(ii) We now study $\det(\omega)$ around the origin. The function admits a third-order continuous derivative; hence, it can be McLaurin expanded at 0 as

$$\det(\omega) = \det(0) + \det{}^{(1)}(0)\omega + \frac{1}{2}\det{}^{(2)}(0)\omega^2 + \frac{1}{6}\det{}^{(3)}(0)\omega^3 + o(\omega^3). \tag{1.40}$$

Assume by contradiction that $\det^{(p)}(0) = 0$ for $p = 0, 1, 2, 3$. Then, (1.40) gives that $\det(\omega) = o(\omega^3)$ around 0. From (1.36), we remark that

$$\frac{\det(\omega)(\mathrm{j}\omega)^3}{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}\widehat{\varphi}_1(\omega) = D(\omega) - B(\omega)\frac{1-\mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}. \tag{1.41}$$

The function $\frac{\det(\omega)(\mathrm{j}\omega)^3}{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}$ vanishes at $\omega = 0$ because det does and $\lim_{\omega\to 0}\frac{(\mathrm{j}\omega)^3}{(1-\mathrm{e}^{-\mathrm{j}\omega})^3} = 1$. Therefore, the left term in (1.41) vanishes when ω goes to 0. Now, by periodicity, around 2π, we have that $\det(\omega) = o((\omega-2\pi)^3)$. Hence, $\frac{\det(\omega)}{(1-\mathrm{e}^{-\mathrm{j}\omega})^3} = o(1)$ around 2π and, again, the left term in (1.41) is also vanishing when ω goes to 2π.

We deduce that the right term in (1.41) vanishes for both $\omega = 0$ and $\omega = 2\pi$. In other terms, we have that

$$0 = D(0) - B(0) = D(2\pi) - 0 \tag{1.42}$$

and, since $D(2\pi) = D(0)$ by periodicity, this implies that $D(0) = B(0) = 0$. A similar reasoning shows that $A(0) = C(0) = 0$. From (1.34) with $\omega = 0$, we obtain that $\widehat{\beta_{\mathrm{D}^4}}(0) = 0$, which is false.

As a consequence, at least one of the derivative of the McLaurin expansion (1.40) is nonzero, from which we easily deduce that $F(\omega) = \frac{(1-\mathrm{e}^{-\mathrm{j}\omega})^3}{\det(\omega)}$ has a limit (possibly 0) at the origin. The function F is well-defined and continuous for $\omega \notin 2\pi\mathbb{Z}$, continuously extendable at 0, and is therefore a continuous periodic function.

At this stage, we obtained that (1.38) is valid for any $\omega \in \mathbb{R}$. The functions $F(\omega)D(\omega)$ and $(1-\mathrm{e}^{-\mathrm{j}\omega})F(\omega)B(\omega)$ are 2π-periodic, hence their inverse Fourier transforms are sums of Dirac impulses located at the integers. It means in particular that we have, in the time domain, that

$$\varphi_1^{(4)}(t) = \sum_{k\in\mathbb{Z}}\left(y[k]\delta(t-k) + z[k]\delta'(t-k)\right), \tag{1.43}$$

where y and z are the inverse Fourier sequences of $(1-\mathrm{e}^{-\mathrm{j}\omega})F(\omega)B(\omega)$ and $F(\omega)D(\omega)$, respectively. Since $\varphi_1^{(4)}$ is compactly supported, like φ_1, only finitely many entries of y and z are non-zero. Then, φ_1 is a compactly supported function whose fourth derivative has a support with integer extreme points (due to (1.43)), and therefore has a support with integer extreme points, too. The same reasoning applies for φ_2, which concludes this part of the proof.

Step 2. We know that the supports of φ_1 and φ_2 are of the form $[a, b]$ with $a < b$, $a, b \in \mathbb{Z}$. By contradiction, we assume that $|\mathrm{supp}(\varphi_1)| + |\mathrm{supp}(\varphi_2)| < 4$. Then, one of the two basis functions has a support of size one, for instance φ_1. We also assume without loss of generality that $\mathrm{supp}(\varphi_1) = [0, 1]$, implying that only $y[0], y[1], z[0], z[1]$ are possibly nonzero in (1.43). Going back to the Fourier domain, one obtains that

$$(\mathrm{j}\omega)^4 \widehat{\varphi}_1(\omega) = y[0] + y[1]\mathrm{e}^{-\mathrm{j}\omega} + \mathrm{j}\omega(z[0] + z[1]\mathrm{e}^{-\mathrm{j}\omega}). \tag{1.44}$$

The function φ_1 is compactly supported. Its Fourier transform is hence infinitely smooth, and we can do the McLaurin expansion of both sides in (1.44) up to order 3. This gives

$$\begin{aligned} o(\omega^3) =&(y[0] + y[1]) + \mathrm{j}\omega(-y[1] + z[0] + z[1]) \\ &+ \omega^2(-y[1]/2 + z[1]) + \mathrm{j}\omega^3(y[1]/6 - z[1]/2) + o(\omega^3). \end{aligned} \tag{1.45}$$

In particular, we obtain the relations

$$y[0] + y[1] = z_[0] + z[1] - y[1] = z[1] - \frac{y[1]}{2} = \frac{y[1]}{6} - \frac{z[1]}{2} = 0. \tag{1.46}$$

This imposes that $y[0] = y[1] = z[0] = z[1] = 0$, which is absurd. Finally, it shows that $|\mathrm{supp}(\varphi_1)| + |\mathrm{supp}(\varphi_2)| \geq 4$, as expected. □

Remark 1.1. *Condition* (1.32) *plays an important role in our proof by imposing some regularity in the Fourier domain. In practice, one even expects that compactly supported basis functions can generate the B-splines β_{D^3} and β_{D^4} with finitely many coefficients, in which case* (1.32) *is automatically satisfied. Later in Section 1.3, we shall prove a more general result (Theorem 1.3) from which we can deduce a stronger version of Theorem 1.1 without the additional constraint* (1.32)*.*

From Theorem 1.1, we easily deduce that Hermite splines have the minimal-support property.

Corollary 1.1. *The Hermite splines* (ϕ_1, ϕ_2) *are of minimal support among the pairs of functions that are able to reproduce both quadratic and cubic B-splines with reproduction sequences satisfying* (1.32)*.*

Proof. From [29, Appendix A], we know that Hermite splines can reproduce both quadratic and cubic B-splines with compactly supported reproduction sequences. Moreover, the supports of ϕ_1 and ϕ_2 are of size two, which implies that $|\mathrm{supp}(\phi_1)| + |\mathrm{supp}(\phi_2)| = 4$. We conclude the proof by invoking Theorem 1.1. □

It is worth noting that the supports of the pair of Hermite splines jointly has the same size as the B-spline β_{D^4}. However, β_{D^4} alone has lesser reproduction properties. Being of class C^2, it is in particular unable to reproduce the quadratic spline β_{D^3}, which only has C^1 transitions at the integers. The simplest way of reproducing $\beta_{\mathrm{D}^3}, \beta_{\mathrm{D}^4}$ is to consider the basis pair $(\beta_{\mathrm{D}^3}, \beta_{\mathrm{D}^4})$ itself, which is not maximally localized since the sum of the supports is 7.

Last but not least, we show that two basis functions are needed to reproduce both cubic and quadratic spline, as formalized in Theorem 1.2.

Theorem 1.2. *There exists no single function* $\varphi \in L_2(\mathbb{R})$ *that is able to reproduce* β_{D^3} *and* β_{D^4} *with summable reproduction sequences.*

Proof. By contradiction, let us assume that there exists φ such that $\beta_{\mathrm{D}^3} = \sum_{k\in\mathbb{Z}} a[k]\varphi(\cdot - k)$ and $\beta_{\mathrm{D}^4} = \sum_{k\in\mathbb{Z}} b[k]\varphi(\cdot - k)$ with $a, b \in \ell_1(\mathbb{Z})$. Then, the Fourier transforms $A(\omega)$ and $B(\omega)$ are continuous 2π-periodic functions. In the Fourier domain, we have that

$$\left(\frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right)^3 = A(\omega)\widehat{\varphi}(\omega), \quad \left(\frac{1 - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right)^4 = B(\omega)\widehat{\varphi}(\omega). \tag{1.47}$$

Set $\omega_0 \in (0, 2\pi)$ and $\omega_1 = \omega_0 + 2\pi$. The relation (1.48) imposes that $A(\omega_i)$, $B(\omega_i)$, and $\widehat{\varphi}(\omega_i)$ are non-zero for $i = 1, 2$, and

$$\frac{1 - \mathrm{e}^{-\mathrm{j}\omega_i}}{\mathrm{j}\omega_i} A(\omega_i)\widehat{\varphi}(\omega_i) = B(\omega_i)\widehat{\varphi}(\omega_i). \tag{1.48}$$

After simplifications, we deduce that

$$\mathrm{j}\omega_i = \frac{(1 - \mathrm{e}^{-\mathrm{j}\omega_i})A(\omega_i)}{B(\omega_i)}. \tag{1.49}$$

The right term in (1.50) is equal for ω_0 and ω_1 by periodicity, while the left term is not. This contradicts our initial assumption and implies Theorem 1.2. □

1.2.4 Summary

When dealing with B-splines, there is a tradeoff between the ability to reproduce smooth functions, which increases with the degree of B-splines, and the possibility to allow for sharper transitions, which decreases with the degree. On one hand, cubic splines can efficiently reproduce smooth functions and are able to capture C^2 transitions, but lack the power to capture C^1 transitions. On the other hand, quadratic splines have a lesser approximation power, but are preferred when dealing with less smooth (C^1) transitions. Hermite splines combine these two strengths in one scheme and are, in terms of support size, better than any two-function scheme, including the one composed of the classical cubic and quadratic B-splines. In addition, we also show that one necessarily requires two generators to achieve this optimality. These results demonstrate that Hermite splines are maximally localized for the purpose of representing piecewise-cubic functions that are continuously differentiable, as exploited, for instance, in image processing for the design of deformable parametric contours [29].

1.3 Multi-Splines

In this section[5], we introduce the notion of multi-splines and we discuss their role in generalized sampling. Generalized sampling, consists in the recovery of a function f from the samples of the responses of a collection of linear shift-invariant systems to the input f. The reconstructed function is typically a member of a finitely generated integer-shift-invariant space that can reproduce polynomials up to a given degree M. While this property allows for an approximation power of order $(M+1)$, it comes with a tradeoff on the length of the support of the basis functions. Specifically, we prove that the sum of the length of the support of the generators is at least $(M+1)$. Following this result, we introduce the notion of shortest basis of degree M, which is motivated by our desire to minimize computational costs. We then demonstrate that any basis of shortest support generates a Riesz basis. Finally, we introduce a recursive algorithm to construct the shortest-support basis for any multi-spline space. It provides a generalization of both polynomial (Section 1.1) and Hermite splines (Section 1.2). This framework paves the way for novel applications such as fast derivative sampling with arbitrarily high approximation power.

1.3.1 Context

Since the formulation of Nyquist-Shannon's celebrated sampling theorem [31], the reconstruction of a function from discrete measurements has been extended in many ways [32, 33]. In particular, Papoulis proposed the framework of generalized sampling [34], where he showed that any bandlimited function f is uniquely determined by the sequences of discrete measurements (generalized samples)

$$g_n(kT) = (h_n * f)(kT) = \langle f, \psi_n(\cdot - kT)\rangle, \quad n = 1, ..., N, \quad k \in \mathbb{Z}, \tag{1.50}$$

where $(g_n(t))_{n=1,...,N}$ are the outcome of N linearly independent systems applied to f. The sampling is assumed to proceed at $1/N$ the Nyquist rate (*i.e.*, $T = NT_{\mathrm{Nyq}} = 2N\pi/\omega_{\max}$, where $\omega_{\max}$ is the maximum frequency of f). The functions $\psi_n(t) = h_n(-t)$, $t \in \mathbb{R}$, are called the analysis functions. They are the time-reversed versions of the impulse responses. The sampling theorem was also generalized

[5] This section is based on our published work [2].

to many different function spaces such as integer-shift-invariant spaces [35, 36], including spline spaces [37, 38, 39]. Following this extension and Papoulis' theory, Unser and Zerubia introduced a framework to perform generalized sampling without the bandlimited constraint [40, 41] which includes important cases such as interlaced and derivative sampling in spline spaces.

Building on top of an impressive body of work from various communities, we propose a systematic study of shortest bases for any multi-spline space. In particular, the main goal is to generalize the concept of B-splines to any multi-spline space. To that end, we first study finitely-generated shift-invariant spaces and the desirable properties of their generating functions. We then focus on multi-spline spaces as a special case, where we present a method to construct their shortest-support bases. Finally, we illustrate the applicability of our introduced framework with some examples.

1.3.2 Finitely Generated Shift-Invariant Spaces

Here, we adopt the framework of generalized sampling and propose to reconstruct a function f from discrete samples $g_n(k), k = 1, ..., N$ in an integer-shift-invariant subspace of $L_2(\mathbb{R})$ as in some recent works [42, 43, 44]. The structure of such reconstruction spaces have been thoroughly studied [13, 45, 46] and there exists theoretical results that lead to the critical choice of relevant generating functions [14].

For a finite collection $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_N)$ of functions in $L_2(\mathbb{R})$, the generated integer-shift-invariant subspace is denoted by $\mathrm{S}(\boldsymbol{\phi})$ and is defined as

$$\mathrm{S}(\boldsymbol{\phi}) = \mathrm{S}(\phi_1) + \mathrm{S}(\phi_2) + \cdots + \mathrm{S}(\phi_N), \tag{1.51}$$

where we recall from (1.14) that

$$\mathrm{S}(\phi_n) = \overline{\mathrm{span}}\left(\{\phi_n(\cdot - k)\}_{k\in\mathbb{Z}}\right) \subseteq L_2(\mathbb{R}), \quad n = 1, \dots, N. \tag{1.52}$$

The space $\mathrm{S}(\boldsymbol{\phi})$ is *integer-shift-invariant* in the sense that $f(\cdot - k)$ is in $\mathrm{S}(\boldsymbol{\phi})$ for $f \in \mathrm{S}(\boldsymbol{\phi})$ and $k \in \mathbb{Z}$ [13, 14, 15, 16]. Ideally, we would like $\boldsymbol{\phi}$ to generate a Riesz basis.

Definition 1.2. *The set of functions* $\{\phi_n(\cdot - k) : k \in \mathbb{Z}, n = 1, \dots, N\} \subset L_2(\mathbb{R})$ *is said to be a Riesz basis with bounds* $A, B \in \mathbb{R}$ *with* $0 < A \leq B < +\infty$ *if, for any vector of square-summable sequences* $\boldsymbol{c} = (c_1, ..., c_N) \in (\ell_2(\mathbb{Z}))^N$*, we have that*

$$A \|\boldsymbol{c}\|_{\ell_2} \leq \left\| \sum_{k\in\mathbb{Z}} \boldsymbol{c}[k]^T \boldsymbol{\phi}(\cdot - k) \right\|_{L_2(\mathbb{R})} \leq B \|\boldsymbol{c}\|_{\ell_2}, \tag{1.53}$$

where $\|\boldsymbol{c}\|_{\ell_2} = \left(\sum_{n=1}^N \|c_n\|_{\ell_2}^2\right)^{\frac{1}{2}}$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_N)$ *and where* A *and* B *are the tightest constants.*

When this property is satisfied, we say that $\boldsymbol{\phi}$ generates a Riesz basis. The Riesz-basis property guarantees that any $f \in \mathrm{S}(\boldsymbol{\phi})$ has the unique and stable representation [47]

$$f(\cdot) = \sum_{k\in\mathbb{Z}} \boldsymbol{c}[k]^T \boldsymbol{\phi}(\cdot - k) = \sum_{k\in\mathbb{Z}} \sum_{n=1}^N c_n[k] \phi_n(\cdot - k). \tag{1.54}$$

This property is well characterized in the Fourier domain via the Gramian matrix-valued function

$$\hat{\boldsymbol{G}}(\omega) = \sum_{k\in\mathbb{Z}} \hat{\boldsymbol{\phi}}(\omega + 2k\pi) \hat{\boldsymbol{\phi}}(\omega + 2k\pi)^H = \sum_{k\in\mathbb{Z}} \langle \boldsymbol{\phi}, \boldsymbol{\phi}^T(\cdot - k)\rangle \mathrm{e}^{-\mathrm{j}\omega k}, \tag{1.55}$$

where the last equality follows from Poisson's formula applied to the sampling at the integers of the matrix-valued autocorrelation function $t \mapsto \langle \boldsymbol{\phi}, \boldsymbol{\phi}^T(\cdot - t)\rangle$[48]. The Fourier equivalent of the Riesz-basis condition is [45]

$$0 < A^2 = \operatorname*{ess\,inf}_{\omega\in[0,2\pi)} \lambda_{\min}(\omega) \leq \operatorname*{ess\,sup}_{\omega\in[0,2\pi)} \lambda_{\max}(\omega) = B^2 < +\infty, \tag{1.56}$$

where $\lambda_{\min}(\omega)$ and $\lambda_{\max}(\omega)$ are the smallest and largest eigenvalues of $\hat{\boldsymbol{G}}(\omega)$.

1.3.3 Shortest Bases of Shift-Invariant Spaces

We now introduce the notion of shortest bases. We recall that for a single generator ϕ such that $\mathrm{S}(\phi)$ reproduces polynomials of degree up to M, Schoenberg stated that $|\operatorname{supp}(\phi)| \geq M + 1$ [5]. We now extend this result to any $N \in \mathbb{N} \setminus \{0\}$.

Theorem 1.3 (Minimal support)**.** *If* $\mathrm{S}(\boldsymbol{\phi}) = \mathrm{S}(\phi_1, \phi_2, \ldots, \phi_N)$ *reproduces polynomials of degree up to* M*, then* $|\operatorname{supp}(\boldsymbol{\phi})| \geq M+1$*, where* $|\operatorname{supp}(\boldsymbol{\phi})| = \sum_{n=1}^{N} |\operatorname{supp}(\phi_n)|$*. In addition, if there is equality, then*

$$\sum_{k\in\mathbb{Z}}\sum_{n=1}^{N} \mathbb{1}_{\operatorname{supp}(\phi_n)}(x+k) = |\operatorname{supp}(\boldsymbol{\phi})| \quad \textit{for almost every } x \in \mathbb{R}. \tag{1.57}$$

Proof. If $\boldsymbol{\phi}$ is not compactly supported, then the inequality is clear. Now, we can assume that $\boldsymbol{\phi}$ is compactly supported. This implies that, for any $x \in \mathbb{R}$, the sum $\sum_{k\in\mathbb{Z}} \boldsymbol{c}[k]^T \boldsymbol{\phi}(x-k) = \sum_{k\in\mathbb{Z}}\sum_{n=1}^{N} c_n[k]\phi_n(x-k)$ has only a finite number of nonzero terms that are identified by the set

$$\Lambda(x) = \{(n,k) \in \{1,\ldots,N\} \times \mathbb{Z} : \quad x \in \operatorname{supp}(\phi_n(\cdot - k))\}, \tag{1.58}$$

and its cardinality

$$\lambda(x) = \sum_{k\in\mathbb{Z}}\sum_{n=1}^{N} \mathbb{1}_{\operatorname{supp}(\phi_n)}(x+k) \in \mathbb{N}. \tag{1.59}$$

Equation (1.60) follows from the fact that $\mathbb{1}_{\operatorname{supp}(\phi_n)}(x+k)$ is 1 if and only if $(n,k) \in \Lambda(x)$ and 0 otherwise. The function $x \mapsto \lambda(x)$ is 1-periodic and bounded because $\operatorname{supp}(\phi_n)$ are compact subsets of $\mathbb{R}$. Its average over one period reads (note that the sums are in fact all finite)

$$\overline{\lambda} = \int_0^1 \sum_{n=1}^{N}\sum_{k\in\mathbb{Z}} \mathbb{1}_{\operatorname{supp}(\phi_n)}(x+k)\mathrm{d}x = \sum_{n=1}^{N}\sum_{k\in\mathbb{Z}} \int_0^1 \mathbb{1}_{\operatorname{supp}(\phi_n)}(x+k)\mathrm{d}x \tag{1.60}$$

$$= \sum_{n=1}^{N} \int_{-\infty}^{\infty} \mathbb{1}_{\operatorname{supp}(\phi_n)}(x)\mathrm{d}x = |\operatorname{supp}(\boldsymbol{\phi})|, \tag{1.61}$$

where we applied Fubini's Theorem in (1.61). Because λ is bounded and takes values in $\mathbb{N}$, it only takes a finite number of values. Consequently, there exists a set $\mathrm{A} \subset [0,1]$ of nonzero measure such that λ is constant on A and no greater than its average, as in

$$\forall x \in A : \quad \lambda(x) = \lambda_A \leq \overline{\lambda} = |\operatorname{supp}(\boldsymbol{\phi})|. \tag{1.62}$$

Noting that A is bounded and that the ϕ_n are compactly supported, the image of A under Λ, denoted by $\Lambda(A)$, is a finite set. Therefore, there exists $\mathrm{B} \subset \mathrm{A} \subset [0, 1]$ of nonzero measure such that Λ is constant on B. This means that the set $\mathrm{S}(\boldsymbol{\phi})_{|B}$ of functions of $\mathrm{S}(\boldsymbol{\phi})$ restricted to B is spanned by λ_A functions $(\phi_n(\cdot - k))_{(n,k)\in\Lambda(B)}$.

Moreover, due to the reproducing property, the polynomials of degree up to M restricted to B form a linear subspace of $\mathrm{S}(\boldsymbol{\phi})_{|B}$ whose dimension is $(M+1)$, because B is infinite. Then, we must have that $\lambda_A \geq M + 1$ and, since $\lambda_A \leq |\operatorname{supp}(\boldsymbol{\phi})|$, we deduce the announced bound $|\operatorname{supp}(\boldsymbol{\phi})| \geq M + 1$.

If λ is not *a.e.* constant, then A can be chosen so that $\lambda_A < \overline{\lambda} = |\operatorname{supp}(\boldsymbol{\phi})|$ and $\mathrm{S}(\boldsymbol{\phi})_{|B}$ is spanned by fewer than $|\operatorname{supp}(\boldsymbol{\phi})|$ functions. The reproduction property implies that $|\operatorname{supp}(\boldsymbol{\phi})| > M + 1$. This means that the equality $|\operatorname{supp}(\boldsymbol{\phi})| = M + 1$ is possible only if λ is *a.e.* constant. □

Following Theorem 1.3, we can introduce the central notion of shortest-support basis.

Definition 1.3. *A collection of functions $\boldsymbol{\phi} \in (L_2(\mathbb{R}))^N$ is said to be a shortest-support basis of degree M if* $\mathrm{S}(\boldsymbol{\phi})$ *reproduces polynomials of degree up to M with the shortest support, i.e. with $|\operatorname{supp}(\boldsymbol{\phi})| = M + 1$.*

The qualifier of basis comes from Theorem 1.4.

Theorem 1.4. *Any shortest-support basis generates a Riesz basis.*

Proof. We define the kth slice of any function f as

$$\forall x \in \mathbb{R}: \quad \mathrm{S}_k\{f\}(x) = \begin{cases} f(x+k), & x \in [0,1) \\ 0, & \text{otherwise}, \end{cases} \tag{1.63}$$

and the set of nonzero slices of all the generating functions as

$$\mathcal{T}(\boldsymbol{\phi}) = \{\mathrm{S}_k\{\phi_n\} : \mathrm{S}_k\{\phi_n\} \not\equiv 0 \text{ and } k \in \mathbb{Z}, n = 1, ..., N\}. \tag{1.64}$$

Step 1. We first show that $\mathcal{F}(\boldsymbol{\phi})$ consists of linearly independent functions. Equivalently, we prove that if $\mathcal{T}(\boldsymbol{\phi})$ is not a set of linearly independent functions, then

$\boldsymbol{\phi}$ is not a shortest-support basis. To that end, suppose that $\mathcal{T}(\boldsymbol{\phi})$ is not a set of linearly independent functions. This means that one can find a slice, say $\mathrm{S}_{k_0}\{\phi_{q_0}\}$, that depends linearly on the others. Now, consider the integer-shift-invariant space generated by the set of functions $\mathcal{T}(\boldsymbol{\phi})\backslash\{\mathrm{S}_{k_0}\{\phi_{q_0}\}\}$. Note that the new generating functions differ now both in size (support size of at most 1) and in number (possibly greater than N). On one hand, the new integer-shift-invariant space is larger than the initial space and, in particular, is still able to reproduce polynomials of degree up to M. On the other hand, the sum of the support size of the generating functions is smaller than $|\operatorname{supp}(\boldsymbol{\phi})|$ because a nonzero slice was removed. So, $\boldsymbol{\phi}$ cannot be of minimal support.

Step 2. To complete the proof, we show that if $\mathcal{T}(\boldsymbol{\phi})$ is a set of linearly independent functions, then $\boldsymbol{\phi}$ generates a Riesz basis. We note that the generating functions can be expressed in terms of their slices as $\phi_n(x) = \sum_{k\in\mathbb{Z}} \mathrm{S}_k\{\phi_n\}(x-k)$. The Riesz-basis property is best characterized in the Fourier domain with the Gramian matrix (note that, $\boldsymbol{\phi}$ being compactly supported, all the sums are in fact finite), which leads to

$$\begin{aligned}
(\hat{\boldsymbol{G}}(\omega))_{mn} &= \sum_{q\in\mathbb{Z}} \langle \phi_m, \phi_n(\cdot - q)\rangle \mathrm{e}^{-\mathrm{j}\omega q} \\
&= \sum_{q\in\mathbb{Z}} \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \langle \mathrm{S}_{k_1}\{\phi_m\}, \mathrm{S}_{k_2}\{\phi_n\}(\cdot - q - (k_2 - k_1))\rangle \mathrm{e}^{-\mathrm{j}\omega q} \\
&= \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \langle \mathrm{S}_{k_1}\{\phi_m\}, \mathrm{S}_{k_2}\{\phi_n\}\rangle \mathrm{e}^{\mathrm{j}\omega(k_2-k_1)} \\
&= \langle \sum_{k_1\in\mathbb{Z}} \mathrm{S}_{k_1}\{\phi_m\}\mathrm{e}^{-\mathrm{j}\omega k_1}, \sum_{k_2\in\mathbb{Z}} \mathrm{S}_{k_2}\{\phi_n\}\mathrm{e}^{-\mathrm{j}\omega k_2}\rangle \\
&= \langle \tilde{\phi}_m(\omega,\cdot), \tilde{\phi}_n(\omega,\cdot)\rangle,
\end{aligned} \tag{1.65}$$

where $\tilde{\phi}_n(\omega,\cdot)$ is the finite weighted sum of slices

$$\tilde{\phi}_n(\omega, x) = \sum_{k\in\mathbb{Z}} \mathrm{S}_k\{\phi_n\}(x) e^{-j\omega k}. \tag{1.66}$$

Due to the linear independence of $\mathcal{T}(\boldsymbol{\phi})$, we deduce that for any $\omega \in \mathbb{R}$, the functions $(\tilde{\phi}_n(\omega,\cdot))_{n=1,\ldots,N}$ are linearly independent because the sums are finite. This means that $\hat{\boldsymbol{G}}(\omega)$ is the Gramian matrix of a linearly independent family of functions, which is known to be equivalent to $\det \hat{\boldsymbol{G}}(\omega) > 0$. In addition $g : \omega \mapsto \det(\hat{\boldsymbol{G}}(\omega))$ is a finite weighted sum of $\mathrm{e}^{\mathrm{j}\omega k}$ since $\boldsymbol{\phi}$ is compactly supported. It is therefore continuous and 2π-periodic. The image of $[0, 2\pi]$ under g is therefore a closed interval such that

$$0 < \min_{\omega \in [0,2\pi]} \det(\hat{\boldsymbol{G}}(\omega)) < \max_{\omega \in [0,2\pi]} \det(\hat{\boldsymbol{G}}(\omega)) < +\infty. \tag{1.67}$$

Noting that $\det(\hat{\boldsymbol{G}}(\omega))$ is the product of the eigenvalues of $\hat{\boldsymbol{G}}(\omega)$, Condition (1.57) is satisfied, which means that $\boldsymbol{\phi}$ is a Riesz basis. □

1.3.4 Multi-Spline Spaces and Their Shortest Bases

A cardinal multi-spline space is defined as the sum of $N \in \mathbb{N}$ spline spaces: $S_{\mathbf{n}} = S_{n_1} + \cdots + S_{n_N}$, $\mathbf{n} = (n_1, ..., n_N)$ and $n_1 < \cdots < n_N \in \mathbb{N}$. It is worth noting that, the multi-spline space $S_{\mathbf{n}}$ inherits the highest approximation power of its spline spaces, *i.e.* its approximation power is $(n_N + 1)$, since $S_{n_N} \subset S_{n_1} + \cdots + S_{n_N}$. Moreover, in the case of consecutive spaces specified by $n_k = n_1 + (k - 1)$, the resulting space is exactly the space of piecewise polynomials of degree n_N that are in $C^{n_1 - 1}(\mathbb{R})$.

Some multi-spline spaces have proved to be of great interest for derivative sampling, where the goal is to reconstruct a signal from the samples of the function and of its first-order derivative. The most well-known example is the pair of cubic Hermite splines (see, Section 1.2). The excellent approximation capabilities and minimal-support property of the Hermite splines [1] give a strong incentive to investigate more general multi-spline spaces. The bicubic Hermite splines are the backbone of many computer-graphics applications and closely linked to Bézier curves [23, 29, 49, 50, 51]. Schoenberg and Lipow also found two fundamental functions to reconstruct any function in $S_4 + S_5$ from its samples and the samples of its first-order derivative. Nonetheless, those functions are not well-suited to practical applications since they are not compactly supported.

With a single generator, the unique shortest basis of degree $n \in \mathbb{N}$ (up to a scaling and a shift operation) is the B-spline of degree n, which is a generator of S_n.

Similar to our findings in Section 1.2, we now prove that multi-spline spaces are not generated optimally by the classical B-splines.

Proposition 1.1. *Let $N \in \mathbb{N} \setminus \{0\}$ and $\mathbf{n} = (n_1, \dots, n_N)$ with $n_1 < \cdots < n_N \in \mathbb{N}$. If $N > 1$, then $\boldsymbol{\beta}_{\mathbf{n}} = (\beta_{\mathrm{D}^{n_1+1}}, ..., \beta_{\mathrm{D}^{n_N+1}})$ is neither a shortest-support basis nor a Riesz basis.*

Proof. We first note that the space $\mathrm{S}(\boldsymbol{\beta}_{\mathbf{n}})$ can reproduce polynomials of degree at most n_N due to the inclusion $S_{n_N} = \mathrm{S}(\beta_{\mathrm{D}^{n_N+1}}) \subset \mathrm{S}(\boldsymbol{\phi})$. Moreover, the sum of the support of $\boldsymbol{\beta}_{\mathbf{n}}$ is $\sum_{m=1}^{N}(n_m + 1) > n_N + 1$, which shows that the basis is not a shortest-support one.

Moreover, we invoke (1.66) to express the Gramian matrix as

$$(\hat{\boldsymbol{G}}(\omega))_{pq} = \langle \tilde{\beta}_{\mathrm{D}^{n_p+1}}(\omega, \cdot), \tilde{\beta}_{\mathrm{D}^{n_q+1|}}(\omega, \cdot) \rangle, \tag{1.68}$$

where $\tilde{\beta}_{\mathrm{D}^{n_p+1}}(\omega, \cdot)$ is the finite weighted sum of slices

$$\tilde{\beta}_{\mathrm{D}^{n_p+1}}(\omega, x) = \sum_{k \in \mathbb{Z}} \mathrm{S}_k\{\beta_{\mathrm{D}^{n_p+1}}\}(x) \mathrm{e}^{-\mathrm{j}\omega k}. \tag{1.69}$$

It is known that $\beta_{\mathrm{D}^{n_p+1}}$ satisfies the partition of unity, meaning that, for any $x \in \mathbb{R}$, we have that $\sum_{k \in \mathbb{Z}} \beta_{\mathrm{D}^{n_p+1}}(x-k) = 1$. In terms of slices, it means that $\tilde{\beta}_{\mathrm{D}^{n_p+1}}(0, x) = \sum_{k \in \mathbb{Z}} \mathrm{S}_k\{\beta_{\mathrm{D}^{n_p+1}}\}(x) = \mathbb{1}_{[0,1)}(x)$. The functions $(\tilde{\beta}_{\mathrm{D}^{n_p+1}}(0, \cdot))_{p=1,\dots,N})$ are therefore not linearly independent (because they are equal) and $\det \hat{\boldsymbol{G}}(0) = 0$. The mapping $\omega \mapsto \det \hat{\boldsymbol{G}}(\omega)$ being continuous (due to the compact support of B-splines), we deduce that

$$\operatorname*{ess\,inf}_{\omega \in [0,2\pi]} \det \hat{\boldsymbol{G}}(\omega) = \min_{\omega \in [0,2\pi]} \det \hat{\boldsymbol{G}}(\omega) = 0. \tag{1.70}$$

Following (1.57), $\boldsymbol{\beta}_{\mathbf{n}}$ cannot be a Riesz basis. □

Definition 1.4. *A finite collection $\boldsymbol{\phi}$ of multi-spline functions is called an mB-spline of degree $\mathbf{n} = (n_1, \dots, n_N)$ with $n_1 < \cdots < n_N \in \mathbb{N}$, if it is a shortest-support basis of the space $S_{\mathbf{n}}$.*

mB-splines are natural extensions of B-splines. Let us recall that the B-spline $\beta_{\mathrm{D}^{n+1}}$ can be constructed recursively by noting that

$$\beta_{\mathrm{D}^{n+1}}(x) = \Delta \left\{ \int_{-\infty}^{x} \beta_{\mathrm{D}^{n}}(t)\mathrm{d}t \right\}, \tag{1.71}$$

where Δ is the finite difference operator $\Delta\{f\}(x) = (f(x) - f(x-1))$. The integration increases the polynomial degree, along with the smoothness at the knots, while Δ ultimately returns a compactly supported function. Inspired from this observation, we propose a recursive algorithm for the construction of mB-splines in any multi-spline spaces.

Theorem 1.5. *Let $n_1 < \cdots < n_N \in \mathbb{N} \setminus \{0\}$. There exists an mB-spline $\boldsymbol{\eta} = (\eta_1, ..., \eta_N) \in (L_2(\mathbb{R}))^N$ of $S_{n_1} + \cdots + S_{n_N}$.*

Proof. We prove the existence by proposing a recursive algorithm that constructs the desired mB-splines. To simplify the explanation, we say that the collection $\boldsymbol{\phi} = (\phi_1, ..., \phi_N) \in (L_2(\mathbb{R}))^N$ of compactly supported functions is standardized if, for $n = 1, \ldots, N$, we have that

1. $\int_{\mathbb{R}} \phi_n(t)\mathrm{d}t \in \{0, 1\}$,
2. $\inf\{t \in \mathbb{R} \colon \phi_n(t) \neq 0\} \in [0, 1)$.

The second condition implies that the generating functions are causal, *i.e.* $\phi_n(t < 0) = 0$. Note that any $\boldsymbol{\phi}$ compactly supported can be standardized without altering $\mathrm{S}(\boldsymbol{\phi})$.

Increment Step: Suppose that $\boldsymbol{\eta} = (\eta_1, ..., \eta_N) \in (L_2(\mathbb{R}))^N$ is an mB-spline of $S_{n_1} + \cdots + S_{n_N}$. The goal is to find an mB-spline of $S_{n_1+1} + \cdots + S_{n_N+1}$ (See, Figure 1.4, for an illustration). It will be a generator with a support size of $(n_N + 2)$, able to reproduce the B-splines of degree $n_1 + 1, ..., n_N + 1$.

The collection of functions $\boldsymbol{\eta}$ is able to reproduce the B-splines of degree $n_1, \ldots, n_N$, that is, for any $s \in \{1, ..., N\}$ there exists a vector sequence $\boldsymbol{c}^s =$

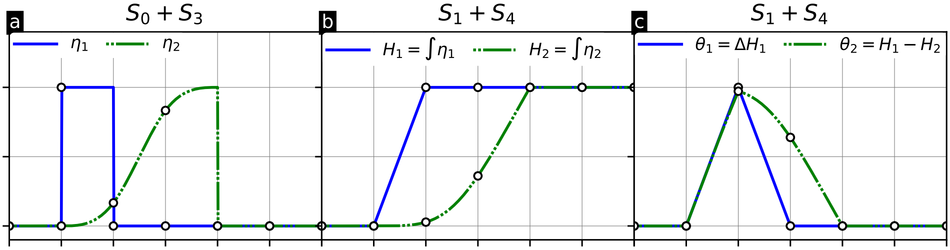


Figure 1.4: Increment step that yields a shortest-support basis of S_1+S_4 starting from S_0+S_3. (a)A shortest-support basis (η_1,η_2) for S_0+S_3 ($|\operatorname{supp}(\boldsymbol{\eta})|=4$). (b)The integration of η_1 and η_2 results in two generators of S_1+S_4, H_1 and H_2. (c) To get compactly supported functions with the same generating properties, we choose $\theta_1=\Delta H_1$ and $\theta_2=(H_1-H_2)$. This yields a shortest support-basis for S_1+S_4 ($|\operatorname{supp}(\boldsymbol{\theta})|=5$).

$(c_1^s,...,c_N^s)$ so that

$$\forall x\in\mathbb{R}:\beta_{\mathrm{D}^{n_s+1}}(x)=\sum_{k\in\mathbb{Z}}\boldsymbol{c}^s[k]^T\boldsymbol{\eta}(x-k). \tag{1.72}$$

To justify the calculations to come, we assume that

$$c_1^s,...,c_N^s \text{ are causal sequences, } i.e.,\ c_n^s[k]=0 \text{ for any } k<0.$$

This assumption is not overly restrictive because it will hold for the starting basis of our algorithm and then will be preserved through the construction process. In the end, all the constructed bases will be able to reproduce the B-splines with causal sequences. Let $\boldsymbol{H}=(H_1,...,H_N)$ be defined as

$$\boldsymbol{H}(x)=\int_{-\infty}^{x}\boldsymbol{\eta}(t)\mathrm{d}t. \tag{1.73}$$

The integration of equation (1.73), followed by the application of the operator Δ, yields

$$\begin{aligned}\beta_{\mathrm{D}^{n_s+2}}(x) &= \Delta\left\{\sum_{k\in\mathbb{Z}} \boldsymbol{c}^s[k]^T \boldsymbol{H}(x-k)\right\} = \sum_{k\in\mathbb{Z}} \boldsymbol{c}^s[k]^T \Delta\{\boldsymbol{H}\}(x-k)\\ &= \sum_{k\in\mathbb{Z}} \boldsymbol{c}^s[k]^T(\boldsymbol{H}(x-k) - \boldsymbol{H}(x-1-k)) \qquad (1.74)\\ &= \sum_{k\in\mathbb{Z}} (\boldsymbol{c}^s[k]^T - \boldsymbol{c}^s[k-1]^T)\boldsymbol{H}(x-k). \qquad (1.75)\end{aligned}$$

The assumption that $c_1^s, ..., c_N^s$ are causal and the fact that $\boldsymbol{H}$ is also causal (because $\boldsymbol{\eta}$ is compactly supported and standardized) implies that, for any $x \in \mathbb{R}$, the sums in (1.76) have a finite number of nonzero terms. This enables us to switch the order of the operations (sum, integral, and Δ). Note that the sequence $(\boldsymbol{c}^s[k]^T - \boldsymbol{c}^s[k-1]^T)_{k\in\mathbb{Z}}$ is causal. In short, $\boldsymbol{H}$ can reproduce B-splines of degree $n_s + 1$ for $s = 1, \ldots, N$ with causal sequences, but it is obviously not a shortest-support basis because its support is infinite.

The aim now is to find a basis with the same reproducing properties as $\boldsymbol{H}$, but with minimal support. To that end, we denote by s_0 the index so that η_{s_0} is the shortest function in $\boldsymbol{\eta}$ that satisfies $\int_{\mathbb{R}} \eta_{s_0} \neq 0$. It must exist; if not, the generating $S(\boldsymbol{\eta})$ would only contains zero-mean functions and could not reproduce the B-splines that are not zero-mean. A shortest-support basis $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$ is then given by

$$\theta_s = \begin{cases} H_s & \text{if } s \neq s_0 \text{ and } \int_{\mathbb{R}} \eta_s(t)\mathrm{d}t = 0\\ H_s - H_{s_0} & \text{if } s \neq s_0 \text{ and } \int_{\mathbb{R}} \eta_s(t)\mathrm{d}t \neq 0\\ \Delta H_{s_0} & s = s_0 \end{cases} \qquad (1.76)$$

Because $\boldsymbol{\eta}$ is compactly supported and standardized, the choice of s_0 ensures that

$$|\operatorname{supp}(\theta_s)| = \begin{cases} |\operatorname{supp}(\eta_s)| & s \neq s_0\\ |\operatorname{supp}(\eta_{s_0})| + 1 & s = s_0 \end{cases} \qquad (1.77)$$

In short, $|\operatorname{supp}(\boldsymbol{\theta})| = 1 + |\operatorname{supp}(\boldsymbol{\eta})| = n_N + 2$. Noting that $H_{s_0} = \sum_{k\in\mathbb{N}} \theta_{s_0}(\cdot - k)$, it is clear that $\boldsymbol{\theta}$ can reproduce $\boldsymbol{H}$ with causal coefficients. It also implies that $\boldsymbol{\theta}$ can reproduce $(\beta_{n_1+1}, ..., \beta_{n_N+1})$ with causal coefficients (see (1.76)), which justifies the original assumption. In conclusion, $\boldsymbol{\theta}$ is a shortest-support basis of $S_{n_1+1} + \cdots + S_{n_N+1}$.

Insertion Step: The present step enables us to add a generator to a shortest-support basis. Suppose $\boldsymbol{\eta} = (\eta_1, ..., \eta_N)$ is a standardized shortest-support basis of $S_{n_1} + \cdots + S_{n_N}$ and let $\boldsymbol{\eta}' = (\delta, \eta_1, ..., \eta_N)$, where δ is the Dirac distribution. The increment step applied to $\boldsymbol{\eta}'$ yields a shortest-support basis for $S_0 + S_{n_1+1} + \cdots + S_{n_N+1}$. Indeed, the shortest function of $\boldsymbol{\eta}$ being δ, the new basis $\boldsymbol{\theta}' = (\theta'_0, ..., \theta'_N)$ is given by

$$\theta'_n : x \mapsto \begin{cases} \Delta\{\int_{-\infty}^{x} \delta(t)\mathrm{d}t\} = \beta_0(x), & n = 0 \\ \int_{-\infty}^{x} \eta_n(t)\mathrm{d}t, & n > 0 \text{ and } \int_{\mathbb{R}} \eta_n(t)\mathrm{d}t = 0 \\ \int_{-\infty}^{x} (\eta_n(t) - \delta(t))\mathrm{d}t, & n > 0 \text{ and } \int_{\mathbb{R}} \eta_n(t)\mathrm{d}t \neq 0. \end{cases} \tag{1.78}$$

Because $\boldsymbol{\eta}$ is compactly supported and standardized, we have that

$$|\operatorname{supp}(\theta'_n)| = \begin{cases} 1, & n = 0 \\ |\operatorname{supp}(\eta_n)|, & \text{otherwise,} \end{cases} \tag{1.79}$$

which means that $|\operatorname{supp}(\boldsymbol{\theta}')| = |\operatorname{supp}(\eta')| + 1 = n_N + 2$. The process also ensures that $\boldsymbol{\theta}'$ is a shortest-support basis of $S_0 + S_{n_1+1} + \cdots + S_{n_N+1}$.

Final procedure: Take $\boldsymbol{\eta_0} = (\beta_{n_N - n_{N-1} - 1})$ a shortest support basis for $S_{n_N - n_{N-1} - 1}$. The insertion step gives a shortest-support basis for $S_0 + S_{n_N - n_{N-1}}$. After $(n_{N-1} - n_{N-2} - 1)$ increment steps and one insertion step, the process gives a shortest-support basis for $S_0 + S_{n_{N-1} - n_{N-2}} + S_{n_N - n_{N-2}}$. By iteration, a shortest-support basis for $S_0 + S_{n_2 - n_1} + \cdots + S_{n_N - n_1}$ is obtained. Applying n_1 increment steps, we finally obtain a shortest-support basis for $S_{n_1} + \cdots + S_{n_N}$ □

We conclude this section with a result on the minimal number of generating functions required to generate multi-spline spaces.

Theorem 1.6. *Let $n_1 < \cdots < n_N \in \mathbb{N} \setminus \{0\}$. The space $S_{\mathbf{n}} = S_{n_1} + \cdots + S_{n_N}$ cannot be generated by fewer than N compactly supported generating functions.*

Proof. From Theorem 1.5, there exists an mB-spline of $S_{\mathbf{n}}$ composed of N functions, say, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N) \in (S_{\mathbf{n}})^N$. Let $\boldsymbol{\psi} = (\psi_1, ..., \psi_M) \in (S_{\mathbf{n}})^M$ be a collection of compactly supported functions able to generate $S_{\mathbf{n}}$. It means that $\boldsymbol{\eta}$ and $\boldsymbol{\psi}$ can reproduce each other and, hence, there exist vector sequences $\mathbf{c}_p : \mathbb{Z} \to \mathbb{R}^M$ such that $\eta_p = \sum_{k\in\mathbb{Z}} \mathbf{c}_p[k]^T \boldsymbol{\psi}(\cdot - k) = \mathbf{c}_p^T * \boldsymbol{\psi}$, which reads in matrix form

$$\boldsymbol{\eta} = \mathbf{C} * \boldsymbol{\psi}, \quad \mathbf{C} : \mathbb{Z} \to \mathbb{R}^{N\times M}. \tag{1.80}$$

Similarly, one can write that

$$\boldsymbol{\psi} = \mathbf{B} * \boldsymbol{\eta}, \quad \mathbf{B} : \mathbb{Z} \to \mathbb{R}^{N\times M}. \tag{1.81}$$

Following the proof of Theorem 1.4, we know that the nonzero slices of $\boldsymbol{\eta}$ are linearly independent (shortest-support basis). This implies that, to generate the compactly supported function $\boldsymbol{\psi}$, the sequence of matrices $\mathbf{B}$ must be compactly supported as well since the only way to generate the zero function on a segment for $\boldsymbol{\eta}$ is to set the active coefficient of $\mathbf{B}$ to 0. Now, one can mix the equations and find that

$$\boldsymbol{\eta} = \mathbf{C} * (\mathbf{B} * \boldsymbol{\eta}) = (\mathbf{C} * \mathbf{B}) * \boldsymbol{\eta}. \tag{1.82}$$

The associativity of the convolution operations is justified by the fact that both $\boldsymbol{\eta}$ and $\mathbf{B}$ are compactly supported, meaning that, for a given argument x, all sums are finite. Because the slices of $\boldsymbol{\eta}$ are linearly independent, $\boldsymbol{\eta}$ can reproduce itself in a unique way, which gives

$$\mathbf{C} * \mathbf{B} = \boldsymbol{\delta}_{N\times N}. \tag{1.83}$$

There exists $s \in \mathbb{N}$ such that $\text{supp}(\boldsymbol{B}) \subset \{-s, ..., s\} \subset \mathbb{N}$. The behavior of $\boldsymbol{C}[k]$ when $|k| \to \infty$ is not known, and it is easier to work with the truncated version $\boldsymbol{C}_m = \mathbb{1}_{\{-s,...,ms\}} \times \boldsymbol{C}$, where $m \in \mathbb{N}$ is a large enough integer $m > 2N + 1$. The sequence of matrices $\boldsymbol{C}_m * \boldsymbol{B}$ is compactly supported and satisfies $\text{supp}(\boldsymbol{C}_m * \boldsymbol{B}) \subset \{-2s, ..., (m+1)s\}$. Following the properties of convolution of compact sequences, we have, for any $k = 0, ..., (m-1)s$, that $\boldsymbol{C}_m * \boldsymbol{B}[k] = \boldsymbol{C} * \boldsymbol{B}[k] = \boldsymbol{\delta}_{N\times N}[k]$. Therefore, one can write that

$$\mathbf{C}_m * \mathbf{B} = \boldsymbol{\delta}_{N\times N} + \sum_{\substack{-2s\leq k<0 \\ (m-1)s+1\leq k\leq (m+1)s}} \mathbf{M}_k \delta[\cdot - k], \tag{1.84}$$

where $\mathbf{M}_k \in \mathbb{R}^{N\times N}$ are matrices that account for the fact that $\mathbf{C}_m$ is a truncated version of $\mathbf{C}$. This then translates into the following z-transform matrix relation

(note that all sequences are compactly supported so the z-transforms are well defined)

$$\begin{aligned}\hat{\mathbf{C}}_m(z)\hat{\mathbf{B}}(z) &= \mathbf{I}_{N\times N} + \sum_{\max(-2s,(m-1)s)}^{\min((m+1)s,-1)} z^{-k}\mathbf{M}_k \\ &= \mathbf{M}_{N\times N} + \sum_{k=-2s}^{-1} z^{-k}\mathbf{M}_k + \sum_{k=(m-1)s+1}^{(m+1)s} z^{-k}\mathbf{M}_k = z^{-2s}\mathbf{A}(z), \quad (1.85)\end{aligned}$$

where $\mathbf{A}(z)$ can be decomposed as

$$\mathbf{A}(z) = z^{2s}\mathbf{I}_{N\times N} + \mathbf{P}(z) + z^{(m+1)s+1}\mathbf{Q}(z), \quad (1.86)$$

where $\mathbf{P}(z)$ and $\mathbf{Q}(z)$ are polynomial matrices of degree $(2s-1)$. The determinant of $\mathbf{A}(z)$ can be expressed in terms of the columns of $\mathbf{I}_{N\times N}, \mathbf{P}(z)$, and $\mathbf{Q}(z)$ (denoted respectively $\mathbf{e}_k, \mathbf{p}_k(z)$, and $\mathbf{q}_k(z)$), so that

$$z \mapsto \det A(z) = \det(z^{2s}\mathbf{e}_1+\mathbf{p}_1(z)+z^{(m+1)s}\mathbf{q}_1(z), \ldots, z^{2s}\mathbf{e}_N+\mathbf{p}_N(z)+z^{(m+1)s}\mathbf{q}_N(z)). \quad (1.87)$$

Knowing that the determinant is n-linear with respect to the columns, $z \mapsto \det \mathbf{A}(z)$ is a polynomial function of degree at most $(m+3)sN$. We now want to prove that it cannot be identically zero. To that end, we expand the determinant with respect to the columns and find that there is a unique term of the form λz^{2sN}. It is obtained by picking for $k = 1, \ldots, N$ the column $\mathbf{e}_k z^{2s}$. The coefficient in front of z^{2sN} is therefore $\det(\mathbf{e}_1, \ldots, \mathbf{e}_N) = 1 \neq 0$. Indeed, for other combinations of columns in the expansion, we would have that

- if at least one column of the form $z^{(m+1)s}\mathbf{q}_k(z)$ is chosen, then it results in a term of degree at least $(m+1)s > (2N+2)s > 2sN$;
- else, at least one column of the form $\mathbf{p}_k(z)$ is chosen. Since the degree of $\mathbf{p}_k(z)$ is lower than $2s$, the resulting term in the expansion has a degree lower than $2sN$.

In the end, we proved that $z \mapsto \det \mathbf{A}(z)$ cannot be identically zero. Therefore, there exists $z_0 \in \mathbb{R}$ so that $\operatorname{rank}(\mathbf{A}(z_0)) = N$. It implies that

$$N = \operatorname{rank}(\hat{\mathbf{C}}_m(z_0)\hat{\mathbf{B}}(z_0)) \leq \min(\operatorname{rank}(\hat{\mathbf{C}_m}(z_0)), \operatorname{rank}(\hat{\mathbf{B}}(z_0))) \leq \min(M, N) \leq M. \quad (1.88)$$

□

Note that N is a lower bound and the number of generating function of a shortest-support basis can exceed N. For instance, take $\boldsymbol{\eta} = (\eta_1, \eta_2)$ with

$$\eta_1 : x \mapsto \beta_0(2x) = \mathbb{1}_{[0,1/2)}(x) \tag{1.89}$$

$$\eta_2 : x \mapsto \beta_0(2(x-1/2)) = \mathbb{1}_{[1/2,1)(x)}. \tag{1.90}$$

Since $\eta_1 + \eta_2 = \beta_0$, $\boldsymbol{\eta}$ can reproduce S_0. In addition, the fact that $|\operatorname{supp}(\boldsymbol{\eta})| = 1$ means that it is a shortest-support basis of degree 0 and now it is composed of two generating functions. (Note that the space they generate is larger than S_0).

1.3.5 Applications

Generalized Sampling in Multi-Spline Spaces

We consider a multi-spline space $S_{\mathbf{n}}$ along with the N-component mB-spline $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)$ and some corresponding analysis functions $\boldsymbol{\psi} = (\psi_1, \dots, \psi_N)$. As we now show, the generalized-sampling formulation presented in [40] can be extended to multiple generators. Let $\mathcal{H}$ be a space considerably larger than $\mathrm{S}(\boldsymbol{\phi})$. Consider $f \in \mathcal{H}$, from which we know only some discrete measurements $(\boldsymbol{g}[n])_{n\in\mathbb{Z}}$ written

$$\boldsymbol{g}[n] = \langle \boldsymbol{\psi}(\cdot - n), f\rangle = (\langle \psi_1(\cdot - n), f\rangle, ..., \langle \psi_N(\cdot - n), f\rangle). \tag{1.91}$$

To construct an approximation $\tilde{f} \in \mathrm{S}(\boldsymbol{\phi})$ of f, a standard way is to enforce consistency [36, 41], in the sense that f and $\tilde{f}$ must give the same measurements. This formulation generalizes the notion of interpolation. For instance, to interpolate the value of f and its derivative at the sampling locations, take $\psi_1 = \delta$ and $\psi_2 = \delta^{'}$. In such a case, consistency simply means that f and $\tilde{f}$ should have the same value and the same derivative at the grid points. In general, the consistency requirement translates into

$$\begin{aligned}\langle \boldsymbol{\psi}(\cdot - n), f\rangle &= \langle \boldsymbol{\psi}(\cdot - n), \tilde{f}\rangle \\ &= \sum_{k\in\mathbb{Z}} \langle \boldsymbol{\psi}(\cdot - n), \boldsymbol{\phi}^T(\cdot - k)\rangle \cdot \boldsymbol{c}[k] \\ &= \sum_{k\in\mathbb{Z}} \langle \boldsymbol{\psi}(\cdot - (n-k)), \boldsymbol{\phi}^T\rangle \cdot \boldsymbol{c}[k] \\ &= (\boldsymbol{A}_{\boldsymbol{\Phi}\boldsymbol{\Psi}} * \boldsymbol{c})[n] \end{aligned} \tag{1.92}$$

where $(\boldsymbol{c}[n])_{n\in\mathbb{Z}}$ is the unique vector sequence representing $\tilde{f} = \sum_{k\in\mathbb{Z}} \mathbf{c}[k]^T \boldsymbol{\phi}(\cdot - k)$ and $\boldsymbol{A}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}[n] = \langle \boldsymbol{\psi}(\cdot - n), \boldsymbol{\phi}^T(\cdot)\rangle$ is the matrix-valued sequence of the measurements of the basis functions. To solve our problem, we rely on the theory of signal and systems, including the z-transform. Indeed, with this framework efficient implementation techniques naturally stand out. When the matrix-valued filter $\boldsymbol{A}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}$ is invertible (see [40, Proposition 1] for the invertibility condition), the vector $\mathbf{c}$ of sequences can be computed from the measurements by applying the matrix-valued inverse filter $\mathbf{Q}$, like in

$$\mathbf{c}[n] = (\mathbf{Q} * \mathbf{g})[n]. \tag{1.93}$$

Its transfer function verifies in the z-domain $\hat{\mathbf{Q}}(z) = \hat{\mathbf{A}}^{-1}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}(z)$. This matrix filter has not necessarily a finite impulse response (FIR) but it can be decomposed as $\hat{\mathbf{Q}}(z) = \frac{1}{\det \hat{\mathbf{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}(z)} \mathrm{com}(\hat{\mathbf{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}(z))^T$, where $\mathrm{com}(\hat{\mathbf{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}})$ denotes the cofactor matrix of $\hat{\mathbf{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}$. For compactly supported analysis functions, the comatrix $\mathrm{com}(\hat{\mathbf{A}}(z))$ is FIR because it is a Laurent polynomial in z, so it is straightforward to implement. On the contrary, $\frac{1}{\det \hat{\mathbf{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}(z)}$ is often not FIR. Nonetheless, it can usually be implemented efficiently too, using the same techniques as in [12].

Derivative Sampling

The derivative sampling problem reads for $f \in \mathcal{H}$

$$\text{find } \tilde{f} \in S_{\mathbf{n}} : \begin{cases} \tilde{f}(k) = f(k) \\ \tilde{f}'(k) = f'(k) \end{cases}, k \in \mathbb{Z}. \tag{1.94}$$

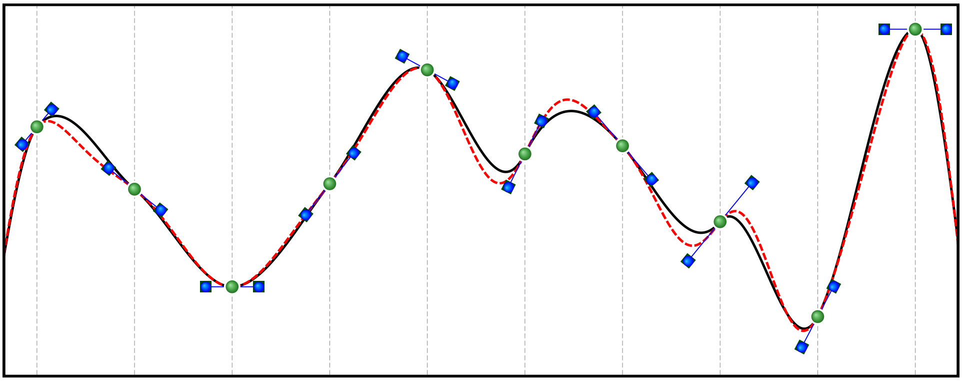


Figure 1.5: Derivative sampling with optimal bases. The solid curve lies in $S_2 + S_3$ (cubic piecewise polynomials with continuous derivative) and the dashed curve lies in $S_4 + S_5$ (quintic piecewise polynomials with continuous third derivative).

The most relevant reconstruction spaces have the form $S_{\mathbf{n}} = S_{2p} + S_{2p+1}$ [52]. The underlying reason is that the filter complexity is the same for the spaces $S_{2p} + S_{2p+1}$ and $S_{2p-1} + S_{2p}$, so, the higher degree is preferred (the filter has $2(p-1)$ roots). Note that the same occurs when one performs classical interpolation with B-splines and odd degrees are usually preferred. To the best of our knowledge, when $p > 1$, no solution based on shortest-support bases and recursive filtering has been proposed so far. Our construction of shortest-bases results in generating functions that have a support size of $(p + 1)$. Due to the symmetry properties of those functions, the entries of $\hat{\boldsymbol{A}}_{\boldsymbol{\Phi}\boldsymbol{\Psi}}(z)$ have poles that come in reciprocal pairs. Consequently, the inverse matrix filter can be implemented with efficient recursive techniques, as detailed in [11, 12]. An example is provided in Figure 1.5.

Bézier Curves

We now use our multi-spline formulation to revisit some Bézier curves and, in particular, the cubic Bézier curves that are popular in computer graphics. Each

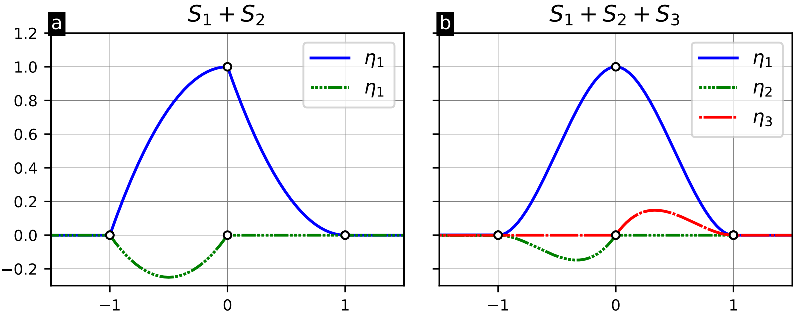


Figure 1.6: Shortest-support bases for application in classical computer-graphics. (a) Shortest basis for $S_1 + S_2$. The function η_1 controls the value of the function on the knots while η_2 controls the left derivative on the knots. These function reproduce any quadratic Bézier curve. (b) Shortest basis for $S_1 + S_2 + S_3$. The function η_1 controls the value of the function on the knots while η_2 and η_3 controls the left and right derivative, respectively, on the knots. These functions can reproduce any cubic Bézier curve with the shortest support. They also give a simple interpretation of such curves.

portion of the curve is a cubic polynomial defined by four control points.

- Starting point and ending point of the portion.
- Two handles that control the tangent of the curve at each extremity of the portion.

Thus, the value of the function and its left and right derivatives are controlled on the knots. From a multi-spline perspective, any cubic Bézier curve lies in the space $S_1 + S_2 + S_3$. With the well chosen generating functions η_1, η_2, and η_3 plotted in Figure 1.6, the interpolation formula is explicit and reads

$$\tilde{f}(x) = \sum_{k\in\mathbb{Z}} f(k)\eta_1(x-k) + \sum_{k\in\mathbb{Z}} f^{'}(k^-)\eta_2(x-k) + \sum_{k\in\mathbb{Z}} f^{'}(k^+)\eta_3(x-k), \quad (1.95)$$

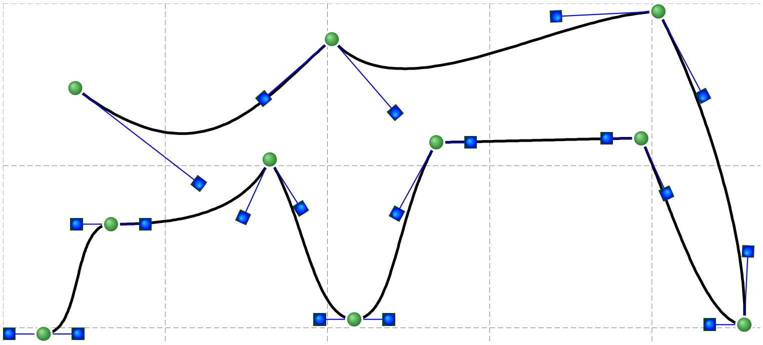


Figure 1.7: The shortest basis of the space $S_1 + S_2 + S_3$ allows one to control the value of the function (green dots) and the left/right derivatives (handles). It yields the same curve as with standard vector-graphics editors relying on cubic Bézier curves. In this figure, the parametric curves are two-dimensional and the interpolation is performed component-wise.

where $f'(k^-)$ and $f'(k^+)$ denote the left and right derivatives at k, respectively. Interestingly, η_2 and η_3 can be obtained from the bi-cubic Hermite splines, by splitting the antisymmetric function into two functions. It gives a simple interpretation to cubic Bézier curves as illustrated in Figure 1.7. Similarly, quadratic Bézier curves are also multi-splines, this time associated to the space $S_1 + S_2$ (Figure 1.6).

Nonconsecutive Bi-Spline Spaces

Nonconsecutive multi-spline spaces are relevant to represent signals that have components of different regularity (see, Chapter 4). For instance, the space $S_0 + S_p$, with $p > 0$, consists of smooth signals with sharp jumps. On Figure 1.8, we show shortest-support bases of $S_0 + S_p$, for $p \in \{2, 3, 4\}$, that were obtained with our construction algorithm.

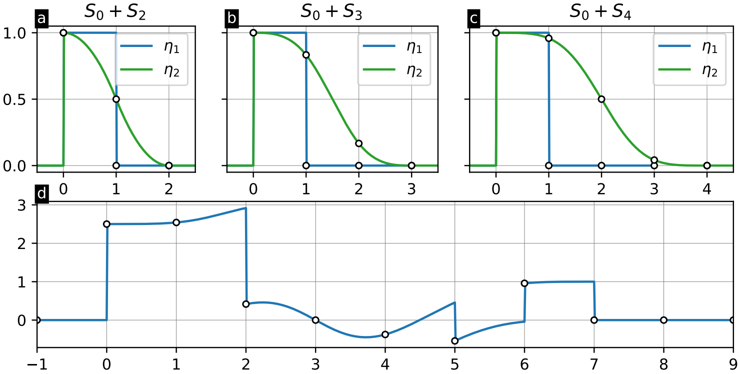


Figure 1.8: (a)(b)(c) Shortest-support bases for the spaces $S_0 + S_2$, $S_0 + S_3$ and $S_0 + S_4$. (d) An example of a hybrid bi-spline that lies in the space $S_0 + S_4$.

1.3.6 Summary

We have introduced the notion of shortest-support bases of degree M. They are the shortest-support collections of functions that generate a reconstruction space with an approximation power of order $(M + 1)$. We proved that shortest-support bases necessarily generate Riesz bases, a minimal requirement for practical applications. With a single generator, the unique shortest-support basis of degree M is the well known B-spline of degree M. We extended this notion to multiple generators and proposed a recursive method that yields shortest bases for any multi-spline space. These new sets of functions helped us transpose the efficient reconstruction techniques developed for B-splines, and perform generalized sampling.

Chapter 2

General Representer Theorem

In this chapter[1], we characterize the solution of a broad class of convex optimization problems that address the reconstruction of a function from a finite number of linear measurements. The underlying hypothesis is that the solution is decomposable as a finite sum of components, where each component belongs to its own prescribed Banach space; moreover, the problem is regularized by penalizing some composite norm of the solution. We establish general conditions for existence and derive the generic parametric representation of the solution components. These representations fall into three categories depending on the underlying regularization norm: (i) a linear expansion in terms of predefined "kernels" when the component space is a reproducing kernel Hilbert space (RKHS), (ii) a non-linear (duality) mapping of a linear combination of measurement functionals when the component Banach space is strictly convex, and, (iii) an adaptive expansion in terms of a small number of atoms within a larger dictionary when the component Banach space is not strictly convex.

[1] This chapter is based on our published work [53].

2.1 Context

2.1.1 From RKHS to Banach Spaces

Reproducing kernel Hilbert spaces (RKHS) play a central role in the classical formulations of machine learning, statistical estimation, and the resolution of linear inverse problems [54, 55]. They go hand in hand with quadratic (or Tikhonov) regularization and Gaussian processes [56, 57]. The popularity of RKHS in machine learning stems from the fact that the minimization of Hilbertian norms results in parametric solutions that are linear combinations of kernels (basis functions) centered on the data points [54, 58, 59, 60], a remarkable property that is supported by the celebrated representer theorem [61].

However, recent works that revolve around the concept of sparsity have demonstrated the advantages of considering Banach spaces instead of Hilbert spaces. In particular, compressed sensing relies on the minimization of ℓ_1-norms. Under suitable conditions, this enables the exact recovery of a signal from a limited number of linear measurements [62, 63, 64, 65, 66, 67]. Researchers have established representer theorems that explain the sparsifying effect of the ℓ_1-norm [68] and of its variants, including its continuous-domain counterpart: the $\mathcal{M}$-norm (a.k.a. the total-variation norm of a measure) [69, 70, 71, 72]. Likewise, it is proven in [73] that non-uniform splines of a type that is matched to the regularization operator are universal solutions of linear inverse problems with generalized total-variation regularization. The main difference with the RKHS (or Tikhonov) framework is that the underlying basis functions—or kernels—are selected in an adaptive fashion and are not necessarily placed on the data points [74]. More recently, it has been shown that the effect of such minimum-norm regularization could be characterized in full generality [75, 76].

2.1.2 From Sums of RKHS to Sums of Banach Spaces

It is a known fundamental property that a convex combination (resp., a tensor product) of reproducing kernels retains the desirable reproducing-kernel property (a.k.a. positive-definiteness) [77]. This has prompted researchers to extend the single-kernel Hilbertian methods of machine learning to a whole range of composite problems that involve direct products or direct sums of RKHS. Examples of practical developments that involve direct product/sums of RKHS are: kernel methods for vector-valued data [78, 79], multi-kernel learning [80, 78], multiscale approximation [81], and semi-parametric models of the form $\widetilde{f} = f + p_0$, where $f \in \mathcal{H}$ (RKHS) and the second component $p_0 \in \mathrm{span}\{p_n\}_{n=1}^{N_0}$ is finite-dimensional [54]. Likewise, the native spaces of variational splines have an inherent direct-sum structure because the underlying regularization functional is a Hilbertian semi-norm [82, 83, 84, 85].

While the Banach counterparts of these methods are still lacking for the most part, there is recent evidence that the use of over-complete dictionaries—in particular, unions of bases—is highly advantageous for the resolution of compressed-sensing problems with sparsity constraints [86, 87, 88, 89, 90, 91, 92]. In the case where the dictionary is a single basis, there is a direct relation between this type of signal recovery and the kind of ℓ_1-regularization problem mentioned in Section 2.1.1 [67]. By taking inspiration from the large body of work already available for RKHS, the next promising step is therefore to investigate this type of reconstruction problem from the unifying perspective of an optimization in a sum of Banach spaces.

2.2 Mathematical Foundations

A Banach space is a complete normed vector space. It is denoted by $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ where $\mathcal{X}$ stands for the vector space and $\|\cdot\|_{\mathcal{X}}$ specifies the underlying norm or, simply, by $\mathcal{X}$ (for short). A Banach space $\mathcal{X}$ has a unique topological dual $\mathcal{X}'$ which is itself a Banach space equipped with the dual norm $\|\cdot\|_{\mathcal{X}'}$ (see (2.3) below). Formally, an element f of the dual space $\mathcal{X}'$ is a continuous linear functional $f : \mathcal{X} \to \mathbb{R}$. Likewise, since $\mathcal{X}$ is embedded in the bidual space $\mathcal{X}'' = (\mathcal{X}')'$, an element $\nu \in \mathcal{X}$, which is therefore also included in $\mathcal{X}''$, can be viewed as a continuous linear functional $\nu : \mathcal{X}' \to \mathbb{R}$. The bilateral character of this association is described by the duality product

$$\mathcal{X} \times \mathcal{X}' \to \mathbb{R} : (\nu, f) \mapsto \langle \nu, f \rangle_{\mathcal{X} \times \mathcal{X}'} = \langle f, \nu \rangle_{\mathcal{X}' \times \mathcal{X}} \in \mathbb{R}, \tag{2.1}$$

which is a map that is linear and continuous in both arguments. To avoid notational overload, we shall henceforth drop the subscript in the specification of the duality product under the understanding that the first argument is a linear functional that acts on the second argument; for instance, $\nu : f \mapsto \langle \nu, f \rangle$, where $f \in \mathcal{X}'$ usually also has a concrete identification as a vector or a function. Mathematically, the continuity of the duality product or, equivalently, the continuity of ν (or of f) viewed as a linear functional—is expressed by the generic duality bound

$$\left| \langle f, \nu \rangle \right| \leq \|\nu\|_{\mathcal{X}} \|f\|_{\mathcal{X}'}, \tag{2.2}$$

which holds for any $(\nu, f) \in \mathcal{X} \times \mathcal{X}'$—for more details, refer to [93, 94]. The upper bound in (2.2) is consistent with the definition of the dual norm

$$\|f\|_{\mathcal{X}'} \stackrel{\Delta}{=} \sup_{\nu \in \mathcal{X} \setminus \{0\}} \frac{\langle f, \nu \rangle}{\|\nu\|_{\mathcal{X}}}. \tag{2.3}$$

In fact, the latter identification suggests that the bound in (2.2) is tight—a property that is embodied in the fundamental notion of duality mapping [95].

Definition 2.1 (Duality mapping)**.** *Let $(\mathcal{X}, \mathcal{X}')$ be a dual pair of Banach spaces. Then, the elements $\nu^* \in \mathcal{X}'$ and $\nu \in \mathcal{X}$ form a $(\mathcal{X}', \mathcal{X})$-conjugate pair if they satisfy:*

1. *Norm preservation: $\|\nu^*\|_{\mathcal{X}'} = \|\nu\|_{\mathcal{X}}$.*
2. *Sharp duality bound: $\langle \nu, \nu^* \rangle = \|\nu\|_{\mathcal{X}} \, \|\nu^*\|_{\mathcal{X}'}$.*

For any given $\nu \in \mathcal{X}$, the set of admissible conjugates defines the duality mapping

$$J(\nu) = \{\nu^* \in \mathcal{X}' : \|\nu^*\|_{\mathcal{X}'} = \|\nu\|_{\mathcal{X}} \text{ and } \langle \nu, \nu^* \rangle = \|\nu\|_{\mathcal{X}} \, \|\nu^*\|_{\mathcal{X}'}\}, \tag{2.4}$$

which is a nonempty subset of $\mathcal{X}'$. Whenever the duality mapping is single-valued (for instance, when $\mathcal{X}'$ is strictly convex), one also defines the duality operator $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$, which is such that $\nu^* = \mathrm{J}_{\mathcal{X}}\{\nu\}$.

Definition 2.2. *A Banach space $\mathcal{X}$ (or its associated norm $\| \cdot \|_{\mathcal{X}}$) is said to be strictly convex if, for all $f_1, f_2 \in \mathcal{X}$ such that $\|f_1\|_{\mathcal{X}} = \|f_2\|_{\mathcal{X}} = 1$ and $f_1 \neq f_2$, one has that $\|\lambda f_1 + (1 - \lambda) f_2\|_{\mathcal{X}} < 1$ for any $\lambda \in (0, 1)$.*

The duality mapping is a powerful mathematical tool that facilitates the investigation of optimization problems in Banach spaces. A primary reference on the topic, which includes the characterization of $\mathrm{J}_{\mathcal{X}}$ for the classical L_p spaces, is [96].

Note that the duality operator $\mathrm{J}_{\mathcal{X}}$ is bijective when $\mathcal{X}$ is reflexive and strictly convex [96, 97], in which case $\mathrm{J}_{\mathcal{X}}^{-1} = \mathrm{J}_{\mathcal{X}'} : \mathcal{X}' \to \mathcal{X}'' = \mathcal{X}$. It can therefore be viewed as the natural generalization of the celebrated Riesz map [98, 94], which describes the linear isometric mapping of a Hilbert space into its dual. The important difference, however, is that the operator $\mathrm{J}_{\mathcal{X}}$ is generally nonlinear. In fact, it is linear if and only if $\mathcal{X}$ is a Hilbert space, in which case it coincides with the Riesz map $\mathcal{X} \to \mathcal{X}'$ [96, 75].

We now define another notion that is used in our generic characterization, specifically when the Banach search space is not strictly convex.

Definition 2.3 (Extremal points). *Let C be a convex set of a Banach space $\mathcal{X}$. The extremal points of C are the points $f \in C$ such that, if there exist $f_1, f_2 \in C$ and $t \in (0,1)$ such that $f = tf_1 + (1-t)f_2$, then it necessarily holds that $f = f_1 = f_2$. The set of these extremal points is denoted by* $\mathrm{Ext}(C)$.

Our final tool is a transformation mechanism that generates application-specific Banach spaces from some primary ones whose basic properties (e.g., the duality mapping and extremal points of the unit ball) are known.

Proposition 2.1 (Isometric isomorphism). *Let $\mathcal{X}$ be a primary Banach space, $\mathcal{Y}$ an arbitrary vector space and $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ a bijective between $\mathcal{X}$ and $\mathcal{Y}$. Then, we have the following properties:*

1. *The vector space $\mathcal{Y}$, equipped with the norm $\|y\|_{\mathcal{Y}} \overset{\Delta}{=} \|\mathrm{T}^{-1}\{y\}\|_{\mathcal{X}}$, is a Banach space that is isometrically isomorphic to $\mathcal{X}$. In other words, the operators $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ and $\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{X}$ are isometries.*
2. *The continuous dual of $\mathcal{Y}$ is $\mathcal{Y}' = \mathrm{T}^{-1*}(\mathcal{X}')$, equipped with the norm $\|y^*\|_{\mathcal{Y}'} = \|\mathrm{T}^*\{y^*\}\|_{\mathcal{X}'}$, where $\mathrm{T}^* : \mathcal{Y}' \to \mathcal{X}'$ is the adjoint operator of* T.
3. *The elements $y^* \in \mathcal{Y}'$ and $y \in \mathcal{Y}$ form a conjugate pair if and only if $x^* = \mathrm{T}^*\{y^*\} \in \mathcal{X}'$ and $x = \mathrm{T}^{-1}\{y\} \in \mathcal{X}$ are themselves $(\mathcal{X}', \mathcal{X})$-Banach conjugates.*

4. *The element $u \in \mathcal{Y}$ is an extremal point of the unit ball in $\mathcal{Y}$ if and only if $e = \mathrm{T}^{-1}\{u\} \in \mathcal{X}$ is an extremal point of the unit ball in $\mathcal{X}$.*

5. *If $\mathcal{X}$ is a Hilbert space, then the spaces $\mathcal{Y}$ and $\mathcal{Y}'$ are Hilbert spaces as well. The corresponding Riesz map is $\mathrm{J}_{\mathcal{Y}} = \mathrm{T}^{-1*}\mathrm{J}_{\mathcal{X}}\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{Y}'$, where $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$ is the Riesz map of the primary space.*

Proof. Since T is linear and one-to-one, the functional $y \mapsto \|\mathrm{T}^{-1}y\|_{\mathcal{X}}$ is a *bona fide* norm on $\mathcal{Y}$. Moreover, from the definition of the $\mathcal{Y}$-norm, we have that

$$\|\mathrm{T}\{x_m\} - \mathrm{T}\{x_n\}\|_{\mathcal{Y}} = \|\mathrm{T}\{x_m - x_n\}\|_{\mathcal{Y}} = \|\mathrm{T}^{-1}\mathrm{T}\{x_m - x_n\}\|_{\mathcal{X}} = \|x_m - x_n\|_{\mathcal{X}}, \quad (2.5)$$

for any $x_m, x_n \in \mathcal{X}$. Together with the bijectivity of T, we deduce that T is an isomorphism between $\mathcal{X}$ and $\mathcal{Y}$. Hence, $\mathcal{Y}$ inherits the topological structure of $\mathcal{X}$. This proves that $\mathcal{Y}$ is indeed a Banach space.

The other properties are immediate consequences of the underlying isometry and the definition of the adjoint, which translate into

$$\langle x_1^*, x_2\rangle_{\mathcal{X}'\times\mathcal{X}} = \langle x_1^*, \mathrm{T}^{-1}\mathrm{T}\{x_2\}\rangle_{\mathcal{X}'\times\mathcal{X}} = \langle \mathrm{T}^{-1*}\{x_1^*\}, \mathrm{T}\{x_2\}\rangle_{\mathcal{Y}'\times\mathcal{Y}} = \langle y_1^*, y_2\rangle_{\mathcal{Y}'\times\mathcal{Y}} \quad (2.6)$$

for any $(x_1^*, x_2) \in \mathcal{X}' \times \mathcal{X}$.

In particular, if $\mathcal{X}$ is a Hilbert space with inner product $(\cdot,\cdot)_{\mathcal{X}}$, then $x^* = \mathrm{J}_{\mathcal{X}}\{x\} \in \mathcal{X}'$ so that $\langle x^*, x\rangle_{\mathcal{X}'\times\mathcal{X}} = \|x\|_{\mathcal{X}}^2 = (x,x)_{\mathcal{X}}$. It follows that $\mathcal{Y} = \mathrm{T}\Big(\mathcal{X}\Big)$ is a Hilbert space equipped with the inner product $(y_1, y_2)_{\mathcal{Y}} = (\mathrm{T}^{-1}y_1, \mathrm{T}^{-1}y_2)_{\mathcal{X}}$. Correspondingly, the dual space $\mathcal{Y}' = \mathrm{T}^{-1*}\Big(\mathcal{X}'\Big)$ is the Hilbert space equipped with the inner product $(y_1^*, y_2^*)_{\mathcal{Y}'} = (\mathrm{T}^* y_1^*, \mathrm{T}^* y_2^*)_{\mathcal{X}'}$. Moreover, we have that $(y_1, y_2)_{\mathcal{Y}} = \langle \mathrm{J}_{\mathcal{Y}}\{y_1\}, y_2\rangle_{\mathcal{Y}'\times\mathcal{Y}} = (\mathrm{J}_{\mathcal{Y}}\{y_1\}, \mathrm{J}_{\mathcal{Y}}\{y_2\})_{\mathcal{Y}'}$, the underlying duality operator (Riesz map) being $\mathrm{J}_{\mathcal{Y}} = \mathrm{T}^{-1*}\mathrm{J}_{\mathcal{X}}\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{Y}'$. □

2.3 Optimization Over Banach Spaces: General Representer Theorem

To set the stage, we now recall the primary results of [75] and [76] and introduce our abstract optimization framework in the form of a single unified theorem.

Theorem 2.1 (General Banach representer theorem)**.** *Let us consider the following setting:*

- *A dual pair $(\mathcal{X}, \mathcal{X}')$ of Banach spaces.*
- *The analysis subspace $\mathcal{N}_{\boldsymbol{\nu}} = \mathrm{span}\{\nu_m\}_{m=1}^M \subset \mathcal{X}$ with the ν_m being linearly independent.*
- *The linear measurement operator $\boldsymbol{\nu} : \mathcal{X}' \to \mathbb{R}^M : f \mapsto \Big(\langle \nu_1, f\rangle, \dots, \langle \nu_M, f\rangle\Big)$.*
- *The proper, lower-semicontinuous, and convex loss functional $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$.*
- *Some arbitrary strictly increasing and convex function $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$.*

Then, for any fixed $\mathbf{y} \in \mathbb{R}^M$, the solution set of the generic optimization problem

$$S = \arg\min_{f \in \mathcal{X}'} \Big(E\Big(\mathbf{y}, \boldsymbol{\nu}(f)\Big) + \psi\left(\|f\|_{\mathcal{X}'}\right)\Big) \tag{2.7}$$

is nonempty, convex, and weak-compact.*

When E is strictly convex, or if it imposes the equality constraint $\mathbf{y} = \boldsymbol{\nu}(f)$, then any solution $f_0 \in S \subset \mathcal{X}'$ is an $(\mathcal{X}', \mathcal{X})$-conjugate of a common $\nu_0 \in \mathcal{N}_{\boldsymbol{\nu}} \subset \mathcal{X}$, so that $S \subseteq J(\nu_0)$. Depending on the type of Banach space, this then results in the following description of the solution(s):

- *If $\mathcal{X}'$ is a Hilbert space and ψ is strictly convex, then the solution is unique* *and admits the linear representation with parameter* $\mathbf{a} \in \mathbb{R}^M$ *given as*

$$f_0 = \sum_{m=1}^{M} a_m \varphi_m, \tag{2.8}$$

with $\varphi_m = \mathrm{J}_{\mathcal{X}}\{\nu_m\} \in \mathcal{X}'$*.*

- *If* $\mathcal{X}'$ *is a strictly convex Banach space and* ψ *is strictly convex, then the solution is unique and admits the parametric representation*

$$f_0 = \mathrm{J}_{\mathcal{X}} \left\{ \sum_{m=1}^{M} a_m \nu_m \right\}. \tag{2.9}$$

- *Otherwise, when* $\mathcal{X}'$ *is not strictly convex, the solution set is the weak*-closure of the convex hull of its extremal points which can all be expressed as*

$$f_0 = \sum_{k=1}^{K_0} c_k e_k \tag{2.10}$$

for some $K_0 \leq M$*,* $c_1, \dots, c_{K_0} \in \mathbb{R}$*, where* $e_1, \dots, e_{K_0} \in \mathcal{X}'$ *are some extremal points of the unit ball* $B_{\mathcal{X}'} = \{x \in \mathcal{X} : \|x\|_{\mathcal{X}'} \leq 1\}$*.*

While the characterization given by (2.10) is always valid, it is practical only when the Banach space $\mathcal{X}'$ has a unit ball $B_{\mathcal{X}'}$ with comparatively much fewer extremal points than boundary points. The prototypical case is $\ell_1(\mathbb{Z})$, whose extremal points $\mathrm{Ext}(B_{\ell_1(\mathbb{Z})}) = \{\pm\delta[\cdot - m]\}_{m\in\mathbb{Z}}$ (the signed Kronecker impulses shifted by m) are indexable, while its boundary points $\{u[\cdot] \in \ell_1(\mathbb{Z}) : \|u\|_{\ell_1} = 1\}$ are uncountable. The extremal points of $B_{\mathcal{X}'}$ can then be interpreted as the elements of a constrained dictionary. This means that the linear expansion in (2.10) is adaptive, meaning that the actual choice of K_0 and of the basis functions $e_k \in \mathcal{X}'$ is data-dependent. This is the main difference with the two other cases for which Theorem 2.1 provides an explicit description of the M-dimensional solution manifold.

We also remark that the representation (2.10) for the case where the solution is non-unique is deducible from Theorem 3.1 of [76] by viewing an extreme point of the solution set as the degenerate case of a face with dimension $j = 0$. Due to the particular importance of the representation (2.10) and since the main result of [76] is more general than what is required here, we now provide an alternative proof for this part.

Proof of (2.10). The proof presented in [76] relies on an earlier theorem by Klee [99], which is itself based on a foundational result by Dubins on the extreme points of the intersection of a convex set and a series of hyperplanes [100]. Here, we have chosen the latter as our starting point in order to simplify the argumentation.

Theorem 2.2 (Main result of [100]). *Consider an arbitrary vector space V over the field of real numbers, a convex set $C \subseteq V$ and M hyperplanes $H_1, \dots, H_M \subset V$. Then, any extreme point of $C \cap \left(\bigcap_{m=1}^{M} H_m\right)$ can be written as a convex combination of at most $M + 1$ extreme points of C.*

For the Banach space $\mathcal{X}'$, we denote the unit ball of size β as $B_{\mathcal{X}',\beta} = \{f \in \mathcal{X}' : \|f\|_{\mathcal{X}'} \leq \beta\}$. It then directly follows from Theorem 2.2 that any extreme point of

$$\mathcal{S} = B_{\mathcal{X}',\beta} \cap \{f \in \mathcal{X}' : \boldsymbol{\nu}(f) = \mathbf{y}\} \tag{2.11}$$

can be written as a convex combination of at most $M + 1$ extreme points of $B_{\mathcal{X}',\beta}$. In what follows, we show that, if

$$\beta = \min_{f \in \mathcal{X}'} \|f\|_{\mathcal{X}'} \quad \text{s.t.} \quad \boldsymbol{\nu}(f) = \mathbf{y}, \tag{2.12}$$

then any extreme points f_0 of $\mathcal{S}$ has the expansion

$$f_0 = \sum_{k=1}^{K} c_k f_k, \quad K \leq M, \tag{2.13}$$

where $f_k \in \mathrm{Ext}(B_{\mathcal{X}',\beta})$, and $c_k > 0$ with $\sum_{k=1}^{K} c_k = 1$. The connection with (2.10) is that (2.13) is obviously also expressible in terms of the basis vectors $e_k = f_k/\beta$ which, due to the homogeneity property of the norm, are extremal points of the unit ball in $\mathcal{X}'$.

Assume by contradiction that $K = M+1$ and that the set $\{f_1, \dots, f_{M+1}\}$ is linearly independent. The set of vectors $\{\boldsymbol{\nu}(f_1), \dots, \boldsymbol{\nu}(f_{M+1})\} \subseteq \mathbb{R}^M$ is clearly linearly dependent. Hence, there exists $(\alpha_m)_{m=1}^{M+1} \neq \mathbf{0}$ such that

$$\boldsymbol{\nu}(\sum_{m=1}^{M+1} \alpha_m \mathbf{f}_m) = \sum_{m=1}^{M+1} \alpha_m \boldsymbol{\nu}(f_m) = \mathbf{0}. \tag{2.14}$$

Denote $A = \sum_{m=1}^{M+1} \alpha_m$ and consider the function $f_\epsilon = f_0 + \epsilon \sum_{m=1}^{M+1} \alpha_m f_m$ for $\epsilon \in \mathbb{R}$. On one hand, for all values of ϵ with $|\epsilon| < \epsilon_{\max} = \frac{\min_m c_m}{\max_m |\alpha_m|}$, the function

$$\frac{f_\epsilon}{1+\epsilon A} = \sum_{m=1}^{M+1} \frac{c_m + \epsilon \alpha_m}{1+\epsilon A} f_m \tag{2.15}$$

is in the convex hull of $\{f_1, \dots, f_{M+1}\}$. Consequently, $\|f_\epsilon\|_{\mathcal{X}'} \leq |1+\epsilon|\beta$. On the other hand, due to (2.14), we have

$$\boldsymbol{\nu}(f_\epsilon) = \boldsymbol{\nu}(f_0) + \epsilon \sum_{m=1}^{M+1} \alpha_m \mathbf{y}_m = \mathbf{y}. \tag{2.16}$$

Hence, due to the optimality of f_0, we deduce that

$$|1+\epsilon A|\beta \geq \|f_\epsilon\|_{\mathcal{X}'} \geq \|f_0\|_{\mathcal{X}'} = \beta, \quad \forall \epsilon \in (-\epsilon_{\max}, \epsilon_{\max}). \tag{2.17}$$

This yields that $|1+\epsilon A| \geq 1$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$, which implies that $A = 0$. Consequently, $f_\epsilon \in \mathcal{S}$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$. Now, since f_0 is an extreme point of $\mathcal{S}$, we deduce from $f_0 = \frac{f_{-\epsilon} + f_\epsilon}{2}$ that $f_0 = f_\epsilon$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$, and hence, $\sum_{m=1}^{M+1} \alpha_m f_m = 0$. This is in contradiction with the linear independence of $\{f_1, \dots, f_{M+1}\}$. □

While the abstract characterization in Theorem 2.1 is remarkably general, it is practical only for the cases in which the duality operator $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$ or the extremal points of the unit ball in $\mathcal{X}'$ are known explicitly, for instance when $\mathcal{X}'$ is a RKHS [77, 8, 61] or when the underlying norm is a variant of the ℓ_1-norm that promotes sparsity [68, 76, 72]. We shall extend the applicability of Theorem 2.1 by starting from basic building blocks (elementary Banach constituents) and by showing how these can be combined via the use of linear transforms and of direct sums to specify more complex regularization norms that can accommodate mixture models.

2.4 Composite Norms and Direct-Sum Spaces

In order to offer flexibility in the specification of direct-product or direct-sum topologies, we introduce the finite-dimensional space $\mathcal{Z} = (\mathbb{R}^N, \|\cdot\|_{\mathcal{Z}})$. The underlying norm is said to be *monotone* if

$$\|(a_1, \dots, a_N)\|_{\mathcal{Z}} \leq \|(b_1, \dots, b_N)\|_{\mathcal{Z}} \tag{2.18}$$

whenever $0 \leq |a_n| \leq |b_n|$ for each $n = 1, \dots, N$, and, *absolute* if $\|\mathbf{z}\|_{\mathcal{Z}} = \|(z_n)\|_{\mathcal{Z}} = \|(|z_n|)\|_{\mathcal{Z}}$ for any $\mathbf{z} \in \mathbb{R}^N$. It is also known that a norm is monotone if and only if it is absolute [101, Theorem 2]. For instance, the latter property is obviously satisfied for $\|\cdot\|_{\mathcal{Z}} = \|\cdot\|_p$ with $p \geq 1$, as well as for any weighted version thereof. Moreover, the dual of an absolute norm is again absolute [101, Theorem 1]. Given a series $\mathcal{X}_1, \dots, \mathcal{X}_N$ of Banach spaces, we then write $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$ for the direct-product space equipped with the composite norm

$$\|(x_1, \dots, x_N)\| = \|(\|x_1\|_{\mathcal{X}_1}, \dots, \|x_N\|_{\mathcal{X}_N})\|_{\mathcal{Z}}. \tag{2.19}$$

Likewise, one can construct (internal) direct-sum spaces via the summation of complemented Banach constituents.

Definition 2.4. *A series $\mathcal{X}_1, \dots, \mathcal{X}_N$ of Banach subspaces of $\mathcal{X}$ is said to be complemented if $\mathcal{X} = \mathcal{X}_1 + \cdots + \mathcal{X}_N = \{x = x_1 + \cdots + x_N : x_n \in \mathcal{X}_n, n = 1, \dots, N\}$ (as a set) and $\mathcal{X}_{n_1} \cap \sum_{n \neq n_1} \mathcal{X}_n = \{0\}$ when $n_1 = 1, \dots, N$.*

In that scenario, any $x \in \mathcal{X}$ has a unique representation as $x = x_1 + \cdots + x_N$ with $x_n = \mathrm{Proj}_{\mathcal{X}_n}\{x\} \in \mathcal{X}_n$, where $\mathrm{Proj}_{\mathcal{X}_n} : \mathcal{X} \to \mathcal{X}_n$ is the corresponding projection operator. We then designate $\mathcal{X} = (\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ as the (internal) direct-sum space equipped with the norm

$$\|x\|_{\mathcal{X}} = \|(\|\mathrm{Proj}_{\mathcal{X}_1}\{x\}\|_{\mathcal{X}_1}, \dots, \|\mathrm{Proj}_{\mathcal{X}_N}\{x\}\|_{\mathcal{X}_N})\|_{\mathcal{Z}}. \tag{2.20}$$

We observe that (2.20) is compatible with (2.19) because $\mathrm{Proj}_{\mathcal{X}_{n_1}}: \mathcal{X} \to \mathcal{X}_{n_1}$ is such that

$$\mathrm{Proj}_{\mathcal{X}_{n_1}}\{x_n\} = \begin{cases} x_{n_1}, & \text{for } n = n_1 \\ 0, & \text{otherwise} \end{cases} \tag{2.21}$$

for any $x_n \in \mathcal{X}_n$. This identification, together with the unicity of the sum decomposition, implies that $(\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ is a Banach space that is isometrically isomorphic to $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$.

We now highlight the key properties of the constructed direct-product/-sum spaces.

Theorem 2.3. *Let $(\mathcal{X}'_1, \mathcal{X}_1), \dots, (\mathcal{X}'_N, \mathcal{X}_N)$ be a series of dual pairs of Banach spaces and $\|\cdot\|_{\mathcal{Z}}$ a norm on $\mathbb{R}^N$ that is absolute. Then, we have the following properties:*

1. *The continuous dual of $\mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$ is the direct-product space $\mathcal{X}' = (\mathcal{X}'_1 \times \cdots \times \mathcal{X}'_N)_{\mathcal{Z}'}$.*
2. *The elements $y = (y_1, \dots, y_N) \in \mathcal{X}'$ and $x = (x_1, \dots, x_N) \in \mathcal{X}$ form a conjugate pair if and only if $y_n = \alpha_n x_n^*$, where $x_n^* \in \mathcal{X}'_n$ is a Banach conjugate of $x_n \in \mathcal{X}_n$ and $\alpha_n \in \mathbb{R}_{\geq 0}$ is given by*

$$\alpha_n = \begin{cases} \frac{z_n^*}{\|x_n\|_{\mathcal{X}_n}} > 0, & x_n \neq 0 \\ 0, & \textit{otherwise} \end{cases} \tag{2.22}$$

   with $\mathbf{z}^* = (z_n^*) \in \mathcal{Z}'$ a Banach conjugate of $\mathbf{z} = (\|x_1\|_{\mathcal{X}_1}, \dots, \|x_N\|_{\mathcal{X}_N}) \in \mathcal{Z}$.
3. *The element $e = (e_1, \dots, e_N) \in \mathcal{X}$ is an extremal point of the unit ball in $\mathcal{X}$ if and only if $(\|e_1\|_{\mathcal{X}_1}, \dots, \|e_N\|_{\mathcal{X}_N})$ is an extremal point of the unit ball in $\mathcal{Z}$, and for each $1 \leq n \leq N$ with $e_n \neq 0$, $\frac{e_n}{\|e_n\|_{\mathcal{X}_n}}$ is an extremal point of the unit ball of $\mathcal{X}_n$.*
4. *If the $\mathcal{X}_n$ are complemented Banach subspaces of the (sum) space $\mathcal{X}_{\mathrm{sum}}$, then the continuous dual of $\mathcal{X}_{\mathrm{sum}} = (\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ is the direct-sum Banach space $\mathcal{X}'_{\mathrm{sum}} = (\mathcal{X}'_1 \oplus \cdots \oplus \mathcal{X}'_N)_{\mathcal{Z}'}$, which is isometrically isomorphic to the direct-product space $\mathcal{X}'$ in Item 1. Consequently, the properties in Item 2 and 3 also apply, with the convention that $x_n = \mathrm{Proj}_{\mathcal{X}_n}\{x\}$ and $y_n = \mathrm{Proj}_{\mathcal{X}'_n}\{y\}$ for $n = 1, \dots, N$.*

Proof. An element $y = (y_1, \dots, y_N)$ of $\mathcal{X}'$ is identified with the linear functional

$$x = (x_1, \dots, x_N) \mapsto \langle y, x\rangle_{\mathcal{X}'\times\mathcal{X}} = \sum_{n=1}^{N} \langle y_n, x_n\rangle_{\mathcal{X}'_n\times\mathcal{X}_n}. \tag{2.23}$$

The first property is a basic result in the theory of Banach spaces [93, Theorem 1.10.13] when the outer norm is Euclidean with $\mathcal{Z} = \mathcal{Z}' = (\mathbb{R}^N, \|\cdot\|_2)$. The present setting is more general so that we need to prove that the dual norm of $y = (y_1, \dots, y_N) \in \mathcal{X}'$ is precisely

$$\|y\|_{\mathcal{X}'} = \sup_{\|x\|_{\mathcal{X}}=1} \langle y, x\rangle = \left\| (\|y_1\|_{\mathcal{X}'_1}, \dots, \|y_N\|_{\mathcal{X}'_N}) \right\|_{\mathcal{Z}'}. \tag{2.24}$$

Since the spaces $(\mathcal{X}'_n, \mathcal{X}_n)$ form dual pairs, we have the generic duality inequalities

$$\langle y_n, x_n\rangle_{\mathcal{X}'_n\times\mathcal{X}_n} \le \left|\langle y_n, x_n\rangle_{\mathcal{X}'_n\times\mathcal{X}_n}\right| \le \|y_n\|_{\mathcal{X}'_n}\|x_n\|_{\mathcal{X}_n} \tag{2.25}$$

with equality if and only if $y_n = \alpha_n x_n^*$ for some $\alpha_n \in \mathbb{R}^+$. This implies that, for any $(y, x) \in \mathcal{X}' \times \mathcal{X}$, we have that

$$\langle y, x\rangle_{\mathcal{X}'\times\mathcal{X}} = \sum_{n=1}^{N} \langle y_n, x_n\rangle_{\mathcal{X}'_n\times\mathcal{X}_n} \le \sum_{n=1}^{N} \left|\langle y_n, x_n\rangle_{\mathcal{X}'_n\times\mathcal{X}_n}\right| \le \sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n}\|x_n\|_{\mathcal{X}_n} \tag{2.26}$$

Likewise, by setting $\mathbf{y} = (\|y_1\|_{\mathcal{X}'_1}, \dots, \|y_N\|_{\mathcal{X}'_N}) \in \mathcal{Z}'$ and $\mathbf{z} = (\|x_1\|_{\mathcal{X}_1}, \dots, \|x_N\|_{\mathcal{X}_N}) \in \mathcal{Z}$, we write the complementary duality inequality

$$\sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n}\|x_n\|_{\mathcal{X}_n} = \langle \mathbf{y}, \mathbf{z}\rangle_{\mathcal{Z}'\times\mathcal{Z}} \le |\langle \mathbf{y}, \mathbf{z}\rangle_{\mathcal{Z}'\times\mathcal{Z}}| \le \|\mathbf{y}\|_{\mathcal{Z}'}\|\mathbf{z}\|_{\mathcal{Z}}. \tag{2.27}$$

By observing that $\|\mathbf{z}\|_{\mathcal{Z}} = \|x\|_{\mathcal{X}}$ and combining these inequalities, we get that

$$\langle y, x\rangle_{\mathcal{X}'\times\mathcal{X}} \le \sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n}\|x_n\|_{\mathcal{X}_n} \le \|\mathbf{y}\|_{\mathcal{Z}'}\|\mathbf{x}\|_{\mathcal{Z}} = \|\mathbf{y}\|_{\mathcal{Z}'}\|x\|_{\mathcal{X}}, \tag{2.28}$$

which shows that $\|y\|_{\mathcal{X}'}$ is upper-bounded by $\|\mathbf{y}\|_{\mathcal{Z}'} = \left\|(\|y_1\|_{\mathcal{X}_1'}, \dots, \|y_N\|_{\mathcal{X}_N'})\right\|_{\mathcal{Z}'}$. To prove that we actually have $\|y\|_{\mathcal{X}'} = \|\mathbf{y}\|_{\mathcal{Z}'}$, for any $\epsilon > 0$, we need to find $x_\epsilon \in \mathcal{X}$ with $\|x_\epsilon\|_{\mathcal{X}} = 1$ such that

$$\langle y, x_\epsilon \rangle_{\mathcal{X}' \times \mathcal{X}} \geq \|\mathbf{y}\|_{\mathcal{Z}'} - \epsilon. \tag{2.29}$$

By definition of the dual norm $\| \cdot \|_{\mathcal{Z}'}$, we have that

$$\|\mathbf{y}\|_{\mathcal{Z}'} = \sup_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^N \\ \|\boldsymbol{\alpha}\|_{\mathcal{Z}} \leq 1}} \mathbf{y}^T \boldsymbol{\alpha}. \tag{2.30}$$

Since $\mathbb{R}^N$ is a finite-dimensional vector-space, the unit ball $B_{\mathcal{Z}} = \{\alpha \in \mathbb{R}^N : \|\boldsymbol{\alpha}\|_{\mathcal{Z}} \leq 1\}$ is compact. Hence, there exists a vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N) \in B_{\mathcal{Z}}$ that attains the supremum in (2.30). In other words,

$$\|\mathbf{y}\|_{\mathcal{Z}'} = \mathbf{y}^T \boldsymbol{\alpha} = \sum_{n=1}^N \|y_n\|_{\mathcal{X}_n'} \alpha_n. \tag{2.31}$$

Similarly, for any $\epsilon > 0$, the definition of the dual norm implies the existence of unit-norm elements $x_n \in \mathcal{X}_n$ for $n = 1, \dots, N$ such that

$$\langle y_n, x_n \rangle_{\mathcal{X}_n' \times \mathcal{X}_n} \geq \|y_n\|_{\mathcal{X}_n'} - \frac{2\epsilon}{N(\alpha_n^2 + 1)}. \tag{2.32}$$

We then set $x_\epsilon = (\alpha_1 x_1, \dots, \alpha_N x_N) \in \mathcal{X}$ and observe that

$$\|x_\epsilon\|_{\mathcal{X}} = \|(\|\alpha_1 x_1\|_{\mathcal{X}_1}, \dots, \|\alpha_N x_N\|_{\mathcal{X}_N})\|_{\mathcal{Z}} = \|(|\alpha_1|, \dots, |\alpha_N|)\|_{\mathcal{Z}} = 1. \tag{2.33}$$

Based on (2.32) and the inequality $\frac{\alpha}{\alpha^2+1} \leq \frac{1}{2}$ for all $\alpha \in \mathbb{R}$, we then deduce that

$$\begin{aligned}\langle y, x_\epsilon \rangle_{\mathcal{X}' \times \mathcal{X}} &= \sum_{n=1}^N \langle y_n, \alpha_n x_n \rangle_{\mathcal{X}_n' \times \mathcal{X}_n} \geq \sum_{n=1}^N \alpha_n \left(\|y_n\|_{\mathcal{X}_n'} - \frac{2\epsilon}{N(\alpha_n^2 + 1)} \right) \\ &= \sum_{n=1}^N \alpha_n \|y_n\|_{\mathcal{X}_n'} - \frac{2\epsilon}{N} \sum_{n=1}^N \frac{\alpha_n}{\alpha_n^2 + 1} = \|\mathbf{y}\|_{\mathcal{Z}'} - \frac{2\epsilon}{N} \sum_{n=1}^N \frac{\alpha_n}{\alpha_n^2 + 1} \geq \|\mathbf{y}\|_{\mathcal{Z}'} - \epsilon,\end{aligned} \tag{2.34}$$

which, in light of the inequality $\|y\|_{\mathcal{X}'} \leq \|\mathbf{y}\|_{\mathcal{Z}'}$, allows us to conclude that $\|y\|_{\mathcal{X}'} = \|\mathbf{y}\|_{\mathcal{Z}'}$.

To prove the second property, we observe that $y \in \mathcal{X}'$ and $x \in \mathcal{X}$ form a conjugate pair if and only if an equality occurs in both (2.26) and (2.27). Inequalities (2.25) and (2.26) are saturated if and only if $y_n = \alpha_n x_n^*$, $\alpha_n \in \mathbb{R}_{\geq 0}$, and (x_n^*, x_n) form a $(\mathcal{X}_n', \mathcal{X}_n)$-conjugate pair. The saturation of (2.27) with $\|y\|_{\mathcal{X}'} = \|\mathbf{y}\|_{\mathcal{Z}'} = \|\mathbf{z}\|_{\mathcal{Z}} = \|x\|_{\mathcal{X}}$ is then equivalent to $\mathbf{y} = \mathbf{z}^* = (z_1^*, \dots, z_N^*)$. Under the assumption that $x_n \neq 0$, this yields $\alpha_n = \frac{z_n^*}{\|x_n^*\|_{\mathcal{X}_n'}}$, which is the announced result since $\|x_n^*\|_{\mathcal{X}_n'} = \|x_n\|_{\mathcal{X}_n}$.

The third property is due to Dowling and Saejung [102]. Finally, the last statement is a direct consequence of the isometric isomorphism between $\mathcal{X}_{\text{sum}} = (\mathcal{X}_1 \oplus \dots, \oplus \mathcal{X}_N)_{\mathcal{Z}}$ and $\mathcal{X} = (\mathcal{X}_1 \times \dots, \times \mathcal{X}_N)_{\mathcal{Z}}$. □

In particular, if $\|\cdot\|_{\mathcal{Z}} = \|\cdot\|_2$ is the usual Euclidean norm, then $\mathbf{z}^*$ in Property 2 is unique and coincides with $\mathbf{z}$, which implies that the Banach conjugate of $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}$ is simply $\mathbf{x}^* = (x_1^*, \dots, x_N^*) \in \mathcal{X}'$.

The combination of these preparatory results and Theorem 2.1 allows us to deduce the following.

Theorem 2.4 (Representer theorem for direct-product spaces)**.** *If the space $\mathcal{X}'$ in Theorem 2.1 has a direct-product decomposition as $\mathcal{X}' = (\mathcal{X}_1' \times \cdots \times \mathcal{X}_N')_{\mathcal{Z}'}$ with predual $\mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$, where $(\mathcal{X}_1', \mathcal{X}_1), \dots, (\mathcal{X}_N', \mathcal{X}_N)$ are dual pairs of Banach spaces and both E and ψ are strictly convex, then the solutions $f_0 = (f_{0,1}, \dots, f_{0,N}) \in S \subset \mathcal{X}'$ of the optimization problem* (2.7) *are $(\mathcal{X}', \mathcal{X})$-Banach conjugates of a common*

$$\nu_0 = (\nu_{0,1}, \dots, \nu_{0,N}) = \sum_{m=1}^{M} a_m \nu_m, \tag{2.35}$$

where $\nu_m = (\nu_{m,1}, \dots, \nu_{m,N}) \in \mathcal{X}$ with $\nu_{m,n} \in \mathcal{X}_n$ and a suitable set of coefficients $\mathbf{a} \in \mathbb{R}^M$. Moreover, depending of the properties of the underlying Banach constituents, the solution components $f_{0,n} \in \mathcal{X}_n'$ have the following characterization *with predefined scaling constants:*

$$\alpha_n = \begin{cases} \frac{y_n}{y_n^*} > 0, & y_n \neq 0 \\ 0, & \textit{otherwise}, \end{cases} \tag{2.36}$$

where $\mathbf{y} = (\|f_{0,1}\|_{\mathcal{X}_1'}, \dots, \|f_{0,N}\|_{\mathcal{X}_N'})$ *and* $\mathbf{y}^* = \mathrm{J}_{\mathcal{Z}'}\{\mathbf{y}\}$*:*

- *If* $\mathcal{X}_n'$ *is a Hilbert space and* $\mathcal{Z}'$ *is strictly convex, then* $f_{0,n}$ *is unique and admits the linear representation*

$$f_{0,n} = \alpha_n \sum_{m=1}^{M} a_m \varphi_{m,n} \tag{2.37}$$

  with $\varphi_{m,n} = \mathrm{J}_{\mathcal{X}_n}\{\nu_{m,n}\} \in \mathcal{X}_n'$*, where* $\mathrm{J}_{\mathcal{X}_n}$ *is the Riesz map* $\mathcal{X}_n \to \mathcal{X}_n'$*.*

- *If* $\mathcal{X}_n'$ *is a strictly convex Banach space and* $\mathcal{Z}'$ *is strictly convex, then the solution component is unique and admits the parametric representation*

$$f_{0,n} = \alpha_n \mathrm{J}_{\mathcal{X}_n}\left\{\sum_{m=1}^{M} a_m \nu_{m,n}\right\} \tag{2.38}$$

  where $\mathrm{J}_{\mathcal{X}_n}$ *is the (nonlinear) duality operator* $\mathcal{X}_n \to \mathcal{X}_n'$*.*

- *If* $\mathcal{X}_n'$ *is a non-strictly convex Banach space, then the subcomponent solution set* $S|_{\mathcal{X}_n'}$ *is the weak*-closure of the convex hull of its extremal points, which can all be expressed as*

$$f_{0,n} = \sum_{k=1}^{K_0} c_{k,n} e_{k,n}, \tag{2.39}$$

  where $e_{1,n}, \dots, e_{K_0,n} \in \mathcal{X}_n'$ *are some extremal points of the unit ball in* $\mathcal{X}_n'$ *and* $c_{1,n}, \dots, c_{K_0,n} \in \mathbb{R}$ *some appropriate weights; the (minimal) number of atoms* $K_0 \leq M$ *is common to all the components associated with non-reflexive Banach spaces.*

In the particular case where $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$, (2.39) *can be replaced by*

$$f_{0,n} = \sum_{k=1}^{K_n} c_{k,n} e_{k,n} \tag{2.40}$$

with $\sum_{n=1}^{N} K_n \leq M$. *In addition,* (2.38) *(resp.* (2.37)*) remains valid for the components for which the space* $\mathcal{X}'_n$ *is strictly convex (resp., Hilbertian), with the caveat that the solution is no longer guaranteed to be unique; this, then, contributes a degenerate version of* (2.40) *with* $K_n = 1$, $c_{1,n} = \|f_{0,n}\|_{\mathcal{X}'_n}$, *and* $e_{1,n} = f_{0,n}/\|f_{0,n}\|_{\mathcal{X}'_n}$.

Proof. The existence of solutions $f_0 \in \mathcal{X}'$ and the property that $S \subseteq J_{\mathcal{X}}(\nu_0)$ for some $\nu_0 = \sum_{m=1}^{M} a_m \nu_m \in \mathcal{N}_{\boldsymbol{\nu}}$ is ensured by Theorem 2.1. We then proceed in three steps.

(i) Constant value of $\psi(\|f_0\|_{\mathcal{X}'})$ for all $f_0 \in S$.
The key here is the strict convexity of $f \mapsto E(\mathbf{y}, \boldsymbol{\nu}(f))$ together with the convexity of $f \mapsto \psi(\|f\|_{\mathcal{X}'})$. By applying a standard argument (by contradiction) that uses the convexity of S, we show that there exist two constants C_1 and C_2 such that $E(\mathbf{y}, \boldsymbol{\nu}(f_0)) = C_1$ and $\psi(\|f_0\|_{\mathcal{X}'}) = C_2$ for all $f_0 \in S$ (see, for instance, the last part of the proof in [74, Appendix B]). By invoking the strict convexity of E, this then implies that all solutions share the same measurement vector $\mathbf{z}_0 = \boldsymbol{\nu}(f_0)$. Likewise, when ψ is strictly convex, we readily deduce that $\|f_0\|_{\mathcal{X}'}$ takes a constant value.

(ii) Uniqueness of $\|f_{0,n}\|_{\mathcal{X}'_n}$ in the strictly-convex case.
To show that $\|f_{0,n}\|_{\mathcal{X}'_n} = y_n$ holds for all $f_0 \in S$, we suppose that there exists another solution $\widetilde{f_0} \in S$ such that $\|\widetilde{f_0}\|_{\mathcal{X}'} = \|f_0\|_{\mathcal{X}'}$ and $\|\widetilde{f}_{0,n}\|_{\mathcal{X}'_n} = \widetilde{y}_n$ with $\widetilde{\mathbf{y}} \neq \mathbf{y}$. Since S is convex, $\lambda\widetilde{f_0} + (1-\lambda)f_0$ with any $\lambda \in (0,1)$ must also be a solution with associated norm $\|\lambda\widetilde{f_0} + (1-\lambda)f_0\|_{\mathcal{X}'} \leq \|\lambda\widetilde{\mathbf{y}} + (1-\lambda)\mathbf{y}\|_{\mathcal{Z}'}$, by the triangle inequality. However, the norm equality $\|\widetilde{f_0}\|_{\mathcal{X}'} = \|\widetilde{\mathbf{y}}\|_{\mathcal{Z}'} = \|\mathbf{y}\|_{\mathcal{Z}'}$ and the strict-convexity of $\|\cdot\|_{\mathcal{Z}'}$ (see Definition 2.2) implies that $\|\lambda\widetilde{\mathbf{y}} + (1-\lambda)\mathbf{y}\|_{\mathcal{Z}'} < \|\mathbf{y}\|_{\mathcal{Z}'} = \|f_0\|_{\mathcal{X}'}$, which results in a contradiction.

(iii) Generic form of the solution component $f_{0,n}$.
We assume that $y_n = \|f_{0,n}\|_{\mathcal{X}'} \neq 0$; otherwise, we simply have that $f_{0,n} = 0$.

From Property 2 in Theorem 2.3, we know that $f_0 = (f_{0,1}, \dots, f_{0,N})$ and $\nu_0 = (\nu_{0,1}, \dots, \nu_{0,N})$ form a conjugate pair if and only if there exists $f^*_{0,n} \in J_{\mathcal{X}'_n}(f_{0,n})$ such that $\nu_{0,n} = (y^*_n/y_n)f^*_{0,n}$, where $\mathbf{y}^* = (y^*_1, \dots, y^*_N) = \mathrm{J}_{\mathcal{X}'}\{\mathbf{y}\}$.

When $\mathcal{X}'_n$ is strictly convex, the duality mapping is single-valued. The representations in (2.37) and (2.39) then directly follow from the primary expansion $\nu_{0,n} = \sum_{m=1}^{M} a_m \nu_{m,n}$ and the homogeneity property of the duality mapping expressed as $\mathrm{J}_{\mathcal{X}}\{\alpha\nu\} = \alpha \mathrm{J}_{\mathcal{X}}\{\nu\}$ for any $\nu \in \mathcal{X}$ and $\alpha \in \mathbb{R}_{\geq 0}$ (see [96]).

Since S is convex and weak*-compact, we can invoke the Krein-Milman theorem, which states that S is the closure of the convex hull of its extremal points. The same holds true for the convex set $S|_{\mathcal{X}'_n}$ (the restriction of S on $\mathcal{X}'_n$) with $\mathrm{Ext}(S|_{\mathcal{X}'_n}) \subseteq \mathrm{Ext}(S)|_{\mathcal{X}'_n}$. By recalling that all points $f_0 \in \mathrm{Ext}(S)$ can be represented as $f_0 = (f_{0,1}, \dots, f_{0,N}) = \sum_{k=1}^{K_0} c_k e_k$, where $e_k = (e_{k,1}, \dots, e_{k,N}) \in \mathrm{Ext}(B_{\mathcal{X}'})$ and $K_0 \leq M$ (by Theorem 2.1), we obtain that

$$f_{0,n} = \sum_{k=1}^{K_0} c_k \|e_{k,n}\|_{\mathcal{X}'_n} \widetilde{e}_{k,n}, \tag{2.41}$$

where $\widetilde{e}_{k,n} = e_{k,n}/\|e_{k,n}\|_{\mathcal{X}'_n}$ are extremal points of the unit ball in $\mathcal{X}'_n$ (by Theorem 2.3, Property 3). The announced statement with $c_{k,n} = \|e_{k,n}\|_{\mathcal{X}'_n} c_k$ then follows from the property that (2.41) is valid for all $f_{0,n} \in \mathrm{Ext}(S)|_{\mathcal{X}'_n} \supseteq \mathrm{Ext}(S|_{\mathcal{X}'_n})$. In fact, Property 3 in Theorem 2.3 tells us that the subset of points $f_{0,n} \in \mathrm{Ext}(S|_{\mathcal{X}'_n})$ are those for which $\mathbf{e}_k = \mathbf{y}/\|\mathbf{y}\|_{\mathcal{Z}'}$ are extremal points of the unit ball in $\mathcal{Z}'$. In particular, when $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$ (outer ℓ_1-norm), the $\mathbf{e}_k$ all have the binary form $(0, 0, \dots, \pm 1, 0, \dots)$ with a single active coefficient at $n = n_k$, which then yields (2.40). □

The outcome of Theorem 2.4 is that the generic form of the solution in Theorem 2.1 is essentially transferred to the direct-product components, with the distribution of the relative energy being controlled by the outer norm $\|\cdot\|_{\mathcal{Z}'}$. The effect of the ℓ_1-norm is significant in that respect because it acts as a threshold that selectively blocks certain solution components and lets others through.

2.5 Convex Optimization in Sums of Banach Spaces

The techniques that we describe next are relevant to inverse problems for which the solution f_0 can be decomposed into a sum of components that have distinct smoothness and/or sparsity properties. The solution then lives in a sum of Banach spaces. Beside the reconstruction of f_0 from the noisy measurement $\mathbf{y} = \boldsymbol{\nu}(f) + \boldsymbol{\epsilon}$, we are now faced with the additional challenge of disambiguating the individual components of the solution.

Let $\mathcal{X}'_1, \dots, \mathcal{X}'_N$ be a series of Banach spaces whose elements are indexed over the same domain. We then define the sum space

$$\mathcal{X}'_1 + \cdots + \mathcal{X}'_N = \{f = f_1 + f_2 + \cdots + f_N : f_n \in \mathcal{X}'_n, n = 1, \dots, N\}. \tag{2.42}$$

Given a linear measurement operator $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M) : \mathcal{X}'_1 + \cdots + \mathcal{X}'_N \to \mathbb{R}^M$ with $\nu_n \in \cap_{n=1}^N \mathcal{X}_n$ and a set of measurements $\mathbf{y} \in \mathbb{R}^M$, we are then interested in the study of the solvability of the convex optimization problem

$$S = \underset{(f_n)_{n=1}^N : f_n \in \mathcal{X}'_n}{\arg\min} \left(E\Big(\mathbf{y}, \boldsymbol{\nu}(\sum_{n=1}^N f_n)\Big) + \psi\left(\left\| (\|f_1\|_{\mathcal{X}'_1}, \dots, \|f_N\|_{\mathcal{X}'_N}) \right\|_{\mathcal{Z}'} \right) \right) \tag{2.43}$$

where the functions E and ψ are the same as in Theorem 2.1, while $\|\cdot\|_{\mathcal{Z}'}$ is a suitable norm that controls the coupling of the components. The idea here is to segregate the components f_n by favouring some "regularized" solutions $f_0 = (f_{0,n})_{n=1}^N$ such that the $\|f_{0,n}\|_{\mathcal{X}'_n}$ are small in an appropriate sense. Problem (2.43) is generally well defined. Its solution can be obtained as a special case of Theorem 2.4. To see this, it suffices to invoke the linearity of ν_m, which yields

$$\nu_m\Big(\sum_{n=1}^N f_n\Big) = \sum_{n=1}^N \langle \nu_m, f_n \rangle_{\mathcal{X}_n \times \mathcal{X}'_n} = \langle \widetilde{\nu}_m, f \rangle_{\mathcal{X} \times \mathcal{X}'} \tag{2.44}$$

with $f = (f_1, \dots, f_N) \in \mathcal{X}' = (\mathcal{X}'_1 \times \cdots \times \mathcal{X}'_N)_{\mathcal{Z}'}$, and $\widetilde{\nu}_m = (\nu_m, \dots, \nu_m) \in \mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$. The multicomponent optimization problem (2.43) is therefore equivalent to (2.7) with $\mathcal{X}'$ being a direct-product space and the specific choice of a "replicated" measurement operator $\widetilde{\boldsymbol{\nu}} = (\boldsymbol{\nu}, \dots, \boldsymbol{\nu})$. Consequently, we get the general form of the solution by simple substitution of $\nu_{m,n}$ by ν_m in Theorem 2.4.

2.6 Minimization of Semi-Norms

We now consider the scenario of a native Banach space $\mathcal{X}'$ that has a direct-sum decomposition as $\mathcal{X}' = \mathcal{U}' \oplus \mathcal{N}_{\mathbf{p}}$, where $\mathcal{U}'$ is the dual of some primary Banach space $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and where the complementary space $\mathcal{N}_{\mathbf{p}}$ is spanned by the finite-dimensional basis $\mathbf{p} = (p_1, \dots, p_{N_0})$. Since $\mathcal{N}_{\mathbf{p}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ is of dimension N_0 and hence also reflexive, the same holds true for its continuous dual $\mathcal{N}'_{\mathbf{p}}$. Moreover, due to the direct-sum property, there exists a unique biorthonormal set of generators $p_1^*, \dots, p_{N_0}^* \in \mathcal{X}$ such that $\mathcal{N}'_{\mathbf{p}} = \mathrm{span}\{p_n^*\}_{n=1}^{N_0} = \mathcal{N}_{\mathbf{p}^*}$ and

$$\begin{aligned}\langle p_m^*, p_n\rangle &= \delta[m-n]\\ \langle p_n^*, s\rangle &= 0\end{aligned}$$

for any $s \in \mathcal{U}'$ and $m, n \in \{1, \dots, N_0\}$. This allows us to specify the canonical projector $\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}} : \mathcal{X}' \to \mathcal{N}_{\mathbf{p}}$ as

$$\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}}\{f\} = \sum_{n=1}^{N_0} \langle p_n^*, f\rangle p_n \tag{2.45}$$

for any $f \in \mathcal{X}'$. This identification also yields the complementary projector $\mathrm{Proj}_{\mathcal{U}'} : \mathcal{X}' \to \mathcal{U}'$ as $\mathrm{Proj}_{\mathcal{U}'} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}})$. Likewise, by interchanging the role of the synthesis and analysis functionals in (2.45), we identify the canonical projector $\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}^*}} : \mathcal{X} \to \mathcal{N}_{\mathbf{p}^*}$ as

$$\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}^*}}\{\nu\} = \sum_{n=1}^{N_0} \langle p_n, \nu\rangle p_n^* \tag{2.46}$$

for any $\nu \in \mathcal{X}$. We now have the means to specify and bound the norm of any $f \in \mathcal{X}'$ as

$$\|f\|_{\mathcal{X}'} = (\|f\|_{\mathcal{U}'}, \|\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}}\{f\}\|_{\mathcal{N}_p})_{\mathcal{Z}'} \leq \|f\|_{\mathcal{U}'} + \|\mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}}\{f\}\|_{\mathcal{N}_p} \tag{2.47}$$

where the functional $f \mapsto \|f\|_{\mathcal{U}'} \stackrel{\Delta}{=} \|\mathrm{Proj}_{\mathcal{U}'} f\|_{\mathcal{U}'}$ is a semi-norm (resp., a norm) over $\mathcal{X}'$ (resp., $\mathcal{U}'$). Because the p_n are linearly independent, we also note that the standard description of $\mathcal{U}'$ as the complement of $\mathcal{N}_{\mathbf{p}}$ in $\mathcal{X}'$, given by

$$\mathcal{U}' = \{s \in \mathcal{X}' : \mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}}\{s\} = \sum_{n=1}^{N_0} \langle p_n^*, s\rangle p_n = 0\},$$

is equivalent to

$$\mathcal{U}' = \{s \in \mathcal{X}' : \mathbf{p}^*(s) = \mathbf{0}\} \tag{2.48}$$

where $\mathbf{p}^*(f) = (\langle p_1^*, f\rangle, \dots, \langle p_{N_0}^*, f\rangle)$ is a vector-valued functional $\mathcal{X}' \to \mathbb{R}^{N_0}$. Likewise, we have that

$$\mathcal{U} = \{u \in \mathcal{X} : \mathbf{p}(u) = \mathbf{0}\}, \tag{2.49}$$

which, once again, capitalizes on the biorthonormality of $(\mathbf{p}^*, \mathbf{p})$.

The consideration of such a direct-sum decomposition of a Banach space $\mathcal{X}'$ is relevant to inverse problems because it suggests that one can substitute the original regularization term $\|f\|_{\mathcal{X}'}$ by the semi-norm $\|f\|_{\mathcal{U}'}$ when one wants to favor solutions with a strong contribution in $\mathcal{N}_{\mathbf{p}}$, the null space of the semi-norm. This is a standard technique in spline theory, albeit within the classical context of RKHS spaces with $\|f\|_{\mathcal{U}'} = \|\mathrm{L}f\|_{L_2}$, where L is a suitable differential operator (e.g., a higher-order derivative or fractional Laplacian) with a null space $\mathcal{N}_{\mathbf{p}}$ that consists of polynomials of degree n [82, 83, 84]. We now show how this technique can be extended in full generality to Banach spaces. The basic requirement for this extension is that the inverse problem be well-posed over $\mathcal{N}_{\mathbf{p}}$. This is made explicit in (2.64), which is equivalent to the fourth condition in Theorem 2.5.

Theorem 2.5 (General representer theorem for Banach semi-norms)**.** *Let us consider the following setting:*

- *A dual pair of Banach spaces* $(\mathcal{X} = \mathcal{U} \oplus \mathcal{N}_{\mathbf{p}^*}, \mathcal{X}' = \mathcal{U}' \oplus \mathcal{N}_{\mathbf{p}})$, *where* $\mathcal{N}_{\mathbf{p}} = \mathcal{N}'_{\mathbf{p}^*}$ *is the vector space spanned by the finite-dimensional basis* $\mathbf{p} = (p_1, \dots, p_{N_0})$.
- *The analysis subspace* $\mathcal{N}_{\boldsymbol{\nu}} = \mathrm{span}\{\nu_m\}_{m=1}^{M} \subset \mathcal{X}$, *with the* ν_m *being linearly independent and* $M > N_0$.
- *The linear measurement operator* $\boldsymbol{\nu} : \mathcal{X}' \to \mathbb{R}^M : f \mapsto \Big(\langle \nu_1, f\rangle, \dots, \langle \nu_M, f\rangle\Big)$.
- *The vectors* $\mathbf{v}_1, \dots, \mathbf{v}_{N_0} \in \mathbb{R}^M$ *with* $[\mathbf{v}_n]_m = \langle \nu_m, p_n\rangle$ *are linearly independent; they admit a complementary set* $\{\mathbf{u}_1, \dots, \mathbf{u}_{M-N_0}\}$ *in* $\mathbb{R}^M$ *such that* $\mathbb{R}^M = \mathrm{span}\{\mathbf{v}_n\}_{n=1}^{N_0} \oplus \mathrm{span}\{\mathbf{u}_m\}_{m=1}^{M-N_0}$.
- *A proper, lower-semicontinuous, coercive, and convex loss functional* $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$.

- *Some arbitrary strictly increasing and convex function* $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$.

Then, for any fixed $\mathbf{y} \in \mathbb{R}^M$*, the solution set of the generic optimization problem*

$$S = \arg \min_{f \in \mathcal{X}'} \Big(E\Big(\mathbf{y}, \boldsymbol{\nu}(f)\Big) + \psi\left(\|f\|_{\mathcal{U}'}\right)\Big) \tag{2.50}$$

is nonempty, convex, and weak-compact.*

If E *is strictly convex, or if it imposes the equality constraint* $\mathbf{y} = \boldsymbol{\nu}(f)$*, then any solution* $f_0 \in S \subset \mathcal{X}'$ *has a unique decomposition as* $f_0 = p_0 + s_0$ *with* $p_0 \in \mathcal{N}_{\mathbf{p}}$ *and* $s_0 \in \mathcal{U}'$ *the* $(\mathcal{U}',\mathcal{U})$*-conjugate of a common* $\widetilde{\nu}_0 \in \mathcal{U}$ *whose generic form is*

$$\widetilde{\nu}_0 = \sum_{m=1}^{M-N_0} a_m \widetilde{\nu}_m \in \mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{U} \tag{2.51}$$

with suitable coefficients $\mathbf{a} = (a_m) \in \mathbb{R}^{M-N_0}$ *and reduced basis functions* $\widetilde{\nu}_m = \widetilde{\mathbf{u}}_m^T \boldsymbol{\nu} \in \mathcal{U}$*, where* $\widetilde{\mathbf{u}}_m \in \mathbb{R}^M$ *is the unique (biorthogonal) vector such that* $\widetilde{\mathbf{u}}_m^T \mathbf{v}_n = 0$ *and* $\widetilde{\mathbf{u}}_m^T \mathbf{u}_{m'} = \delta[m - m']$ *for any* $m, m' \in \{1, \dots, M - N_0\}$ *and* $n \in \{1, \dots, N_0\}$*. Depending on the Banach characteristics of* $\mathcal{U}'$*, this then results in the following explicit description of the solution(s):*

- *If* $\mathcal{U}'$ *is a Hilbert space and* ψ *is strictly convex, then the solution* f_0 *is unique and admits the linear representation*

$$f_0 = \sum_{m=1}^{M-N_0} a_m \varphi_m + \sum_{n=1}^{N_0} b_n p_n, \tag{2.52}$$

  with coefficients $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^M$ *and basis functions* $p_n \in \mathcal{N}_{\mathbf{p}}$, $\varphi_m = \mathrm{J}_{\mathcal{U}}\{\widetilde{\nu}_m\} \in \mathcal{U}'$*, where* $\mathrm{J}_{\mathcal{U}}$ *is the Riesz map* $\mathcal{U} \to \mathcal{U}'$.

- *If* $\mathcal{U}'$ *is a strictly convex Banach space and* ψ *is strictly increasing, then the solution is unique and admits the parametric representation*

$$f_0 = \mathrm{J}_{\mathcal{U}} \left\{ \sum_{m=1}^{M-N_0} a_m \widetilde{\nu}_m \right\} + \sum_{n=1}^{N_0} b_n p_n \tag{2.53}$$

  where $\mathrm{J}_{\mathcal{U}}$ *is the (nonlinear) duality operator* $\mathcal{U} \to \mathcal{U}'$.

- *Otherwise, when $\mathcal{U}'$ is not strictly convex, the solution set is the weak*-closure of the convex hull of its extremal points, which can all be expressed as*

$$f_0 = \sum_{k=1}^{K_0} c_k e_k + \sum_{n=1}^{N_0} b_n p_n \tag{2.54}$$

for some $K_0 \leq (M - N_0)$, $c_1, \dots, c_{K_0} \in \mathbb{R}$, where $e_1, \dots, e_{K_0} \in \mathcal{U}'$ are some extremal points of the unit ball $B_{\mathcal{U}'} = \{s \in \mathcal{U} : \|s\|_{\mathcal{U}'} \leq 1\}$. The vector $\mathbf{b} = (b_n) \in \mathbb{R}^{N_0}$ that characterizes the null-space component of f_0 is unique and common to all solutions whenever E is strictly convex and $\mathcal{N}_{\mathbf{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$.

Before proceeding with the proof of Theorem 2.5, we detail the way in which the reduced basis $\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \dots, \widetilde{\nu}_{M-N_0})$ in (2.51) is constructed. To that end, we first define the cross-correlation matrix $\mathbf{V} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_{N_0}] = \boldsymbol{\nu}(\mathbf{p}) \in \mathbb{R}^{M \times N_0}$ with $[\mathbf{V}]_{m,n} = \langle \nu_m, p_n \rangle$. Let us highlight the following notation that shall be used throughout this part: for any matrix $\mathbf{A} = [a_{m,n}] \in \mathbb{R}^{M \times N}$ and any vector $\boldsymbol{\nu} = (\nu_n) \in \mathcal{X}^N$, the vector $\tilde{\boldsymbol{\nu}} = (\tilde{\nu}_m) = \mathbf{A}\boldsymbol{\nu} \in \mathbb{R}^M$ is defined as

$$\widetilde{\nu}_m = \sum_{n=1}^{N} a_{m,n} \nu_n. \tag{2.55}$$

Proposition 2.2 (Direct-sum decomposition of the measurement space)**.** *Let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M) \in \mathcal{X}^M$ with $\mathcal{X} = \mathcal{U} \oplus \mathcal{N}_{\mathbf{p}^*}$ and $\mathbf{p} = (p_1, \dots, p_{N_0}) \in (\mathcal{X}')^{N_0}$ be two vectors of linear functionals such that the matrix $\mathbf{V} = \boldsymbol{\nu}(\mathbf{p}) \in \mathbb{R}^{M \times N_0}$ is of rank N_0. Then, one can always find three matrices $\mathbf{U} \in \mathbb{R}^{M \times (M-N_0)}$, $\widetilde{\mathbf{V}} \in \mathbb{R}^{M \times N_0}$, and $\widetilde{\mathbf{U}} \in \mathbb{R}^{M \times (M-N_0)}$ such that*

$$\begin{bmatrix} \widetilde{\mathbf{U}}^T \\ \widetilde{\mathbf{V}}^T \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} = \mathbf{I}_M. \tag{2.56}$$

Based on these matrices, $\boldsymbol{\nu} \in \mathcal{X}^M$ has a unique and reversible decomposition as

$$\boldsymbol{\nu} = \mathbf{U}\widetilde{\boldsymbol{\nu}} + \mathbf{V}\widetilde{\mathbf{p}}^*, \tag{2.57}$$

where

$$\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \dots, \widetilde{\nu}_{M-N_0}) = \widetilde{\mathbf{U}}^T \boldsymbol{\nu} \in \mathcal{U}^{M-N_0} \tag{2.58}$$

$$\widetilde{\mathbf{p}}^* = (\widetilde{p}_1^*, \dots, \widetilde{p}_{N_0}^*) = \widetilde{\mathbf{V}}^T \boldsymbol{\nu} \in \mathcal{X}^{N_0}. \tag{2.59}$$

In effect, this yields a decomposition of the measurement space $\mathcal{N}_{\boldsymbol{\nu}} = \mathrm{span}\{\nu_m\}_{m=1}^{M}$ *as* $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\widetilde{\mathbf{p}}^*} \oplus \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$ *with* $\mathcal{N}_{\widetilde{\boldsymbol{\nu}}} = \mathrm{span}\{\widetilde{\nu}_m\}_{m=1}^{M-N_0} \subset \mathcal{U}$. *In particular, if* $\mathcal{N}_{\mathbf{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, *then there is a unique matrix* $\widetilde{\mathbf{V}} \in \mathbb{R}^{M \times N_0}$ *of rank* N_0 *such that* $\widetilde{\mathbf{V}}^T \boldsymbol{\nu} = \mathbf{p}^*$ *and such that the decomposition still applies with* $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\mathbf{p}^*} \oplus \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$.

Proof. Since the vectors $\mathbf{v}_1, \dots, \mathbf{v}_{N_0} \in \mathbb{R}^M$ are linearly independent, they can always be completed by adding some vectors $\mathbf{v}_{N_0+1} = \mathbf{u}_1, \dots, \mathbf{v}_M = \mathbf{u}_{M-N_0}$ to form a basis of $\mathbb{R}^M$. The linear independence of the resulting family (basis property) is equivalent to the existence of a unique dual basis $\widetilde{\mathbf{v}}_1, \dots, \widetilde{\mathbf{v}}_M \in \mathbb{R}^M$ such that

$$\langle \widetilde{\mathbf{v}}_m, \mathbf{v}_n \rangle = \widetilde{\mathbf{v}}_m^T \mathbf{v}_n = \delta_{m,n} \tag{2.60}$$

for $m, n \in \{1, \dots, M\}$ (biorthonormality property). This means that any vector $\mathbf{y} \in \mathbb{R}^M$ has a unique (and reversible) decomposition as $\mathbf{y} = \sum_{m=1}^{M} \langle \widetilde{\mathbf{v}}_m, \mathbf{y} \rangle \mathbf{v}_m$. By collecting the expansion coefficients in the two vectors $\widetilde{\mathbf{b}} = (\langle \widetilde{\mathbf{v}}_1, \mathbf{y} \rangle, \dots, \langle \widetilde{\mathbf{v}}_{N_0}, \mathbf{y} \rangle)$ and $\widetilde{\mathbf{y}} = (\langle \widetilde{\mathbf{v}}_{N_0+1}, \mathbf{y} \rangle, \dots, \langle \widetilde{\mathbf{v}}_M, \mathbf{y} \rangle)$ and identifying the matrices $\widetilde{\mathbf{V}} = [\widetilde{\mathbf{v}}_1 \ \cdots \ \widetilde{\mathbf{v}}_{N_0}]$, $\mathbf{U} = [\mathbf{v}_{N_0+1} \ \cdots \ \mathbf{v}_M]$, and $\widetilde{\mathbf{U}} = [\widetilde{\mathbf{v}}_{N_0+1} \ \cdots \ \widetilde{\mathbf{v}}_M]$, we then observe that the decomposability of $\mathbf{y} \in \mathbb{R}^M$ is equivalent to

$$\mathbf{y} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{y}} \\ \widetilde{\mathbf{b}} \end{bmatrix} \tag{2.61}$$

with

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{U}}^T \mathbf{y} \in \mathbb{R}^{M-N_0}, \quad \widetilde{\mathbf{b}} = \widetilde{\mathbf{V}}^T \mathbf{y} \in \mathbb{R}^{N_0}. \tag{2.62}$$

Likewise, the enabling biorthonormality property (2.60) is equivalent to the invertibility condition (2.56).

By substituting $\mathbf{y} \in \mathbb{R}^M$, $\widetilde{\mathbf{y}} \in \mathbb{R}^{M-N_0}$, and $\widetilde{\mathbf{b}} \in \mathbb{R}^{N_0}$ by $\boldsymbol{\nu}(f), \widetilde{\boldsymbol{\nu}}(f)$, and $\widetilde{\mathbf{p}}^*(f)$, respectively, we then rephrase (2.61) and (2.62) in term of functionals, which yields the reversible decomposition described by (2.57), (2.58), and (2.59). To prove that $\mathcal{N}_{\widetilde{\boldsymbol{\nu}}} \subset \mathcal{U} = \{u \in \mathcal{X} : \mathbf{p}(u) = 0\}$, we invoke the invertibility condition (2.56), which yields $\mathbf{p}(\widetilde{\boldsymbol{\nu}}) = \Big(\widetilde{\boldsymbol{\nu}}(\mathbf{p})\Big)^T = (\widetilde{\mathbf{U}}^T \mathbf{V})^T = \mathbf{0}^T$. Since, for any $\mathbf{a} \in \mathbb{R}^M$, we have that $\mathbf{a}^T \mathbf{U} \widetilde{\boldsymbol{\nu}} \in \mathrm{span}\{\widetilde{\nu}_n\}_{n=1}^{M-N_0}$ and $\mathbf{a}^T \mathbf{V} \widetilde{\mathbf{p}}^* \in \mathrm{span}\{\widetilde{p}_n^*\}_{n=1}^{N_0}$ with the linear expansion of

$\mathbf{a}^T\boldsymbol{\nu} \in \mathcal{N}_{\boldsymbol{\nu}}$ in the corresponding basis being reversible, we can interpret (2.57) as the direct-sum decomposition $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\widetilde{\boldsymbol{\nu}}} \oplus \mathcal{N}_{\widetilde{\mathbf{p}}^*}$.

The inclusion $\mathcal{N}_{\mathbf{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, together with the linear independence of the ν_m, is equivalent to the existence of a unique transformation matrix $\widetilde{\mathbf{V}}^T$ of rank N_0 such that $\mathbf{p}^* = \widetilde{\mathbf{V}}^T\boldsymbol{\nu}$. While this sets the matrix $\widetilde{\mathbf{V}} \in \mathbb{R}^{M\times N_0}$, one is still left with sufficiently many degrees of freedom to select the complementary matrices $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ such that (2.56) holds. □

We note that, irrespective of whether we fix $\widetilde{\mathbf{V}}$ (second part of Proposition 2.2) or not, there are generally infinitely many admissible choices for $\mathbf{U}$ in (2.56) and, hence, for the construction of the reduced basis $\widetilde{\boldsymbol{\nu}}$ defined by (2.58). This does not contradict the unicity of (2.51). Indeed, different choices of extension correspond to different biorthogonal bases $\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \dots, \widetilde{\nu}_{M-N_0})$ of the same subspace.

Proof of Theorem 2.5.

(i) Existence: The classical conditions that ensure the existence of a minimizer of the functional $J(f) = E\Big(\mathbf{y}, \boldsymbol{\nu}(f)\Big) + \psi\left(\|f\|_{\mathcal{U}'}\right)$ are that $J(f)$ should be proper, convex, (weak$*$-)lower-semi-continuous, and coercive over $\mathcal{X}'$. These higher-level properties also imply that the solution set S is convex, and weak*-compact.

The first three conditions follow from the listed assumptions and the general properties of a (semi-)norm—see argumentation in the proof of Theorem 2.1 in [103]. To establish coercivity, we recall that the hypothesis $\nu_m \in \mathcal{X}$ implies the continuity of $\nu_m : \mathcal{X}' \to \mathbb{R}$ due to the continuous embedding of $\mathcal{X}$ in its bidual $\mathcal{X}'' = (\mathcal{X}')'$. Consequently, there exists some constant $A > 0$ such that

$$\|\boldsymbol{\nu}(f)\|_2 \leq A\|f\|_{\mathcal{X}'} \tag{2.63}$$

for all $f \in \mathcal{X}'$. Likewise, the linear independence of the $\mathbf{v}_n$ and the property that all finite-dimensional norms are equivalent implies the existence of $B > 0$ such that, for any $p \in \mathcal{N}_{\mathbf{p}}$,

$$B\|p\|_{\mathcal{N}_{\mathbf{p}}} \leq \|\boldsymbol{\nu}(p)\|_2. \tag{2.64}$$

By using the direct-sum decomposition $f = s + p$ with $(s,p) \in \mathcal{U}' \times \mathcal{N}_{\mathbf{p}}$, $\|f\|_{\mathcal{X}'} \leq \|s\|_{\mathcal{U}'} + \|p\|_{\mathcal{N}_{\mathbf{p}}}$, and $\|s\|_{\mathcal{U}'} = \|f\|_{\mathcal{U}'}$, we readily deduce that

$$\|\boldsymbol{\nu}(f)\|_2 \geq \|\boldsymbol{\nu}(p)\|_2 - \|\boldsymbol{\nu}(s)\|_2 \geq B\|p\|_{\mathcal{N}_{\mathbf{p}}} - A\|s\|_{\mathcal{U}'} \geq B\|f\|_{\mathcal{X}'} - (A+B)\|f\|_{\mathcal{U}'}, \tag{2.65}$$

where we have made use of the triangle inequality and the two previous bounds. Let us now consider some sequence (f_m) in $\mathcal{X}'$ with $f_m = (s_m, q_m) \in \mathcal{U}' \times \mathcal{N}_{\mathbf{p}}$ such that $\|f_m\|_{\mathcal{X}'} \geq \|f_n\|_{\mathcal{X}'}$ for $m \geq n$ and $\lim_{m\to\infty} \|f_m\|_{\mathcal{X}'} = \infty$. Then, there are two possible asymptotic behaviors for the norm of $s_m = \mathrm{Proj}_{\mathcal{U}'} f_m$:

1. The quantity $\|s_m\|_{\mathcal{X}'} = \|f_m\|_{\mathcal{U}'} \to \infty$ as $m \to \infty$, in which case $J(f_m) \to \infty$ due to the unboundedness of $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ at infinity.
2. There exists a constant C such that $\|f_m\|_{\mathcal{U}'} \leq C$ for all m. By invoking (2.65), we get that $\|\boldsymbol{\nu}(f_m)\|_2 \to \infty$ as $m \to \infty$, which, in turn, gives $J(f_m) \to \infty$, due to the coercivity of $E(\cdot, \mathbf{y})$.

In summary, $J(f) \to \infty$ as $\|f\|_{\mathcal{X}'} \to \infty$, which is the required coercivity property.

(ii) Representation of a solution: The underlying direct-sum property implies that any $f_0 \in S \subset \mathcal{X}'$ has a unique decomposition as $f_0 = s_0 + p_0$ with $(s_0, p_0) \in \mathcal{U}' \times \mathcal{N}_{\mathbf{p}}$. To derive the parametric form of a solution, we momentarily assume that p_0 (and, hence, $\boldsymbol{\nu}(p_0) \in \mathbb{R}^M$) and $\mathbf{y}_0 = \boldsymbol{\nu}(f_0) \in \mathbb{R}^M$ are known. By making use of the decomposition of the measurement space in Proposition 2.2, we observe that the penalized component $s_0 \in \mathcal{U}'$ solves the equivalent constrained-optimization problem

$$s_0 \in S_{p_0,\mathbf{y}_0} = \arg\min_{s\in\mathcal{U}'} \|s\|_{\mathcal{U}'} \text{ s.t. } \mathbf{y}_0 - \boldsymbol{\nu}(p_0) = \boldsymbol{\nu}(s) = \mathbf{U}\widetilde{\boldsymbol{\nu}}(s) + \mathbf{V}\widetilde{\mathbf{p}}^*(s), \tag{2.66}$$

where $\widetilde{\boldsymbol{\nu}} = \widetilde{\mathbf{U}}^T\boldsymbol{\nu} \in \mathcal{U}^{M-N_0}$ and $\widetilde{\mathbf{p}}^* = \widetilde{\mathbf{V}}^T\boldsymbol{\nu} \in \mathcal{X}^{N_0}$. We now show that the effective number of linear constraints in (2.66) is actually $(M - N_0)$ and not M, as may be thought on first inspection. To that end, we multiply the linear constraint by $\widetilde{\mathbf{U}}^T \in \mathbb{R}^{(M-N_0)\times M}$ on both sides and use the properties that $\widetilde{\mathbf{U}}^T\mathbf{U} = \mathbf{I}_{M-N_0}$ and $\widetilde{\mathbf{U}}^T\mathbf{V} = \mathbf{0}$ (see (2.56) in Proposition 2.2). This yields

$$s_0 \in S_{p_0,\mathbf{y}_0} = \arg\min_{s\in\mathcal{U}'} \|s\|_{\mathcal{U}'} \text{ s.t. } \widetilde{\boldsymbol{\nu}}(s) = \widetilde{\mathbf{y}}_0, \tag{2.67}$$

where $\widetilde{\mathbf{y}}_0 = \widetilde{\mathbf{U}}^T\Big(\mathbf{y}_0 - \boldsymbol{\nu}(p_0)\Big) = \widetilde{\mathbf{U}}^T\mathbf{y}_0 \in \mathbb{R}^{M-N_0}$. This latter simplification occurs because $\widetilde{\mathbf{U}}^T\boldsymbol{\nu}(p) = \widetilde{\boldsymbol{\nu}}(p) = 0$ for all $p \in \mathcal{N}_{\mathbf{p}}$ by construction. The theoretical significance of the cancellation of $\widetilde{\mathbf{U}}^T\boldsymbol{\nu}(p_0)$ is that the above manipulation does not depend on p_0, so that $S_{p_0,\mathbf{y}_0} = S_{\mathbf{y}_0}$. In effect, this means that the characterization of the optimal s_0 in (2.67) holds for all solutions that share the same measurements $\mathbf{y}_0$. This is true, in particular, when E is strictly convex, by a standard argument in convex analysis. The description of $s_0 \in \mathcal{U}'$ as the $(\mathcal{U}',\mathcal{U})$-conjugate of a common $\widetilde{\nu}_0 \in \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$, as well as (2.52), (2.53), and (2.54), then follow from Theorem 2.1.

For the special scenario $\mathcal{N}_{\mathbf{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, we select $\widetilde{\mathbf{V}}$ such that $\mathbf{p}^* = \widetilde{\mathbf{V}}^T\boldsymbol{\nu}$ (see the second part of Proposition 2.2) and are then able to obtain the expansion of coefficients of $p_0 = \mathrm{Proj}_{\mathcal{N}_{\mathbf{p}}}\{f_0\}$ directly from the measurements as $\mathbf{b} = \mathbf{p}^*(f_0) = \widetilde{\mathbf{V}}^T\mathbf{y}_0$, which establishes this part of the solution as well for all $f_0 \in S$. □

When the (semi)-norm $\|\cdot\|_{\mathcal{U}'}$ is strictly convex, Theorem 2.5 states that the unique solution f_0 lives in a finite-dimensional manifold that is parameterized by $\mathbf{b} \in \mathbb{R}^{N_0}$ (for the null-space component p_0) and $\mathbf{a} \in \mathbb{R}^{M-N_0}$ (for the preimage $\widetilde{\nu}_0$ of the penalized component $s_0 = (\widetilde{\nu}_0)^*$). While the two primary expansions are linear, the high-level ingredient of the representation is the duality mapping, which introduces a nonlinearity in the non-Hilbert scenario. At any rate, the main point is that the intrinsic dimensionality of the solution space is still M, as in the case of Theorem 2.1, except that the repartition is now very different, with the contribution of the null-space component $p_0 \in \mathcal{N}_{\mathbf{p}}$ being maximized since it is no longer penalized.

An important outcome of Theorem 2.5 is that it becomes possible to characterize the full solution set S via the specification of a single pair $(p_0, \widetilde{\nu}_0) \in \mathcal{N}_{\mathbf{p}} \times \mathcal{U}$. For the challenging cases where there are multiple solutions (third scenario), this requires the additional assumption that $\mathcal{N}_{\mathbf{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, which has the desirable effect of decoupling the determination p_0 from that of $\widetilde{\nu}_0$. It turns out that this decoupling is applicable to most practical problems that involve a semi-norm regularization. The key is that it is generally possible to adapt the semi-norm topology to the problem at hand by selecting a biorthonormal system $(\mathbf{p}^*, \mathbf{p})$ with $p_1^*, \dots, p_{N_0}^* \in \mathrm{span}\{\nu_m\}_{m=1}^M \subset \mathcal{X}$ (see, for instance, [73, 104]).

2.7 Summary

In this chapter, we presented a refinement of Theorem 2.1 for the cases where $\mathcal{X}'$ admits the decomposition $\mathcal{X}' = \mathcal{X}_1' \times \cdots \times \mathcal{X}_N'$ (direct product of Banach spaces) or $\mathcal{X}' = \mathcal{X}_1' \oplus \cdots \oplus \mathcal{X}_N'$ (direct sum of Banach spaces). Our main result is Theorem 2.4, which explicitly tells us how the underlying direct-product duality mappings and extremal points can be determined from the knowledge of the same entities for the simpler constituent spaces $\mathcal{X}_n'$. We then extend Theorem 2.1 by replacing the original regularizing norm by a semi-norm that has a finite-dimensional null space $\mathcal{N}_{\mathbf{p}} = \mathrm{span}\{p_1, \dots, p_{N_0}\} \subset \mathcal{X}'$. The main result expressed by Theorem 2.5 is that this adds a null-space component p_0 to the generic solution(s) of Theorem 2.1, which is the desired outcome. At the same time, it reduces the intrinsic dimension of the complementary component $s_0 = f_0 - p_0$ from M to $(M - N_0)$. The mathematical analysis amounts to making sure that the solution exists and to then properly split the problem in order to decouple the determination of the two solution components. The significance of our new Theorem 2.5 is to show that the traditional techniques of spline approximation [82, 83, 8, 84], which involve semi-reproducing-kernel Hilbert spaces [85], are extendable to Banach spaces in general. Likewise, the non-strictly convex scenario in Theorem 2.5 is consistent with a number of recent results that have appeared in the literature for sparsity-promoting functionals [73, 76, 72], although the overlap is only partial due to the generality of our present formulation.

Chapter 3

Supervised Learning with Sparsity Prior

In this chapter[1], we employ our theoretical findings in Chapter 2 to develop novel supervised learning schemes in the nonparametric setting. The common thread throughout all of our proposed methods is the existence of the concept of *sparsity*, which facilitates learning simpler and, hence, more interpretable models.

We begin with a brief introduction to supervised learning and common methods used in this domain (Section 3.1). We then present a novel kernel-based regression scheme that proposes a sparse and adaptive kernel expansion (Section 3.2). After that, we focus on the problem of learning univariate models under joint sparsity and Lipschitz constraint (Section 3.3). This paves the way to develop a functional framework for learning activation functions of deep neural networks (Section 3.4). Last but not least, we introduce a novel seminorm—the Hessian-Schatten total variation—and propose its use as a regularization functional for learning continuous and piecewise linear models (Section 3.5).

[1] This chapter is based on our published [105, 106, 107, 108, 109, 110] and submitted [111] works.

3.1 Overview on Supervised Learning

The determination of an unknown function from a series of samples is a classical problem in machine learning. It falls under the category of "supervised learning," which has been abundantly described in the literature (see [112, 113, 8] for classical textbooks, as well as [114, 115, 116, 117] for more recent ones). The goal of supervised learning is to recover a target function $f : \mathbb{R}^d \to \mathbb{R}$ from its M noisy samples $y_m = f(\boldsymbol{x}_m) + \epsilon_m$, $m = 1, 2, \ldots, M$. The disturbance terms ϵ_m are typically assumed to be i.i.d. samples of a zero-mean probability law (*e.g.*, additive Gaussian noise) while the input vectors $\boldsymbol{x}_m$ are either assumed to be in the random or fixed design [116, Section 1.9].

A common way of carrying out this task is to solve a minimization problem of the form

$$\min_{f \in \mathcal{F}(\mathbb{R}^d)} \left(\sum_{m=1}^{M} E\left(f(\mathbf{x}_m), y_m\right) + \mathcal{R}(f) \right), \tag{3.1}$$

where $\mathcal{F}(\mathbb{R}^d)$ is the underlying search space, the convex loss function $E : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ enforces the consistency of the learned mapping with the given data points, and the regularization functional $\mathcal{R} : \mathcal{F}(\mathbb{R}^d) \to \mathbb{R}_{\geq 0}$ injects prior knowledge on the form of the mapping f, which is designed to alleviate the problem of overfitting.

3.1.1 Nonparametric Regression

In some cases, the optimization can be performed over an infinite-dimensional function space. A prominent example is the family of reproducing-kernel Hilbert spaces (RKHS). The Hilbert space $\mathcal{H}(\mathbb{R}^d)$ consisting of functions from $\mathbb{R}^d$ to $\mathbb{R}$ is called an RKHS if there exists a bivariate symmetric and positive-definite function $\mathrm{k} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that, for all $\boldsymbol{x} \in \mathbb{R}^d$, $\mathrm{k}(\boldsymbol{x}, \cdot) \in \mathcal{H}(\mathbb{R}^d)$ and $f(\boldsymbol{x}) = \langle \mathrm{k}(\boldsymbol{x}, \cdot), f(\cdot) \rangle_{\mathcal{H}}$ [77]. The function $\mathrm{k}(\cdot, \cdot)$ is unique and is called the reproducing kernel of $\mathcal{H}(\mathbb{R}^d)$.

Supervised learning over the RKHS $\mathcal{H}(\mathbb{R}^d)$ can be formulated through the minimization

$$\min_{f\in\mathcal{H}(\mathbb{R}^d)}\left(\sum_{m=1}^{M} E(f(\boldsymbol{x}_m), y_m) + \lambda\|f\|_{\mathcal{H}}^2\right). \tag{3.2}$$

The kernel representer theorem states that the solution of (3.2) admits the form

$$f(\cdot) = \sum_{m=1}^{M} a_m \mathrm{k}(\cdot, \boldsymbol{x}_m) \tag{3.3}$$

for some appropriate weights $a_m \in \mathbb{R}$, where $m = 1, 2, \ldots, M$ [118, 61]. The expansion (3.3) is the key element of kernel-based schemes in machine learning [54, 119, 120] and, in particular, support-vector machines (SVM) [121, 117]. Moreover, optimal rates have been derived for learning using the expansion (3.3) in several setups [122, 123, 124], particularly for Gaussian kernels [125]. A central element in these analyses is that the regularization functional $\mathcal{R}(\cdot)$ (in this case, the underlying Hilbertian norm) directly controls the complexity of the learned mapping [126, Section 2.4].

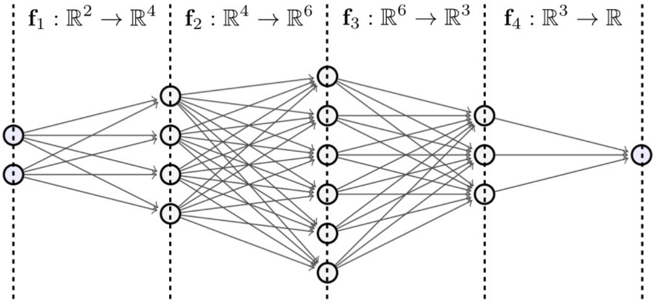
Computing the RKHS norm of a function f of the form (3.3) results in $\|f\|_{\mathcal{H}}^2 = \boldsymbol{a}^T\mathbf{G}\boldsymbol{a}$, where $\mathbf{G} \in \mathbb{R}^{M\times M}$ is a symmetric and positive-definite matrix with $[\mathbf{G}]_{m,n} = \mathrm{k}(\boldsymbol{x}_m, \boldsymbol{x}_n)$. It is called the Gram matrix of the kernel $\mathrm{k}(\cdot,\cdot)$. The practical outcome of this observation is that the infinite-dimensional problem (3.2) over the space of functions $\mathcal{H}(\mathbb{R}^d)$ becomes equivalent to the finite-dimensional problem [61]

$$\min_{\boldsymbol{a}\in\mathbb{R}^M}\left(\sum_{m=1}^{M} E([\mathbf{G}\boldsymbol{a}]_m, y_m) + \lambda\boldsymbol{a}^T\mathbf{G}\boldsymbol{a}\right), \tag{3.4}$$

which is of size M and can be computed numerically.

3.1.2 Parametric Regression

In cases when (3.1) cannot be recast as a finite-dimensional optimization problem, another common approach is to restrict the search space $\mathcal{F}$ to a given family of parametric functions $f_{\boldsymbol{\Theta}}$, where $\boldsymbol{\Theta}$ denotes the vector of the underlying parameters.


$\mathbf{f}_1 : \mathbb{R}^2 \to \mathbb{R}^4$
$\mathbf{f}_2 : \mathbb{R}^4 \to \mathbb{R}^6$
$\mathbf{f}_3 : \mathbb{R}^6 \to \mathbb{R}^3$
$\mathbf{f}_4 : \mathbb{R}^3 \to \mathbb{R}$

$$\mathbf{f}_{\text{deep}} = \mathbf{f}_4 \circ \mathbf{f}_3 \circ \mathbf{f}_2 \circ \mathbf{f}_1 : \mathbb{R}^2 \to \mathbb{R}$$

Figure 3.1: Schematic view of a neural network with the layer descriptor $(2, 4, 6, 3, 1)$. Each layer consists of linear weights (arrows) and point-wise nonlinearities (circles).

A celebrated example of this approach is deep learning, which has become state-of-the-art for image classification [127], inverse problems [128], and image segmentation [129]. The underlying principle is the construction of an overall map $f_{\boldsymbol{\Theta}} : \mathbb{R}^d \to \mathbb{R}$ built as a neural network via the composition of parameterized affine mappings and pointwise nonlinearities known as activation functions. The attribute “deep” refers to the high number of such module compositions (layers), which is instrumental to improve the approximation power of the network [130, 131, 132] and its generalization ability [133].

More precisely, an L-layer fully connected feed forward neural network $\mathbf{f}_{\text{deep}} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ with the layer descriptor $(N_0, N_1, \dots, N_L)$ is the composition of the vector-valued functions $\mathbf{f}_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ for $l = 1, \dots, L$ as

$$\mathbf{f}_{\text{deep}} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L} : \mathbf{x} \mapsto \mathbf{f}_L \circ \cdots \circ \mathbf{f}_1(\mathbf{x}). \tag{3.5}$$

Each vector-valued function $\mathbf{f}_l$ is a layer of the neural network $\mathbf{f}_{\text{deep}}$ and consists of two elementary operations: linear transformations and point-wise nonlinearities. In other words, for the lth layer, there exists weight vectors $\mathbf{w}_{n,l} \in \mathbb{R}^{N_{l-1}}$ and nonlinear

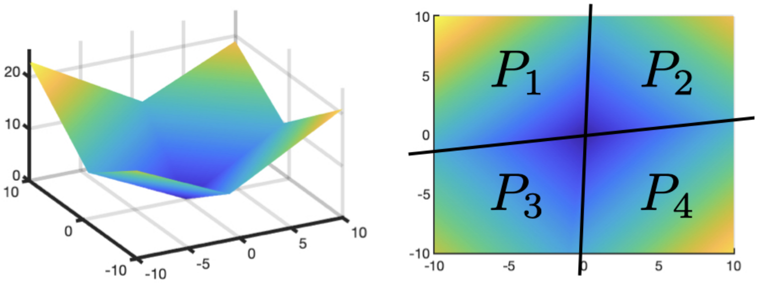


Figure 3.2: Illustration of a CPWL function $f : \mathbb{R}^2 \to \mathbb{R}$. Left: 3D view. Right: 2D partitioning.

activation functions $\sigma_{n,l} : \mathbb{R} \to \mathbb{R}$ for $n = 1, 2, \dots, N_l$ such that

$$\mathbf{f}_l(\mathbf{x}) = \left(\sigma_{1,l}(\mathbf{w}_{1,l}^T\mathbf{x}), \sigma_{2,l}(\mathbf{w}_{2,l}^T\mathbf{x}), \dots, \sigma_{N_l,l}(\mathbf{w}_{N_l,l}^T\mathbf{x})\right). \tag{3.6}$$

One can also consider an alternative representation of the lth layer by defining the matrix $\mathbf{W}_l = \begin{bmatrix}\mathbf{w}_{1,l} & \mathbf{w}_{2,l} & \cdots & \mathbf{w}_{N_l,l}\end{bmatrix}^T$ and the vector-valued nonlinear function $\boldsymbol{\sigma}_l : \mathbb{R}^{N_l} \to \mathbb{R}^{N_l}$ as the mapping

$$(x_1, \dots, x_{N_l}) \mapsto (\sigma_{1,l}(x_1), \sigma_{2,l}(x_2), \dots, \sigma_{N_l,l}(x_{N_l})). \tag{3.7}$$

With this notation, the lth layer has simply the form $\mathbf{f}_l = \boldsymbol{\sigma}_l \circ \mathbf{W}_l$.

The classical choice for the activation function is the sigmoid function due to its biological interpretation and universal approximation property [134]. However, neural networks with sigmoidal activation functions suffer from vanishing gradients which essentially makes training difficult and slow. This stems from the fact that the sigmoid function is bounded and horizontally asymptotic at large positive and negative values. Currently, the popular activation functions are rectified linear unit $\text{ReLU}(x) = \max(x, 0)$ [20] and its variants such as LeakyReLU, defined as $\text{LReLU}(x) = \max(x, ax)$ for some $a \in (0, 1)$ [135]. ReLU-based activation functions have a wide range, which prevents the network from having vanishing gradients.

Neural networks with ReLU activation functions have been considered thoroughly in the literature [136]. In particular, they are intimately connected to the family of continuous and piecewise-linear (CPWL) functions. A function $f : \Omega \to \mathbb{R}$ is said to be continuous and piecewise-linear if

1. It is continuous.
2. There exists a finite partitioning $\Omega = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_N$ such that, for any $n = 1, \ldots, N$, P_n is a convex polytope with the property that the restricted function $f\Big|_{P_n}$ is an affine mapping of the form $f\Big|_{P_n}(\boldsymbol{x}) = \boldsymbol{a}_n^T \boldsymbol{x} + b_n$ for all $\boldsymbol{x} \in P_n$.

An example of a CPWL function is shown in Figure 3.2. The vector-valued extension is quite straightforward; the mapping $\mathbf{f} = (f_i) : \Omega \to \mathbb{R}^d$ is CPWL if, $f_i : \Omega \to \mathbb{R}$ is CPWL for $i = 1, \ldots, d$.

Following the above definition, it can be shown that the CPWL structure is preserved through addition, scalar multiplication, and composition. Moreover, the univariate CPWL functions are indeed linear splines which includes the ReLU activation function. Putting these together, one readily observes that the input-output mapping of any feed-forward neural network with ReLU activation functions is CPWL [130]. Interestingly, the converse of the latter also holds: any CPWL function can be represented *exactly* by a deep ReLU neural network [137]. This establishes a direct link with spline theory, as first highlighted by Poggio *et al.* [138] and then further explored in various works [103, 139, 140, 141, 142].

3.2 Multi-Kernel Regression

In this section[2], we provide a Banach-space formulation of supervised learning with generalized total-variation (gTV) regularization. We identify the class of kernel functions that are admissible in this framework. Then, we propose a variation of supervised learning in a continuous-domain hybrid search space with gTV regularization. We show that the solution admits a multi-kernel expansion with adaptive positions. In this representation, the number of active kernels is upper-bounded by the number of data points while the gTV regularization imposes an ℓ_1 penalty on the kernel coefficients. Finally, we illustrate numerically the outcome of our theory.

3.2.1 Context

Toward Sparse Kernel Expansions

In the solution form (3.3), the kernels are shifted to the location of the data samples. This is elegant but can become cumbersome when the number of samples M grows large. Several schemes have been developed to reduce the number of active kernels. One proposed approach is to use a sparsity-enforcing loss such as the ϵ-insensitive norm of SVM regression [143, 144, 145]. Another approach is to replace the quadratic regularization $\boldsymbol{a}^T\mathbf{G}\boldsymbol{a}$ in the reduced finite-dimensional problem (3.4) by a sparsity-promoting penalty such as $\|\boldsymbol{a}\|_1 = \sum_{m=1}^{M} |a_m|$. This results in (3.4) becoming

$$\min_{\boldsymbol{a}\in\mathbb{R}^M} \left(\sum_{m=1}^{M} E([\mathbf{G}\boldsymbol{a}]_m, y_m) + \lambda\|\boldsymbol{a}\|_1 \right), \tag{3.8}$$

which is called the generalized LASSO [146]. The properties of this estimator have been studied both from a statistical [147] and approximation-theoretical point of view [148]

In this work, we consider a Banach-space formulation of supervised learning. We choose the generalized total-variation (gTV) norm as the regularization term in order

[2] This section is based on our published work [105].

to promote sparsity in the continuous domain. The effect of gTV regularization has been extensively studied in the context of linear inverse problems (see, Chapter 4). For an invertible operator L (see Definition 3.1), the gTV norm is defined as

$$\mathrm{gTV}(f) = \|\mathrm{L}\{f\}\|_{\mathcal{M}}, \tag{3.9}$$

where $\mathcal{M}(\mathbb{R}^d)$ is the space of bounded Radon measures (see (3.19) for a precise definition) and $\|\cdot\|_{\mathcal{M}}$ is the total-variation norm in the sense of measures [149].

One can formulate supervised learning with gTV regularization through the minimization

$$\min_{f \in \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)} \left(\sum_{m=1}^{M} E(f(\boldsymbol{x}_m), y_m) + \lambda \|\mathrm{L}\{f\}\|_{\mathcal{M}} \right), \tag{3.10}$$

where $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is the native Banach space of the operator $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ equipped with the gTV norm (see Definition 3.2). The fact that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$ is a Banach space (*i.e.*, a complete normed space) follows from the invertibility of L (see Theorem 3.2). A consequence of the general representer theorem of [73] is that there is always a solution of (3.10) that admits a linear kernel expansion of the form

$$f(\cdot) = \sum_{l=1}^{M_0} a_l \mathrm{k}(\cdot, \boldsymbol{z}_l), \tag{3.11}$$

for some unknown integer $M_0 \leq M$, non-zero kernel weights $a_l \in \mathbb{R}$, and some distinct adaptive kernel positions $\boldsymbol{z}_l \in \mathbb{R}^d$ [74]. There, the function $\mathrm{k}(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the shift-invariant kernel associated to the Green's function of the operator L. In other words, we have that $\mathrm{k}(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y})$, where $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta\}$.

There exist works on supervised learning over Banach spaces, especially via the concept of reproducing-kernel Banach spaces (RKBS) [150, 151, 152]. However, there are several differences between RKBS and our proposed scheme of learning with gTV regularization. Firstly, as highlighted in [152], the RKBS representer theorem yields a nonlinear kernel expansion for the optimal solution. Secondly, its kernel positions necessarily coincide with the data points. Last but not least, the Banach spaces in the RKBS theory are restricted to reflexive one, which excludes the case of learning with gTV regularization that is known to enforce sparsity in the continuous domain.

Let us also mention that a formulation with strong link to (3.10) has been presented in [153] for learning a function $f : \mathbb{R}^d \to \mathbb{R}$ from a continuously indexed family of atoms $\{\mathrm{k}_{\boldsymbol{z}}\}_{\boldsymbol{z}\in\mathcal{V}}$, where $\mathcal{V}$ is a compact topological space. Putting it in a similar form as (3.10), the proposed formulation in [153] for supervised learning is equivalent to the minimization

$$\min_{\mu\in\mathcal{M}(\mathcal{V})} \left(\sum_{m=1}^{M} E\left(\int_{\mathcal{V}} \mathrm{k}_{\boldsymbol{z}}(\boldsymbol{x}_m)\mathrm{d}\mu(\boldsymbol{z}), y_m \right) + \lambda\|\mu\|_{\mathcal{M}} \right), \tag{3.12}$$

where $\mathcal{M}(\mathcal{V})$ is the space of Radon measures over $\mathcal{V}$. The relevant property there is that the minimization of (3.12) introduces an atomic measure $\mu = \sum_{l=1}^{M_0} a_l \delta(\cdot - \boldsymbol{z}_l)$. It hence suggests the parametric form (3.11) with $\mathrm{k}(\cdot, \boldsymbol{z}_l) = \mathrm{k}_{\boldsymbol{z}_l}(\cdot)$ for the learned function.

The minimization problem (3.12) is a synthetic-based formulation for supervised learning where the basis functions are known *a priori*, in contrary to (3.10) which is an analysis-based formalism that relies on regularization theory in Banach spaces. Interestingly, the two formulations are equivalent when the family of atoms in (3.12) coincides with the class of shifted Green's function of the regularization operator L; that is, $\mathrm{k}_{\boldsymbol{z}}(\cdot) = \rho_{\mathrm{L}}(\cdot - \boldsymbol{z})$.

Last but not least, we discuss the connection between (3.10) and generalized LASSO. One readily verifies that the gTV norm enforces an ℓ_1 penalty on the kernel coefficients a_l. More precisely, the expansion (3.11) translates the original problem (3.10) into the discrete minimization

$$\min_{\boldsymbol{a}\in\mathbb{R}^M, \mathrm{Z}\in\mathbb{R}^{d\times M_0}} \left(\sum_{m=1}^{M} E([\mathbf{G}_{\mathrm{Z}}\boldsymbol{a}]_m, y_m) + \lambda\|\boldsymbol{a}\|_{\ell_1} \right), \tag{3.13}$$

where $\mathrm{Z} = \big(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{M_0}\big)$ is the kernel-position matrix and $\mathbf{G}_{\mathrm{Z}} \in \mathbb{R}^{M\times M_0}$ is a matrix with $[\mathbf{G}_{\mathrm{Z}}]_{m,l} = \mathrm{k}(\boldsymbol{x}_m, \boldsymbol{z}_l)$. The reduced problem (3.13) can be seen as an extended version of the generalized LASSO in (3.8). The fundamental difference is that the minimization is through the positions as well.

Multikernel Schemes

The solution forms (3.3) and (3.11) heavily depend on the kernel function $\mathrm{k}(\cdot,\cdot)$. Hence, choosing the proper kernel is a challenging task that requires careful consideration. One can use a cross-validation scheme in order to compare the performance of several kernel estimators and select the best one for the desired application [80]. Another approach is to learn a new kernel function $\mathrm{k}_{\boldsymbol{\mu}} = \sum_{n=1}^{N} \mu_n \mathrm{k}_n$ from a family of given kernels $\mathrm{k}_1, \mathrm{k}_2, \dots, \mathrm{k}_N$ [154, 155, 156, 78]. This transforms the original problem (3.4) into the joint optimization

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^N, \boldsymbol{a}\in\mathbb{R}^M} \left(\sum_{m=1}^{M} E([\mathbf{G}_{\boldsymbol{\mu}}\boldsymbol{a}]_m, y_m) + \lambda \boldsymbol{a}^T \mathbf{G}_{\boldsymbol{\mu}} \boldsymbol{a} + \mathrm{R}(\boldsymbol{\mu}) \right), \tag{3.14}$$

where $\mathbf{G}_{\boldsymbol{\mu}}$ is the Gram matrix of the learned kernel $\mathrm{k}_{\boldsymbol{\mu}}$ and $\mathrm{R}(\cdot)$ regularizes the coefficient vector $\boldsymbol{\mu}$, for example like in $\mathrm{R}(\boldsymbol{\mu}) = \|\boldsymbol{\mu}\|_{\ell_p} = \left(\sum_{n=1}^{N} |\mu_n|^p\right)^{\frac{1}{p}}$ for $1 \leq p \leq 2$ [157, 158, 159, 160, 161]. The learned function will then take the generic form

$$f(\cdot) = \sum_{n=1}^{N} \sum_{m=1}^{M} \mu_n a_m \mathrm{k}(\cdot, \boldsymbol{x}_m). \tag{3.15}$$

In this work, we propose a multi-kernel extension of supervised learning with gTV regularization. To that end, we consider the minimization

$$\min_{\substack{f_n \in \mathcal{M}_{\mathrm{L}_n}(\mathbb{R}^d),\\ f=\sum_{n=1}^{N} f_n}} \left(\sum_{m=1}^{M} E(f(\boldsymbol{x}_m), y_m) + \lambda \sum_{n=1}^{N} \|\mathrm{L}_n\{f_n\}\|_{\mathcal{M}} \right). \tag{3.16}$$

In this formulation, the target function f is decomposed into N additive components, where the smoothness of each component has been expressed by its corresponding regularization operator. Our main result, which follows from Theorem 3.4, is the existence of a solution of (3.16) that yields a multi-kernel expansion of the target function and that takes the form

$$f(\cdot) = \sum_{n=1}^{N} \sum_{l=1}^{M_n} a_{n,l} \mathrm{k}_n(\cdot, \boldsymbol{z}_{n,l}), \quad \|\boldsymbol{a}\|_{\ell_0} \leq M, \tag{3.17}$$

where $\|\boldsymbol{a}\|_{\ell_0}$ is called the ℓ_0 norm of $\boldsymbol{a}$ and is equal to the number of nonzero elements of $\boldsymbol{a}$, and k_n is the shift-invariant kernel associated to the operator L_n. Moreover, the total number of nonzero coefficients is upper-bounded by the number M of data points and, hence, is not growing with the number N of components. We also illustrate numerically the effect of using multiple kernels.

3.2.2 Mathematical Background

Function Spaces

All the derivatives of a rapidly decaying function decay faster than the inverse of any polynomial at infinity. Then, a smooth and slowly growing function is an element of $\mathcal{C}^\infty(\mathbb{R}^d)$ such that all of its derivatives have asymptotic growth controlled by a polynomial. Finally, a heavy-tailed function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies $f(\boldsymbol{x}) \geq C(1+\|\boldsymbol{x}\|)^\alpha$ for some finite constants $C, \alpha > 0$.

For $p \in [1, \infty)$, we denote by $L_p(\mathbb{R}^d)$, the Banach space of measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ with finite L_p norm, *i.e.*

$$L_p(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ measurable} : \|f\|_{L_p} \triangleq \left(\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^p \mathrm{d}\boldsymbol{x} \right)^{\frac{1}{p}} < +\infty \right\}. \tag{3.18}$$

The Schwartz space of smooth and rapidly decaying functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ is denoted by $\mathcal{S}(\mathbb{R}^d)$. Its topological dual is $\mathcal{S}'(\mathbb{R}^d)$, the space of tempered distributions [162]. We remark that any smooth and slowly growing function $f : \mathbb{R}^d \to \mathbb{R}$ specifies the continuous linear functional $\varphi \mapsto \int_{\mathbb{R}^d} f(\boldsymbol{x})\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ over $\mathcal{S}(\mathbb{R}^d)$ and, hence, is an element of $\mathcal{S}'(\mathbb{R}^d)$.

The space of continuous functions over $\mathbb{R}^d$ that vanish at infinity is $\mathcal{C}_0(\mathbb{R}^d)$. It is a Banach space equipped with the supremum norm $\|\cdot\|_{L_\infty}$. The space of Schwartz functions $\mathcal{S}(\mathbb{R}^d)$ is densely embedded in $\mathcal{C}_0(\mathbb{R}^d)$. Hence, the topological dual of

$\mathcal{C}_0(\mathbb{R}^d)$ can be defined as

$$\mathcal{M}(\mathbb{R}^d) = \left\{ w \in \mathcal{S}'(\mathbb{R}^d) : \quad \|w\|_{\mathcal{M}} \overset{\Delta}{=} \sup_{\substack{\varphi \in \mathcal{S}(\mathbb{R}^d) \\ \|\varphi\|_\infty = 1}} |\langle w, \varphi \rangle| < +\infty \right\}. \tag{3.19}$$

In fact, $\mathcal{M}(\mathbb{R}^d)$ is the Banach space of bounded Radon measures over $\mathbb{R}^d$ equipped with the total-variation norm $\|\cdot\|_{\mathcal{M}}$ [149]. It includes the shifted Dirac impulses $\delta(\cdot - \boldsymbol{x}_0)$, with $\|\delta(\cdot - \boldsymbol{x}_0)\|_{\mathcal{M}} = 1$. Moreover, $L_1(\mathbb{R}^d) \subseteq \mathcal{M}(\mathbb{R}^d)$ with the relation $\|f\|_{L_1} = \|f\|_{\mathcal{M}}$ for all $f \in L_1(\mathbb{R}^d)$. This allows one to interpret $(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathcal{M}})$ as a generalization of $(L_1(\mathbb{R}^d), \|\cdot\|_{L_1})$. Finally, for any sequence $\boldsymbol{a} = (a_n) \in \ell_1(\mathbb{Z})$ and distinct locations $x_n, n \in \mathbb{Z}$, we have that

$$w_{\boldsymbol{a}} = \sum_{n \in \mathbb{Z}} a_n \delta(\cdot - x_n) \in \mathcal{M}(\mathbb{R}) \quad \text{and} \quad \|w_{\boldsymbol{a}}\|_{\mathcal{M}} = \|\boldsymbol{a}\|_{\ell_1}. \tag{3.20}$$

This property establishes a tight link between the total-variation norm and the discrete ℓ_1 norm which is known to promote sparsity and is the key element in the field of compressed sensing [62, 163, 66]. This enabled researchers to interpret the total-variation norm as a sparsity-promoting norm in the continuous domain. Since then, additional connections have been drawn between optimization problems that involve the total-variation norm and many areas of research such as super resolution [164, 69, 70], kernel methods, [105, 153], and splines [73, 3, 71, 72, 141]. The computational aspects of this framework have also been investigated, leading to the development of practical algorithms in various settings [165, 166, 167].

For a Banach space $\mathcal{X}$, we consider two topologies for its continuous dual space $\mathcal{X}'$. The first one is the strong topology. It is induced from the dual norm in the sense that a sequence $\{w_n\}_{n=0}^{\infty} \in \mathcal{X}'$ is said to converge in the strong topology to $w^* \in \mathcal{X}'$ if $\lim_{n\to\infty} \|w_n - w^*\|_{\mathcal{X}'} = 0$. The second one is the weak*-topology that comes from the predual space $\mathcal{X}$ in the sense that a sequence $\{w_n\}_{n=0}^{\infty}$ is said to converge in the weak*-topology to w^* if, for any element $\varphi \in \mathcal{X}$, $\{\langle w_n, \varphi \rangle\}_{n=0}^{\infty}$ converges to $\langle w^*, \varphi \rangle$. Consequently, a functional $\nu : \mathcal{X}' \to \mathbb{R}$ is weak*-continuous if $\nu(w_n) \to \nu(w_{\lim})$ for any sequence $\{w_n\}_{n \in \mathbb{N}} \in \mathcal{X}'$ that converges in the weak*-topology to $w_{\lim}$. This can be shown to be equivalent to the inclusion $\nu \in \mathcal{X}$. In other words, the predual space $\mathcal{X}$ is isometrically isomorphic to the space of weak*-continuous functionals over $\mathcal{X}'$.

Finally, let us mention that any Banach space $\mathcal{X}$ is isometrically isomorphic to a closed subspace of its second dual $\mathcal{X}'' = (\mathcal{X}')'$ (see, for example, [94, pp. 95]). For the sake of simplicity, we make the possible embedding mappings implicit in our framework. This leads to writing the latter proposition simply, via the inclusion $\mathcal{X} \subseteq \mathcal{X}''$. In this regard, a Banach space is reflexive if we have that $\mathcal{X} = \mathcal{X}''$. Typical examples of reflexive Banach spaces are $L_p(\mathbb{R}^d)$ spaces for $p \in (1, \infty)$. By contrast, the space $\mathcal{C}_0(\mathbb{R}^d)$ and, consequently, its dual $\mathcal{M}(\mathbb{R}^d)$, are not reflexive.

Linear Operators

The linear operator $\mathrm{L} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is called shift-invariant if, for any function $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and any shift value $\boldsymbol{x}_0 \in \mathbb{R}^d$, we have that

$$\mathrm{L}\{\varphi(\cdot - \boldsymbol{x}_0)\} = \mathrm{L}\{\varphi\}(\cdot - \boldsymbol{x}_0). \tag{3.21}$$

We recall a variant of the celebrated Schwartz kernel theorem for linear and shift-invariant (LSI) operators (see [168] for a “simple” proof of the general case).

Theorem 3.1 (Schwartz kernel theorem)**.** *For any LSI operator* $\mathrm{L} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$*, there exists a unique distribution* $h \in \mathcal{S}'(\mathbb{R}^d)$*, known as the impulse response of* L*, such that*

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d) : \mathrm{L}\{\varphi\}(\cdot) = \int_{\mathbb{R}^d} h(\cdot - \boldsymbol{y})\varphi(\boldsymbol{y})\mathrm{d}\boldsymbol{y}. \tag{3.22}$$

In this work, we restrict ourselves to the class of continuous LSI operators that have an extended domain and are defined over the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$. One can fully characterize this class in the Fourier domain. The Fourier transform is a well-defined and continuous operator over $\mathcal{S}'(\mathbb{R}^d)$ and is denoted by $\mathcal{F} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$. Consequently, the frequency response of the LSI operator $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is defined as the Fourier transform of its impulse response

$$\widehat{\mathrm{L}}(\boldsymbol{\omega}) \triangleq \mathcal{F}\{\mathrm{L}\{\delta\}\}(\boldsymbol{\omega}). \tag{3.23}$$

It is known that the frequency response of any continuous LSI operator over $\mathcal{S}'(\mathbb{R}^d)$ is a smooth and slowly growing function [169]. Additionally, any smooth and slowly growing function $\widehat{\mathrm{L}}(\cdot)$ defines an LSI and continuous operator $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ via

$$\mathrm{L}\{f\} = \mathcal{F}^{-1}\{\widehat{\mathrm{L}}\widehat{f}\}. \tag{3.24}$$

Typical examples of such operators are polynomials of derivative in dimension $d = 1$ and polynomials of the Laplacian operator for $d > 1$ [83].

3.2.3 Banach-Space Kernels

We now introduce our Banach-space framework of learning with gTV regularization. We start by defining the class of kernel-admissible operators.

Definition 3.1. *The linear operator* $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *is called kernel-admissible (or simply admissible) if*

1. *it is shift-invariant*[3]*;*
2. *it is an isomorphism over* $\mathcal{S}'(\mathbb{R}^d)$*, meaning that it is continuous and invertible, its inverse being the continuous operator* $\mathrm{L}^{-1} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$*;*
3. *the sampling functional* $\delta_{\boldsymbol{x}_0} : f \mapsto f(\boldsymbol{x}_0)$ *is weak*-continuous in the topology of its native space (see Definition 3.2 and Theorem 3.2).*

The restriction to LSI operators is not crucial to our framework. However, it lends itself to the convenience of an analysis in the Fourier domain. It also allows us to provide necessary and sufficient conditions to characterize the class of admissible operators (see Theorem 3.3). The invertibility assumption, on the other hand, is essential to have decaying kernels; that is, to have $\mathrm{k}(\boldsymbol{x} - \boldsymbol{y}) \to 0$ whenever $\|\boldsymbol{x} - \boldsymbol{y}\| \to \infty$. In fact, it is known that the Green's function of any LSI operator with a nontrivial null space necessarily has a singularity in the Fourier domain at the origin [48]. Finally, the assumption of the (weak*) continuity of the sampling functional is a natural choice in learning theory. The main motivation here is to

[3] Although the notion of shift-invariant operators in (3.21) is defined for operators acting on Schwartz functions, one can extend it by duality to those whose domain is $\mathcal{S}'(\mathbb{R}^d)$. For more details on extension by duality, we refer to [48, Section 3.3.2].

guarantee the (weak*) lower semicontinuity of the global cost functional in (3.16). This can be used, together with the generalized Weierstrass theorem, to prove the existence of solutions. Let us note that the definition of weak*-continuity depends on the Banach structure of the native space. In the sequel, we first properly define native spaces and then specify their underlying Banach structures.

Definition 3.2. *The native space of the LSI isomorphism* $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *is the pre-image of* L *over the space of bounded Radon measures; that is, the space* $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) = \mathrm{L}^{-1}\{\mathcal{M}(\mathbb{R}^d)\}$.

Theorem 3.2 summarizes the important properties of the native spaces.

Theorem 3.2. *Let* $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *be an LSI isomorphism over* $\mathcal{S}'(\mathbb{R}^d)$*. Then, its native space is a topological vector space with the following properties:*

1. *It is a Banach space equipped with the generalized total-variation norm*

$$\mathrm{gTV}(f) = \|f\|_{\mathcal{M}_{\mathrm{L}}} \triangleq \|\mathrm{L}\{f\}\|_{\mathcal{M}}. \tag{3.25}$$

2. *The restriction of* L *to its native space results in the isomorphism* $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$.

3. *The adjoint operator* L^* *is well-defined over* $\mathcal{C}_0(\mathbb{R}^d)$ *and its image is the Banach space* $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ *with the norm* $\|f\|_{\mathcal{C}_{\mathrm{L}}} \triangleq \|\mathrm{L}^{-1*}\{f\}\|_{L_\infty}$.

4. *The space* $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ *is the predual of* $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$*, meaning that* $(\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d))' = \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$.

5. *The space of Schwartz functions is embedded in the native space. Moreover, the native space itself is densely embedded in the space of tempered distributions. The embedding hierarchy is indicated as*

$$\mathcal{S}(\mathbb{R}^d) \hookrightarrow \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \overset{d.}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^d). \tag{3.26}$$

Proof. Item 1: The linearity and invertibility of L implies that the native space together with the gTV norm is a *bona fide* Banach space.

Item 2: The restriction of L over its native space is injective (inherited from L) and is continuous due to the definition of the gTV norm. For all $w \in \mathcal{M}(\mathbb{R}^d)$, the relation $\mathrm{L}\{\mathrm{L}^{-1}\{w\}\} = w$ implies that it is surjective as well and that its inverse is the restriction of L^{-1} over $\mathcal{M}(\mathbb{R}^d)$ which continuously maps $\mathcal{M}(\mathbb{R}^d) \to \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. This ensures that $\mathrm{L} : \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ is an isomorphism.

Item 3: The isomorphism of Part 2 implies the existence of the adjoint operator over $\big(\mathcal{M}(\mathbb{R}^d)\big)'$. By restricting the adjoint operator to $\mathcal{C}_0(\mathbb{R}^d)$, we obtain the operator $\mathrm{L}^* : \mathcal{C}_0(\mathbb{R}^d) \to \mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$, where the space $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ is the image of L^* over $\mathcal{C}_0(\mathbb{R}^d)$. This space, equipped with the norm $\|f\|_{\mathcal{C}_{\mathrm{L}}} \triangleq \|\mathrm{L}^{-1*}\{f\}\|_{L_\infty}$, is a Banach space due to the linearity and invertibility of L^{-1*}.

Item 4: Similarly to Part 2, we readily verify that the adjoint operator $\mathrm{L}^* : \mathcal{C}_0(\mathbb{R}^d) \to \mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ is indeed an isomorphism. Therefore, the double-adjoint operator is the isomorphism $\mathrm{L}^{**} : (\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d))' \to \mathcal{M}(\mathbb{R}^d)$. Consequently, the domains of L and L^{**} must be equal, which implies that $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$ is the predual of the native space.

Item 5: First, we show that the operator L is closed over the space of Schwartz functions. It is known that the impulse response of $\mathrm{L}^* : \mathcal{S} \to \mathcal{S}$ is the flipped version of the one of L [48]. In other words, the application of L^* on a Schwartz function can be expressed by

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d) : \mathrm{L}^*\{\varphi\}(\cdot) = \int_{\mathbb{R}^d} h(\boldsymbol{x} - \cdot)\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{3.27}$$

where $h \in \mathcal{S}'(\mathbb{R}^d)$ is the impulse response of L, described in (3.22). By the change of variable $\boldsymbol{y} = (-\boldsymbol{x})$, one verifies that, for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, we have that

$$\mathrm{L}\{\varphi\} = \mathrm{L}^*\{\varphi^\vee\}^\vee, \tag{3.28}$$

where $\varphi^\vee$ is the flipped version of $\varphi \in \mathcal{S}(\mathbb{R}^d)$ with $\varphi^\vee(\boldsymbol{x}) = \varphi(-\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$. In effect, (3.28) shows that $\mathrm{L}\{\varphi\} \in \mathcal{S}(\mathbb{R}^d)$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Now, from the inclusions $\mathrm{L}\{\mathcal{S}(\mathbb{R}^d)\} \subseteq \mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{M}(\mathbb{R}^d)$, we deduce that $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. Moreover, $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$ by Definition 3.2. This verifies the inclusion $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{M}(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$. To complete the proof, we need to show that the identity operators $id_1 : \mathcal{S}(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ and $id_2 : \mathcal{M}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ are continuous.

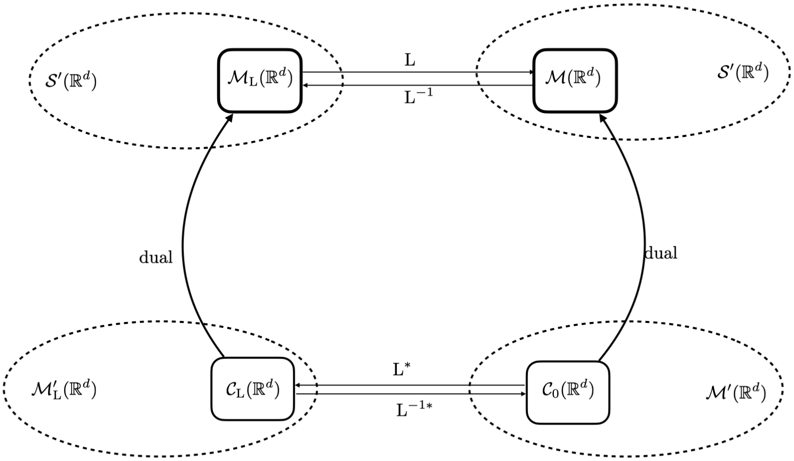

$\mathcal{S}'(\mathbb{R}^d)$
$\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$
L
L^{-1}
$\mathcal{M}(\mathbb{R}^d)$
$\mathcal{S}'(\mathbb{R}^d)$
dual
dual
$\mathcal{M}'_{\mathrm{L}}(\mathbb{R}^d)$
$\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$
L^*
L^{-1*}
$\mathcal{C}_0(\mathbb{R}^d)$
$\mathcal{M}'(\mathbb{R}^d)$

Figure 3.3: A Schematic diagram that illustrates the Banach spaces of interest.

For a converging sequence of Schwartz functions $\varphi_n \overset{\mathcal{S}}{\to} \varphi$, the continuity of L implies that $\mathrm{L}\{\varphi_n\} \overset{\mathcal{S}}{\to} \mathrm{L}\{\varphi\}$. Since $\mathcal{S}(\mathbb{R}^d)$ is continuously embedded in $\mathcal{M}(\mathbb{R}^d)$, we have that $\mathrm{L}\{\varphi_n\} \overset{\mathcal{M}}{\to} \mathrm{L}\{\varphi\}$ and, consequently, that $\varphi_n \overset{\mathcal{M}_{\mathrm{L}}}{\to} \varphi$. This proves that the embedding is continuous, which is denoted by $\mathcal{S}(\mathbb{R}^d) \hookrightarrow \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. Moreover, since the space $\mathcal{M}(\mathbb{R}^d)$ is continuously embedded in $\mathcal{S}'(\mathbb{R}^d)$, the convergence $\mathrm{L}\{\varphi_n\} \overset{\mathcal{M}}{\to} \mathrm{L}\{\varphi\}$ implies that $\mathrm{L}\{\varphi_n\} \overset{\mathcal{S}'}{\to} \mathrm{L}\{\varphi\}$. This proves that $\mathcal{M}_{\mathrm{L}}(\mathbb{R}^d) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$. The latter continuous embedding is also dense due to the denseness of $\mathcal{S}(\mathbb{R}^d)$ in $\mathcal{S}'(\mathbb{R}^d)$ and the inclusion $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{M}_{\mathrm{L}}(\mathbb{R}^d)$. □

We have summarized the Banach spaces and the mappings between them in Figure 3.3. Due to Theorem 3.2, the weak*-continuity of the sampling functional (Condition 3 in Definition 3.1) is equivalent to the inclusion of the shifted Dirac impulses in the predual of the native space. In other words, for all $\boldsymbol{x}_0 \in \mathbb{R}^d$, one should have that $\delta(\cdot - \boldsymbol{x}_0) \in \mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$.

We now define the shift-invariant kernel associated to an admissible operator.

Definition 3.3. *The shift-invariant kernel associated to the admissible operator* $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *is the bivariate function* $\mathrm{k} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *with* $\mathrm{k}(\boldsymbol{x}, \boldsymbol{y}) = \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{y})$*, where* $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta\}$ *is the Green's function of* L*.*

In Theorem 3.3, we provide the necessary and sufficient conditions that characterize the class of admissible LSI operators.

Theorem 3.3. *Let* L *be an admissible operator. Then, its associated Green's function* $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta\} : \mathbb{R}^d \to \mathbb{R}$ *satisfies the following properties:*

1. *It is a continuous function that vanishes at infinity. In other words,* $\rho_{\mathrm{L}} \in \mathcal{C}_0(\mathbb{R}^d)$*.*
2. *Its Fourier transform* $\widehat{\rho_{\mathrm{L}}}(\boldsymbol{\omega})$ *is a smooth, non-vanishing, slowly growing, and heavy-tailed function of* $\boldsymbol{\omega}$*.*

Additionally, any function $\rho : \mathbb{R}^d \to \mathbb{R}$ *that satisfies these properties can be appointed to an admissible operator* $\mathrm{L} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ *defined as*

$$\mathrm{L}\{f\} = \mathcal{F}^{-1}\left\{\frac{\widehat{f}(\boldsymbol{\omega})}{\widehat{\rho}(\boldsymbol{\omega})}\right\}. \tag{3.29}$$

Proof. Assume that L is a kernel-admissible operator. The weak*-continuity of the sampling functional implies that the shifted Dirac impulses $\delta(\cdot - \boldsymbol{x}_0)$ should be included in the predual space $\mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$. Therefore, $\mathrm{L}^{-1*}\{\delta(\cdot - \boldsymbol{x}_0)\}$ should be in $\mathcal{C}_0(\mathbb{R}^d)$. Since, the Green's functions of L and L* are flipped version of each other, we deduce that $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta(\cdot - \boldsymbol{x}_0)\} \in \mathcal{C}_0(\mathbb{R}^d)$. For the second property, we recall that the continuity of $\mathrm{L}^{-1} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ implies the smoothness and slow growth of the Fourier transform of its frequency response. Hence, $\widehat{\rho_{\mathrm{L}}}(\omega)$ is smooth and slowly growing. Similarly, the continuity of L implies that $\frac{1}{\widehat{\rho_{\mathrm{L}}}(\omega)}$ is a smooth and slowly growing function as well. Thus, $\widehat{\rho_{\mathrm{L}}}(\omega)$ is non-vanishing and heavy-tailed.

For the converse, assume that the function ρ satisfies Properties 1 and 2 in Theorem 3.3. First, note that, if $f, g : \mathbb{R}^d \to \mathbb{R}$ are smooth and slowly growing functions and, moreover, g is nonzero and heavy-tailed, then

$$\frac{\partial}{\partial x_i}\left(\frac{f}{g}\right) = \frac{\frac{\partial f}{\partial x_i} g - \frac{\partial g}{\partial x_i} f}{g^2} \tag{3.30}$$

is a quotient whose numerator is a smooth and slowly growing function and whose denominator g^2 is a nonzero, heavy-tailed, smooth, and slowly growing function. Hence, the quotient itself is a smooth function whose growth is bounded by a polynomial. Based on this observation, one can deduce from induction that all the arbitrary-order derivatives of $\frac{1}{\widehat{\rho}(\boldsymbol{\omega})}$ can be expressed by a quotient with a slowly growing nominator and a heavy-tailed denominator. This shows that $\frac{1}{\widehat{\rho}(\boldsymbol{\omega})}$ is a smooth and slowly growing function as well. These properties ensure the existence of continuous LSI operators $\mathrm{L}, \tilde{\mathrm{L}} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ with the frequency responses $\frac{1}{\widehat{\rho}(\omega)}$ and $\widehat{\rho}(\omega)$, respectively. The one-to-one correspondence between an operator and its frequency response then yields that $\tilde{\mathrm{L}} = \mathrm{L}^{-1}$, from which we conclude that L is an isomorphism over $\mathcal{S}'(\mathbb{R}^d)$. Moreover, due to Property 1, we know that the Green's function of L is in $\mathcal{C}_0(\mathbb{R}^d)$. Hence, the Green's function of L^* is also in $\mathcal{C}_0(\mathbb{R}^d)$ so that, for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, we have that

$$\mathrm{L}^{-1*}\{\delta(\cdot - \boldsymbol{x}_0)\} = \mathrm{L}^{-1*}\{\delta\}(\cdot - \boldsymbol{x}_0) \in \mathcal{C}_0(\mathbb{R}^d). \tag{3.31}$$

In other words, $\delta(\cdot - \boldsymbol{x}_0) \in \mathrm{L}^*(\mathcal{C}_0(\mathbb{R}^d)) = \mathcal{C}_{\mathrm{L}}(\mathbb{R}^d)$, which shows that the sampling functionals are weak*-continuous. □

Remark 3.1. *We have previously stated a well-known result by Schwartz that determines the general family of LSI operators (not necessarily invertible) over $\mathcal{S}'(\mathbb{R}^d)$. In Theorem 3.3, particularly via Condition 2, we are excluding the noninvertible members of this family. Hence, Condition 2 fully characterizes the class of linear isomorphisms over $\mathcal{S}'(\mathbb{R}^d)$.*

Using Theorem 3.3, we now draw a connection to the well-known class of reproducing kernels, which are constrained to be symmetric (because of their positive-definiteness).

Corollary 3.1. *Any symmetric admissible kernel (in the sense of Theorem 3.3) is a shift-invariant reproducing kernel up to multiplication by a sign factor.*

Proof. Let $\mathrm{k}(\cdot,\cdot)$ be a symmetric and shift-invariant admissible kernel. Then, the corresponding Green's function ρ_{L} is also a symmetric function and, hence, its Fourier transform $\widehat{\rho}_{\mathrm{L}}(\boldsymbol{\omega})$ is a real function that is also smooth and non-vanishing. Hence, the sign of $\widehat{\rho}_{\mathrm{L}}(\boldsymbol{\omega})$ is constant everywhere. By multiplying with a sign factor, we can then assume that $\widehat{\rho}_{\mathrm{L}}(\boldsymbol{\omega})$ is positive everywhere. Now by invoking Bochner's theorem (see, for example, [48, Appendix B]), we deduce that ρ_{L} is a positive-definite function which, together with the symmetric assumption, implies that $\mathrm{k}(\cdot,\cdot)$ is indeed a reproducing kernel. □

The practical implication of Theorem 3.3 is that it yields Fourier-domain criteria to determine the admissibility of an operator L. In particular, and due to the Rieman-Lebesgue lemma, if $\widehat{\rho_{\mathrm{L}}}$ is an absolutely integrable function then condition 1 holds.

As the last part of this section, we use this characterization to introduce some families of admissible kernels. Our first example is made of super-exponential kernels defined as

$$\mathrm{k}_\alpha(\boldsymbol{x},\boldsymbol{y}) = \exp(-\|\boldsymbol{x}-\boldsymbol{y}\|_\alpha^\alpha), \quad \alpha \in (0,2), \tag{3.32}$$

where $\|\boldsymbol{x}\|_\alpha = \left(\sum_{i=1}^d |x_i|^\alpha\right)^{\frac{1}{\alpha}}$ for any $\boldsymbol{x} = (x_i) \in \mathbb{R}^d$. These functions are known to be positive-definite [48, Appendix B]. Their inverse Fourier transforms (the so-called α-stable distributions) are heavy-tailed and infinitely smooth, with algebraically decaying derivatives of any order [170, Chapter 5]. Hence, they satisfy the conditions of Theorem 3.3. Note that the classical Gaussian kernels are excluded because their frequency responses are not heavy-tailed. However, one can get arbitrarily close by letting α tend to its critical value 2. Moreover, there are arguments in regularized RKHS that support the use of Gaussian kernels. For example, in [171, 172, 173], the Gaussian RKHS has been implicitly characterized by using the Taylor expansion of the corresponding regularization operator. Further, [174] uses the notion of holomorphic functions to explicitly characterize Gaussian RKHS. We conjecture that the present Banach-space formulation can be extended to cover Gaussian kernels as well. However, this requires one to consider a space larger than $\mathcal{S}'(\mathbb{R})$.

Our second example is made of Bessel potentials used in kernel estimation [175]. For a positive real number $s > d$, we consider the operator $(\mathrm{I} - \Delta)^{\frac{s}{2}} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$, where Δ is the Laplacian operator. The Bessel potentials are the Green's function of such operators. They correspond to the shift-invariant kernels

$$G_s(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{F}^{-1}\left\{\frac{1}{(1 + \|\boldsymbol{\omega}\|_2^2)^{\frac{s}{2}}}\right\}(\boldsymbol{x} - \boldsymbol{y}). \tag{3.33}$$

Clearly, the function $\frac{1}{(1+\|\boldsymbol{\omega}\|_2^2)^{\frac{s}{2}}}$ is in $L_1(\mathbb{R}^d)$ for $s > d$. By invoking the Riemann-Lebesgue lemma, we deduce that its inverse Fourier transform is a continuous function that vanishes at infinity. Hence, the kernel function $G_s(\cdot, \cdot)$ satisfies Property 1 of Theorem 3.3. Moreover, from the Fourier-domain definition (3.33) of $G_s(\cdot, \cdot)$, it can be seen that Property 2 also holds. Together, we deduce the admissibility of these kernels. We remark that the Bessel potential kernels are rotation-invariant as well.

Our final example is a general class of separable shift-invariant kernels of the form

$$\mathrm{k}(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{d} \rho_{\mathrm{L}}(x_i - y_i), \tag{3.34}$$

where $\mathrm{L} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is a stable rational operator whose frequency response is of the form $\widehat{\mathrm{L}}(\omega) = \frac{P(\omega)}{Q(\omega)}$, where P and Q are polynomials with no real roots such that $\deg(P) \geq \deg(Q) + 2$. Since $\widehat{\mathrm{L}}(\omega)$ is real, we conclude that that the tail of $\widehat{\mathrm{L}}(\omega)^{-1} = \frac{Q(\omega)}{P(\omega)}$ behaves like ω^{-2} and is absolutely integrable which, together with the Riemann-Lebesgue lemma, implies that $\rho_{\mathrm{L}} \in \mathcal{C}_0(\mathbb{R})$. The other conditions of Theorem 3.3 can be readily shown to be true so that any separable kernel of the form (3.34) is admissible to our theory.

It is worth to mention that one can rotate and dilate any admissible kernel by considering an invertible mixture matrix $\mathbf{A}$ and by defining the transformed kernel as $\mathrm{k}(\mathbf{A}\boldsymbol{x}, \mathbf{A}\boldsymbol{y})$. One readily verifies that the transformed kernel also satisfies the conditions of Theorem 3.3 and, hence, is also admissible. In Figure 3.4, we have plotted the super-exponential and Bessel-potential kernels in dimension $d = 1$ for different sets of parameters. It can be seen that the width and regularity of these

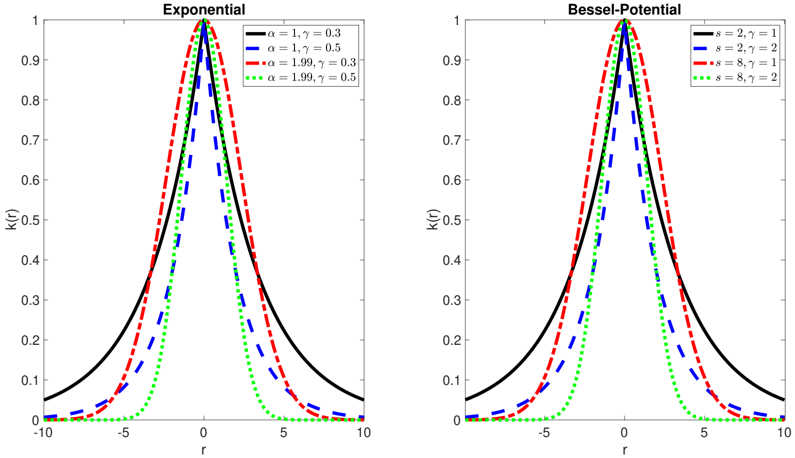


Figure 3.4: Super-exponential kernels $\mathrm{k}_\alpha(\boldsymbol{r}) = \exp(-\gamma\|\boldsymbol{r}\|_\alpha^\alpha)$ (left) and Bessel-potential kernels $G_s(\gamma\boldsymbol{r})$ (right), where $\boldsymbol{r} = (\boldsymbol{x} - \boldsymbol{y})$. The plots are in the special case $d = 1$. The parameters ($\alpha \in (0, 2)$ and $s > 2$) and $\gamma > 0$ adjust smoothness and width of the kernel, respectively.

kernels can be adjusted through their parameters. This can be exploited in our framework of learning with multiple kernels to benefit from this diversity.

3.2.4 Multiple-Kernel Regression

We now invoke our general results in Chapter 2 to prove a representer theorem for multiple-kernel regression with gTV regularization. In effect, the gTV norm will force the learned function to use the fewest active kernels.

Theorem 3.4 (Multiple-kernel regression with gTV)**.** *Given a training dataset that consists of M distinct pairs $(\boldsymbol{x}_m, y_m)$ for $m = 1, 2, \dots, M$, we consider the* *minimization problem*

$$\min_{\substack{f_n \in \mathcal{M}_{\mathrm{L}_n}(\mathbb{R}^d),\\ f=\sum_{n=1}^N f_n}} \left(\sum_{m=1}^{M} E(f(\boldsymbol{x}_m), y_m) + \lambda \sum_{n=1}^{N} \|\mathrm{L}_n\{f_n\}\|_{\mathcal{M}} \right), \tag{3.35}$$

where $E(\cdot, y)$ is a strictly convex nonnegative function and L_n is a kernel-admissible operator in the sense of Definition 3.1 for $n = 1, 2, \dots, N$. Then, the solution set of this problem is nonempty, convex, and weak-compact. For any of its extreme points $(f_1, f_2, \dots, f_N)$, we have the kernel expansions*

$$f_n = \sum_{l=1}^{M_n} a_{n,l} \mathrm{k}_n(\cdot, \boldsymbol{z}_{n,l}), \quad n = 1, 2, \dots, N \tag{3.36}$$

for its components, where $a_{n,l} \in \mathbb{R}$ are kernel weights, $\boldsymbol{z}_{n,l} \in \mathbb{R}^d$ are adaptive kernel positions, and $\mathrm{k}_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the shift-invariant kernel associated to the regularization operator L_n for $n = 1, 2, \dots, N$. Moreover, the number of active kernels is upper-bounded by the number of data points, so that $\sum_{n=1}^N M_n \leq M$.

Proof. It is known that the extreme points of the unit ball of the total-variation norm are of the form $\pm\delta(\cdot - \boldsymbol{x}_0)$ with $\boldsymbol{x}_0 \in \mathbb{R}^d$ [76]. Using Item 4 of Proposition 2.1, we deduce that the extreme points of the unit ball of the gTV norm $\|\mathrm{L}_n\{\cdot\}\|_{\mathcal{M}}$ are of the form $\pm\mathrm{k}_n(\cdot - x_0)$. Finally, we note that from Definition 3.1, we have that $\delta_{\mathbf{x}_m} \in \mathcal{C}_{\mathrm{L}_n}(\mathbb{R}^d)$ for $m = 1, \dots, M$ and $n = 1, \dots, N$. All these together allow us to invoke Theorem 2.4 (in particular, (2.40)) with $\tilde{\boldsymbol{\nu}}_m = (\delta_{\mathbf{x}_m}, \dots, \delta_{\mathbf{x}_m}) \in \prod_{n=1}^N \mathcal{C}_{\mathrm{L}_n}(\mathbb{R}^d)$ for $m = 1, \dots, M$, $\psi = id$, and $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$, which completes the proof. □

The practical outcome of Theorem 3.4 is that any extreme point of (3.35) maps into a solution of the form

$$f(\cdot) = \sum_{n=1}^{N} \sum_{l=1}^{M_n} a_{n,l} \mathrm{k}_n(\cdot, \boldsymbol{z}_{n,l}) \tag{3.37}$$

for the learned function. The solution form (3.37) has the following important properties:

- The number of active kernels is upper-bounded by the number of samples M. This justifies the use of multiple kernels since the flexibility of the model will be increased while the problem remains well-posed.
- The gTV norm enforces an ℓ_1 penalty on the kernel coefficients. Practically, this will result in an ℓ_1-minimization problem that is reminiscent of the generalized LASSO.
- The kernel positions are adaptive and will be chosen such that the solution becomes sparse. In other words, the adaptiveness of the kernel positions, together with the ℓ_1 regularization on the kernel coefficients, favors solutions with a small number of nonzero terms in the expansion (3.35).

To conclude this section, let us mention that the existence of the kernel locations $\mathbf{z}_{n,l}$ in (3.37) is guaranteed by our representer theorem. However, unlike in RKHS methods, these locations do not necessarily coincide with the data points. The adaptiveness comes from the fact that the kernel positions become part of the reduced finite-dimensional optimization problem (see (3.13) for the single-kernel scenario). Hence, an optimization scheme is required in order to "learn" these unknown parameters along with the kernel weights.

3.2.5 Numerical Illustration

Here, we provide a numerical example in the case $d = 1$. We would like to emphasize that the computational aspects of our framework (*e.g.*, the derivation of efficient algorithms in high dimensions) is left to future works. The sole purpose of our example is to illustrate the use of Theorem 3.4 and highlight two important features, namely, adaptivity and sparsity. In our example, we compare the performance of five kernel estimators:

1. **RKHS $\mathbf{L_2}$:** RKHS regularization (3.4).
2. **RKHS $\mathbf{L_1}$:** Generalized LASSO (3.8).

3. **SimpleMKL:** Multiple-kernel learning (MKL) using the SimpleMKL algorithm [156].

4. **Single gTV:** Single-kernel gTV regularized learning (3.13).

5. **Multi gTV:** Learning with multiple kernels and gTV regularization (3.35).

Setup

The ground-truth signal for our experiment is a piecewise linear function with four segments that connects five points, located at $\{(0,0),(0.45,0),(\frac{7}{15},-2),(\frac{8}{15},2),(1,2)\}$. We then sample data from the model $y_m = f(x_m)+\epsilon_m, m = 1,\dots,M$, where $\epsilon_m \sim \mathcal{N}(0,\sigma^2)$ is i.i.d. Gaussian with $\sigma = 0.1$. We formed two training datasets of size $M = 100$. In the first one, x_m are i.i.d. samples of a uniform distribution over $[0,1]$. In the second case, we put a gap in the training dataset by sampling x_m uniformly over $[0,1]\backslash[0.6,0.8]$.

We use Gaussian kernels in the RKHS-based methods and super-exponential kernels with $\alpha = 1.99$ in the gTV-based methods. We have set $\alpha = 1.99$ to have similar (near-Gaussian) kernel shapes in all cases. All methods have access to ten different width parameters from 10 to 10^5 in log scale.

To avoid the difficulty of optimizing over the data centers in the gTV-based methods, which would result in a nonconvex problem, we use a convex proxy in which a redundant set of centers is placed on a grid and the excess ones are suppressed with the help of ℓ_1-minimization. With this grid-based approach, the search for the kernel positions is reduced to a large-scale ℓ_1-minimization problem for which robust algorithms are known to exist—specifically, we have used fast iterative shrinkage-thresholding algorithm (FISTA) [176] in our example. This scheme will obviously only work when the input dimension is very low, such as $d = 1$ in the present example. In these cases, we have also used the multiresolution strategy of [166] to control the accuracy. More precisely, we start by considering 16 equi-spaced kernels and we then use FISTA to solve the convex problem of finding the corresponding kernel coefficients. The solution is propagated as initialization of a finer grid (with 32 kernels) and we continue until we reach to the finest scale, with 1,024 kernels.

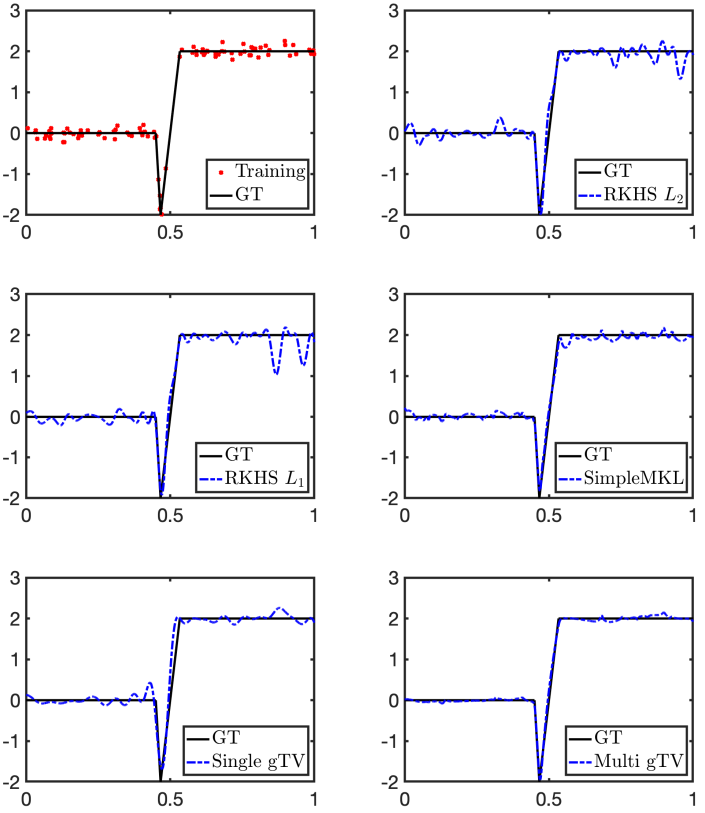


Figure 3.5: Performance of the kernel estimators in full dataset. Solid line: ground-truth (GT) function. Dash-dotted line: reconstructed functions. Dots: noisy data points.

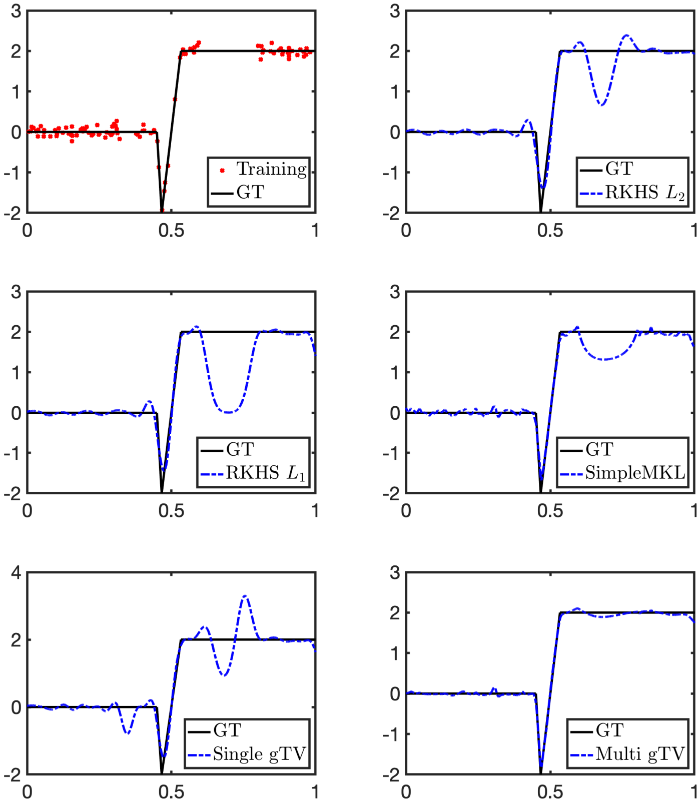


Figure 3.6: Performance of the kernel estimators in missing dataset. Solid line: ground-truth (GT) function. Dash-dotted line: reconstructed functions. Dots: noisy data points.

Table 3.1: MSE and sparsity of the kernel estimators. The results are averaged over 10 runs.

Quantity	Dataset	L2-RKHS	L1-RKHS	SimpleMKL	Single gTV	Multi gTV
Sparsity	Full data	64.7	44.1	54.4	32.5	**20.0**
	Missing data	66.1	39.3	56.0	32.9	**31.1**
MSE (dB)	Full data	-17.2	-16.1	-15.2	-16.7	**-18.1**
	Missing data	-2.6	-2.7	-10.9	-3.9	**-17.3**

We set the data fidelity to be the quadratic term $E(x, y) = (x - y)^2$ in all cases except for MKL, since the SimpleMKL toolbox [156] uses the ϵ-insensitive SVM loss. The other methods are implemented using the GlobalBioIm library [177] and the codes are all available online[4]. To have a fair comparison, we optimize the hyper-parameters of each method by following a standard K-fold cross-validation scheme, setting $K = 5$ in our example. This includes a tuning of the regularization parameter λ for all methods. In addition, we tune the width of the kernel function in single-kernel schemes so that all methods have access to the same family of kernel functions. For computing the test error, we consider a very fine grid with stepsize 10^{-4} over $[0, 1]$ and we compute the MSE between the learned function and the ground-truth signal.

Results

We consider the reconstruction of a function from its noisy samples in two scenarios: full data *versus* missing data. The results are depicted in Figures 3.5 and 3.6, respectively. As we can see in Figure 3.5, due to the presence of a non-smooth region in the target function, the single-kernel methods are forced to use narrow kernels with a small width which creates undesirable oscillations in the smoother regions. By contrast, our multi-kernel scheme uses both narrow and wide kernels, hence providing the reconstruction with the least fluctuation. In the presence of missing data, we observe in Figure 3.6 that the reconstructed function of RKHS-based methods, exhibit an undesirable dip. This is due to the fact that, in the RKHS-based methods, the kernel functions are located on the data points and their

[4] https://github.com/Biomedical-Imaging-Group/Multi-Kernel-Regression-gTV-

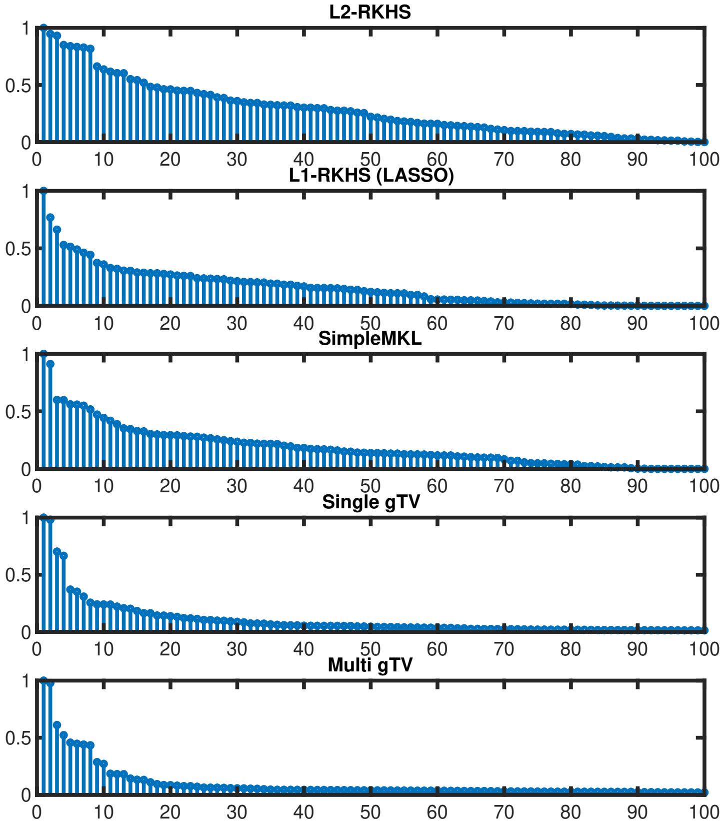


Figure 3.7: 100 largest coefficients of each expansion in the full-data case.

width is too short to fill the gap in the data. By contrast, the kernel locations are adaptive in our scheme, which yields a decent reconstruction in this case as well.

Finally, we have plotted the 100 largest kernel coefficients of each expansion in the full-data experiment in Figure 3.7. This plot highlights that the gTV-based methods are providing the sparsest representation for the target function, as expected. Our visual observations are also supported quantitatively in Table 1, where we report the mean-squared error (MSE) error and sparsity (number of coefficients that are larger than one tenth of the maximum coefficient) of each method in the two scenarios.

3.2.6 Summary

We have provided a theoretical foundation for multiple-kernel regression with gTV regularization. We have studied the Banach structure of our search space and identified the class of kernel functions that are admissible. Then, we have derived a representer theorem that shows that the learned function can be written as a linear combination of kernels with adaptive centers. Our representer theorem also provides an upper bound to the number of active elements, which allows us to use as many kernels as convenient. We have illustrated numerically the effect of using multiple kernels with a sparsity constraint.

3.3 Sparsest Univariate Learning Models Under Lipschitz Constraint

3.3.1 Context

Besides the minimization of the prediction error, two of the most desirable properties of a regression scheme are *stability* and *interpretability*. Driven by these principles, we introduce[5] two variational formulations for regressing one-dimensional data that favor "stable" and "simple" regression models. Similar to RKHS theory, the latter are nonparametric continuous-domain problems. Inspired by the stability principle, we focus on the development of regression schemes with controlled Lipschitz regularity. This is motivated by the observation that many analyses in deep learning require assumptions on the Lipschitz constant of the learned mapping [178, 179, 180]. Likewise, in the context of so-called "plug-and-play" methods—*i.e.* when a trainable module is inserted into an iterative-reconstruction framework—, the rate of convergence of the overall scheme often depends on the Lipschitz constant of this module [181, 182, 183, 184, 185, 186].

In our first formulation, we use the Lipschitz constant of the learned mapping as a regularization term. Specifically, we consider the minimization problem

$$\min_{f\in\mathrm{Lip}(\mathbb{R})}\left(\sum_{m=1}^{M} E\left(f(x_m), y_m\right) + \lambda L(f)\right), \tag{3.38}$$

where $\mathrm{Lip}(\mathbb{R})$ is the space of Lipschitz-continuous real functions and $L(f)$ denotes the Lipschitz constant of $f \in \mathrm{Lip}(\mathbb{R})$. In this formulation, one can implicitly control the Lipschitz regularity of the learned function by varying the regularization parameter λ. We prove a representer theorem that characterizes the solution set of (3.38). In particular, we prove that the global minimum is achieved by a continuous and piecewise-linear (CPWL) mapping. Next, motivated by the simplicity principle, we find the mapping with the minimal number of linear regions. Note that many previous works study problems similar to (3.38) in more general settings, typically using the Lipschitz constant of the nth derivative $\mathcal{R}(f) = L(f^{(n)})$ with $n \geq 0$ as the

[5] This section is based on our published work [110].

regularization term [187, 188, 189, 190, 191, 192]. More recently, [193] has studied the classification problem over metric spaces and derived a parametric form for a solution of this problem. Our work complements this interesting line of research by providing an in-depth analysis of the $n = 0$ case which is related to second-order total-variation minimization, and by focusing more on computational aspects of (3.38). More precisely, we propose a two-step algorithm to reach the sparsest CPWL solution of (3.38). The first step consists in solving a discrete problem with ℓ_∞ regularization, and the second is a sparsification step proposed in [194] that reaches the sparsest solution.

In the second scenario, we explicitly control the Lipschitz constant of the learned mapping by imposing a hard constraint. Inspired by the theoretical insights of the first problem, we add a second-order total-variation (TV) regularization term that is known to promote sparse CPWL functions [103, 194]. This leads to the minimization problem

$$\min_{f \in \mathrm{BV}^{(2)}(\mathbb{R})} \left(\sum_{m=1}^{M} E\left(f(x_m), y_m\right) + \lambda \mathrm{TV}^{(2)}(f) \right), \quad \text{s.t.} \quad L(f) \leq \overline{L}, \tag{3.39}$$

where $\mathrm{BV}^{(2)}(\mathbb{R})$ is the space of functions with bounded second-order TV and $\overline{L}$ is the user-defined upper-bound for the desired Lipschitz regularity of the learned mapping. The interesting aspect of (3.39) is that the simplicity and stability of the learned mapping can be adjusted by tuning the parameters $\lambda > 0$ and $\overline{L} > 0$, respectively. In this case as well, we prove a representer theorem which guarantees the existence of CPWL solutions. We propose a two-step algorithm to find the sparsest CPWL solution which is similar to that of the first scenario. The main difference is the first step, where the discrete problem has a ℓ_1 regularization term and a ℓ_∞ constraint.

Another major motivation for this work is to further elucidate the tight connection between CPWL functions and neural networks. When it comes to shallow networks, the connection with our framework becomes even more explicit. It is well known in the literature that the standard training (*i.e.,* with weight decay) of a two-layer univariate ReLU network is equivalent to solving a TV-based variational problem such as (3.39) without the Lipschitz constraint [195, 140]. This is due to the fact that the weight-decay penalty can be shown to be equal to the second-order TV of the input-output mapping of the full network at the optimum. As we demonstrate, these results can be readily extended to prove the equivalence between the training of a Lipschitz-constrained two-layer univariate ReLU network and our formulation (3.39). Our description of the solution set of Problem (3.39) thus provides insights into the training of Lipschitz-aware neural networks.

3.3.2 Mathematical Background

Weak Derivatives

Schwartz' space of smooth and compactly supported test functions is denoted by $\mathcal{D}(\mathbb{R})$. It is known that the nth-order derivative is a continuous mapping over $\mathcal{D}(\mathbb{R})$, which we denote as $\mathrm{D}^n : \mathcal{D}(\mathbb{R}) \to \mathcal{D}(\mathbb{R})$ [169]. By duality, this allows one to extend the derivative operator to the whole class $\mathcal{D}'(\mathbb{R})$ of distributions. The extended operator is called the nth-order weak derivative and will be denoted by $\mathrm{D}^n : \mathcal{D}'(\mathbb{R}) \to \mathcal{D}'(\mathbb{R})$. For any $w \in \mathcal{D}'(\mathbb{R})$, the distribution $\mathrm{D}^n\{w\} \in \mathcal{D}'(\mathbb{R})$ is defined via its action on a generic test function $\varphi \in \mathcal{D}(\mathbb{R})$ as $\langle \mathrm{D}^n w, \varphi\rangle = (-1)^n \langle w, \mathrm{D}^n \varphi\rangle$. The fundamental property is that the weak derivative of any Schwartz test function $\varphi \in \mathcal{D}(\mathbb{R}) \subseteq \mathcal{D}'(\mathbb{R})$ is well-defined and coincides with the classical notion of derivative (see [48, Section 3.3.2.] for more details on the extension by duality).

Lipschitz Constant

Given generic Banach spaces $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$, a function $f : \mathcal{X} \to \mathcal{Y}$ is said to be Lipschitz-continuous if there exists a finite constant $C > 0$ such that

$$\|f(x_1) - f(x_2)\|_{\mathcal{Y}} \leq C\|x_1 - x_2\|_{\mathcal{X}}, \quad \forall x_1, x_2 \in \mathcal{X}. \tag{3.40}$$

The minimal value of C is called the Lipschitz constant of f and is denoted by $L(f)$. In particular, we denote by $\mathrm{Lip}(\mathbb{R})$, the space of Lipschitz-continuous functions $f : \mathbb{R} \to \mathbb{R}$ with a finite Lipschitz constant, satisfying

$$L(f) = \sup_{x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} < +\infty. \tag{3.41}$$

Following Rademacher's theorem, any Lipschitz-continuous function $f \in \mathrm{Lip}(\mathbb{R})$ is differentiable almost everywhere with a measurable and essentially bounded derivative. The Lipschitz constant of the function then corresponds to the essential supremum of its derivative, so that

$$L(f) = \|\mathrm{D}\{f\}\|_{L_\infty} = \operatorname*{ess\,sup}_{x \in \mathbb{R}} |f'(x)|. \tag{3.42}$$

Conversely, any distribution $f \in \mathcal{D}'(\mathbb{R})$ whose weak derivative lies in $L_\infty(\mathbb{R})$ is indeed a Lipschitz-continuous function [196, Theorem 1.36]. In other words, we have that

$$\mathrm{Lip}(\mathbb{R}) = \{f \in \mathcal{D}'(\mathbb{R}) : \mathrm{D}\{f\} \in L_\infty(\mathbb{R})\}. \tag{3.43}$$

Second-Order Total-Variation

Finally, we introduce the space $\mathrm{BV}^{(2)}(\mathbb{R})$ of functions with finite second-order total-variation, defined as

$$\mathrm{TV}^{(2)}(f) = \|\mathrm{D}^2\{f\}\|_{\mathcal{M}} = \sup_{\substack{\varphi \in \mathcal{D}(\mathbb{R}) \\ \|\varphi\|_{L_\infty}=1}} \langle \mathrm{D}^2 f, \varphi \rangle = \sup_{\substack{\varphi \in \mathcal{D}(\mathbb{R}) \\ \|\varphi\|_{L_\infty}=1}} \int_{\mathbb{R}} f(x)\varphi''(x)\mathrm{d}x. \tag{3.44}$$

Let us mention that the second-order total variation $\mathrm{TV}^{(2)}(f) \triangleq \|\mathrm{D}^2 f\|_{\mathcal{M}}$ is only a semi-norm in this space, since the null space of the linear operator D^2 is nontrivial and consists of degree-one polynomials (affine mappings in $\mathbb{R}$). However, it can become a *bona fide* Banach space with the $\mathrm{BV}^{(2)}$ norm

$$\|f\|_{\mathrm{BV}^{(2)}} \triangleq \mathrm{TV}^{(2)}(f) + |f(0)| + |f(1)|. \tag{3.45}$$

This space has been extensively studied in [103] and in a more general setting in [73]. Remarkably, the sampling functionals $\delta_{x_0} : f \mapsto f(x_0)$ for $x_0 \in \mathbb{R}$ are weak*-continuous in the topology of $\mathrm{BV}^{(2)}(\mathbb{R})$ [103, Theorem 1]. Moreover, any function $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ can be uniquely represented as

$$f(x) = \int_{\mathbb{R}} h(x,y)u(y)\mathrm{d}y + b_1 + b_2 x, \tag{3.46}$$

where $h(x,y) = (x-y)_+ - (1-x)(-y)_+ - x(1-y)_+$, $u = \mathrm{D}^2 f$, $b_1 = f(0)$, and $b_2 = f(1) - f(0)$. This result is a special case of Theorem 5 in [73].

Analogous to the famous total-variation regularization of Rudin-Osher-Fatemi [197], which promotes piecewise-constant functions and causes the notorious staircase effect, the second-order total variation favors CPWL functions. In dimension $d = 1$, this coincides with the known class of nonuniform linear splines which has been extensively studied from an approximation-theoretical point of view [10, 6]. Precisely, it has been shown [194, 73] that the extreme points of the solution set of

$$\min_{f \in \mathrm{BV}^{(2)}(\mathbb{R})} \left(\sum_{m=1}^{M} \mathrm{E}(f(x_m), y_m) + \lambda \mathrm{TV}^{(2)}(f) \right) \tag{3.47}$$

are linear splines with the generic form

$$f(x) = b_0 + b_1 x + \sum_{k=1}^{K} a_k \mathrm{ReLU}(x - z_k), \tag{3.48}$$

where $\mathbf{a} = (a_k) \in \mathbb{R}^K$ and $\mathbf{b} = (b_0, b_1) \in \mathbb{R}^2$ are expansion parameters, $K < M$, and $\{z_k\}_{k=1}^K$ are adaptive (learnable) knots. The regularization term for these functions has the simple expression given by $\mathrm{TV}^{(2)}(f) = \|\mathbf{a}\|_1$. As a result, the original infinite-dimensional problem can be recast as a finite-dimensional one, parameterized by $K, \mathbf{a}, \mathbf{b}$, and $\{z_k\}_{k=1}^K$ [74].

3.3.3 Lipschitz-Aware Formulation for Supervised Learning

We now introduce our formulations for supervised learning that are based on controlling the Lipschitz constant of the learned mapping. Let us first mention that the Lipschitz constant can be indirectly controlled using a $\mathrm{TV}^{(2)}$-type regularizer. Indeed, the two seminorms are connected, as demonstrated in Theorem 3.5.

Theorem 3.5. *Any function with second-order bounded-variation is Lipschitz continuous. Moreover, for any $f \in \mathrm{BV}^{(2)}(\mathbb{R})$, we have the upper-bound*

$$L(f) \leq \mathrm{TV}^2(f) + \ell(f) \tag{3.49}$$

for the Lipschitz constant of f, where

$$\ell(f) = \inf_{x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \geq 0. \tag{3.50}$$

Finally, (3.49) *is saturated if and only if f is monotone and convex/concave.*

Proof. For any $h > 0$ and $\mathbf{p} = (p_1, p_2) \in \mathbb{R}^2$ with $p_1 < p_2$, let us first define the test function $\varphi_h(\cdot; \mathbf{p}) \in \mathcal{C}_0(\mathbb{R})$ as

$$\varphi_h(x; \mathbf{p}) = h^{-1}\Big(\mathrm{ReLU}\left(x - (p_1 - h)\right) - \mathrm{ReLU}(x - p_1) + \mathrm{ReLU}\left(x - (p_2 + h)\right) - \mathrm{ReLU}(x - p_2)\Big). \tag{3.51}$$

This function will be used on several occasions throughout the proof. In particular, we use the explicit form of its second-order derivative given by

$$\mathrm{D}^2\varphi_h(\cdot; \mathbf{p}) = h^{-1}\Big(\delta\left(\cdot - (p_1 - h)\right) - \delta(\cdot - p_1) + \delta\left(\cdot - (p_2 + h)\right) - \delta(\cdot - p_2)\Big). \tag{3.52}$$

Upper-Bound: Similar to (3.42), we have that $\ell(f) = \operatorname{ess\,inf}_{x \in \mathbb{R}} |f'(x)|$. For a fixed $\epsilon > 0$, by definition of the essential supremum and infimum, there exist $\bar{x}, \underline{x} \in \mathbb{R}$ at which f is differentiable with $|f'(\bar{x})| \geq (L(f) - \epsilon)$ and $|f'(\underline{x})| \leq (\ell(f) + \epsilon)$. Without loss of generality, we assume that $\bar{x} < \underline{x}$. Following the limit definition of the derivative, we then consider a small radius $h > 0$ such that

$$\left|\frac{f(\bar{x} + h) - f(\bar{x})}{h}\right| \geq |f'(\bar{x})| - \epsilon \geq L(f) - 2\epsilon, \qquad \left|\frac{f(\underline{x} + h) - f(\underline{x})}{h}\right| \leq |f'(\underline{x})| + \epsilon \leq \ell(f) + 2\epsilon. \tag{3.53}$$

Now, let us consider the test function $\varphi = \varphi_h\left(\cdot; (\bar{x} + h, \underline{x})\right)$. Following the definition of the total-variation norm (as the dual of the supremum norm) together with $\|\varphi\|_\infty = 1$, we deduce that $\mathrm{TV}^{(2)}(f) \geq |\langle \mathrm{D}^2 f, \varphi\rangle| = |\langle f, \mathrm{D}^2\varphi\rangle|$, where the last equality follows from the self-adjointness of the second-order derivative. Using

(3.52), we thus have that

$$\begin{aligned}\mathrm{TV}^{(2)}(f) &\geq h^{-1}|f(\bar{x}) - f(\bar{x}+h) + f(\underline{x}+h) - f(\underline{x})| \\ &\geq \frac{|f(\bar{x}+h) - f(\bar{x})|}{h} - \frac{|f(\underline{x}+h) - f(\underline{x})|}{h} \\ &\geq L(f) - 2\epsilon - \ell(f) - 2\epsilon = L(f) - \ell(f) - 4\epsilon. \end{aligned} \tag{3.54}$$

Finally, by letting $\epsilon \to 0$, we deduce the desired upper-bound.

Saturation—Sufficient Conditions: Assume that $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ is convex and increasing; we denote its second-order weak derivative by $w = \mathrm{D}^2 f$. Note that, in this case, the functions $(-f(\cdot))$, $f(-\cdot)$, and $(-f(-\cdot))$ are concave/decreasing, convex/decreasing, and concave/increasing, respectively. Hence, we only need to prove the saturation for f and the other cases immediately follow.

For a fixed $\epsilon > 0$, from (3.44) there exists a test function $\psi \in \mathcal{D}(\mathbb{R})$ with compact support $K = \mathrm{supp}(\psi)$ such that $\|\psi\|_{L_\infty} = 1$ and $\langle w, \psi\rangle \geq \left(\mathrm{TV}^{(2)}(f) - \epsilon\right)$. For any $T > 0$, we consider the test function $\psi_T = \varphi_1\left(\cdot; (-T, T)\right)$. From (3.52), we obtain that

$$\begin{aligned}\langle w, \psi_T\rangle &= \langle f, \mathrm{D}^2\psi_T\rangle \\ &= (f(T+1) - f(T)) - (f(-T) - f(-T-1)) \\ &\leq L(f) - \ell(f), \end{aligned} \tag{3.55}$$

where we have used the increasing assumption to deduce that $f(T+1) \geq f(T)$ and $f(-T) \geq f(-T-1)$. By choosing T large enough so that $K \subseteq [-T, T]$, we ensure that $(\psi_T - \psi)$ is a nonnegative function, since for all $x \in K$, we will have that $\psi_T(x) = 1 = \|\psi\|_{L_\infty} \geq \psi(x)$. Next, the convexity of f implies that $w = \mathrm{D}^2 f$ is a positive measure. Hence,

$$0 \leq \langle w, \psi_T - \psi\rangle \leq L(f) - \ell(f) - \mathrm{TV}^{(2)}(f) + \epsilon. \tag{3.56}$$

By letting $\epsilon \to 0$, we deduce that $\mathrm{TV}^{(2)}(f) \leq (L(f) - \ell(f))$, which implies the saturation of (3.49).

Saturation—Necessary Conditions: Let $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ be a function for which (3.49) is saturated.

Monotonicity: Assume by contradiction that f is not monotone. Hence, there exists $x_n \in \mathbb{R}$ such that $f'(x_n) < 0$. Indeed, if f' were a positive distribution, then for any $a, b \in \mathbb{R}$ with $a < b$, we would have that $(f(b) - f(a)) = \langle f', \mathbb{1}_{[a,b]} \rangle \geq 0$, which contradicts the assumption of non-monotonicity. Similarly, there exists $x_p \in \mathbb{R}$ such that $f'(x_p) > 0$.

Next, consider a point $x_L \in \mathbb{R}$, distinct from x_n and x_p, such that $|f'(x_L)| > (L(f) - \epsilon) > 0$, where $0 < \epsilon < \frac{\min(-f'(x_n), f'(x_p))}{3}$ is a small constant. Without loss of generality, let us assume that $f'(x_L) > 0$ and $x_n < x_L$. (For $f'(x_L) < 0$, the same arguments can be applied to x_p instead of x_n.) There exists a small radius $h \in (0, \frac{|x_L - x_n|}{2})$ such that

$$\frac{f(x_n + h) - f(x_n)}{h} \leq f'(x_n) + \epsilon < 0, \qquad \frac{f(x_L + h) - f(x_L)}{h} \geq f'(x_L) - \epsilon > 0. \tag{3.57}$$

By considering the test function

$$\varphi = \begin{cases} \varphi_h(\cdot; (x_n + h, x_L)), & x_n < x_L \\ \varphi_h(\cdot; (x_L, x_n + h)), & x_n > x_L \end{cases} \tag{3.58}$$

and using (3.44) once again, we deduce that

$$\begin{aligned} \mathrm{TV}^{(2)}(f) &\geq h^{-1} \left| f(x_n) - f(x_n + h) + f(x_L + h) - f(x_L) \right| \\ &= \frac{f(x_L + h) - f(x_L)}{h} - \frac{f(x_n + h) - f(x_n)}{h} \\ &\geq f'(x_L) - \epsilon - f'(x_n) - \epsilon \geq L(f) - f'(x_n) - 3\epsilon > L(f), \end{aligned} \tag{3.59}$$

which contradicts the original assumption that (3.49) is saturated. For the case $f'(x_L) < 0$, the same arguments can be applied to x_p instead of x_n. This proves that f is monotone. In the following, we consider the case where f is an increasing function; the decreasing case can be deduced by symmetry.

Convexity/Concavity: We first consider the canonical decomposition $f = \mathrm{D}_\phi^{-2} w + p$, where $w = \mathrm{D}^2 f$, D_ϕ^{-2} is a right inverse of the second-order derivative, and $p(x) = ax + b$ is an affine term [103, Proposition 9]. We then use the Jordan decomposition of $w = \mathrm{D}^2 f$ as $w = (w_+ - w_-)$, where $w_+, w_- \in \mathcal{M}(\mathbb{R})$ are positive measures such that $\|w\|_{\mathcal{M}} = \|w_+\|_{\mathcal{M}} + \|w_-\|_{\mathcal{M}}$. This allows us to form the decomposition $f = (f_+ - f_-)$, where $f_s = \mathrm{D}_\phi^{-2} w_s + p_s, \quad s \in \{+,-\}$, $p_+(x) = (A+a)x + b$, and $p_-(x) = Ax$ with $A > 0$ being a sufficiently large constant such that the functions f_+ and f_- are both convex and strictly increasing. Hence, they both satisfy the sufficient conditions for saturation, which implies that $\mathrm{TV}^{(2)}(f_s) = (L(f_s) - \ell(f_s))$ for $s \in \{+,-\}$.

Assume by contradiction that $w_s \neq 0$ for $s \in \{+,-\}$ and let $\epsilon < \frac{\min(\mathrm{TV}^{(2)}(f_+), \mathrm{TV}^{(2)}(f_-))}{2}$ be a small positive constant and let $\bar{x}, \underline{x} \in \mathbb{R}$ such that $f'(\bar{x}) \geq (L(f) - \epsilon)$ and $f'(\underline{x}) \leq (\ell(f) + \epsilon)$. Using these inequalities, we deduce that

$$\mathrm{TV}^{(2)}(f) = L(f) - \ell(f) \leq f'(\bar{x}) - f'(\underline{x}) + 2\epsilon = \big(f_+'(\bar{x}) - f_-'(\bar{x})\big) - \big(f_+'(\underline{x}) - f_-'(\underline{x})\big) + 2\epsilon = \tag{3.60}$$

where $A_s = (f_s'(\bar{x}) - f_s'(\underline{x}))$ for $s \in \{+,-\}$. We now consider two cases:

Case I: $\bar{x} > \underline{x}$**.** The convexity of f_- implies that $A_- \geq 0$. Moreover, we have that $A_+ = \big(f_+'(\bar{x}) - f_+'(\underline{x})\big) \leq (L(f_+) - \ell(f_+)) = \mathrm{TV}^{(2)}(f_+)$. Using (3.60), this yields that $\mathrm{TV}^{(2)}(f) \leq \mathrm{TV}^{(2)}(f_+) + 2\epsilon$, which can be rewritten as $2\epsilon \geq \mathrm{TV}^{(2)}(f_-)$. However, our original choice of ϵ implies that $\epsilon < \mathrm{TV}^{(2)}(f_-)/2$ which is a contradiction and, hence, $w_- = 0$ or $w_+ = 0$, which implies that f is either convex or concave.

Case II: $\bar{x} \leq \underline{x}$**.** Following a similar treatment, we deduce that $A_+ \leq 0$ and $-A_- \leq \mathrm{TV}^{(2)}(f_-)$. Hence, we obtain that $2\epsilon \geq \mathrm{TV}^{(2)}(f_+)$ which is in contradiction with $\epsilon < \mathrm{TV}^{(2)}(f_+)/2$. □

Lipschitz Regularization

We first consider the Lipschitz constant as a regularizer and study the minimization problem

$$\mathcal{V}_{\mathrm{Lip}} = \underset{f \in \mathrm{Lip}(\mathbb{R})}{\arg\min} \left(\sum_{m=1}^{M} E(f(x_m), y_m) + \lambda L(f) \right), \tag{3.61}$$

where $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a strictly convex and coercive function and where $\lambda > 0$ is the regularization parameter. We also assume, without loss of generality, that the data points x_m are sorted in the increasing order $x_1 < x_2 < \cdots < x_M$. In Theorem 3.6, we state our main theoretical contributions regarding the minimization problem (3.61).

Theorem 3.6. *Regarding the minimization problem* (3.61)*, the following statements hold.*

1. *The solution set $\mathcal{V}_{\text{Lip}}$ is a nonempty, convex and weak*-compact subset of* $\text{Lip}(\mathbb{R})$*.*
2. *There exists a unique vector $\mathbf{z} = (z_m) \in \mathbb{R}^M$ such that*
$$\mathcal{V}_{\text{Lip}} = \operatorname*{arg\,min}_{f \in \text{Lip}(\mathbb{R})} L(f), \quad s.t. \quad f(x_m) = z_m, \quad \forall m. \tag{3.62}$$
3. *The optimal Lipschitz constant has the closed-form expression*
$$L_{\min} = \max_{2 \le m \le M} \left| \frac{z_m - z_{m-1}}{x_m - x_{m-1}} \right|. \tag{3.63}$$
Consequently, any $L_{\min}$-Lipschitz function f that satisfies $f(x_m) = z_m, m = 1, \ldots, M$ is a solution of (3.61)*.*
4. *Let $\mathcal{E} \subseteq \mathbb{R}^2$ be the union of the graphs of all solutions of* (3.61)*, defined as*
$$\mathcal{E} = \left\{ (x, y) \in \mathbb{R}^2 : \exists f \in \mathcal{V}_{\text{Lip}}, y = f(x) \right\}. \tag{3.64}$$
Let us also define the right and left planar cones $\mathcal{R}, \mathcal{L} \subseteq \mathbb{R}^2$ as
$$\mathcal{R} = \{ \alpha_1(1, L_{\min}) + \alpha_2(1, -L_{\min}) : \alpha_1, \alpha_2 \ge 0 \}, \tag{3.65}$$
and $\mathcal{L} = -\mathcal{R}$. With the convention that $\mathcal{R}_0 = \mathcal{L}_{M+1} = \mathbb{R}^2$, we have that
$$\mathcal{E} = \bigcup_{m=1}^{M+1} \left(\mathcal{R}_{m-1} \cap \mathcal{L}_m \right), \tag{3.66}$$
where the $\mathcal{R}_m$ and $\mathcal{L}_m$ are shifted versions of $\mathcal{R}$ and $\mathcal{L}$, with
$$\mathcal{R}_m = (x_m, z_m) + \mathcal{R}, \quad \mathcal{L}_m = (x_m, z_m) + \mathcal{L}, \forall m. \tag{3.67}$$

5. *Any solution of the constrained minimization problem*

$$\min_{f\in \mathrm{BV}^{(2)}(\mathbb{R})} \mathrm{TV}^{(2)}(f), \quad \textit{s.t.} \quad f(x_m)=z_m, 1\leq m\leq M \tag{3.68}$$

is included in $\mathcal{V}_{\mathrm{Lip}}$. In particular, the solution set of (3.61) *always includes a continuous and piecewise-linear function.*

Proof. **Items 1 and 2:** The first step is to show that the sampling functional $\delta(\cdot - x_0) : f \mapsto f(x_0)$ is weak*-continuous in $\mathrm{Lip}(\mathbb{R})$. To that end, we identify the predual Banach space $\mathcal{X}$ such that $\mathrm{Lip}(\mathbb{R}) = \mathcal{X}'$ and then show that shifted Dirac impulses are included in $\mathcal{X}$, which is equivalent to weak*-continuity. We recall that following (3.43), we can view $\mathrm{Lip}(\mathbb{R})$ as the native Banach space associated to the pair $(L_\infty(\mathbb{R}), \mathrm{D})$. This allows us to deploy the machinery of [104] to identify its predual space. In short, it follows from [104] that the predual space has the direct-sum structure $\mathcal{X} = \mathrm{D}\left(L_1(\mathbb{R})\right) \oplus \mathrm{span}\left(\mathrm{e}^{-(\cdot)^2}\right)$. In other words, any function $f \in \mathcal{X}$ can be decomposed as $f = \mathrm{D}\{g\} + c\mathrm{e}^{-(\cdot)^2}$, where $g \in L_1(\mathbb{R})$ and $c \in \mathbb{R}$. One can formally verify that $\delta = \mathrm{D}\{\mathrm{sgn} - \mathrm{erf}\} + \frac{2}{\sqrt{\pi}}\mathrm{e}^{-(\cdot)^2}$, where sgn is the sign function and erf is the Gauss error function. Due to the rapid decay of the erf function at $t = -\infty$ and the symmetry of $(\mathrm{sgn} - \mathrm{erf})$, we deduce that $\mathrm{sgn} - \mathrm{erf} \in L_1(\mathbb{R})$ and, hence, that $\delta \in \mathcal{X}$. Finally, due to the shift-invariant structure of $\mathcal{X}$, we deduce the weak*-continuity of the sampling functional $\delta(\cdot - x_0)$ for any $x_0 \in \mathbb{R}$.

We now invoke Theorem 2.5 to deduce that the solution set $\mathcal{V}_{\mathrm{Lip}}$ of (3.61) is a nonempty, convex, weak*-compact set whose elements all pass through a fixed set of points. Put differently, the vector $\mathbf{z} = (z_m)$ with $z_m = f(x_m)$ is invariant to the choice of $f \in \mathcal{V}_{\mathrm{Lip}}$. Consequently, we can represent $\mathcal{V}_{\mathrm{Lip}}$ as a solution set of a constrained problem of the form (3.62).

Item 3: Let us first define the canonical CPWL interpolant of a collection of 1D data points.

Definition 3.4. *For a series of data points $(x_m, z_m), m = 1, \dots, M$, the canonical interpolant $f_{\mathrm{cano}} : \mathbb{R} \to \mathbb{R}$ is the unique CPWL function that passes through these points and is differentiable over $\mathbb{R}\backslash\{x_2, \dots, x_{M-1}\}$.*

We first prove that f_{cano} is a solution of (3.62). Clearly, the Lipschitz constant of f_{cano} is equal to $L(f_{\mathrm{cano}}) = L_{\min}$, where $L_{\min}$ is given in (3.63). Moreover, any function f that passes through the data points (x_m, z_m) necessarily has a Lipschitz constant greater than or equal to $L_{\min}$. This implies that f_{cano} is a solution of (3.62) and $L_{\min}$ is the minimal value of the Lipschitz constant. Consequently, any function that satisfies the interpolation constraints and is $L_{\min}$-Lipschitz is a solution of (3.62).

Item 4: Consider a generic point $(x, y) \in \mathcal{E}$, and let m be such that $x \in (x_{m-1}, x_m)$. By definition of $\mathcal{E}$, there exists a function $f \in \mathcal{V}_{\mathrm{Lip}}$ such that $y = f(x)$. From Item 3, we deduce that $L(f) = L_{\min}$. Hence, we have the inequalities

$$\left|\frac{y - z_{m-1}}{x - x_{m-1}}\right|, \left|\frac{y - z_m}{x - x_m}\right| \leq L_{\min}. \tag{3.69}$$

These inequalities can readily be translated into the inclusion $(x, y) \in \mathcal{R}_{m-1} \cap \mathcal{L}_m$, which implies that $\mathcal{E} \subseteq \bigcup_{m=1}^{M} (\mathcal{R}_{m-1} \cap \mathcal{L}_m)$. To show the reverse inclusion, consider a point in $(x, y) \in \mathcal{R}_{m-1} \cap \mathcal{L}_m$ for some $m \in \{1, \dots, M+1\}$ and denote by $\tilde{f}_{\mathrm{cano}}$ the canonical interpolant of $\{(x_m, z_m)\}_{m=1}^{M} \cup \{(x, y)\}$. Following Item 3, the Lipschitz constant of $\tilde{f}_{\mathrm{cano}}$ is given by

$$L(\tilde{f}_{\mathrm{cano}}) = \max\left(L_{\min}, \left|\frac{y - z_{m-1}}{x - x_{m-1}}\right|, \left|\frac{y - z_m}{x - x_m}\right|\right) = L_{\min}, \tag{3.70}$$

where we establish the last equality by translating the inclusion $(x, y) \in \mathcal{R}_{m-1} \cap \mathcal{L}_m$ into the inequalities in (3.69). This implies that $\tilde{f}_{\mathrm{cano}}$ is a solution of (3.62) and so, by definition, we have that $(x, y) \in \mathcal{E}$.

Item 5: By [194, Proposition 5], f_{cano} is also a solution of (3.68). We therefore need to prove that any solution f_{opt} of (3.68) has the same Lipschitz constant $L(f_{\mathrm{opt}}) = L(f_{\mathrm{cano}}) = L_{\min}$. Due to the interpolation constraints, we necessarily have that $L(f_{\mathrm{opt}}) \geq L(f_{\mathrm{cano}})$; we must now prove the reverse inequality $L(f_{\mathrm{opt}}) \leq L(f_{\mathrm{cano}})$. By [194, Theorem 2], f_{opt} must follow f_{cano} in $\mathbb{R} \backslash [x_2, x_{M-1}]$. Moreover, in each interval $[x_m, x_{m+1}]$ for $m \in \{2, \dots, M-2\}$, f_{opt} either follows f_{cano} or is concave or convex over the interval $[x_{m-1}, x_{m+2}]$. Hence, it suffices to prove that, for any $m \in \{2, \dots, M-2\}$, we have that $L_m(f_{\mathrm{opt}}) \leq L(f_{\mathrm{cano}})$, where $L_m(f)$ denotes the Lipschitz constant of f restricted to the interval $[x_m, x_{m+1}]$.

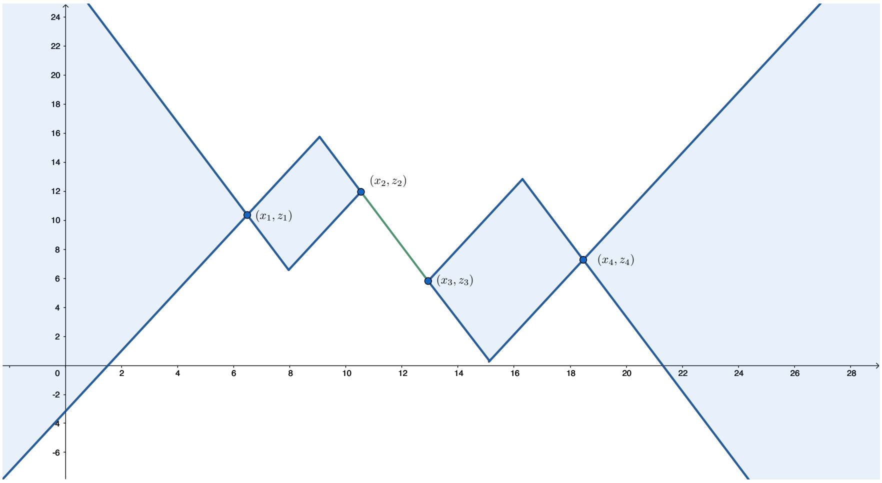


Figure 3.8: The union of the graphs of all solutions in a simple example with four data points. Note that all solutions must directly connect (x_2, z_2) to (x_3, z_3), since the slope of this segment is $L_{\min}$ whose formula is given in (3.63).

Let m be an index for which f_{opt} need not follow f_{opt} in $[x_m, x_{m+1}]$. (If no such index exists, then the result is trivially true.) Assume that f_{opt} is convex in the interval $[x_{m-1}, x_{m+2}]$; the concave scenario is derived in a similar fashion. This implies that, in this interval, the function $(\tilde{x}_1, \tilde{x}_2) \mapsto \frac{f_{\text{opt}}(\tilde{x}_2) - f_{\text{opt}}(\tilde{x}_1)}{\tilde{x}_2 - \tilde{x}_1}$ is increasing in both its variables.

Hence, for any $\tilde{x}_1, \tilde{x}_2 \in [x_m, x_{m+1}]$ with $\tilde{x}_1 \neq \tilde{x}_2$, we have that $\frac{z_m - z_{m-1}}{x_m - x_{m-1}} \leq \frac{f_{\text{opt}}(\tilde{x}_2) - f_{\text{opt}}(\tilde{x}_1)}{\tilde{x}_2 - \tilde{x}_1} \leq \frac{z_{m+2} - z_{m+1}}{x_{m+2} - x_{m+1}}$. This directly implies the desired result $L_m(f_{\text{opt}}) \leq L(f_{\text{cano}})$. □

We remark that the result that has the greatest practical relevance is stated in Item 5 which creates an interesting link with $\text{TV}^{(2)}$ minimization problems and hence guarantees the existence of CPWL solutions.

Lipschitz Constraint

While the first formulation is interesting on its own right and results in learning CPWL mappings with tunable Lipschitz constants, it does not necessarily yield a sparse (and, hence, interpretable) solution. In fact, the learned mapping can have undesirable oscillations as illustrated in Figure 3.11. This observation motivates us to propose a second formulation that combines $\mathrm{TV}^{(2)}$ regularization with a constraint over the Lipschitz constant, as expressed by

$$\mathcal{V}_{\mathrm{hyb}} = \underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min} \left(\sum_{m=1}^{M} E(f(x_m), y_m) + \lambda \mathrm{TV}^{(2)}(f) \right), \quad \text{s.t.} \quad L(f) \leq \bar{L}. \tag{3.71}$$

The quantity $\bar{L}$ is the maximal value allowed for the Lipschitz constant of the learned mapping. In this way, the stability is directly controlled by the user, while the regularization term removes undesired oscillations (tunable with $\lambda > 0$). The solution set $\mathcal{V}_{\mathrm{hyb}}$ is characterized in Theorem 3.7, from which we also deduce the existence of CPWL solutions.

Theorem 3.7. *The solution set* $\mathcal{V}_{\mathrm{hyb}}$ *of Problem* (3.71) *is a nonempty, convex, and weak*-compact subset of* $\mathrm{BV}^{(2)}(\mathbb{R})$ *whose extreme points are linear splines with at most* $(M-1)$ *linear regions. Moreover, there exists a unique vector* $\mathbf{z} = (z_m)$ *such that*

$$\mathcal{V}_{\mathrm{hyb}} = \underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min} \ \mathrm{TV}^{(2)}(f), \quad s.t. \quad f(x_m) = z_m, 1 \leq m \leq M. \tag{3.72}$$

Finally, the optimal $\mathrm{TV}^{(2)}$ *cost has the closed-form expression*

$$\mathrm{TV}_{\min} = \sum_{m=2}^{M-1} \left| \frac{z_m - z_{m-1}}{x_m - x_{m-1}} - \frac{z_m - z_{m+1}}{x_m - x_{m+1}} \right|. \tag{3.73}$$

Proof. **Existence:** We rewrite the problem in (3.71) as an unconstrained minimization

$$\mathcal{V}_{\text{hyb}} = \underset{f \in \mathcal{M}_{\mathrm{D}^2}(\mathbb{R})}{\arg\min} \sum_{m=1}^{M} E(f(x_m), y_m) + \lambda \mathrm{TV}^{(2)}(f) + i_{L(f) \le \bar{L}}, \tag{3.74}$$

where i_A denotes the characteristic function of the set A and is defined as

$$i_A(f) = \begin{cases} 0, & f \in A \\ +\infty, & \text{otherwise}. \end{cases} \tag{3.75}$$

To prove the existence of a minimizer, we use a standard technique in convex analysis which involves the generalized Weierstrass theorem [198] to show that the cost functional of (3.74) is coercive and lower semicontinuous (in the weak*-topology), which is a sufficient condition for the existence of a solution.

The cost functional in (3.71) consists of three terms: (i) an empirical loss term $H(f) = \sum_{m=1}^{M} E(f(x_m), y_m)$; (ii) a second-order total-variation regularization term $R(f) = \lambda \mathrm{TV}^{(2)}(f)$; and (iii) a Lipschitz constraint i_A, where $A = \{L(f) \le \bar{L}\}$. It is known (see [74] for a more general statement) that the functional $H(f) + R(f)$ is coercive and weak*-lowersemincontinuous. This, together with the non-negativity of i_E, yields the coercivity of the total cost. The only missing item is the weak*-lowersemicontinuity of i_A, for which it is sufficient to prove that A is a closed set for the weak*-topology.

Let $f_n \in \mathrm{BV}^{(2)}(\mathbb{R})$ be a sequence of functions with $L(f_n) \le \bar{L}$ converging in the weak*-topology to $f_{\lim} \in \mathrm{BV}^{(2)}(\mathbb{R})$. To prove the weak*-closedness of A, we need to show that $L(f_{\lim}) \le \bar{L}$, which is equivalent to $|f_{\lim}(a) - f_{\lim}(b)| \le \bar{L}|a - b|$ for any $a, b \in \mathbb{R}$.

For any $n \in \mathbb{N}$, we have that

$$|f_{\lim}(a) - f_{\lim}(b)| \le |f_{\lim}(a) - f_n(a)| + |f_n(a) - f_n(b)| + |f_n(b) - f_{\lim}(b)|. \tag{3.76}$$

Using the weak*-continuity of the sampling functionals $\delta(\cdot - a)$ and $\delta(\cdot - b)$ in $\mathrm{BV}^{(2)}(\mathbb{R})$ (see, for example, [103]), we deduce that $f_n(a) \to f_{\lim}(a)$ and $f_n(b) \to f_{\lim}(b)$. Moreover, we have the estimate $|f_n(a) - f_n(b)| \le \bar{L}|a - b|$ for any $n \in \mathbb{N}$. Using these and letting the limit $n \to +\infty$ in (3.76), we get the desired bound.

Form of the Solution Set: Now that we have proved the existence of a solution $f_0^* \in \mathcal{V}_{\mathrm{hyb}}$, we can apply a standard argument based on the strict convexity of $E(\cdot,\cdot)$ (see, for example, [199, Lemma 1]) to deduce that for any $f^* \in \mathcal{V}_{\mathrm{hyb}}$, we have that $f^*(x_m) = f_0^*(x_m)$ for $m = 1, \dots, M$. Hence, the original Problem (3.71) is equivalent to

$$\mathcal{V}_{\mathrm{hyb}} = \underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min}\ \mathrm{TV}^{(2)}(f), \quad \text{s.t.} \quad \begin{cases} L(f) \leq \bar{L}, \\ f(x_m) = f_0^*(x_m), \quad m = 1, \dots, M. \end{cases} \tag{3.77}$$

Since $f_0^* \in \mathcal{V}_{\mathrm{hyb}}$, we deduce that

$$L_0 \triangleq \max_{2 \leq m \leq M} \left| \frac{f_0^*(x_m) - f_0^*(x_{m-1})}{x_m - x_{m-1}} \right| \leq L(f_0^*) \leq \bar{L}.$$

Yet, Item 5 in Theorem 3.6 implies that any solution f^* of the problem

$$\underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min}\ \mathrm{TV}^{(2)}(f), \quad \text{s.t.} \quad f(x_m) = f_0^*(x_m),\ m = 1, \dots, M, \tag{3.78}$$

is a solution of (3.62) with $z_m = f_0^*(x_m)$. Hence, by Item 3 of Theorem 3.6, we have that $L(f^*) = L_0 \leq \bar{L}$. This means that adding the Lipschitz constraint $L(f) \leq \bar{L}$ does not change the solution set of Problem (3.78). Hence, we have that

$$\mathcal{V}_{\mathrm{hyb}} = \underset{f \in \mathrm{BV}^{(2)}(\mathbb{R})}{\arg\min}\ \mathrm{TV}^{(2)}(f), \quad \text{s.t.} \quad f(x_m) = f_0^*(x_m),\ m = 1, \dots, M. \tag{3.79}$$

The solution set of (3.79) has been fully described in [194], which yields the announced characterization. □

Let us remark that the Lipschitz constraint only affects the vector $\mathbf{z}$ in (3.72), which forces its entries to satisfy the inequalities

$$\left| \frac{z_m - z_{m-1}}{x_m - x_{m-1}} \right| \leq \overline{L}, \quad m = 2, \dots, M. \tag{3.80}$$

Connection to Neural Networks

In this part, we show that our second formulation (3.71) is equivalent to training a two-layer neural network with weight decay and a Lipschitz constraint. Let us recall that a univariate ReLU network with two layers and skip connections is a mapping $f_{\boldsymbol{\theta}} : \mathbb{R} \to \mathbb{R}$ of the form

$$f_{\boldsymbol{\theta}}(x) = c_0 + c_1 x + \sum_{k=1}^{K} v_k \mathrm{ReLU}(w_k x - b_k), \tag{3.81}$$

where $c_1 \in \mathbb{R}$ is the weight of the skip connection, $K \in \mathbb{N}$ is the width of the network, $v_k, w_k \in \mathbb{R}, k = 1, \ldots, K$ are the linear weights and $b_k \in \mathbb{R}$, $k = 1, \ldots, K$ and $c_0 \in \mathbb{R}$ are the bias terms of the first and second layers, respectively. These parameters are concatenated in a single vector $\boldsymbol{\theta} = (K, \mathbf{v}, \mathbf{w}, \mathbf{b}, \mathbf{c})$, and we denote by $\boldsymbol{\Theta}$ the set of all possible parameter vectors $\boldsymbol{\theta}$. Thus, the training problem with Lipschitz constraint and weight decay is formulated as

$$\mathcal{V}_{NN} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\sum_{m=1}^{M} \mathrm{E}(f_{\boldsymbol{\theta}}(x_m), y_m) + \lambda \mathrm{R}(\boldsymbol{\theta}) \right), \quad \text{s.t.} \quad L(f_{\boldsymbol{\theta}}) \leq \bar{L}, \tag{3.82}$$

where $\mathrm{R}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \left(\frac{|v_k|^2 + |w_k|^2}{2} \right)$ is the regularization term corresponding to weight decay. In Proposition 3.1, we show the equivalence between this training problem and our Lipschitz-constrained formulation (3.71).

Proposition 3.1. *For any solution $\boldsymbol{\theta}^*$ of* (3.82)*, $f_{\boldsymbol{\theta}^*}$ is a CPWL solution of* (3.71)*. Moreover, any CPWL solution of* (3.71) *can be expressed as a two-layer ReLU network $f_{\boldsymbol{\theta}^*}$ with skip connections whose parameter vector is optimal in the sense of* (3.82)*,* i.e., $\boldsymbol{\theta}^* \in \mathcal{V}_{NN}$.

Before going to the proof, we first state a useful lemma.

Lemma 3.1. *For any $\boldsymbol{\theta}^* = (K^*, \mathbf{v}^*, \mathbf{w}^*, \mathbf{b}^*, \mathbf{c}^*) \in \mathcal{V}_{\mathrm{NN}}$, we have that $|v_k^*| = |w_k^*|$ for any $k = 1, \ldots, K$.*

Proof. Let $\boldsymbol{\theta}^* = (K^*, \mathbf{v}^*, \mathbf{w}^*, \mathbf{b}^*, \mathbf{c}^*) \in \mathcal{V}_{NN}$ and $1 \leq k \leq K$. For any $\epsilon \in (-1, 1)$, we define a perturbed parameter vector $\boldsymbol{\theta}_\epsilon = (K^*, \mathbf{v}_\epsilon, \mathbf{w}_\epsilon, \mathbf{b}_\epsilon, \mathbf{c}^*)$, where for any

$k' = 1, \dots, K$ we have that

$$v_{\epsilon,k'} = \begin{cases} v^*_{k'}, & k' \neq k \\ (1+\epsilon)^{\frac{1}{2}} v^*_k, & k' = k \end{cases}, \tag{3.83}$$

$$w_{\epsilon,k'} = \begin{cases} w^*_{k'}, & k' \neq k \\ (1+\epsilon)^{-\frac{1}{2}} w^*_k, & k' = k \end{cases}, \tag{3.84}$$

$$b_{\epsilon,k'} = \begin{cases} b^*_{k'}, & k' \neq k \\ (1+\epsilon)^{-\frac{1}{2}} b^*_k, & k' = k. \end{cases} \tag{3.85}$$

Due to the positive homogeneity of the ReLU, one readily deduces from (3.81) that $f_{\boldsymbol{\theta}^*} = f_{\boldsymbol{\theta}_\epsilon}$ for any $\epsilon \in (-1, 1)$. This together with the optimality of $\boldsymbol{\theta}^*$ in Problem (3.82) implies that

$$v^{*2}_k + w^{*2}_k \leq (1+\epsilon) v^{*2}_k + (1+\epsilon)^{-1} w^{*2}_k, \quad \forall \epsilon \in (-1, 1). \tag{3.86}$$

Multiplying both sides of the above inequality by $(1+\epsilon) > 0$ yields

$$\epsilon w^{*2}_k \leq \epsilon (1+\epsilon) v^{*2}_k, \quad \forall \epsilon \in (-1, 1). \tag{3.87}$$

Letting $\epsilon \to 0^+$ yields $w^{*2}_k \leq v^{*2}_k$ and $\epsilon \to 0^-$ yields $w^{*2}_k \geq v^{*2}_k$, which proves that $|w^*_k| = |v^*_k|$. □

Proof of Proposition 3.1. Using Lemma 3.1, we observe that for any $\boldsymbol{\theta}^* \in \mathcal{V}_{\mathrm{NN}}$, we have that

$$\mathrm{R}(\boldsymbol{\theta}^*) = \frac{1}{2} \sum_{k=1}^{K} (v^{*2}_k + w^{*2}_k) = \sum_{k=1}^{K} |v^*_k||w^*_k| = \mathrm{TV}^{(2)}(f_{\boldsymbol{\theta}^*}), \tag{3.88}$$

where the last inequality comes from the simple observation that $\mathrm{TV}^{(2)}(v\mathrm{ReLU}(w \cdot -b)) = |v||w|$ for any $v, w, b \in \mathbb{R}$. Hence, one can rewrite the solution set $\mathcal{V}_{\mathrm{NN}}$ as

$$\mathcal{V}_{\mathrm{NN}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathrm{red}}}{\arg\min} \left(\sum_{m=1}^{M} \mathrm{E}(f_{\boldsymbol{\theta}}(x_m), y_m) + \lambda \mathrm{TV}^{(2)}(f_{\boldsymbol{\theta}}) \right), \quad \text{s.t.} \quad L(f_{\boldsymbol{\theta}}) \leq \bar{L}, \tag{3.89}$$

where $\mathbf{\Theta}_{\text{red}} = \{\boldsymbol{\theta} \in \mathbf{\Theta} : \mathrm{R}(\boldsymbol{\theta}) = \mathrm{TV}^{(2)}(f_{\boldsymbol{\theta}})\}$ is the reduced parameter space. To prove the announced equivalence, it remains to show that the mapping $\mathbf{\Theta}_{\text{red}} \to \mathrm{BV}^{(2)}(\mathbb{R}) : \boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}$ is a bijection onto the CPWL members of $\mathrm{BV}^{(2)}(\mathbb{R})$ with finitely many linear regions.

For any $\boldsymbol{\theta} \in \mathbf{\Theta}_{\text{red}}$, the function $f_{\boldsymbol{\theta}}$ is a CPWL member of $\mathrm{BV}^{(2)}(\mathbb{R})$ with finitely many linear regions. To prove the converse, let $f \in \mathrm{BV}^{(2)}(\mathbb{R})$ be a CPWL function with finitely many linear regions. Using the canonical representation of f, there exist $c_0, c_1 \in \mathbb{R}$, $K \in \mathbb{N}$ and $a_k, \tau_k \in \mathbb{R}$ with $a_k \neq 0$ for $k = 1, \dots, K$ such that

$$f(x) = c_0 + c_1 x + \sum_{k=1}^{K} a_k \mathrm{ReLU}(x - \tau_k). \tag{3.90}$$

Now by defining $v_k = \frac{a_k}{\sqrt{|a_k|}}$, $w_k = \sqrt{|a_k|}$ and, $b_k = \sqrt{|a_k|}\tau_k$ for $k = 1, \dots, K$, the homogeneity of the ReLU yields $f = f_{\boldsymbol{\theta}}$ with $\theta = (K, \mathbf{c}, \mathbf{v}, \mathbf{w}, \mathbf{b}) \in \mathbf{\Theta}_{\text{red}}$, where the latter inclusion is due to the equalities $|v_k| = |w_k|$ for $k = 1, \dots, K$. □

Proposition 3.1 is an extension of the results of [195, 140], where this equivalence is proved in the absence of a Lipschitz constraint. These works rely on a result (*e.g.*, [195, Corollary C.2]) that describes the energy propagation in the training of feed-forward neural networks with weight decay, which can easily be extended to the Lipschitz-constrained case (Lemma 3.1). Proposition 3.1 provides a functional framework to study the training of Lipschitz-aware neural networks, which is a nontrivial task. To this end, Proposition 3.1 allows us to deploy our proposed algorithm. Our description of the solution set of Problem (3.71) (Theorem 3.7) and of its sparsest solutions (Theorem 3.8) also provides interesting insights on ReLU neural networks.

3.3.4 Finding the Sparsest CPWL Solution

Here, we propose an algorithm to find the sparsest CPWL solution of Problems (3.61) and (3.71). To that end, we first compute the vector $\mathbf{z}$ of the value of the optimal function at the data points $x_1, \dots, x_m$. Using this vector, we then deploy the sparsification algorithm of [194], whose use in the present method is motivated by the following theorem.

Theorem 3.8. *Let $(x_m, z_m) \in \mathbb{R}^2, m = 1, \dots, M$ be a collection of ordered data points with $x_1 < \cdots < x_M$. Then, the output f_{sparse} of the sparsification algorithm of Debarre et al. in [194] is the sparsest linear-spline interpolator of the data points. In other words, f_{sparse} is the CPWL interpolator with the fewest number of linear regions.*

Proof. Let f^* be the output of [194, Algorithm 1]. It is thus a CPWL solution of Problem (3.62) with the minimum number of linear regions. We prove that *any* CPWL interpolant f of the data points $\mathrm{P}_m = (x_m, z_m), m = 1, \dots, M$—not necessarily a minimizer of $\mathrm{TV}^{(2)}(f)$—has at least as many linear regions as f^*. Our proof is based on induction over the number M of data points. The initialization $M = 2$ trivially holds, since f^* then has a single linear region—it is simply the line connecting the two data points. Next, let $M > 2$ and assume that Theorem 3.8 holds for $(M-1)$ or less data points (the induction hypothesis). The canonical interpolatant f_{cano} introduced in Definition 3.4 can be expressed as

$$f_{\mathrm{cano}}(x) = \alpha_1 x + \alpha_2 + \sum_{m=2}^{M-1} a_m (x - x_m)_+ \tag{3.91}$$

for some coefficients $\alpha_1, \alpha_2, a_m \in \mathbb{R}$. There are three possible scenarios:

1. all a_m's are positive (or negative);
2. at least one of them is zero;
3. there are two consecutive coefficients with opposite signs, so that $a_m a_{m+1} < 0$ for some m.

We analyze each case separately and use the induction hypothesis to deduce the desired result. In this proof, we refer to singularities of CPWL functions (*i.e.,* the boundary points between linear regions) as *knots.*

Case 1: In this case, it is known that f^* has $K = \big(\lceil \frac{M}{2} \rceil - 1\big)$ knots [194, Theorem 4]. Assume by contradiction that there exists a CPWL interpolant f with fewer knots and consider the K disjoint intervals (x_{2k-1}, x_{2k+1}) for $1 \leq k \leq \big(\lceil \frac{M}{2} \rceil - 1\big) = K$. We deduce that there exists an interval (x_{2k-1}, x_{2k+1}) in which f has no knots. This in turn implies that the data points P_{2k-1}, P_{2k}, and P_{2k+1} are aligned, and so that $a_{2k} = 0$, which yields a contradiction.

Case 2: Let $m \in \{2, M-1\}$ be such that $a_m = 0$. Consider the collection of $m < M$ data points $(\mathrm{P}_{m'})_{1 \leq m' \leq m}$; by the induction hypothesis, f^* interpolates them with the minimal number K_1 of knots. The same applies to the collection of $(M - m + 1) < M$ points $(\mathrm{P}_{m'})_{m \leq m' \leq M}$ with K_2 knots. Let f be a CPWL interpolant of all the M data points with the minimal number of knots. By definition of the K_i, f must have at least K_1 knots in the interval (x_1, x_m) and K_2 knots in the interval (x_m, x_M). Since these intervals are disjoint, f must have at least $K_1 + K_2$ knots in total. Yet, f^* has exactly $(K_1 + K_2)$ knots: indeed, f^* follows f_{cano} in the interval $[x_{m-1}, x_{m+1}]$, which has no knot at x_m since $a_m = 0$ (the points P_{m-1}, P_m, and P_{m+1} are aligned). This concludes that f^* has the minimum number of knots.

Case 3: Let $m \in \{2, M-2\}$ be such that $a_m a_{m+1} < 0$. Consider the collection of $(m+1) < M$ data points $(\mathrm{P}_{m'})_{1 \leq m' \leq m+1}$; by the induction hypothesis, f^* interpolates them with the minimal number K_1 of knots. Similarly, f^* interpolates the $(M - m + 1) < M$ points $(\mathrm{P}_{m'})_{m \leq m' \leq M}$ with the minimal number K_2 of knots. Let f be a CPWL interpolant of all the M data points with the minimal number of knots. We now state a useful lemma whose proof is given below.

Lemma 3.2. *Let $m \in \{2, \dots, M-2\}$ be such that $a_m a_{m+1} < 0$. Then, any CPWL interpolant f of the data points $(\mathrm{P}_{m'})_{1 \leq m' \leq M}$ can be modified to become another CPWL interpolant $\tilde{f}$ with as many (or fewer) knots such that $\tilde{f}$ has no knot in the interval (x_m, x_{m+1}).*

By Lemma 3.2, it can be modified to become another interpolant $\tilde{f}$ with the same total number of knots and none in the interval (x_m, x_{m+1}). By definition of the K_i, $\tilde{f}$ must have at least K_1 knots in the interval (x_1, x_{m+1}) and K_2 knots in the interval (x_m, x_M). Yet, $\tilde{f}$ has no knots in the interval (x_m, x_{m+1}), so it must have at least K_1 knots in $(x_1, x_m]$ and K_2 knots in $[x_{m+1}, x_M)$. Since these intervals are disjoint, $\tilde{f}$ must have at least $(K_1 + K_2)$ knots in total. Yet, f^* follows f_{cano} in the interval $[x_{m-1}, x_{m+2}]$ and thus also has no knot in the interval (x_1, x_{m+1}). Therefore, by the induction hypothesis, f^* has K_1 knots in $(x_1, x_m]$ and K_2 knots in $[x_{m+1}, x_M)$, for a total of $(K_1 + K_2)$ knots. Since this is no more than $\tilde{f}$, f^* has the minimal number of knots, which proves the induction. □

Proof of Lemma 3.2. Let f be a CPWL interpolant of the data points $(\mathrm{P}_{m'})_{1 \leq m' \leq M}$ with P knots. In what follows, we consider a CPWL function $\tilde{f}$ that follows f outside this interval and (x_{m-1}, x_{m+2}), and we modify it inside this interval in order to remove all knots in (x_m, x_{m+1}) without increasing the total number of knots.

We consider the case $a_m > 0$ and $a_{m+1} < 0$ without loss of generality. Let $s^- = f'(x_{m-1}^-)$ and $s^+ = f'(x_{m+2}^+)$ be the slopes of f before and after the interval of interest (x_{m-1}, x_{m+2}), respectively, and we let $s_{\text{cano}}^- = f'_{\text{cano}}(x_{m-1}^-)$ and $s_{\text{cano}}^+ = f'_{\text{cano}}(x_{m+2}^+)$ be those of f_{cano}. We also introduce the linear functions $f^-(x) = z_{m-1} + s^-(x - x_{m-1})$ and $f^+(x) = z_{m+2} + s^+(x - x_{m+2})$. They prolong f in a straight line after P_{m-1} and before P_{m+2}, respectively. We now distinguish cases based on s^- and s^+.

Case I: $s^- \leq s_{\text{cano}}^-$ and $s^+ \leq s_{\text{cano}}^+$. Graphically, this corresponds to f lying in none of the gray regions in Figure 3.9. In this case, the line $(\mathrm{P}_m \mathrm{P}_{m+1})$ intersects the linear function f^- at some point $\mathrm{P}^- = (x^-, z^-)$ where $x^- \in (x_{m-1}, x_m)$, and with f^+ at some point $\mathrm{P}^+ = (x^+, z^+)$ with $x^+ \in (x_{m+1}, x_{m+2})$. This is obvious graphically (see Figure 3.9 as an illustration for P^-), and is due to the fact that $a_m > 0$ and $a_{m+1} < 0$. Hence, by taking an $\tilde{f}$ that connects the points P_{m-1}, P^-, P^+, and P_{m+2}, then $\tilde{f}$ has two knots in $[x_{m-1}, x_{m+2}]$ and its knots satisfy $x^-, x^+ \notin (x_m, x_{m+1})$. Since f clearly cannot have fewer than two knots in this interval, this proves the desired result.

Case II: $s^+ > s_{\text{cano}}^+$ and $s^- > s_{\text{cano}}^-$. In this case, f lies in both gray regions in Figure 3.9. To pass through P_m, f must have at least one knot in $[x_{m-1}, x_m)$; let $\mathrm{P}^- = (x^-, z^-)$ be the first of those knots (with $x^- < x_m$). Similarly, to pass through P_{m+1}, f must have a knot in $(x_{m+1}, x_{m+2}]$; let $\mathrm{P}^+ = (x^+, z^+)$ be the last of those knots (with $x^+ > x_{m+1}$). Then, f must pass through the points P^-, P_m, P_{m+1}, P^+. Yet, the lines $(\mathrm{P}^- \mathrm{P}_m)$ and $(\mathrm{P}_{m+1} \mathrm{P}^+)$ clearly cannot intersect in the interval $[x_m, x_{m+1}]$, which implies that at least two knots are needed in the interval (x^-, x^+). We conclude that f must have at least four knots in the interval $[x_{m-1}, x_{m+2}]$. Hence, we take an $\tilde{f}$ that simply connects the points P_{m-1}, P_m, P_{m+1}, and P_{m+2} and follows f elsewhere; the latter has four knots in $[x_{m-1}, x_{m+2}]$, which is no more than f and thus fulfills the requirements of the proof.

Case III: $s^+ > s^+_{\text{cano}}$ **and** $s^- \leq s^-_{\text{cano}}$. This case is illustrated in Figure 3.9: f is outside the gray region on the left, and inside the one on the right. With a similar argument as in Case II, f must have a least three knots in the interval $[x_{m-1}, x_{m+2}]$. The fact that $a_m > 0$ implies that the line $(\mathrm{P}_m\mathrm{P}_{m+1})$ intersects the linear function f^- at some point $\mathrm{P}^- = (x^-, z^-)$ where $x^- \in (x_{m-1}, x_m)$. We then take an $\tilde{f}$ that connects the points P_{m-1}, P^-, P_{m+1}, and P_{m+2} and follows f elsewhere. The interpolant $\tilde{f}$ has three knots at x^-, x_{m+1}, and x_{m+2} in $[x_{m-1}, x_{m+2}]$ and thus satisfies the requirements of the proof.

Case IV: $s^+ \leq s^+_{\text{cano}}$ **and** $s^- > s^-_{\text{cano}}$**.** This is similar to Case III, and can be readily deduced by symmetry, thus completing the proof of Lemma 3.2.

□

Theorem 3.8 is a strong enhancement of [194, Theorem 4] where it is merely established that f_{sparse} is the sparsest CPWL solution of (3.68). In Theorem 3.8, we prove that f_{sparse} is in fact the sparsest of *all* CPWL interpolants of the data points (x_m, z_m), without restricting the search to the solutions of (3.68). This is a remarkable result in its own right, as it gives a nontrivial answer to the seemingly simple question: how to interpolate data points with the minimum number of lines? Here, we invoke Theorem 3.8 to deduce that, with the vector $\mathbf{z}$ defined in Item 2 of Theorem 3.6, f_{sparse} is the sparsest CPWL solution of (3.62). Similarly, with the vector $\mathbf{z}$ defined in Theorem 3.7, f_{sparse} is the sparsest CPWL solution of (3.71).

In the remaining part of this section, we detail our computation of the vectors $\mathbf{z}$ defined in Theorems 3.6 and 3.7. Let us define the empirical loss function $F : \mathbb{R}^M \to \mathbb{R}_{\geq 0}$ as

$$F(\mathbf{z}) = \sum_{m=1}^{M} E(z_m, y_m). \tag{3.92}$$

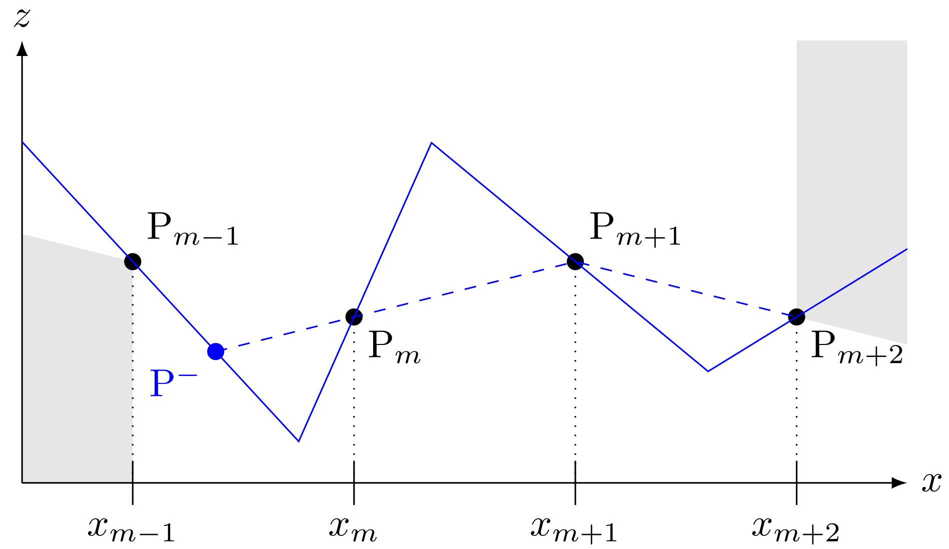


Figure 3.9: Illustration of Lemma 3.2 in the case $a_m > 0$ and $a_{m+1} < 0$. The interpolant f (solid line) satisfies $s^+ > s^+_{\text{cano}}$ and $s^- \leq s^-_{\text{cano}}$. The modified interpolant $\tilde{f}$ (dashed line) also has three knots P^-, P_{m+1}, and P_{m+2}, but none in (x_m, x_{m+1}).

For simplicity, we assume that F is differentiable; the prototypical example is the quadratic loss $F(\mathbf{z}) = \frac{1}{2}\sum_{m=1}^{M}(z_m - y_m)^2$. Following this notation and using (3.63), the vector $\mathbf{z}$ in Problem (3.62) is solution to the minimization problem

$$\min_{\mathbf{z}\in\mathbb{R}^M} \left(F(\mathbf{z}) + \lambda\|\mathbf{L}_{\inf}\mathbf{z}\|_\infty\right), \tag{3.93}$$

where the matrix $\mathbf{L}_{\inf} \in \mathbb{R}^{(M-1)\times M}$ is given by

$$[\mathbf{L}_{\inf}]_{m,n} = \begin{cases} -v_{m+1}, & n = m \\ v_{m+1}, & n = m+1 \\ 0, & \text{otherwise} \end{cases} \tag{3.94}$$

where $v_m = (x_m - x_{m-1})^{-1}, m = 2, \dots, M$. To solve (3.93), we use the well-known alternating-direction method of multipliers (ADMM) [200] by defining the augmented Lagrangian as

$$J(\mathbf{z}, \mathbf{u}, \mathbf{w}) = F(\mathbf{z}) + \lambda \|\mathbf{u}\|_\infty + \frac{\rho}{2} \|\mathbf{L}_{\text{inf}}\mathbf{z} - \mathbf{u}\|_2^2 + \mathbf{w}^T (\mathbf{L}_{\text{inf}}\mathbf{z} - \mathbf{u}), \tag{3.95}$$

where $\rho > 0$ is a tunable parameter. The principle of ADMM is to sequentially update the unknown variables $\mathbf{z} \in \mathbb{R}^M$ and $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{M-1}$. Precisely, its kth iteration is given explicitly by

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z} \in \mathbb{R}^M} J(\mathbf{z}, \mathbf{u}^{(k)}, \mathbf{w}^{(k)}), \tag{3.96}$$

$$\mathbf{u}^{(k+1)} = \arg\min_{\mathbf{u} \in \mathbb{R}^{M-1}} J(\mathbf{z}^{(k+1)}, \mathbf{u}, \mathbf{w}^{(k)}), \tag{3.97}$$

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \rho \left(\mathbf{L}_{\text{inf}}\mathbf{z}^{(k+1)} - \mathbf{u}^{(k+1)} \right). \tag{3.98}$$

The benefit of these sequential updates is that Problem (3.96) has a differentiable cost and hence, can be efficiently solved using gradient-based methods. (In the case of the quadratic loss $E(z, y) = \frac{1}{2}(z - y)^2$, one can even obtain a closed-form solution.) Unfortunately, the cost in (3.97) is not differentiable. However, one can rewrite the augmented Lagrangian as

$$J(\mathbf{z}_k, \mathbf{u}, \mathbf{w}_k) = \frac{\rho}{2} \left\| \mathbf{u} - \mathbf{L}_{\text{inf}}\mathbf{z}_k - \frac{1}{\rho}\mathbf{w}_k \right\|_2^2 + \lambda \|\mathbf{u}\|_\infty + \text{Cnst.}, \tag{3.99}$$

where the constant term accounts for all terms that do not depend on $\mathbf{u}$. Then, by defining the vector $\mathbf{v}_k = \left(\mathbf{L}_{\text{inf}}\mathbf{z}_k + \frac{1}{\rho}\mathbf{w}_k \right)$, we rewrite (3.97) as

$$\mathbf{u}^{(k+1)} = \arg\min_{\mathbf{u} \in \mathbb{R}^{M-1}} \left(\frac{1}{2} \|\mathbf{u} - \mathbf{v}_k\|_2^2 + \frac{\lambda}{\rho} \|\mathbf{u}\|_\infty \right) = \text{prox}_{\frac{\lambda}{\rho}\|\cdot\|_\infty}(\mathbf{v}_k), \tag{3.100}$$

by definition of the proximal operator. The proximal operator of the ℓ_∞-norm has computationally cheap implementations (see, for example, [201, Section 6.5.2]), which can be used to update $\mathbf{u}$ via (3.100).

Similarly and using (3.73), we formulate the search for the vector $\mathbf{z}$ associated to the Problem (3.71) as

$$\min_{\mathbf{z} \in \mathbb{R}^M} \left(F(\mathbf{z}) + \lambda \|\mathbf{L}_1\mathbf{z}\|_1 + i_{\|\mathbf{L}_{\text{inf}}\mathbf{z}\|_\infty \leq \overline{L}} \right), \tag{3.101}$$

where $\mathbf{L}_1 \in \mathbb{R}^{(M-2)\times M}$ with

$$[\mathbf{L}_1]_{m,n} = \begin{cases} -v_{m+1}, & n = m \\ (v_{m+1} + v_{m+2}), & n = m+1 \\ -v_{m+2}, & n = m+2, \\ 0, & \text{otherwise} \end{cases} \tag{3.102}$$

for all $m = 1, \ldots, M-2$ and $n = 1, \ldots, M$. In this case, the augmented Lagrangian takes the form

$$\begin{aligned} J(\mathbf{z}, \mathbf{u}_1, \mathbf{u}_{\text{inf}}, \mathbf{w}_1, \mathbf{w}_{\text{inf}}) = F(\mathbf{z}) &+ \frac{\rho_1}{2}\|\mathbf{L}_1\mathbf{z} - \mathbf{u}_1\|_2^2 + \mathbf{w}_1^T(\mathbf{L}_1\mathbf{z} - \mathbf{u}_1) + \|\mathbf{u}_1\|_1 \\ &+ \frac{\rho_{\text{inf}}}{2}\|\mathbf{L}_{\text{inf}}\mathbf{z} - \mathbf{u}_{\text{inf}}\|_2^2 + \mathbf{w}_{\text{inf}}^T(\mathbf{L}_{\text{inf}}\mathbf{z} - \mathbf{u}_{\text{inf}}) + i_{\|\mathbf{u}_{\text{inf}}\|_\infty \leq \overline{L}}. \end{aligned} \tag{3.103}$$

At the kth iteration, we then solve sequentially the following optimization problems

$$\mathbf{z}^{(k+1)} = \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^M} J(\mathbf{z}, \mathbf{u}_1^{(k)}, \mathbf{u}_{\text{inf}}^{(k)}, \mathbf{w}_1^{(k)}, \mathbf{w}_{\text{inf}}^{(k)}), \tag{3.104}$$

$$\mathbf{u}_1^{(k+1)} = \operatorname*{arg\,min}_{\mathbf{u}_1\in\mathbb{R}^{M-2}} J(\mathbf{z}^{(k+1)}, \mathbf{u}_1, \mathbf{u}_{\text{inf}}^{(k)}, \mathbf{w}_1^{(k)}, \mathbf{w}_{\text{inf}}^{(k)}), \tag{3.105}$$

$$\mathbf{u}_{\text{inf}}^{(k+1)} = \operatorname*{arg\,min}_{\mathbf{u}_{\text{inf}}\in\mathbb{R}^{M-1}} J(\mathbf{z}^{(k+1)}, \mathbf{u}_1^{(k+1)}, \mathbf{u}_{\text{inf}}, \mathbf{w}_1^{(k)}, \mathbf{w}_{\text{inf}}^{(k)}), \tag{3.106}$$

$$\mathbf{w}_1^{(k+1)} = \mathbf{w}_1^{(k)} + \rho_1\left(\mathbf{L}_1\mathbf{z}^{(k+1)} - \mathbf{u}_1^{(k+1)}\right), \tag{3.107}$$

$$\mathbf{w}_{\text{inf}}^{(k+1)} = \mathbf{w}_{\text{inf}}^{(k)} + \rho_{\text{inf}}\left(\mathbf{L}_{\text{inf}}\mathbf{z}^{(k+1)} - \mathbf{u}_{\text{inf}}^{(k+1)}\right). \tag{3.108}$$

The cost function of Problem (3.104) is differentiable and so, we can solve it using gradient-based methods. For Problem (3.105), we invoke the proximal operator of the ℓ_1-norm that is known to be soft-thresholding [201, Section 6.5.2.]. Finally and for (3.106), the proximal operator of the indicator function $i_{\|\cdot\|_\infty \leq \overline{L}}$ is the projection over the ℓ_∞ ball which has the simple separable expression

$$[\text{prox}_{i_{\|\cdot\|_\infty \leq \overline{L}}}(\mathbf{v})]_n = \begin{cases} \overline{L}, & v_n > \overline{L} \\ v_n, & |v_n| \leq \overline{L} \\ -\overline{L}, & v_n < -\overline{L}. \end{cases} \tag{3.109}$$

3.3.5 Numerical Illustration

In all our experiments, we consider the standard quadratic loss $E(y,z) = \frac{1}{2}(y-z)^2$. We draw the data-point locations x_m randomly in the interval $[0,1]$. The values y_m are then generated as $y_m = f_0(x_m) + n_m$, where f_0 is some known CPWL function (gold standard) and n_m is drawn i.i.d. from a zero-mean normal distribution with variance σ^2.

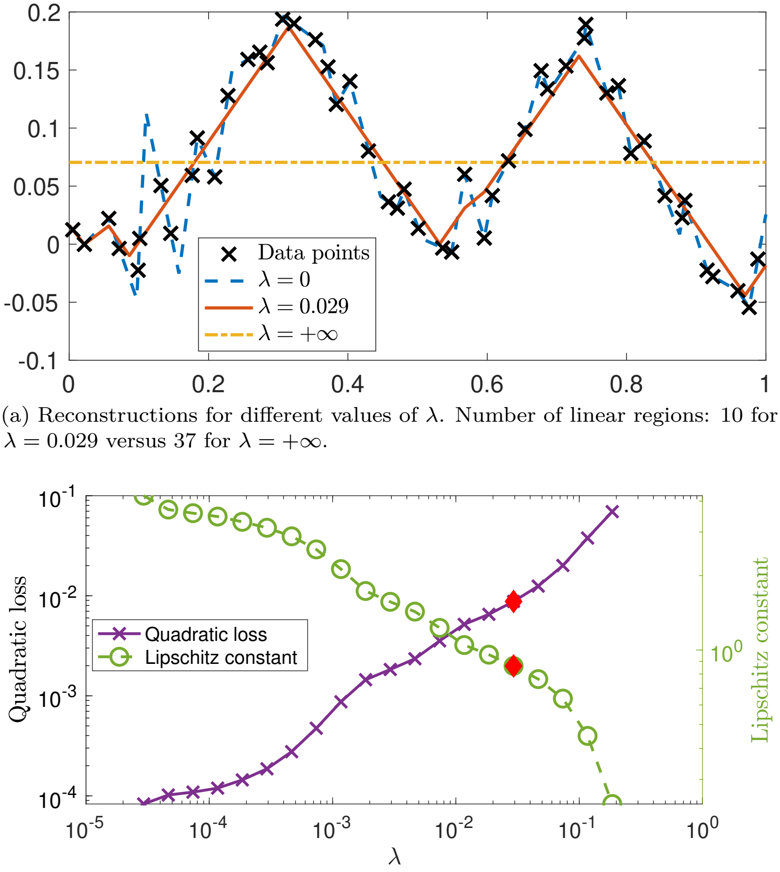
Lipschitz Regularization

In this first experiment, we illustrate our first formulation (3.61). We take $M = 50$ data points, a CPWL ground-truth f_0 with 6 linear regions, and a noise level $\sigma = 0.02$.

The results are shown in Figure 3.10. In Figure 3.10a, we show the reconstructions for extreme values of λ. On one hand, $\lambda \to 0$ corresponds to the exact interpolation Problem (3.62). On the other hand, $\lambda = +\infty$ corresponds to constant regression. Obviously, neither is very satisfactory: interpolation leads to overfitting (the reconstruction has 37 linear regions), and the constant regression to underfitting. We show an example of a more satisfactory reconstruction for $\lambda = 0.029$ (10 linear regions), which is visually acceptable. In Figure 3.10b, we show the evolution of the quadratic loss $\frac{1}{2}\sum_{m=1}^{M}(f^*(x_m) - y_m)^2$ and the Lipschitz constant $L(f^*)$, for various values of λ. With the aid of such curves, the user can choose what is considered acceptable for either of these costs and select a suitable value of λ.

Limitations of Lipschitz-Only Regularization

Despite its interesting theoretical properties, Problem (3.61) does not always yield satisfactory reconstructions. This is because it does not enforce a sparse reconstruction in the problem formulation, despite the fact that our algorithm reconstructs (one of) the sparsest elements of $\mathcal{V}_{\text{lip}}$. This leads to learned mappings with too many linear regions and, consequently, poor interpretability.

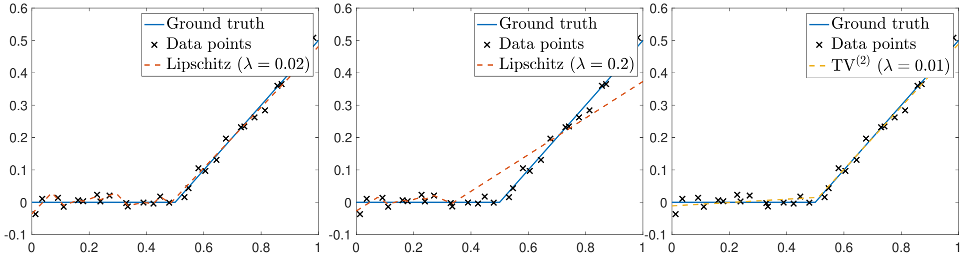


(a) Reconstructions for different values of λ. Number of linear regions: 10 for $\lambda = 0.029$ versus 37 for $\lambda = +\infty$.

(b) Evolution of the training error and the Lipschitz regularity with respect to λ. The diamond corresponds to $\lambda = 0.029$ (shown in Figure 3.10a).

Figure 3.10: Example of our first formulation (3.61) for $M = 50$ data points.

One such example is shown in Figure 3.11, where we consider the shifted ReLU function $f_0(\cdot) = (\cdot - \frac{1}{2})_+$ as the ground-truth mapping. We also fix the standard deviation of the noise to $\sigma = 0.02$. Figure 3.11a shows a reconstruction that



(a) Lipschitz regularization. Number of linear regions: 13.

(b) Lipschitz regularization. Number of linear regions: 9.

(c) TV$^{(2)}$ regularization. Number of linear regions: 2.

Figure 3.11: Reconstructions with a ReLU ground truth and $M = 30$ data points.

solves Problem (3.61) with the regularization parameter $\lambda = 0.02$. Although the reconstruction is satisfactory in the active section ($x > 1/2$), it has many linear regions in the flat section ($x < 1/2$) that are not present in f_0. This is due to the fact that the active section forces the Lipschitz constant of the reconstruction to be around 1, while oscillations with a slope smaller than 1 in the flat section are not penalized by the regularization. This problem clearly cannot be fixed by a simple increase in the regularization parameter: with $\lambda = 0.2$ (Figure 3.11b), not only there are still too many linear regions in the flat section (the reconstruction has 9 linear regions in total), but also the active section is poorly reconstructed because the Lipschitz constant is penalized too heavily by the regularization.

Hence, to reconstruct such a ground truth accurately, it is necessary to enforce the sparsity of the reconstruction, which is exactly the purpose of the TV$^{(2)}$ regularization. The reconstruction result of the TV$^{(2)}$-regularized problem (*i.e.*, Problem (3.71) with a relatively large Lipschitz bound) with $\lambda = 0.01$ is also shown in Figure 3.11c; it is clearly much more satisfactory than any of the Lipschitz-penalized reconstructions since it is very close to the ground truth and has the same sparsity (two linear regions).

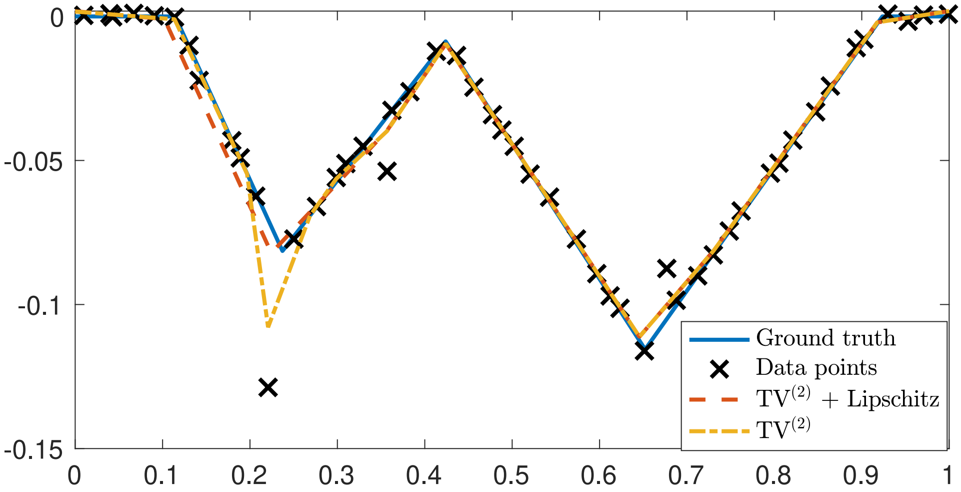


Figure 3.12: Reconstruction of $M = 50$ data points for $\lambda = 10^{-4}$. Our second formulation with $\bar{L} = 0.66$ produces 9 linear regions. We compare it to that of TV$^{(2)}$ which produces 12 linear regions.

Lischitz-Constraint Formulation and Robustness to Outliers

In this final experiment, we demonstrate the pertinence of our second formulation (Problem (3.71)). More precisely, we examine the increased robustness to outliers of our second formulation (3.71) with respect to TV$^{(2)}$ regularization. To that end, we generate the CPWL ground truth f_0 with 6 linear regions and $M = 50$ data points. We then consider an additive Gaussian-noise model with low standard deviation $\sigma = 10^{-3}$ for 90% of the data, and a much stronger $\sigma' = 3.5 * 10^{-2}$ for the remaining 10%, which can be considered outliers.

We show in Figure 3.12 the reconstruction results using our second formulation with $\lambda = 10^{-4}$ and $\bar{L} = 0.66$. The latter is quite satisfactory despite the presence of a strong outlier around $x_m = 0.22$. This is due to the fact that the Lipschitz constant is constrained. When using TV$^{(2)}$-regularization alone, at same regularization parameter, the reconstruction is very similar in most regions but is much more sensitive to this outlier which leads to an unwanted sharp peak and to the high Lipschitz constant $L(f^*) = 2.21$. Moreover, our reconstruction is more satisfactory in terms of sparsity (9 linear regions compared to 12, which is closer to the 6 linear regions of the target function f_0).

3.3.6 Summary

We have proposed two schemes for the learning of one-dimensional continuous and piecewise-linear (CPWL) mappings with tunable Lipschitz constant. In the first scheme, we directly use the Lipschitz constant as a regularization term. We establish a representer theorem that allows us to deduce the existence of a CPWL solution for this continuous-domain optimization problem. In the second scheme, we use the second-order total-variation seminorm as the regularization term to which we add a Lipschitz constraint. Again, we proved the existence of a CPWL solution for this problem. Finally, we proposed an efficient algorithm to find the sparsest CPWL solution of each problem. We illustrated the outcome of each scheme via numerical examples.

3.4 Learning Activation Functions of Deep Neural Networks

We introduce[6] a variational framework to learn the activation functions of deep neural networks. Our aim is to increase the capacity of the network while controlling an upper-bound of the actual Lipschitz constant of the input-output relation. To that end, we first establish a global bound for the Lipschitz constant of neural networks. Based on the obtained bound, we then formulate a variational problem for learning activation functions. Our variational problem is infinite-dimensional and not computationally tractable. However, we prove that there always exists a solution that has continuous and piecewise-linear (linear-spline) activations. This reduces the original problem to a finite-dimensional minimization where an ℓ_1 penalty on the parameters of the activations favors the learning of sparse nonlinearities. We propose a scalable algorithm for training activation functions that is based on B-spline representation of cardinal splines. Finally, we numerically compare our scheme with standard ReLU network and its variations, PReLU and LeakyReLU and we empirically demonstrate the practical aspects of our framework.

3.4.1 Context

Although the ReLU networks are favorable, both from a theoretical and practical point of view, one may want to go even farther and learn the activation functions as well. The minimal attempt is to learn the parameter a in LeakyReLU activations, which is known as the parametric ReLU (PReLU) [202]. More generally, one can consider a parametric form for the activations and learn the parameters in the training step. There is a rich literature on the learning of activations represented by splines, a parametric form characterized by optimality and universality [10, 6]. Examples are perceptive B-splines [203], Catmull-Rom cubic splines [204, 205], and adaptive piecewise linear splines [206], to name a few.

In theoretical analyses of deep neural networks, the Lipschitz-continuity of the network and the control of its regularity is of great importance and is crucial in

[6] This section is based on our published works [106, 107].

several schemes of deep learning, for example in Wasserstein GANs [178], in providing compressed sensing type guarantees for generative models [179], in showing the convergence of CNN-based projection algorithms to solve inverse problems [181], and in understanding the generalization property of deep neural networks [180]. Moreover, the Lipschitz regularity drives the stability of neural networks, a matter that has been tackled recently [207, 208, 209].

In this work, we propose a variational framework to learn the activation functions with the motivation of increasing the capacity of the network while controlling its Lipschitz regularity. To that end, we first provide a global bound for the Lipschitz constant of the input-output relation of neural networks that have second-order bounded-variation activations. Based on the minimization of this bound, we propose an optimization scheme in which we learn the linear weights and the activation functions jointly. We show that there always exists a global solution of our proposed minimization made of linear spline activations. We also demonstrate that our proposed regularization has a sparsity-promoting effect on the parameters of the spline activations. Let us remark that our regularization, which is based on an upper-bound, does not ensure that the actual Lipschitz constant of the neural network is minimized—it only prevents it from exceeding a certain range.

3.4.2 Second-Order Bounded-Variation Activation Functions

We consider activations from the space of second-order bounded-variation functions $\mathrm{BV}^{(2)}(\mathbb{R})$. This ensures that the corresponding neural network satisfies several desirable properties which we discuss in this section. The key feature of these activations is their Lipschitz continuity (see, Theorem 3.9). Lipschitz functions are known to be continuous and differentiable almost everywhere [210]. Moreover, in Proposition 3.2, we show that any element of $\mathrm{BV}^{(2)}(\mathbb{R})$ has well-defined right and left derivatives at any point. This is an important property for activation functions, since it is the minimum requirement for performing gradient-based algorithms that take advantage of the celebrated back-propagation scheme in the training step [211].

Proposition 3.2. *For any function $\sigma \in \mathrm{BV}^{(2)}(\mathbb{R})$ and any $x_0 \in \mathbb{R}$, the left and right derivatives of σ at the point $x = x_0$ exist and are finite.*

Proof. We prove the existence of the right derivative at $x = 0$ and deduce the existence of the left derivative by symmetry. Moreover, since $\mathrm{BV}^{(2)}(\mathbb{R})$ is a shift-invariant function space, the existence of left and right derivatives will be ensured at any point $x_0 \in \mathbb{R}$.

Let us denote

$$\Delta\sigma(a,b) = \frac{\sigma(a) - \sigma(b)}{a-b}, \quad a, b \in \mathbb{R}, a \neq b. \tag{3.110}$$

From Theorem 3.9, we have that

$$-\|\sigma\|_{\mathrm{BV}^{(2)}} \leq \Delta\sigma(a,b) \leq \|\sigma\|_{\mathrm{BV}^{(2)}}, \quad a, b \in \mathbb{R}, a \neq b. \tag{3.111}$$

Define quantities $M_{\sup}$ and $M_{\inf}$ as

$$M_{\sup} = \limsup_{h \to 0^+} \Delta\sigma(h, 0), \qquad M_{\inf} = \liminf_{h \to 0^+} \Delta\sigma(h, 0). \tag{3.112}$$

The finiteness of $|M_{\sup}|$ and $|M_{\inf}|$ is guaranteed by (3.111). Now, it remains to show that $M_{\sup} = M_{\inf}$ to prove the existence of the right derivative. Assume, by contradiction, that $M_{\sup} > M_{\inf}$. Consider a small value of $0 < \epsilon < \frac{M_{\sup} - M_{\inf}}{3}$ and define the constants $C_1 = (M_{\sup} - \epsilon)$ and $C_2 = (M_{\inf} + \epsilon)$. Clearly, we have $C_1 - C_2 \geq \epsilon > 0$. Moreover, due to the definition of $\limsup$ and $\liminf$, there exist sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ that are monotically decreasing to 0 and are such that

$$\Delta\sigma(a_n, 0) > C_1, \quad \Delta\sigma(b_n, 0) < C_2, \quad a_n > b_n > a_{n+1}, \quad \forall n \in \mathbb{N}. \tag{3.113}$$

One then has that

$$\sigma(a_n) > C_1 a_n + \sigma(0), \quad \sigma(b_n) < C_2 b_n + \sigma(0) \tag{3.114}$$

and, consequently,

$$\Delta\sigma(a_n, b_n) \geq \frac{C_1 a_n - C_2 b_n}{a_n - b_n} > C_1, \tag{3.115}$$

$$\Delta\sigma(b_n, a_{n+1}) \leq \frac{C_2 b_n - C_1 a_{n+1}}{b_n - a_{n+1}} < C_2. \tag{3.116}$$

From the definition of second-order total variation and using (3.115) and (3.116), we obtain that

$$\begin{aligned}\|\mathrm{D}^2\sigma\|_{\mathcal{M}} &\geq \sum_{n=0}^{\infty} |\Delta\sigma(a_n, b_n) - \Delta\sigma(b_n, a_{n+1})| \\ &\geq \sum_{n=0}^{\infty} \left(\frac{C_1 a_n - C_2 b_n}{a_n - b_n} - \frac{C_2 b_n - C_1 a_{n+1}}{b_n - a_{n+1}} \right) \\ &\geq \sum_{n=1}^{\infty} (C_1 - C_2) = \sum_{n=0}^{\infty} \epsilon = +\infty, \end{aligned} \tag{3.117}$$

which contradicts the original assumption $\sigma \in \mathrm{BV}^{(2)}(\mathbb{R})$. Hence, $M_{\sup} = M_{\inf}$ and the right derivative exists. □

Let us mention that Lipschitz functions in general do not have one-sided derivatives at all points; it is a property that is specific of $\mathrm{BV}^{(2)}$ functions. As an example, consider the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = \begin{cases} x \sin(\log(x)), & x > 0 \\ 0, & x \leq 0. \end{cases} \tag{3.118}$$

One readily verifies that f is Lipschitz-continuous with the constant $C = \sqrt{2}$. However, for positive values of h, the function $\frac{f(h)}{h} = \sin(-\log(h))$ oscillates between (-1) and 1 as h goes to zero. Hence, f does not have a right derivative at the point $x_0 = 0$.

In Theorem 3.9, we prove that any neural network with activations from $\mathrm{BV}^{(2)}(\mathbb{R})$ specifies a Lipschitz-continuous input-output relation. Moreover, we provide an upper-bound for its Lipschitz constant. Before stating the theorem, let us define the $(\mathrm{BV}^{(2)}, p)$-norm of the nonlinear layer $\boldsymbol{\sigma}_l$ for any $p \in [1, +\infty)$ as

$$\|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},p} = \left(\sum_{n=1}^{N_l} \|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}}^p \right)^{\frac{1}{p}}. \tag{3.119}$$

Theorem 3.9. *Any feed forward fully connected deep neural network* $\mathbf{f}_{\text{deep}} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ *with second-order bounded-variation activations* $\sigma_{n,l} \in \mathrm{BV}^{(2)}(\mathbb{R})$ *is Lipschitz-continuous. Moreover, if we consider the* ℓ_p *for* $p \in [1, \infty]$ *topology in the input and output spaces, the neural network satisfies the global Lipschitz bound*

$$\|\mathbf{f}_{\text{deep}}(\mathbf{x}_1) - \mathbf{f}_{\text{deep}}(\mathbf{x}_2)\|_p \leq C\|\mathbf{x}_1 - \mathbf{x}_2\|_p \tag{3.120}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{N_0}$*, where*

$$C = \left(\prod_{l=1}^{L} \|\mathbf{W}_l\|_{q,\infty}\right) \cdot \left(\prod_{l=1}^{L} \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},p}\right), \tag{3.121}$$

$q \in [1, \infty]$ *is such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and* $\|\mathbf{W}_l\|_{q,\infty} = \max_n \|\mathbf{w}_{n,l}\|_q$ *is the mixed norm* ($\ell_q - \ell_\infty$) *of the lth linear layer.*

Proof. From Theorem 3.9, for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^{N_{l-1}}$, we have that

$$\left|\sigma_{n,l}(\mathbf{w}_{n,l}^T\mathbf{x}_1) - \sigma_{n,l}(\mathbf{w}_{n,l}^T\mathbf{x}_2)\right| \leq \|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}} \left|\mathbf{w}_{n,l}^T(\mathbf{x}_1 - \mathbf{x}_2)\right|. \tag{3.122}$$

Now, by using Hölder's inequality, we bound the Lipschitz constant of the linear layers as

$$\left|\mathbf{w}_{n,l}^T(\mathbf{x}_1 - \mathbf{x}_2)\right| \leq \|\mathbf{w}_{n,l}\|_q\|\mathbf{x}_1 - \mathbf{x}_2\|_p. \tag{3.123}$$

By combining (3.122) and (3.123) and using the fact that $\|\mathbf{w}_{n,l}\|_q \leq \|\mathbf{W}_l\|_{q,\infty}$, we obtain that

$$\left|\sigma_{n,l}(\mathbf{w}_{n,l}^T\mathbf{x}_1) - \sigma_{n,l}(\mathbf{w}_{n,l}^T\mathbf{x}_2)\right|^p \leq \|\mathbf{W}_l\|_{q,\infty}^p\|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}}^p\|\mathbf{x}_1 - \mathbf{x}_2\|_p^p, \tag{3.124}$$

which is a Lipschitz bound for the (n, l)th neuron of the neural network. By summing up over the neurons of layer l, we control the output of this layer as

$$\|\mathbf{f}_l(\mathbf{x}_1) - \mathbf{f}_l(\mathbf{x}_2)\|_p \leq \|\mathbf{W}_l\|_{q,\infty}\|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},p}\|\mathbf{x}_1 - \mathbf{x}_2\|_p. \tag{3.125}$$

Now, the composition of the layer inequalities results in the inequality (3.120) with the constant introduced in (3.121). □

When the standard Euclidean topology is assumed for the input and output spaces, Proposition 3.3 provides an alternative bound for the Lipschitz constant of the neural network.

Proposition 3.3. *Let* $\mathbf{f}_{\mathrm{deep}} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ *be a fully connected feed forward neural network with activations selected from* $\mathrm{BV}^{(2)}(\mathbb{R})$*. For all* $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{N_0}$ *we have that*

$$\|\mathbf{f}_{\mathrm{deep}}(\mathbf{x}_1) - \mathbf{f}_{\mathrm{deep}}(\mathbf{x}_2)\|_2 \leq C_E \|\mathbf{x}_1 - \mathbf{x}_2\|_2, \tag{3.126}$$

where

$$C_E = \left(\prod_{l=1}^{L} \|\mathbf{W}_l\|_F\right) \cdot \left(\prod_{l=1}^{L} \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},1}\right). \tag{3.127}$$

Proof. Following Theorem 3.9 and using Cauchy–Schwarz' inequality, we obtain that

$$\left|\sigma_{n,l}(\mathbf{w}_{n,l}^T \mathbf{x}_1) - \sigma_{n,l}(\mathbf{w}_{n,l}^T \mathbf{x}_2)\right| \leq \|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}} \|\mathbf{w}_{n,l}\|_2 \|\mathbf{x}_1 - \mathbf{x}_2\|_2. \tag{3.128}$$

Combining it with the known hierarchy between the discrete ℓ_p norms and, in particular, the inequality $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ for any $\mathbf{x} \in \mathbb{R}^{N_l}$, we obtain that

$$\begin{aligned}
\|\mathbf{f}_l(\mathbf{x}_1) - \mathbf{f}_l(\mathbf{x}_2)\|_2 &\leq \|\mathbf{f}_l(\mathbf{x}_1) - \mathbf{f}_l(\mathbf{x}_2)\|_1 \\
&= \sum_{n=1}^{N_l} \left|\sigma_{n,l}(\mathbf{w}_{n,l}^T \mathbf{x}_1) - \sigma_{n,l}(\mathbf{w}_{n,l}^T \mathbf{x}_2)\right| \\
&\leq \sum_{n=1}^{N_l} \|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}} \|\mathbf{w}_{n,l}\|_2 \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \\
&\leq \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},2} \|\mathbf{W}_l\|_F \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2.
\end{aligned} \tag{3.129}$$

Note that in the last inequality of (3.129), we have again used Cauchy-Schwarz' inequality. Combining with $\|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},2} \leq \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},1}$, we have that

$$\|\mathbf{f}_l(\mathbf{x}_1) - \mathbf{f}_l(\mathbf{x}_2)\|_2 \leq \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},1} \|\mathbf{W}_l\|_F \|\mathbf{x}_1 - \mathbf{x}_2\|_2. \tag{3.130}$$

Finally by composing (3.130) through the layers, we obtain the announced bound. □

Proposition 3.3 will take a particular relevance where (3.127) will allow us to design a joint-optimization problem to learn the linear weights and activations. Interestingly, the proposed minimization is compatible with the use of weight decay [212] in the training of neural networks (see (3.139) with $\mathrm{R}(\mathbf{W}) = \|\mathbf{W}\|_F^2$).

3.4.3 Representer Theorem

We now propose our variational formulation of learning Lipschitz-continuous activation functions in a deep neural network. We select $\mathrm{BV}^{(2)}(\mathbb{R})$ as our search space to ensure the Lipschitz continuity of the input-output relation of the global network (see Theorem 3.9).

Similarly to the RKHS theory, the (weak*) continuity of the sampling functional is needed to guarantee the well-posedness of the learning problem. We say that a sequence of neural networks converges in the weak*-topology if

- the networks all have the same layer descriptor (architecture);
- for any neuron in the architecture, the corresponding linear weights converge in the Euclidean topology and the corresponding activations converge in the weak*-topology of $\mathrm{BV}^{(2)}(\mathbb{R})$.

Theorem 3.10. *For any $\boldsymbol{x}_0 \in \mathbb{R}^{N_0}$, the sampling functional $\delta_{\boldsymbol{x}_0} : \mathbf{f}_{\mathrm{deep}} \mapsto \mathbf{f}_{\mathrm{deep}}(\boldsymbol{x}_0)$ is weak*-continuous in the space of neural networks with second-order bounded-variation activations.*

Before going to the proof, we first prove a lemma. We recall that a sequence of functions $f_t : \mathcal{X} \to \mathcal{Y}$, $t \in \mathbb{N}$, converges pointwise to $f_{\mathrm{lim}} : \mathcal{X} \to \mathcal{Y}$ if, for all $x \in \mathcal{X}$,

$$f_{\mathrm{lim}}(x) = \lim_{t \to +\infty} f_t(x). \tag{3.131}$$

Lemma 3.3. *Given the Banach spaces $\mathcal{X}$,$\mathcal{Y}$, and $\mathcal{Z}$, consider the two sequences of functions $f_t : \mathcal{X} \to \mathcal{Y}$ and $g_t : \mathcal{Y} \to \mathcal{Z}$ such that they converge pointwise to the functions $f_{\mathrm{lim}} : \mathcal{X} \to \mathcal{Y}$ and $g_{\mathrm{lim}} : \mathcal{Y} \to \mathcal{Z}$, respectively. Moreover, assume that the* *functions g_t are all Lipschitz-continuous with a shared constant $C > 0$, so that, for any $y_1, y_2 \in \mathcal{Y}$, one has that*

$$\|g_t(y_1) - g_t(y_2)\|_{\mathcal{Z}} \leq C\|y_1 - y_2\|_{\mathcal{Y}}, \quad \forall t \in \mathbb{N}. \tag{3.132}$$

Then, the composed sequence $h_t : \mathcal{X} \to \mathcal{Z}$ with $h_t = g_t \circ f_t$ converges pointwise to $h_{\text{lim}} = g_{\text{lim}} \circ f_{\text{lim}}$.

Proof. We use the triangle inequality to obtain that

$$\|h_t(x) - h_{\text{lim}}(x)\|_{\mathcal{Z}} \leq \|h_t(x) - g_t(f_{\text{lim}}(x))\|_{\mathcal{Z}} + \|g_t(f_{\text{lim}}(x)) - h_{\text{lim}}(x)\|_{\mathcal{Z}} \tag{3.133}$$

for all $x \in \mathcal{X}$. The uniform Lipschitz-continuity of g_t then yields that

$$\|h_t(x) - g_t(f_{\text{lim}}(x))\|_{\mathcal{Z}} = \|g_t(f_t(x)) - g_t(f_{\text{lim}}(x))\|_{\mathcal{Z}} \leq C\|f_t(x) - f_{\text{lim}}(x)\|_{\mathcal{Y}} \to 0 \tag{3.134}$$

as $t \to +\infty$. This is due to the pointwise convergence of $\{f_t\} \to f_{\text{lim}}$. Similarly, the pointwise convergence $\{g_t\} \to g_{\text{lim}}$ implies that

$$\|g_t(f_{\text{lim}}(x)) - g_{\text{lim}}(f_{\text{lim}}(x))\|_{\mathcal{Z}} \to 0, \quad t \to +\infty \tag{3.135}$$

which, together with (3.134) and (3.133), proves the pointwise convergence of $h_t \to h_{\text{lim}}$ as $t \to +\infty$. □

Proof of Theorem 3.10. Assume that the sequence $\{\mathbf{f}_{\text{deep}}^{(t)}\}$ of neural networks with layers $\mathbf{f}_l^{(t)} = \boldsymbol{\sigma}_l^{(t)} \circ \mathbf{W}_l^{(t)} : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ for $l = 1, \dots, L$ (described in (3.5)) converges in the weak*-topology to

$$f_{\text{deep}}^{\text{lim}} = \mathbf{f}_L^{\text{lim}} \circ \cdots \circ \mathbf{f}_1^{\text{lim}}, \quad \forall l : \mathbf{f}_l^{\text{lim}} = \boldsymbol{\sigma}_l^{\text{lim}} \circ \mathbf{W}_l^{\text{lim}}. \tag{3.136}$$

By definition, every element of $\boldsymbol{\sigma}_l^{(t)}$ converges in the weak*-topology to the corresponding element in $\boldsymbol{\sigma}_l^{\text{lim}}$. The convergence is also pointwise due to the weak*-continuity of the sampling functional in the space of activation functions $\text{BV}^{(2)}(\mathbb{R})$. The conclusion is that $\mathbf{f}_l^{(t)}$ also converges pointwise to $\mathbf{f}_l^{\text{lim}}$.

In addition, knowing that any norm is weak*-continuous in its corresponding Banach space, one can find the uniform constant $T_1 > 0$ such that, for all $t > T_1$ and for all $l = 1, \dots, L$, we have that

$$\|\boldsymbol{\sigma}_l^{(t)}\|_{\mathrm{BV}^{(2)},2} \leq C_1 = 2 \max_l \|\boldsymbol{\sigma}_l^{\lim}\|_{\mathrm{BV}^{(2)},2}. \tag{3.137}$$

Similarly, from the convergence $\mathbf{W}_l^{(t)} \to \mathbf{W}_l^{\lim}$, one deduces that there exists a constant $T_2 > 0$ such that, for all $t > T_2$ and for all $l = 1, \dots, L$, we have that

$$\|\mathbf{W}_l^{(t)}\|_{2,\infty} \leq C_2 = 2 \max_l \|\mathbf{W}_l^{\lim}\|_{2,\infty}. \tag{3.138}$$

Now, from (3.138) with $p = q = 2$ and using (3.125), (3.137), we deduce that, for $t > \max(T_1, T_2)$, each layer of $\mathbf{f}_{\mathrm{deep}}^{(t)}$ is Lipschitz-continuous with the shared constant $C = C_1 C_2$. Combining it with the pointwise convergence $\mathbf{f}_l^{(t)} \to \mathbf{f}_l^{\lim}$, one completes the proof by sequentially using the outcome of Lemma 3.3. □

Given the data-set (X, Y) of size M that consists in the pairs $(\mathbf{x}_m, \mathbf{y}_m) \in \mathbb{R}^{N_0} \times \mathbb{R}^{N_L}$ for $m = 1, 2, \dots, M$, we then consider the following cost functional

$$\mathcal{J}(\mathbf{f}_{\mathrm{deep}}; X, Y) = \sum_{m=1}^{M} E\Big(\mathbf{y}_m, \mathbf{f}_{\mathrm{deep}}(\mathbf{x}_m)\Big) + \sum_{l=1}^{L} \mu_l \mathrm{R}_l(\mathbf{W}_l) + \sum_{l=1}^{L} \lambda_l \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},1}, \tag{3.139}$$

where $\mathbf{f}_{\mathrm{deep}}$ is a neural network with linear layers $\mathbf{W}_l$ and nonlinear layers $\boldsymbol{\sigma}_l = (\sigma_{1,l}, \dots, \sigma_{N_l,l})$, as specified in (3.5) and (3.6), $E(\cdot, \cdot)$ is an arbitrary loss function, and $\mathrm{R}_l : \mathbb{R}^{N_l \times N_{l-1}} \to \mathbb{R}$ is a regularization functional for the linear weights of the lth layer. The standard choice for weight regularization is the Frobenius norm $\mathrm{R}(\mathbf{W}) = \|\mathbf{W}\|_F^2$, which corresponds to weight decay scheme in deep learning. Finally, the positive constants $\mu_l, \lambda_l > 0$ balance the regularization effect in the training step.

Theorem 3.11 states that, under some natural conditions, there always exists a solution of (3.139) with continuous piecewise-linear activation functions, which we refer to as a *deep-spline* neural network.

Theorem 3.11. *Consider the training of a deep neural network via the minimization*

$$\min_{\substack{\mathbf{w}_{n,l}\in\mathbb{R}^{N_{l-1}},\\ \sigma_{n,l}\in\mathrm{BV}^{(2)}(\mathbb{R})}} \mathcal{J}(\mathbf{f}_{\mathrm{deep}};X,Y), \tag{3.140}$$

where $\mathcal{J}(\cdot;X,Y)$ *is defined in* (3.139)*. Moreover, assume that the loss function* $E(\cdot,\cdot)$ *is proper, lower semi-continuous, and coercive. Assume that the regularization functionals* R_l *are continuous, and coercive. Then, there always exists a solution* $\mathbf{f}^*_{\mathrm{deep}}$ *of* (3.139) *with activations* $\sigma_{n,l}$ *of the form*

$$\sigma_{n,l}(x)=\sum_{k=1}^{K_{n,l}} a_{n,l,k}\mathrm{ReLU}(x-\tau_{n,l,k})+b_{1,n,l}x+b_{2,n,l}, \tag{3.141}$$

where $K_{n,l}\leq M$ *and,* $a_{n,l,k},\tau_{n,l,k},b_{\cdot,n,l}\in\mathbb{R}$ *are adaptive parameters.*

Proof. We divide the proof in two parts. First, we show the existence of the solution of (3.139) and, then, we show the existence of a solution with activations of the form (3.141).

Existence of Solution Consider an arbitrary neural network $\mathbf{f}_{\mathrm{deep},0}$ with the cost

$$A=\mathcal{J}(\mathbf{f}_{\mathrm{deep},0};X,Y). \tag{3.142}$$

The coercivity of R_l guarantees the existence of the constants B_l for $l=1,2,\dots,L$ such that

$$\|\mathbf{w}_{n,l}\|_2\geq B_l \Rightarrow \mathrm{R}_l(\mathbf{w}_{n,l})\geq\frac{A}{\mu_l}. \tag{3.143}$$

This allows us to transform the unconstrained problem (3.139) into the equivalent constrained minimization

$$\min_{\substack{\mathbf{w}_{n,l}\in\mathbb{R}^{N_{l-1}},\\ \sigma_{n,l}\in\mathrm{BV}^{(2)}(\mathbb{R})}} \mathcal{J}(\mathbf{f}_{\mathrm{deep}}),\quad \text{s.t.}\quad \begin{cases}\|\mathbf{w}_{n,l}\|_2\leq B_l,\\ \mathrm{TV}^{(2)}(\sigma_{n,l})\leq A/\lambda_l,\\ |\sigma_{n,l}(0)|\leq A/\lambda_l,\\ |\sigma_{n,l}(1)|\leq A/\lambda_l.\end{cases} \tag{3.144}$$

The equivalence is due to the fact that any neural network that does not satisfy the constraints of (3.145) has a strictly bigger cost than $\mathbf{f}_{\mathrm{deep},0}$ and, hence, is not in the solution set. Due to the decomposition (3.46), we can rewrite (3.144) as

$$\min_{\substack{\mathbf{w}_{n,l}\in\mathbb{R}^{N_{l-1}},\\ u_{n,l}\in\mathcal{M}(\mathbb{R})\\ b_{\cdot,n,l}\in\mathbb{R}}} \mathcal{J}(\mathbf{f}_{\mathrm{deep}}), \quad \text{s.t.,} \quad \begin{cases} \|\mathbf{w}_{n,l}\|_2 \leq B_l, \\ \|u_{n,l}\|_{\mathcal{M}} \leq A/\lambda_l, \\ |b_{1,n,l}|, |b_{2,n,l}| \leq 2A/\lambda_l. \end{cases} \tag{3.145}$$

Due to the Banach-Anaoglu theorem [94], the feasible set in (3.145) is weak*-compact. Moreover, the cost functional defined in (3.139) is a composition and sum of lower-semicontinuous functions and weak*-continuous functionals (see Theorem 3.10). Hence, it is itself weak*-lower semicontinuous. This guarantees the existence of a minimizer of (3.145) (and, consequently, of (3.140)), due to the generalized Weierstrass theorem [198].

Optimal Activations Let $\tilde{\mathbf{f}}_{\mathrm{deep}}$ be a solution of (3.140) with

$$\tilde{\mathbf{f}}_{\mathrm{deep}} = \tilde{\mathbf{f}}_L \circ \cdots \circ \tilde{\mathbf{f}}_1, \quad \forall l : \tilde{\mathbf{f}}_l = \tilde{\boldsymbol{\sigma}}_l \circ \tilde{\mathbf{W}}_l. \tag{3.146}$$

For any input vector $\boldsymbol{x}_m$ in the dataset $\boldsymbol{X}$, we then define the vectors $\boldsymbol{z}_{l,m} = (z_{1,l,m}, \dots, z_{N_l,l,m}), \boldsymbol{s}_{l,m} = (s_{1,l,m}, \dots, s_{N_l,l,m}) \in \mathbb{R}^{N_l}$ as

$$\boldsymbol{z}_{l,m} = \tilde{\mathbf{f}}_l \circ \cdots \circ \tilde{\mathbf{f}}_1(\boldsymbol{x}_m), \tag{3.147}$$

$$\boldsymbol{s}_{l,m} = \tilde{\mathbf{W}}_l \circ \tilde{\mathbf{f}}_{l-1} \circ \cdots \circ \tilde{\mathbf{f}}_1(\boldsymbol{x}_m). \tag{3.148}$$

Now, we show that the activation $\tilde{\sigma}_{n,l}$ of the neuron indexed by (n, l) is indeed a solution of the minimization

$$\min_{\sigma\in\mathrm{BV}^{(2)}(\mathbb{R})} \mathrm{TV}^{(2)}(\sigma) = \|\mathrm{D}^2\sigma\|_{\mathcal{M}} \quad s.t. \quad \begin{cases} \sigma(s_{n,l,m}) = z_{n,l,m}, & m = 1, 2, \dots, M, \\ \sigma(x) = \tilde{\sigma}_{n,l}(x), & x \in \{0, 1\}. \end{cases} \tag{3.149}$$

Assume by contradiction that there exists a function $\sigma \in \mathrm{BV}^{(2)}(\mathbb{R})$ that satisfies the feasiblity conditions (3.149) and is such that $\mathrm{TV}^{(2)}(\sigma) < \mathrm{TV}^{(2)}(\tilde{\sigma}_{n,l})$. Then, we have that

$$\begin{aligned}\|\sigma\|_{\mathrm{BV}^{(2)}} &= \mathrm{TV}^{(2)}(\sigma) + |\sigma(0)| + |\sigma(1)| = \mathrm{TV}^{(2)}(\sigma) + |\tilde{\sigma}_{n,l}(0)| + |\tilde{\sigma}_{n,l}(1)| \\ &< \mathrm{TV}^{(2)}(\tilde{\sigma}_{n,l}) + |\tilde{\sigma}_{n,l}(0)| + |\tilde{\sigma}_{n,l}(1)| \\ &= \|\tilde{\sigma}_{n,l}\|_{\mathrm{BV}^{(2)}}. \end{aligned} \tag{3.150}$$

In addition, due to the feasiblity assumptions $\sigma(s_{n,l,m}) = z_{n,l,m}$ for $m = 1, \dots, M$, one readily verifies that, by replacing $\tilde{\sigma}_{n,l}$ by σ in the optimal neural network $\tilde{\mathbf{f}}_{\mathrm{deep}}$, the data fidelity term $\sum_{m=1}^{M} E\Big(\mathbf{y}_m, \tilde{\mathbf{f}}_{\mathrm{deep}}(\mathbf{x}_m)\Big)$ in (3.139) remains untouched. The same holds for the weight regularization term $\sum_{l=1}^{L} \mathrm{R}_l(\mathbf{W}_l)$. However, from (3.150), one gets a strictly smaller overall $\mathrm{BV}^{(2)}$ penalty with σ that contradicts the optimality of $\tilde{\mathbf{f}}_{\mathrm{deep}}$. With a similar argument, one sees that, for any solution $\sigma \in \mathrm{BV}^{(2)}$ of (3.149), the substitution of $\tilde{\sigma}_{n,l}$ by σ yields another solution of (3.140). Due to Lemma 1 of [103], Problem (3.149) has a solution that is a linear spline of the form (3.141) with $K_{n,l} \leq \big(\tilde{M} - 2\big)$, where $\tilde{M} = M + 2$ is the number of constraints in (3.149). By using this result for every neuron (n, l), we verify the existence of a deep-spline solution of (3.140). □

Theorem 3.11 suggests an optimal ReLU-based parametric to learn activations. This is a remarkable property as it translates the original infinite-dimensional problem (3.140) into a finite-dimensional parametric optimization, where one only needs to determine the ReLU weights $a_{n,l,k}$ and positions $\tau_{n,l,k}$ together with the affine terms $b_{1,n,l}, b_{2,n,l}$. Let us also mention that the baseline ReLU network and its variations (PReLU, LeakyReLU) are all included in this scheme as special cases of an activation of the form (3.141) with $K = 1$.

A similar result has been shown in the deep-spline representer theorem of Unser in [103]. However, there are three fundamental differences. Firstly, we relax the assumption of having normalized weights due to the practical considerations and the optimization challenges it brings. Secondly, we slightly modify the regularization functional that enables us to control the global Lipschitz constant of the neural network. Lastly, we show the existence of a minimizer in our proposed variational formulation that is, to the best of our knowledge, the first result of existence in this framework.

We remark that the choice of our regularization restrains the coefficients $b_{1,n,l}$ and $b_{2,n,l}$ from taking high values. This enables us to obtain the global bound (3.121) for the Lipschitz constant of the network, as opposed to the framework of [103], where only a semi-norm has been used for the regularization. The payoff is that, in [103], the activations have at most $(M-2)$ knots, which are the junctions between the consecutive linear pieces of a piecewise linear function, while our bound is $K_{n,l} \leq M$. This is the price to pay for controlling the Lipschitz regularity of the network. However, this is inconsequential in practice since there are usually much fewer knots than data points, because of the regularization penalty. The latter is justified through the computation of the $\mathrm{BV}^{(2)}$ norm of an activation of the form (3.141). It yields

$$\|\sigma_{n,l}\|_{\mathrm{BV}^{(2)}} = \|\boldsymbol{a}_{n,l}\|_1 + |\sigma(1)| + |\sigma(0)|, \tag{3.151}$$

where $\boldsymbol{a}_{n,l} = (a_{n,l,1}, \dots, a_{n,l,K_{n,l}})$ is the vector of ReLU coefficients. This shows that the $\mathrm{BV}^{(2)}$-regularization imposes an ℓ_1 penalty on the ReLU weights in the expansion (3.141), thus promoting sparsity [213]. Another interesting property of the variational formulation (3.140) is the relation between the energy of consecutive linear and nonlinear layers. In Theorem 3.12, we exploit this relation.

Theorem 3.12. *Consider Problem* (3.140) *with the weight regularization* $\mathrm{R}_l(\mathbf{W}_l) = \|\mathbf{W}_l\|_F^2$ *and positive parameters* $\mu_l, \lambda_l > 0$ *for all* $l = 1, 2, \dots, L$. *Then, for any of its local minima with the linear layers* $\mathbf{W}_l$ *and nonlinear layers* $\boldsymbol{\sigma}_l$, *we have that*

$$\lambda_l \|\boldsymbol{\sigma}_l\|_{\mathrm{BV}^{(2)},1} = 2\mu_{l+1}\|\mathbf{W}_{l+1}\|_F^2, \quad l = 1, 2, \dots, L-1. \tag{3.152}$$

Proof. For any local minima $\mathbf{f}_{\mathrm{deep}}$ of (3.140) with linear weights $\mathbf{W}_l$ and nonlinear layers $\boldsymbol{\sigma}_l$ and for any layer $l^* \neq L$, consider the perturbed network $\mathbf{f}_{\mathrm{deep},\epsilon}$ with the linear layers

$$\mathbf{W}_{l,\epsilon} = \begin{cases} \mathbf{W}_l, & l \neq l^* + 1 \\ (1+\epsilon)\mathbf{W}_{l^*}, & l = l^* + 1 \end{cases} \tag{3.153}$$

and the nonlinear layers

$$\boldsymbol{\sigma}_{l,\epsilon} = \begin{cases} \boldsymbol{\sigma}_l, & l \neq l^* \\ (1+\epsilon)^{-1}\boldsymbol{\sigma}_{l^*}, & l = l^* \end{cases} \tag{3.154}$$

for any $\epsilon \in (-1, 1)$. One readily verifies that, for any $\boldsymbol{x} \in \mathbb{R}^{N_0}$ and any $\epsilon \in \mathbb{R}$, we have that $\mathbf{f}_{\text{deep},\epsilon}(\boldsymbol{x}) = \mathbf{f}_{\text{deep}}(\boldsymbol{x})$ and, hence, both networks have the same data-fidelity penalty in the global cost (3.139). In fact, the only difference between their overall cost is associated to the regularization terms of the $(l^* + 1)$th linear layer and the lth nonilnear layer. For those, the scaling property of norms yields that

$$\|\mathbf{W}_{l^*+1,\epsilon}\|_F^2 = (1+\epsilon)^2 \|\mathbf{W}_{l^*+1}\|_F^2, \tag{3.155}$$

$$\|\boldsymbol{\sigma}_{l^*,\epsilon}\|_{\text{BV}^{(2)},1} = (1+\epsilon)^{-1} \|\boldsymbol{\sigma}_{l^*}\|_{\text{BV}^{(2)},1}. \tag{3.156}$$

Due to the (local) optimality of $\mathbf{f}_{\text{deep}}$, there exists a constant $\epsilon_{\max}$ such that, for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$, we have that

$$\mathcal{J}(\mathbf{f}_{\text{deep}}) \leq \mathcal{J}(\mathbf{f}_{\text{deep},\epsilon}). \tag{3.157}$$

Now, from (3.155) and (3.156), we have that

$$\mu_{l^*+1}\|\mathbf{W}_{l^*+1}\|_F^2 + \lambda_l \|\boldsymbol{\sigma}_{l^*}\|_{\text{BV}^{(2)},1} \leq \mu_{l^*+1}(1+\epsilon)^2 \|\mathbf{W}_{l^*+1}\|_F^2 + \lambda_l (1+\epsilon)^{-1} \|\boldsymbol{\sigma}_{l^*}\|_{\text{BV}^{(2)},1}, \tag{3.158}$$

for any $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$. By simplifying the latter inequality, we get that

$$0 \leq \epsilon g(\epsilon), \quad \forall \epsilon \in (-\epsilon_{\max}, \epsilon_{\max}), \tag{3.159}$$

where $g(\epsilon) = \mu_{l^*+1}\|\mathbf{W}_{l^*+1}\|_F^2(\epsilon + 2) - \lambda_l \|\boldsymbol{\sigma}\|_{\text{BV}^{(2)}}(1+\epsilon)^{-1}$ is a continuous function of ϵ in the interval $(-1, 1)$. This yields that $g(\epsilon)$ is nonnegative for positive values of ϵ and is nonpositive for negative values of ϵ. Hence, we get that $g(0) = 0$ and, consequently, that

$$\lambda_{l^*} \|\boldsymbol{\sigma}_{l^*}\|_{\text{BV}^{(2)}} = 2\mu_{l^*+1}\|\mathbf{W}_{l^*+1}\|_F^2. \tag{3.160}$$

□

Theorem 3.12 shows that the regularization constants μ_l and λ_l provide a balance between the linear and nonlinear layers. In our experiments, we use the outcome of this theorem to determine the value of λ_l. More precisely, we select λ such that (3.152) holds in the initial setup. This is relevant in practice as it reduces the number of hyper-parameters that one needs to tune and results in a faster training scheme. We also show experimentally that this choice of λ is desirable.

3.4.4 Using B-Splines to Learn Activation Functions

According to Theorem 3.11, we can rewrite the original infinite-dimensional problem (3.140) as

$$\min_{\substack{\mathbf{w}_{n,l}\in\mathbb{R}^{N_{l-1}},\\ \mathbf{a}_{n,l}\in\mathbb{R}^{K_{n,l}}\\ b_{i,n,l}\in\mathbb{R}}} \sum_{m=1}^{M} E\Big(\mathbf{y}_m, \mathbf{f}_{\text{deep}}(\mathbf{x}_m)\Big) + \sum_{l=1}^{L} \mu_l \sum_{n=1}^{N_l} \|\mathbf{w}_{n,l}\|_2^2$$

$$+ \sum_{l=1}^{L} \lambda_l \sum_{n=1}^{N_l} \left(\|\mathbf{a}_{n,l}\|_1 + |\sigma_{n,l}(1)| + |\sigma_{n,l}(0)|\right), \tag{3.161}$$

where $\mathbf{f}_{\text{deep}}$ is the global input-output mapping and $\sigma_{n,l}$ follows the parametric form given in (3.141). Thus, we now optimize over a set of finitely many variables—the linear weights $\mathbf{w}_{n,l}$ and the unknown parameters of $\sigma_{n,l}$ for each neuron (n,l).

The major difficulty in optimizing the DNN with respect to the spline parameters is that the number $K = K_{n,\ell}$ of knots is unknown and that the activation model is nonlinear with respect to the knot locations $\tau_k = \tau_{k,n,\ell}$. Our workaround is to place a fixed but highly redundant set of knots on a uniform grid with a step size T. We then rely on the sparsifying effect of ℓ_1-minimization to nullify the coefficients of $\mathbf{a} = (a_k)$ that are not needed. This amounts to representing the spline activation functions by

$$\sigma(x) = b_0 + b_1 x + \sum_{k=k_{\min}}^{k_{\max}} a_k (x - kT)_+, \tag{3.162}$$

with $\mathrm{TV}^{(2)}(\sigma) = \|\mathbf{a}\|_1$. The consideration of the linear model (3.162), thereafter referred to as "gridded ReLU," gives rise to a classical ℓ_1-optimization problem that can be handled by most neural-network software frameworks. In the case of a shallow network with $L = 1$, it even results in a convex problem that is reminiscent of the LASSO [214]. We also note that (3.162) can be made arbitrarily close to (3.141) by taking T sufficiently small. While the solution $\mathbf{a}$ is expected to be sparse, with few active knots, the downside of the approach is that the underlying representation is cumbersome and badly conditioned due to the exploding behavior of the basis functions $(\cdot - kT)_+$ at infinity.

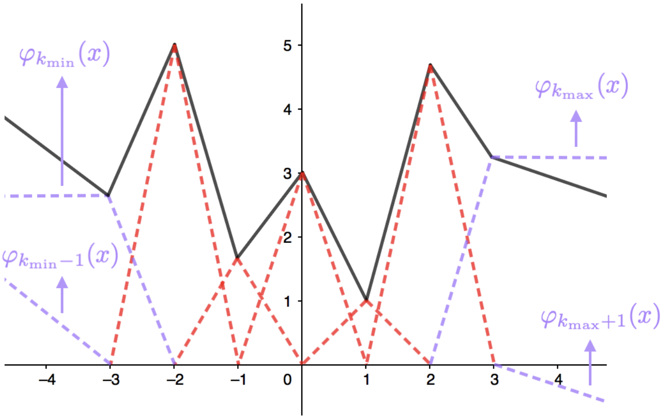


Figure 3.13: Decomposition of a deep spline activation function (solid line) in terms of B-spline basis functions (dashed lines), as expressed by (3.163) with $T = 1$. The basis is composed of $(K - 2)$ triangular functions, which are compactly supported and shifted replicates of each other, plus 4 one-sided outside functions. The key property is that the evaluation of $\sigma(x)$ for any fixed $x \in \mathbb{R}$ involves no more than two basis functions.

While the direct connection with ℓ_1-minimization in (3.162) is very attractive, the less favorable aspect of the model is that its computational cost is proportional to the underlying number of ReLUs (or spline knots); that is, $K = (k_{\max} - k_{\min} + 1)$, which can be arbitrarily large depending on the value of T. Here, we propose a way to bypass this limitation by switching to another equivalent but maximally localized basis: the B-splines. Our model takes the form

$$\sigma(x) = \sum_{k=k_{\min}-1}^{k_{\max}+1} c_k \varphi_k \left(\frac{x}{T} \right), \tag{3.163}$$

which involves triangular-shaped basis functions that are rescaled versions of B-splines defined on an integer grid. As illustrated in Figure 1, the central bases for $k = (k_{\min} + 1)$ to $(k_{\max} - 1)$ are shifted replicates of the compactly supported linear

B-spline

$$\varphi_k(x) = \beta_{\mathrm{D}^2}(x-k), \text{ for } k_{\min} < k < k_{max}, \tag{3.164}$$

where we recall from Section 1.1.6 that

$$\beta_{\mathrm{D}^2}(x) = (x+1)_+ - 2(x)_+ + (x-1)_+ = \begin{cases} 1-|x|, & x \in [-1,1] \\ 0, & \text{otherwise.} \end{cases} \tag{3.165}$$

The four remaining boundary basis functions are one-sided splines that allow the activation function defined in (3.163) to exhibit a linear behavior at both ends, for $x < k_{\min}T$ as well as for $x > k_{\max}T$. Specifically, we have that

$$\varphi_{k_{\min}-1}(x) = (-x+k_{\min})_+, \tag{3.166}$$

$$\varphi_{k_{\min}}(x) = (-x+k_{\min}+1)_+ - (-x+k_{\min})_+, \tag{3.167}$$

$$\varphi_{k_{\max}}(x) = (x-k_{\max}+1)_+ - (x-k_{\max})_+, \tag{3.168}$$

$$\varphi_{k_{\max}+1}(x) = (x-k_{\max})_+. \tag{3.169}$$

The B-spline model defined in (3.163) has the same knots as those of the gridded ReLU representation given by (3.162). It also has the same number of degrees of freedom; namely, $K+2 = (k_{\max}+1) - (k_{\min}-1) + 1$. By using the property that the φ_k can all be expanded in terms of integer shifts of ReLUs (see (3.164)-(3.169)), we can show that the two sets of basis functions span the same subspace. In doing so, we obtain a formula for the retrieval of the a_k and, hence, the $\mathrm{TV}^{(2)}(\sigma)$—in terms of the second-order difference of the c_k. Specifically, the relationship between the ReLU coefficients $\mathbf{a} \in \mathbb{R}^K$ and the B-spline coefficients $\mathbf{c} \in \mathbb{R}^{K+2}$ is given by

$$\begin{bmatrix} a_{k_{\min}} \\ \vdots \\ a_{k_{\max}} \end{bmatrix} = \frac{1}{T} \underbrace{\begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}}_{\mathbf{L} \in \mathbb{R}^{(K+2)\times K}} \begin{bmatrix} c_{k_{\min}-1} \\ c_{k_{\min}} \\ \vdots \\ c_{k_{\max}} \\ c_{k_{\max+1}} \end{bmatrix}, \tag{3.170}$$

while the linear-term parameters b_0, b_1 can be determined from $c_{k_{\min-1}}$ and $c_{k_{\min}}$. From (3.170), we see that the $\mathrm{TV}^{(2)}$ regularization of σ can also be computed from the B-spline coefficients as $\mathrm{TV}^{(2)}(\sigma) = \|\mathbf{Lc}\|_1$.

While the gridded ReLU and B-spline models (3.162) and (3.163) are mathematically equivalent, the advantage of (3.163) is that there are at most two active basis functions at any given point $x = x_0$, independently of the step size T. This has important implications for the efficiency and scalability of both the evaluation of the DNN at a given point $\mathbf{x}_m$ and the computation of its gradient with respect to c_k (as opposed to a_k in the equivalent ReLU representation).

3.4.5 Numerical Illustration

Here, we discuss the practical aspects of our framework and conduct numerical experiments in which we compare the performance of our method to ReLU neural networks and its variants: LeakyReLU and PReLU activations.

Setup

We designed a simple experiment in which the goal is to classify points that are inside a circle of area 2 centred at the origin. This is a classical two-dimensional supervised-learning problem, where the target function is

$$\mathbb{1}_{\mathrm{Circle}}(x_1, x_2) = \begin{cases} 1, & x_1^2 + x_2^2 \leq \frac{2}{\pi} \\ 0, & \text{otherwise.} \end{cases} \tag{3.171}$$

The training dataset is obtained by generating $M = 1000$ random points from a uniform distribution on $[-1, 1]^2$. The points that lie inside and outside of the circle are then labeled as 1 and 0, respectively.

To illustrate the effect of our proposed scheme, we consider a family of fully connected architectures with layer descriptors of the form $(2, 2W, 1)$, where the width parameter $W \in \mathbb{N}$ governs the complexity of the architecture. We follow the classical choice of using a sigmoid activation in the last layer, together with the binary cross-entropy loss

$$E(y, \widehat{y}) = -y \log(\widehat{y}) - (1 - y) \log(1 - \widehat{y}). \tag{3.172}$$

We take $\mu_1 = \mu_2 = \mu$ and, in each scheme, we tune the single hyper-parameter $\mu > 0$.

In our Lipschitz-based design, we set $K = 21$ spline knots per activation functions. Moreover, we use Xavier's rule [215] to initialize the linear weights. For the activations, we consider the simple piecewise-linear functions *absolute value* and *soft-thresholding*, defined as

$$f_{\mathrm{abs}}(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0, \end{cases} \tag{3.173}$$

$$f_{\mathrm{soft}}(x) = \begin{cases} x - \frac{1}{2}, & x \geq \frac{1}{2} \\ 0, & x \in (-\frac{1}{2}, \frac{1}{2}) \\ x + \frac{1}{2}, & x \leq -\frac{1}{2}. \end{cases} \tag{3.174}$$

We then initialize half of the activations with f_{abs} and the other half with f_{soft}. Intuitively, such initializations may allow the network to be flexible to both even and odd functions.

Moreover, we deploy Theorem 3.12 to tune the parameter λ. A direct calculation reveals that

$$\|f_{\mathrm{abs}}\|_{\mathrm{BV}^{(2)}} = 3, \qquad \|f_{\mathrm{soft}}\|_{\mathrm{BV}^{(2)}} = \frac{5}{2}. \tag{3.175}$$

This allows us to tune λ so that the optimality condition (3.152) holds in the initial setup. Due to the Xavier initialization, the linear weights of the second layer have variance $\sigma^2 = 2/(2W + 1)$. Therefore, we obtain that

$$\lambda = \frac{16}{11(2W + 1)}\mu. \tag{3.176}$$

For an informed comparison, we also count the total number of parameters that is used in each scheme to represent the learned function. More specifically, with the layer descriptor $(2, 2W, 1)$, there are $6W$ linear weights and one bias for the last (sigmoidal) activation. In addition, there are parameters that depend on the specific activation we are using: There is a bias parameter in ReLU and LeakyReLU activations. In addition to bias, PReLU activation has an extra parameter (the slope in the negative part of the real line) as well and finally, the number of parameters in our scheme is adaptive and is equal to the number of active ReLUs plus the null-space coefficients in the representation (3.141).

Comparison with ReLU-Based Activation Functions

We display in Figure 3.14 the learned function $f : \mathbb{R}^2 \to \mathbb{R}$ in each case. We also disclose in Table 3.2 the performance and the number of active parameters of each scheme. One verifies that our scheme, already in the simplest configuration with layer descriptor $(2, 2, 1)$, outperforms all other methods, even when they are deployed over the richer architecture $(2, 10, 1)$. Moreover, there are fewer parameters in the final representation of the target function in our scheme, as compared to the other methods. This experiment shows that the learning of activations in simple architectures is beneficial as it compensates the low capacity of the network and contributes to the generalization power of the global learning scheme.

In the minimal case $W = 1$, we expect the network to learn parabola-type activations. This is due to the fact that the target function can be represented as

$$\mathbb{1}_{\mathrm{Circle}}(x_1, x_2) = \mathbb{1}_{[0, \frac{2}{\pi}]}(x_1^2 + x_2^2), \tag{3.177}$$

which is the composition of the sum of two parabolas and a threshold function. To verify this intuition, we have also plotted the learned activations for the case $W = 1$ in Figure 3.15.

Sparsity-Promoting Effect of $\mathrm{BV}^{(2)}$-Regularization

Despite allowing a large number of ReLUs in the expansion (3.141) ($K = 21$), the learned activations (see Figure 3.15) have sparse expansion in the ReLU basis. This is due to the sparsity-promoting effect of the $\mathrm{BV}^{(2)}$-norm on the ReLU coefficients and also the thresholding step that we added at the end of training.

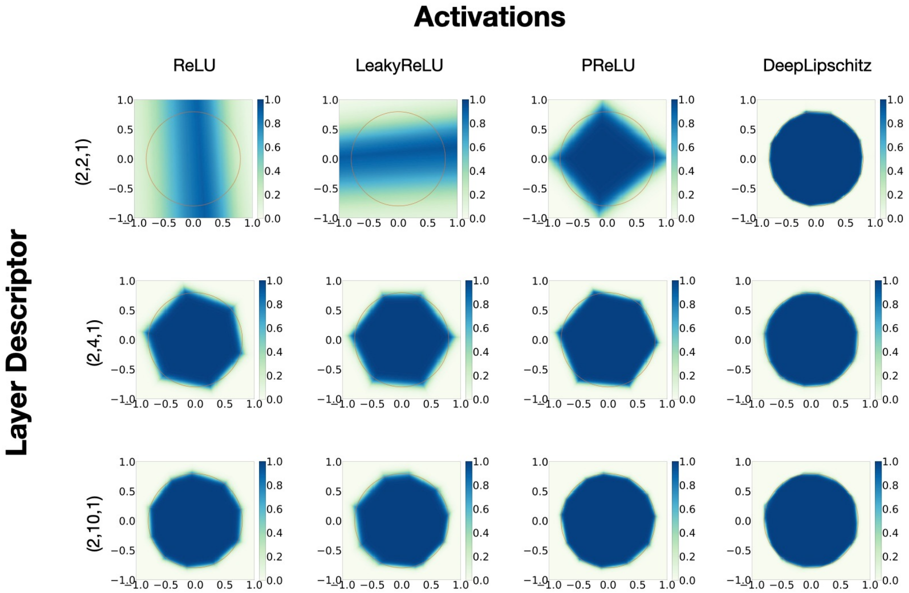


Figure 3.14: Area classification with different activations, namely ReLU, LeakyReLU, PReLU, and our proposed scheme, which we refer to as *Deep Lipschitz*. In each case, we consider $W = 1, 2, 5$ hidden neurons.

Effect of the Parameter λ

To investigate the effect of the parameter λ in our experiments, we have set the weight decay parameter to $\mu = 10^{-4}$ and plotted in Figure 3.16 the error rate, our proposed Lipschitz bound, and the total number of active ReLUs versus λ. As expected, the sparsity and Lipschitz regulariy of the network increases with λ. Consequently, one can control the overall regularity/complexity of the network by tuning this parameter.

As for the error rate, a definite transition occurs as λ varies. This suggests a

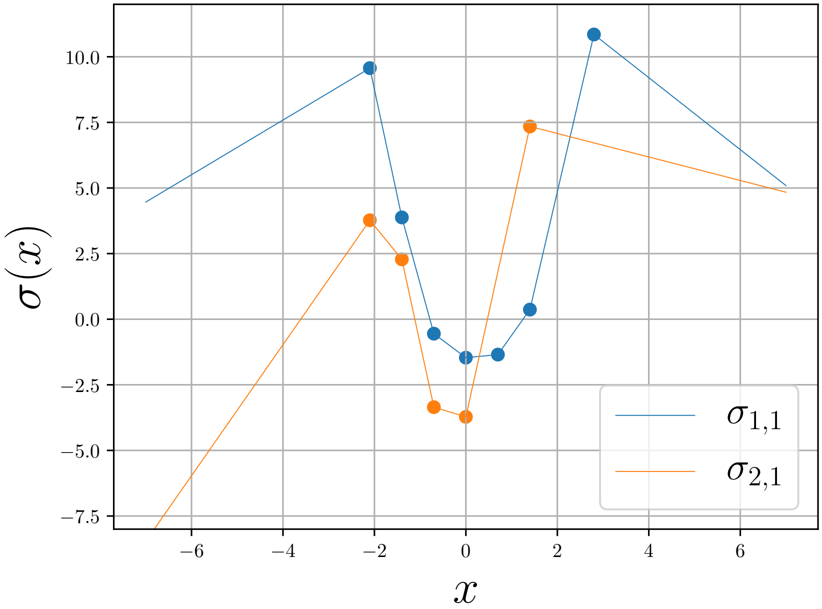


Figure 3.15: Learned activations in the area classification experiment for a simple network with layer descriptor $(2, 2, 1)$.

range of "proper" values of λ (in this case, $\lambda < 10^{-3}$) in which the error would not change much. The critical value $\lambda = 10^{-3}$ is certainly the best choice, since it has a small error and, in addition, the overall network is maximally regularized in the sense of Lipschitz. However, one is required to compute these curves for each value of μ to find the optimal λ, which can be time consuming. A heuristic (but faster) approach is to honor (3.176). In this case, it yields $\lambda \approx 0.5 \times 10^{-5}$, which lies within the favourable range of each plot.

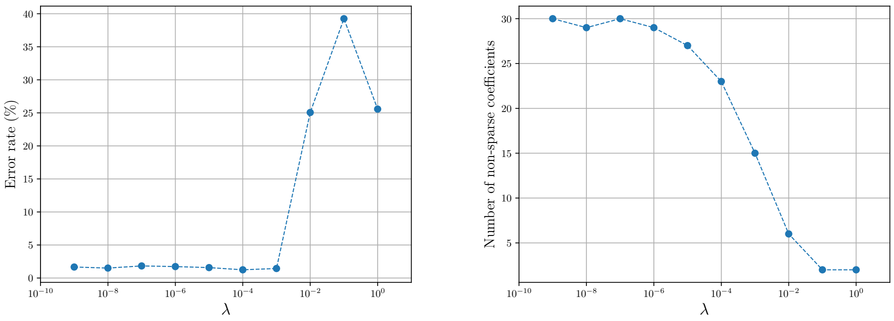


Figure 3.16: Error rate (left) and the total number of nonzero ReLU coefficients (right) versus λ in the simple architecture with layer descriptor $(2, 2, 1)$.

Table 3.2: Number of parameters and Performance in the area classification experiment.

	Architecture	N_{param}	Performance
ReLU	$(2, 10, 1)$	41	98.15
LeakyReLU	$(2, 10, 1)$	41	98.12
PReLU	$(2, 10, 1)$	51	98.19
Deep Lipschitz	$(2, 2, 1)$	**23**	**98.54**

Effect of the Parameter K

Until now, we have performed all experiments with $K = 21$ spline knots. In this section, we let K vary and examine how this effects. We consider the area classification problem and train a simple neural network with layer descriptor $(2, 2, 1)$.

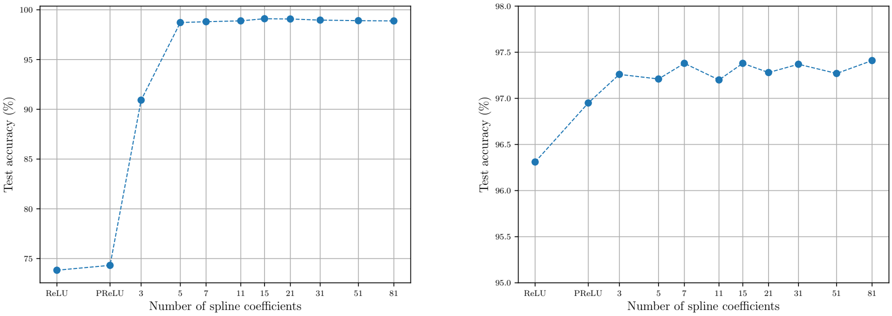


Figure 3.17: The performance versus the number K of spline knots of each activation functions in the area classification (left) and in the MNIST experiment (right).

We also perform this experiment on the MNIST dataset [216] that consists of 28×28 grayscale images of digits from 0 to 9. In this case, we used a neural network that consists of three blocks. The first two are each composed of three layers: 1) a convolutional layer with a filter of size 5×5 and two output channels; 2) a nonlinear layer that has shared activations across each output channels (two activations in each layer are being learned); and 3) a max-pooling layer with kernel and stride of size 2. The third block is composed of a fully connected layer with output of size 10 followed by soft-max. The output of the network represents the probability of each digit.

The results are depicted in Figure 3.17. In both cases, we also indicated the performance of ReLU and PReLU for comparison. We observe that the performance monotonically increases with K until it reaches saturation. We conclude that, although finding the best value for K is challenging, suboptimal value still leads to substantial improvements in the performance of the network and typically to better performances than ReLU networks and its variants.

3.4.6 Summary

We have introduced a variational framework to learn the activations of a deep neural network while controlling its global Lipschitz regularity. We have considered neural networks with second-order bounded-variation activations and we provided a global bound for their Lipschitz constants. We have showed that the solution of our proposed variational problem exists and is in the form of a deep-spline network with continuous piecewise linear activation functions. Finally, we proposed an efficient algorithm based on B-splines that is scalable in both time and memory.

3.5 Learning Multivariate CPWL Functions

Our main objective in this section is to propose a functional framework for learning multivariate CPWL functions. Our learning scheme is based on a novel Hessian-based seminorm that quantifies the total "rugosity" of multivariate functions. Our contributions in this section are threefold:

1. We fully characterize the duality mapping over the space of matrices that are equipped with Schatten norms (Section 3.5.1). This is motivated by the centrality of the family of Schatten norms in our proposed framework. Our approach is based on the analysis of the saturation of the Hölder inequality for Schatten norms. We prove in our main result that, for $p \in (1, \infty)$, the duality mapping over the space of real-valued matrices with Schatten-p norm is a continuous and single-valued function. Moreover, we provide an explicit form for its computation. For the special case $p = 1$, the mapping is set-valued; by adding a rank constraint, we show that it can be reduced to a Borel-measurable single-valued function for which we also provide a closed-form expression.

2. We define the Hessian-Schatten total variation (HTV) seminorm by specifying the adequate matrix-valued Banach spaces that are equipped with suitable classes of mixed norms (Section 3.5.2). We demonstrate that the HTV properly assesses the complexity of supervised-learning schemes. In particular, we show that the HTV is invariant to rotations, scalings, and translations. Additionally, its minimum value is achieved for linear mappings, which supports the common intuition that linear regression is the least complex learning model. We also present closed-form expressions of the HTV of CPWL functions, where we show that the HTV reflects the total change in slopes between linear regions that have a common facet. Hence, it can be viewed as a convex relaxation (ℓ_1-type) of the number of linear regions (ℓ_0-type) of CPWL mappings.

3. Finally, we demonstrate the practical aspects of learning with HTV seminorm by focusing on 2D mappings (Section 3.5.3). Motivated by the nature of the regularizer, we restrict the search space to the span of piecewise-linear box splines shifted on a lattice. Our formulation of the infinite-dimensional problem on this search space allows us to recast it exactly as a finite-dimensional one

that can be solved using standard methods in convex optimization. We validate our framework in various experimental setups and compare it with neural networks.

3.5.1 Schatten Duality Mapping

In linear algebra and matrix analysis, Schatten norms are a family of spectral matrix norms that are defined via the singular-value decomposition [217]. They have appeared in many applications such as image reconstruction [218, 219], image denoising [220], and tensor decomposition [221], to name a few.

Generally, the Schatten-p norm of a matrix is the ℓ_p norm of its singular values [222]. The family contains some well-known matrix norms: the Frobenius and the spectral (operator) norms are special cases in the family, with $p = 2$ and $p = \infty$, respectively. The case $p = 1$ (trace or nuclear norm) is of particular interest for applications as it can be used to recover low-rank matrices [223]. This is the current paradigm in matrix completion, where the goal is to recover an unknown matrix given some of its entries [224]. Prominent examples of applications that can be reduced to low-rank matrix-recovery problems are phase retrieval [225], sensor-array processing [226], system identification [227], and index coding [228, 229].

In addition to their many applications in data science, Schatten norms have been extensively studied from a theoretical point of view. Various inequalities concerning Schatten norms have been proven [230, 231, 232, 233, 234, 235, 236, 237, 238]; sharp bounds for commutators in Schatten spaces have been given [239, 240]; moreover, facial structure [241], Fréchet differentiablity [242], and various other aspects [243, 244] have been studied already.

Our objective in this work[7] is to investigate the duality mapping in spaces of matrices that are equipped with Schatten norms. We first recall the notion of duality mapping for finite-dimensional vector spaces (see, Chapter 2 for a more general definition).

Definition 3.5. *Let V be a finite-dimensional vector space and let $(\|\cdot\|_X, \|\cdot\|_{X'})$ be a pair of dual norms that are defined over V. The pair $(\mathbf{u}, \mathbf{v}) \in V \times V$ is said to be a (X, X')-conjugate, if*

- $\langle \mathbf{v}, \mathbf{u} \rangle = \|\mathbf{v}\|_{X'} \|\mathbf{u}\|_X$,

[7] From our published work [108].

- $\|\mathbf{v}\|_{X'} = \|\mathbf{u}\|_X$.

For any $\mathbf{u} \in V$, the set of all elements $\mathbf{v} \in V$ such that $(\mathbf{u}, \mathbf{v})$ forms an (X, X')-conjugate is denoted by $\mathcal{J}_X(\mathbf{u}) \subseteq V$. We refer to the set-valued mapping $\mathcal{J}_X : V \to 2^V$ as the duality mapping. If, for all $\mathbf{u} \in V$, the set $\mathcal{J}_X(\mathbf{u})$ is a singleton, then we indicate the duality mapping for the X-norm via the single-valued function $\mathrm{J}_X : V \to V$ with $\mathcal{J}_X(\mathbf{u}) = \{\mathrm{J}_X(\mathbf{u})\}$.

The duality mapping is a powerful tool to understand the topological structure of Banach spaces [95, 96]. It has been used to derive powerful characterizations of the solution of variational problems in function spaces [191, 75] (*e.g.*, Theorem 2.1) and also to determine generalized linear inverse operators [245]. Here, we prove that the duality mapping over Schatten-p spaces with $p \in (1, +\infty)$ is a single-valued and continuous function which, in fact, highlights the strict convexity of these spaces. Although the provided characterization is intuitive, we could not find it in the literature and this is, to the best of our knowledge, the first work which provides a direct way of computing this mapping in this case. For the special case $p = 1$, the mapping is set-valued. However, we prove that, by adding a rank constraint, it reduces to a single-valued Borel-measurable function. In both cases, we also derive closed-form expressions that allow one to compute them explicitly.

Duality Mapping for ℓ_p-Norms

We recall that for any $p \in [1, +\infty]$, the ℓ_p-norm of a vector $\mathbf{u} = (u_i) \in \mathbb{R}^n$ is defined as

$$\|\mathbf{u}\|_p = \begin{cases} \left(\sum_{i=1}^n |u_i|^p\right)^{\frac{1}{p}}, & p < +\infty \\ \max_i |u_i|, & p = +\infty. \end{cases} \tag{3.178}$$

It is widely known that the dual norm of ℓ_p is the ℓ_q-norm, where (p, q) are Hölder conjugates (*i.e.*, $1/p + 1/q = 1$) [94]. This stems from the Hölder inequality which states that

$$\langle \mathbf{v}, \mathbf{u} \rangle \leq \|\mathbf{u}\|_p \|\mathbf{v}\|_q, \tag{3.179}$$

for all $\mathbf{u} = (u_i), \mathbf{v} = (v_i) \in \mathbb{R}^n$. In the sequel, we exclude the trivial cases $\mathbf{u} = \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$ to avoid unnecessary complexities in our statements.

When $1 < p < +\infty$, Inequality (3.179) is saturated if and only if $u_i v_i \geq 0$ for $i = 1, \dots, n$ and there exists a constant $c > 0$ such that $|\mathbf{u}|^p = c|\mathbf{v}|^q$, where $|\mathbf{u}|^p = (|u_i|^p)$. This ensures that the duality mapping is single-valued and also yields the map

$$\mathrm{J}_p(\mathbf{u}) = \mathrm{sign}(\mathbf{u}) \frac{|\mathbf{u}|^{p-1}}{\|\mathbf{u}\|_p^{p-2}}. \tag{3.180}$$

For $p = 1$, one can verify that the equality happens if and only if, for any index $i = 1, \dots, n$ with $u_i \neq 0$, one has that

$$v_i = \mathrm{sign}(u_i)\|\mathbf{v}\|_\infty. \tag{3.181}$$

In other words, the vector $\mathbf{v}$ should attain its extreme values at places where $\mathbf{u}$ has nonzero values, with the sign being determined by the corresponding element in $\mathbf{u}$.

Due to (3.181), the set $\mathcal{J}_1(\mathbf{u})$ is not necessarily a singleton. However, if we add an additional sparsity constraint, then the mapping becomes single-valued. This leads us to introduce the new notion of *sparse duality mapping* in Definition 3.6.

Definition 3.6. *Let V be a finite-dimensional vector space and let $s_0 : V \to \mathbb{N}$ be an integer-valued function that acts as a sparsity measure. Assuming a pair $(\|\cdot\|_X, \|\cdot\|_{X'})$ of dual norms over V, we call the pair $(\mathbf{u}, \mathbf{v}) \in V \times V$ a sparse (X, X')-conjugate if*

- $(\mathbf{u}, \mathbf{v})$ *forms an (X, X')-conjugate pair. In other words, $\mathbf{v} \in \mathcal{J}(\mathbf{u})$.*
- *The quantity $s_0(\mathbf{v})$ attains its minimal value over the set $\mathcal{J}(\mathbf{u})$.*

We denote the set of sparse conjugates of $\mathbf{u}$ by $\mathcal{J}_{X,s_0}(\mathbf{u})$. Whenever $\mathcal{J}_{X,s_0}(\mathbf{u})$ is a singleton for any $\mathbf{u} \in V$, we refer to the single-valued function $\mathrm{J}_{X,s_0} : V \to V$ with $\mathcal{J}_{X,s_0}(\mathbf{u}) = \{\mathrm{J}_{X,s_0}(\mathbf{u})\}$ as the sparse duality mapping.

Following Definition 3.6, if we use the ℓ_0-norm as the sparsity measure, that is

$s_0(\mathbf{u}) = \|\mathbf{u}\|_0 = \operatorname{Card}(\{i : u_i \neq 0\})$[8], then we have the sparse duality mapping

$$\begin{aligned} &\mathrm{J}_{1,0} : \mathbb{R}^n \to \mathbb{R}^n : \mathbf{u} = (u_i) \mapsto \mathbf{v} = (v_i) = \mathrm{J}_{1,0}(\mathbf{u}), \\ &v_i = \begin{cases} \operatorname{sign}(u_i)\|\mathbf{u}\|_1, & u_i \neq 0 \\ 0, & u_i = 0. \end{cases} \end{aligned} \tag{3.182}$$

Finally, we mention that, for $p = +\infty$, the reduced set $\mathcal{J}_{\infty,0}$ is not single-valued. Indeed, let us define $I_{\max}(\mathbf{u}) = \{i : |u_i| = \|\mathbf{u}\|_\infty\} \subseteq \{1, \ldots, n\}$. We readily deduce from (3.181) that $\mathbf{v} = (v_1, \ldots, v_n) \in \mathcal{J}_\infty(\mathbf{u})$ if and only if $v_i = 0$ whenever $i \notin I_{\max}(\mathbf{u})$ and $\operatorname{sign}(v_i) = \operatorname{sign}(u_i)$ for $i \in I_{\max}(\mathbf{u})$ with $\sum_{i \in I_{\max}(\mathbf{u})} |v_i| = \|\mathbf{u}\|_\infty$. This shows that $\mathcal{J}_\infty(\mathbf{u})$ is a convex set with $\mathcal{J}_{\infty,0}(\mathbf{u})$ being its extreme points, where $\mathcal{J}_{\infty,0}(\mathbf{u}) = \{u_i \mathbf{e}_i : i \in I_{\max}(\mathbf{u})\}$.

Schatten Matrix Norms

It is widely known that any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{3.183}$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\mathbf{S}$ is an m by n rectangular diagonal matrix with nonnegative real entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ sorted in descending order [246]. In the literature, (3.183) is known as the singular-value decomposition (SVD) and the entries σ_i are the singular values of $\mathbf{A}$. In general, the SVD of a matrix $\mathbf{A}$ is not unique. However, the diagonal matrix $\mathbf{S}$ and, consequently, its entries, are fully determined from $\mathbf{A}$. In other words, the values of σ_i are invariant to a specific choice of decomposition. This is why one can refer to the diagonal entries of $\mathbf{S}$ as the "singular values" of $\mathbf{A}$.

When $\mathbf{A}$ is not full rank, one can obtain a reduced version of (3.183). Indeed, if we denote the rank of $\mathbf{A}$ by r, then we have that

$$\mathbf{A} = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T, \tag{3.184}$$

[8] Although this functional does not satisfy the homogeneity property of a norm, it has been widely referred to as the ℓ_0-norm.

where $\mathbf{U}_r \in \mathbb{R}^{m\times r}$ and $\mathbf{V}_r \in \mathbb{R}^{n\times r}$ are (sub)-orthogonal matrices such that $\mathbf{U}_r^T\mathbf{U}_r = \mathbf{V}_r^T\mathbf{V}_r = \mathbf{I}_r$ and $\mathbf{S}_r = \mathrm{diag}(\boldsymbol{\sigma})$ is a diagonal matrix that contains positive singular values $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_r) \in \mathbb{R}^r$ of $\mathbf{A}$.

For any $p \in [1, +\infty]$, the Schatten-p norm of $\mathbf{A}$ is defined as

$$\|\mathbf{A}\|_{S_p} = \begin{cases} \left(\sum_{i=1}^r \sigma_i^p\right)^{\frac{1}{p}}, & p < +\infty \\ \sigma_1, & p = +\infty. \end{cases} \tag{3.185}$$

We remark that (3.185) defines a family of quasi norms for $p \in (0, 1)$. In the extreme case $p = 0$, the Schatten-0 norm actually coincides with the rank of the matrix, *i.e.* $\|\mathbf{A}\|_{S_0} = \mathrm{rank}(\mathbf{A})$. The Schatten quasi norms have also been studied in the literature from both theoretical and practical point of views (see, [247, 248, 249, 250], and references therein).

For any $p \in [1, \infty]$, the dual of the Schatten-p norm is the Schatten-q norm, where $q \in [1, \infty]$ is such that $\frac{1}{p} + \frac{1}{q} = 1$ [217]. This is due to the generalized version of Hölder's inequality for Schatten norms, as stated in Proposition 3.4.

Proposition 3.4. *For any pair* $(p, q) \in [1, +\infty]^2$ *of Hölder conjugates with* $\frac{1}{p} + \frac{1}{q} = 1$ *and any pair of matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$*, we have that*

$$\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{Tr}\left(\mathbf{A}^T\mathbf{B}\right) \le \|\mathbf{A}\|_{S_p}\|\mathbf{B}\|_{S_q}. \tag{3.186}$$

While this is a known result (see, for example, [251]), it is also the basis for the present work, which is the reason why we provide its proof for the sake of completeness.

Proof. Let us recall the reduced SVD of the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T, \tag{3.187}$$

where $r = \mathrm{rank}(\mathbf{A})$, $\mathbf{U}_r = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathbb{R}^{m\times r}$, $\mathbf{V}_r = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathbb{R}^{n\times r}$, and $\mathbf{S} = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$. Similarly, for the matrix $\mathbf{B}$, we have that

$$\mathbf{B} = \tilde{\mathbf{U}}_{\tilde{r}}\tilde{\mathbf{S}}_{\tilde{r}}\tilde{\mathbf{V}}_{\tilde{r}}^T, \tag{3.188}$$

where $\tilde{r} = \operatorname{rank}(\mathbf{A})$, $\tilde{\mathbf{U}}_{\tilde{r}} = [\tilde{\mathbf{u}}_1 \cdots \tilde{\mathbf{u}}_{\tilde{r}}] \in \mathbb{R}^{m\times\tilde{r}}$, $\tilde{\mathbf{V}}_{\tilde{r}} = [\tilde{\mathbf{v}}_1 \cdots \tilde{\mathbf{v}}_r] \in \mathbb{R}^{n\times\tilde{r}}$, and $\tilde{\mathbf{S}} = \operatorname{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{\tilde{r}})$. A direct computation then reveals that

$$\operatorname{Tr}\left(\mathbf{A}^T\mathbf{B}\right) = \sum_{i=1}^{r}\sum_{j=1}^{\tilde{r}} \sigma_i\tilde{\sigma}_j \mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j. \tag{3.189}$$

By using the weighted Hölder inequality for vectors [252], we obtain for $p \neq 1$ that

$$\sum_{i=1}^{r}\sum_{j=1}^{\tilde{r}} \sigma_i\tilde{\sigma}_j \mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j \leq \left(\sum_{i=1}^{r}\sigma_i^p\sum_{j=1}^{\tilde{r}}|\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j|\right)^{\frac{1}{p}}\left(\sum_{j=1}^{\tilde{r}}\sigma_j^p\sum_{i=1}^{r}|\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j|\right)^{\frac{1}{q}} \tag{3.190}$$

and for $p = 1$ that

$$\sum_{i=1}^{r}\sum_{j=1}^{\tilde{r}} \sigma_i\tilde{\sigma}_j \mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j \leq \left(\sum_{i=1}^{r}\sigma_i\sum_{j=1}^{\tilde{r}}|\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j|\right)\|\tilde{\boldsymbol{\sigma}}\|_\infty. \tag{3.191}$$

Finally, by invoking Cauchy-Schwartz and the orthonormality of the matrices $\mathbf{U}_r, \mathbf{V}_r, \tilde{\mathbf{U}}_{\tilde{r}}, \tilde{\mathbf{V}}_{\tilde{r}}$, we deduce for $i = 1, \dots, r$ that

$$\sum_{j=1}^{\tilde{r}}|\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j| \leq \left(\sum_{j=1}^{\tilde{r}}(\mathbf{u}_i^T\tilde{\mathbf{u}}_j)^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{\tilde{r}}(\mathbf{v}_i^T\tilde{\mathbf{v}}_j)^2\right)^{\frac{1}{2}} \leq \|\mathbf{u}_i\|_2\|\mathbf{v}_i\|_2 = 1, \tag{3.192}$$

For $j = 1, \dots, \tilde{r}$, we deduce that

$$\sum_{i=1}^{r}|\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j| \leq \left(\sum_{i=1}^{r}(\mathbf{u}_i^T\tilde{\mathbf{u}}_j)^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{r}(\mathbf{v}_i^T\tilde{\mathbf{v}}_j)^2\right)^{\frac{1}{2}} \leq \|\tilde{\mathbf{u}}_i\|_2\|\tilde{\mathbf{v}}_i\|_2 = 1. \tag{3.193}$$

The combination of these inequalities completes the proof. □

Main Results

In Proposition 3.5, we investigate the case where the Hölder inequality is saturated, in the sense that

$$\operatorname{Tr}\left(\mathbf{A}^T\mathbf{B}\right) = \|\mathbf{A}\|_{S_p}\|\mathbf{B}\|_{S_q}. \tag{3.194}$$

This saturation is central to our work, as it is tightly linked to the notion of duality mapping.

Proposition 3.5. *Let (p, q) be a pair of Hölder conjugates and let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ be a pair of nonzero matrices with reduced SVDs of the form*

$$\mathbf{A} = \mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T, \quad \mathbf{B} = \tilde{\mathbf{U}}_{\tilde{r}} \mathrm{diag}(\tilde{\boldsymbol{\sigma}})\tilde{\mathbf{V}}_{\tilde{r}}^T. \tag{3.195}$$

- *If $p \in (1, \infty)$, then the Hölder inequality is saturated if and only if we have that*

$$\mathbf{B} = c\mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T \tag{3.196}$$

  or, equivalently,

$$\mathbf{A} = c^{-1}\tilde{\mathbf{U}}_{\tilde{r}} \mathrm{diag}(\mathrm{J}_q(\tilde{\boldsymbol{\sigma}}))\tilde{\mathbf{V}}_{\tilde{r}}^T, \tag{3.197}$$

  where $c = \frac{\|\mathbf{B}\|_{S_q}}{\|\mathbf{A}\|_{S_p}}$ and $\mathrm{J}_p(\cdot)$ and $\mathrm{J}_q(\cdot)$ are the duality mappings for the ℓ_p and ℓ_q norms, respectively (see (3.180)*).*

- *If $p = 1$, then a necessary condition for the saturation of the Hölder inequality is that*

$$\mathrm{rank}(\mathbf{A}) \leq r_1 \leq \mathrm{rank}(\mathbf{B}), \tag{3.198}$$

  where $r_1 = \mathrm{Card}\left(\{i : \tilde{\sigma}_i = \tilde{\sigma}_1\}\right)$ is the multiplicity of the first singular value of $\mathbf{B}$. Moreover, if we denote the first r_1 singular vectors of $\mathbf{B}$ in (3.195) *by $\tilde{\mathbf{U}}_1 \in \mathbb{R}^{m\times r_1}$ and $\tilde{\mathbf{V}}_1 \in \mathbb{R}^{n\times r_1}$, then the Hölder inequality is saturated if and only if there exists a symmetric matrix $\mathbf{X} \in \mathbb{R}^{r_1\times r_1}$ such that*

$$\mathbf{A} = \tilde{\mathbf{U}}_1\mathbf{X}\tilde{\mathbf{V}}_1^T. \tag{3.199}$$

  Finally in the rank-equality case $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$*, we have saturation if and only if*

$$\mathbf{B} = c\mathbf{U}_r\mathbf{V}_r^T, \tag{3.200}$$

  where $c = \|\mathbf{B}\|_{S_\infty}$ and the matrices $\mathbf{U}_r$ and $\mathbf{V}_r$ are defined in (3.195)*.*

Proof. We separate the two cases and analyze each one independently.

Case 1: $1 < p < +\infty$. We prove (3.196) and deduce (3.197) by symmetry. Following the proof of Proposition 3.5 and considering the reduced SVD of the matrices **A** and **B** given in (3.187) and (3.188), we immediately see that the inequalities (3.190), (3.192), and (3.193) should all be saturated. The equality condition of the weighted Hölder implies the existence of a positive constant $\alpha > 0$ such that, for all $(i, j) \in \{1, \dots, r\} \times \{1, \dots, \tilde{r}\}$, we have one of the following conditions:

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j = 0, \quad \text{or} \tag{3.201}$$

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j > 0 \quad \text{and} \quad \tilde{\sigma}_j^q = \alpha \sigma_i^p. \tag{3.202}$$

Moreover, the saturation of (3.192) implies that

$$\mathbf{u}_i \in \text{Range}(\tilde{\mathbf{U}}_{\tilde{r}}), \quad \mathbf{v}_i \in \text{Range}(\tilde{\mathbf{V}}_{\tilde{r}}) \qquad \forall i = 1, \dots, r \tag{3.203}$$

and also that there exists a positive constant $\beta_i > 0$ (positivity follows from (3.202) and (3.201)) such that

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j = \beta_i \mathbf{v}_i^T \tilde{\mathbf{v}}_j, \quad \forall j = 1, \dots, \tilde{r}. \tag{3.204}$$

However, from the normality of $\mathbf{u}_i$ and (3.203), we have that

$$1 = \|\mathbf{u}_i\|_2^2 = \sum_{j=1}^{\tilde{r}} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j|^2 = \beta_i^2 \sum_{j=1}^{\tilde{r}} |\mathbf{v}_i^T \tilde{\mathbf{v}}_j|^2 = \beta_i^2 \|\mathbf{v}\|_2^2 = \beta_i^2 \tag{3.205}$$

which, together with the positivity of β_i, leads to the conclusion that $\beta_i = 1$ for $i = 1, \dots, r$. Using this, we rewrite (3.204) in matrix form as

$$\mathbf{U}_r^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T \tilde{\mathbf{V}}_{\tilde{r}}. \tag{3.206}$$

Similarly, the saturation of (3.193) implies that

$$\tilde{\mathbf{u}}_j \in \text{Range}(\mathbf{U}_r), \quad \tilde{\mathbf{v}}_i \in \text{Range}(\mathbf{V}_r), \tag{3.207}$$

for all $j = 1, \dots, \tilde{r}$. Putting together (3.203) and (3.207), we deduce that $r = \tilde{r}$ and

$$\text{Range}(\mathbf{U}_r) = \text{Range}(\tilde{\mathbf{U}}_{\tilde{r}}), \quad \text{Range}(\mathbf{V}_r) = \text{Range}(\tilde{\mathbf{V}}_{\tilde{r}}). \tag{3.208}$$

This implies the existence of two orthogonal matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{r \times r}$ such that

$$\tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{U}_r \mathbf{P}, \quad \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{V}_r \mathbf{Q}. \tag{3.209}$$

However, replacing (3.209) in (3.206), we conclude that

$$\mathbf{P} = \mathbf{U}_r^T \mathbf{U}_r \mathbf{P} = \mathbf{U}_r^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{V}_r^T \mathbf{V}_r \mathbf{Q} = \mathbf{Q}. \tag{3.210}$$

This implies that the matrix $\mathbf{B}$ can be represented as

$$\mathbf{B} = \mathbf{U}_r \mathbf{P} \tilde{\mathbf{S}}_{\tilde{r}} \mathbf{P}^T \mathbf{V}_r^T = \mathbf{U}_r \mathbf{S}_0 \mathbf{V}_r^T, \tag{3.211}$$

where $\mathbf{S}_0 = \mathbf{P}\tilde{\mathbf{S}}_{\tilde{r}}\mathbf{P}^T$. We now show that $\mathbf{S}_0$ is a diagonal matrix. Indeed, by denoting the (i,j)-th entry of $\mathbf{P}$ as $p_{i,j}$ such that $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_r] = [p_{i,j}]$, we rewrite (3.201) and (3.202) as

$$p_{i,j} = 0, \quad \text{or} \tag{3.212}$$

$$p_{i,j} > 0 \quad \text{and} \quad \tilde{\sigma}_j^q = \alpha \sigma_i^p, \tag{3.213}$$

for all $(i,j) \in \{1,\dots,r\}^2$. Moreover, by expanding the (i,j)-th entry of the matrix $\mathbf{S}_0$, we have that

$$\begin{aligned}
[\mathbf{S}_0]_{i,j} = [\mathbf{P}\tilde{\mathbf{S}}_{\tilde{r}}\mathbf{P}^T]_{i,j} &= \sum_{k=1}^r p_{i,k}\tilde{\sigma}_k p_{j,k} = \sum_{k=1}^r p_{i,k}\sigma_i^{\frac{p}{q}}\alpha^{\frac{1}{q}} p_{j,k} \\
&= \sigma_i^{\frac{p}{q}}\alpha^{\frac{1}{q}}\mathbf{p}_i^T\mathbf{p}_j = [\mathrm{J}_p(\boldsymbol{\sigma})]_i c_{\mathbf{B}}\delta[i-j],
\end{aligned} \tag{3.214}$$

where $\delta[\cdot]$ denotes the Kronecker delta and $c_{\mathbf{B}} = \alpha^{\frac{1}{q}} > 0$ is a positive constant. Finally, we obtain the announced expression in (3.196) by replacing the above characterization of $\mathbf{S}_0$ in (3.211).

For the converse, we note that, if the matrix $\mathbf{B}$ is in the form of (3.196), then we have that

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{A}^T\mathbf{B}\right) &= \mathrm{Tr}\left(\mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T \left(\mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T\right)^T\right) \\
&= c_{\mathbf{B}}\mathrm{Tr}\left(\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T\mathbf{V}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{U}_r^T\mathbf{U}_r\right) \\
&= c_{\mathbf{B}}\boldsymbol{\sigma}^T\mathrm{J}_p(\boldsymbol{\sigma}) \\
&= c_{\mathbf{B}}\|\boldsymbol{\sigma}\|_p\|\mathrm{J}_p(\boldsymbol{\sigma})\|_q = \|\mathbf{A}\|_{S_p}\|\mathbf{B}\|_{S_q},
\end{aligned} \tag{3.215}$$

which shows that the equality is indeed saturated in this case.

Case 2: $p = 1$. In this case, the saturation of the weighted Hölder inequality implies that, for all $(i, j) \in \{1, \ldots, r\} \times \{1, \ldots, \tilde{r}\}$, we have that

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j = 0, \quad \text{or} \tag{3.216}$$

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j > 0 \quad \text{and} \quad \tilde{\sigma}_j = \tilde{\sigma}_1. \tag{3.217}$$

For equality, we also need to have the saturation of (3.192), which we showed to be equivalent to (3.203) and (3.206). From (3.203), we deduce the existence of matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{\tilde{r} \times r}$ such that

$$\mathbf{U}_r = \tilde{\mathbf{U}}_{\tilde{r}} \mathbf{P}_1, \qquad \mathbf{V}_r = \tilde{\mathbf{V}}_{\tilde{r}} \mathbf{P}_2. \tag{3.218}$$

The replacement of these in (3.206) implies that

$$\mathbf{P}_1^T = \mathbf{P}_1^T \tilde{\mathbf{U}}_{\tilde{r}}^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{U}_r^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{P}_2^T \tilde{\mathbf{V}}_{\tilde{r}}^T \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{P}_2^T, \tag{3.219}$$

and, hence, that $\mathbf{P}_1 = \mathbf{P}_2 = [p_{i,j}] \in \mathbb{R}^{\tilde{r} \times r}$. Now, one can rewrite the conditions (3.216) and (3.217) and deduce that, for any $j = 1, \ldots, \tilde{r}$, we have that

$$p_{j,i} = 0, \quad \forall i = 1, \ldots, r \quad \text{or} \quad \tilde{\sigma}_j = \tilde{\sigma}_1. \tag{3.220}$$

From Conditions (3.220) and following the definition of r_1 (the multiplicity of the largest singular value), we deduce that

$$\mathbf{P}_1 = \begin{bmatrix} \mathbf{P} \\ \mathbf{0}_{r_{\text{res}} \times r} \end{bmatrix}, \tag{3.221}$$

where $\mathbf{P} \in \mathbb{R}^{r_1 \times r}$ and $r_{\text{res}} = (\tilde{r} - r_1)$. Using this form and the definition of $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{V}}_1$ (given in the statement of the proposition), we rewrite (3.218) as

$$\mathbf{U}_r = \tilde{\mathbf{U}}_1 \mathbf{P}, \qquad \mathbf{V}_r = \tilde{\mathbf{V}}_1 \mathbf{P}. \tag{3.222}$$

Therefore,

$$\mathbf{I}_r = \mathbf{U}_r^T \mathbf{U}_r = \mathbf{P}^T \tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_1 \mathbf{P} = \mathbf{P}^T \mathbf{P}. \tag{3.223}$$

Hence, $\mathbf{P}$ is a sub-orthogonal matrix and

$$\text{rank}(\mathbf{B}) = \tilde{r} \geq r_1 \geq \text{rank}(\mathbf{P}) \geq r = \text{rank}(\mathbf{A}). \tag{3.224}$$

The replacement of (3.222) in the reduced SVD of $\mathbf{A}$ yields the announced expression with $\mathbf{X} = \mathbf{P}\mathbf{S}\mathbf{P}^T$.

Based on the definitions of r_1, $\tilde{\mathbf{U}}_1$, and $\tilde{\mathbf{V}}_1$, we note that one can rewrite the reduced SVD of $\mathbf{B}$ as

$$\mathbf{B} = \tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T + \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T, \tag{3.225}$$

where $\tilde{\mathbf{U}}_{\text{res}} \in \mathbb{R}^{m \times r_{\text{res}}}$, $\tilde{\mathbf{S}}_{\text{res}} \in \mathbb{R}^{r_{\text{res}} \times r_{\text{res}}}$, and $\tilde{\mathbf{V}}_{\text{res}} \in \mathbb{R}^{n \times r_{\text{res}}}$ are the remaining singular values and vectors such that

$$\tilde{\mathbf{U}} = [\tilde{\mathbf{U}}_1 \quad \tilde{\mathbf{U}}_{\text{res}}], \quad \tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_1 \quad \tilde{\mathbf{V}}_{\text{res}}], \quad \tilde{\mathbf{S}} = \begin{bmatrix} \tilde{\sigma}_1 \mathbf{I}_{r_1} & \mathbf{0} \\ \mathbf{0} & \tilde{S}_{\text{res}} \end{bmatrix}. \tag{3.226}$$

Now, if $\mathbf{A}$ admits the form (3.199) and if we consider the SVD of $\mathbf{X} = \mathbf{P}\mathbf{S}\mathbf{P}^T$ (the assumption that $\mathbf{X}$ is symmetric ensures that is has an orthogonal eigendecomposition), then

$$\begin{aligned}
\operatorname{Tr}\left(\mathbf{A}^T\mathbf{B}\right) &= \operatorname{Tr}\left(\tilde{\mathbf{V}}_1 \mathbf{P}\mathbf{S}\mathbf{P}^T \tilde{\mathbf{U}}_1^T \left(\tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T + \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T\right)\right) \\
&= \tilde{\sigma}_1 \operatorname{Tr}\left(\tilde{\mathbf{V}}_1 \mathbf{P}\mathbf{S}\mathbf{P}^T \tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T\right) + \operatorname{Tr}\left(\tilde{\mathbf{V}}_1 \mathbf{P}\mathbf{S}\mathbf{P}^T \tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T\right) \\
&= \tilde{\sigma}_1 \operatorname{Tr}\left(\tilde{\mathbf{V}}_1 \mathbf{P}\mathbf{S}\mathbf{P}^T \mathbf{I}_{r_1} \tilde{\mathbf{V}}_1^T\right) + \operatorname{Tr}\left(\tilde{\mathbf{V}}_1 \mathbf{P}\mathbf{S}\mathbf{P}^T \mathbf{0}_{r_1 \times r_{\text{res}}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T\right) \\
&= \tilde{\sigma}_1 \operatorname{Tr}\left(\mathbf{S}\mathbf{P}^T \tilde{\mathbf{V}}_1^T \tilde{\mathbf{V}}_1 \mathbf{P}\right) + 0 \\
&= \tilde{\sigma}_1 \operatorname{Tr}\left(\mathbf{S}\mathbf{P}^T \mathbf{P}\right) \\
&= \tilde{\sigma}_1 \operatorname{Tr}\left(\mathbf{S}\right) = \|\mathbf{B}\|_{S_\infty} \|\mathbf{A}\|_{S_1},
\end{aligned} \tag{3.227}$$

which establishes the sufficiency in this case.

Finally, assuming that $r = r_1 = \tilde{r}$, we deduce that $\mathbf{P} \in \mathbb{R}^{r \times r}$ is an orthogonal matrix and, hence, that $\mathbf{P}^{-1} = \mathbf{P}^T$. Now, using (3.222) and the rank assumption, we can simplify the expansion (3.225) as

$$\mathbf{B} = \tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{P}^T \left(\mathbf{V}_r \mathbf{P}^T\right)^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{P}^T \mathbf{P} \mathbf{V}_r^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{V}_r^T. \tag{3.228}$$

□

Remark 3.2. *Note that even though the reduced SVD is not unique (i.e. there are multiple choices for the sub-orthogonal matrices in* (3.195)*), the parametric forms* *given in Proposition 3.5 do not depend on a specific decomposition and the results are invariant to any arbitrary choice of these reduced SVDs, primarily due to the "only if" parts of the statements.*

We observe that, in the case $p \in (1, \infty)$, the saturation of Hölder inequality provides a very tight link between the two matrices: If we know one of them, then the other lies in a one-dimensional ray that is parameterized by the constant $c > 0$. However, in the special case $p = 1$, the identification is not as simple. There again, for a given matrix $\mathbf{B}$, one can fully characterize the set of admissible matrices $\mathbf{A}$. However, for the reverse direction, an additional rank-equality constraint is essential to reduce the set of admissible matrices $\mathbf{B}$ to just one ray.

Inspired from Proposition 3.5, we now propose our main result in Theorem 3.13, where we explicitly characterize the duality mapping for the Schatten p-norms.

Theorem 3.13. *Let $p, q \in [1, +\infty]$ be a pair of Hölder conjugates with $\frac{1}{p} + \frac{1}{q} = 1$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$ a matrix whose reduced SVD is specified in* (3.184).

- *If $1 < p < +\infty$, then the single-valued duality mapping $\mathrm{J}_{S_p} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is well-defined and can be expressed as*

$$\mathrm{J}_{S_p} : \mathbf{A} = \mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma}) \mathbf{V}_r^T \mapsto \mathbf{A}^* = \mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma})) \mathbf{V}_r^T. \tag{3.229}$$

- *If $p = 1$ and if we consider the rank function as the sparsity measure in Definition 3.6, then the sparse duality mapping $\mathrm{J}_{S_1,\mathrm{rank}} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is well-defined (singleton) and is given as*

$$\mathrm{J}_{S_1,\mathrm{rank}} : \mathbf{A} = \mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma}) \mathbf{V}_r^T \mapsto \mathbf{A}^* = \|\boldsymbol{\sigma}\|_1 \mathbf{U}_r \mathbf{V}_r^T. \tag{3.230}$$

- *If $p = +\infty$, then the set-valued mapping $\mathcal{J}_{S_\infty}(\cdot)$ can be described as*

$$\mathcal{J}_{S_\infty}(\mathbf{A}) = \left\{ \sigma_1 \mathbf{U}_1 \mathbf{X} \mathbf{V}_1^T : \mathbf{X} \in \mathbb{R}^{r_1 \times r_1} \textit{ is symmetric and } \|\mathbf{X}\|_{S_1} = 1 \right\}, \tag{3.231}$$

  where r_1 denotes the multiplicity of the first singular value σ_1 of $\mathbf{A}$ and $\mathbf{U}_1, \mathbf{V}_1$ are singular vectors that correspond to σ_1 in (3.184)*. Finally, the set of sparse* *dual conjugates is the collection of rank-1 elements of $\mathcal{J}_{S_\infty}(\mathbf{A})$ which can be characterized as*

$$\mathcal{J}_{S_\infty,\mathrm{rank}}(\mathbf{A}) = \{\sigma_1 \mathbf{U}_1 \mathbf{p}\mathbf{p}^T \mathbf{V}_1^T : \mathbf{p} \in \mathbb{R}^{r_1}, \|\mathbf{p}\|_2 = 1\}. \tag{3.232}$$

Proof. **Case I:** $1 < p < +\infty$**.** Assume that $(\mathbf{A}, \mathbf{B})$ forms an (S_p, S_q)-conjugate pair. Hence, we have that $\langle \mathbf{A}, \mathbf{B}\rangle = \|\mathbf{A}\|_{S_p} \|\mathbf{B}\|_{S_q}$ which, together with Proposition 3.5, implies that $\mathbf{B}$ admits the form

$$\mathbf{B} = \frac{\|\mathbf{B}\|_{S_q}}{\|\mathbf{A}\|_{S_p}} \mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T = \mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T. \tag{3.233}$$

Case II: $p = 1$**.** Similarly to the previous case, consider $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathcal{J}_{S_1,\mathrm{rank}}(\mathbf{A})$. We have that

$$\mathrm{Tr}\left(\mathbf{A}^T\mathbf{B}\right) = \|\mathbf{A}\|_{S_1}\|\mathbf{B}\|_{S_\infty} \tag{3.234}$$

$$\|\mathbf{A}\|_{S_1} = \|\mathbf{B}\|_{S_\infty} \tag{3.235}$$

$$\mathrm{rank}(\mathbf{B}) \leq \mathrm{rank}(\mathbf{C}), \quad \forall \mathbf{C} \in \mathcal{J}_{S_1}(\mathbf{A}). \tag{3.236}$$

From (3.234) and using Proposition 3.5, we deduce that rank$(\mathbf{B}) \geq$ rank$(\mathbf{A})$ which, together with (3.236), implies that $\mathbf{B}$ should be equal to

$$\mathbf{B} = \|\mathbf{B}\|_{S_\infty} \mathbf{U}_r \mathbf{V}_r^T = \|\boldsymbol{\sigma}\|_1 \mathbf{U}_r \mathbf{V}_r^T, \tag{3.237}$$

where the last equality is obtained using (3.235).

Case III: $p = +\infty$**.** Following Proposition 3.5, any matrix $\mathbf{B} \in \mathcal{J}_{S_\infty}(\mathbf{A})$ can be expressed as $\mathbf{B} = \mathbf{U}_1 \tilde{\mathbf{X}} \mathbf{V}_1^T$, where $\tilde{\mathbf{X}} \in \mathbb{R}^{r_1 \times r_1}$ is a symmetric matrix. By defining $\mathbf{X} = \sigma_1^{-1}\tilde{\mathbf{X}}$, one readily verifies that $\mathbf{B} = \sigma_1 \mathbf{U}_1 \mathbf{X} \mathbf{V}_1^T$. By recalling the normalization constraint $\|\mathbf{A}\|_{S_\infty} = \|\mathbf{B}\|_{S_1}$, we therefore obtain that

$$\sigma_1 = \|\mathbf{A}\|_{S_\infty} = \|\mathbf{B}\|_{S_1} = \sigma_1 \|\mathbf{X}\|_{S_1}, \tag{3.238}$$

which implies that $\|\mathbf{X}\|_{S_1} = 1$. To show that $\mathcal{J}_{S_\infty}(\mathbf{A})$ is convex, consider two symmetric matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ in the unit ball of Schatten-1 norm and define

$$\mathbf{B}_\alpha = \sigma_1 \mathbf{U}_1 \mathbf{X}_\alpha \mathbf{V}_1^T, \quad \mathbf{X}_\alpha = \alpha \mathbf{X}_0 + (1-\alpha)\mathbf{X}_1 \tag{3.239}$$

for $\alpha \in [0, 1]$. On one hand, from the linearity of traces, we have that

$$\mathrm{Tr}(\mathbf{A}^T\mathbf{B}_\alpha) = \mathrm{Tr}\left(\mathbf{A}^T\left(\alpha\mathbf{B}_1 + (1-\alpha)\mathbf{B}_0\right)\right) = \alpha\mathrm{Tr}(\mathbf{A}^T\mathbf{B}_1) + (1-\alpha)\mathrm{Tr}(\mathbf{A}^T\mathbf{B}_0).$$

On the other hand, from the definition of $\mathbf{X}_0$ and $\mathbf{X}_1$, we deduce that $\mathbf{B}_0, \mathbf{B}_1 \in \mathcal{J}_{S_\infty}(\mathbf{A})$. Hence,

$$\mathrm{Tr}(\mathbf{A}^T\mathbf{B}_\alpha) = \alpha\|\mathbf{A}\|_{S_\infty}^2 + (1-\alpha)\|\mathbf{A}\|_{S_\infty}^2 = \|\mathbf{A}\|_{S_\infty}^2. \tag{3.240}$$

However, from the Hölder inequality and the convexity of norms, we have that

$$\mathrm{Tr}(\mathbf{A}^T\mathbf{B}_\alpha) \leq \|\mathbf{A}\|_{S_\infty}\|\mathbf{B}_\alpha\|_{S_1} \leq \|\mathbf{A}\|_{S_\infty}\left(\alpha\|\mathbf{B}_1\|_{S_1} + (1-\alpha)\|\mathbf{B}_0\|_{S_1}\right) = \|\mathbf{A}\|_{S_\infty}. \tag{3.241}$$

This implies that the Hölder inequality is saturated and also that $\|\mathbf{B}_\alpha\|_{S_1} = \|\mathbf{A}\|_{S_\infty}$ which, altogether, implies that $\mathbf{B}_\alpha \in \mathcal{J}_{S_\infty}(\mathbf{A})$ for all $\alpha \in [0, 1]$.

Finally, we observe that the set $\mathcal{J}_{S_\infty}(\mathbf{A})$ contains all matrices of the form $\mathbf{B} = \mathbf{U}_1\mathbf{p}\mathbf{p}^T\mathbf{V}_1^T$ for any vector $\mathbf{p} \in \mathbb{R}^{r_1}$ with $\|\mathbf{p}\|_2 = 1$. These are indeed all the rank-1 elements of $\mathcal{J}_{S_\infty}(\mathbf{A})$ which, due to the Definition 3.6, forms the set of sparse dual conjugates. □

Discussion

Theorem 3.13 provides an interesting characterization of the duality mapping in three scenarios: The first case is $1 < p < +\infty$ which is the most straightforward one. Theorem 3.13 tells us that the mapping is single-valued and also gives a formula to compute the dual conjugate $\mathbf{A}^*$ of any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. . We use this result to deduce the continuity of the duality mapping as well as the strict convexity of the Schatten space in this case (see Corollary 3.2). In the second case, with $p = 1$, the mapping is not single-valued. However, there is a unique element in the set of dual conjugates with the minimal rank (that is equal to the rank of $\mathbf{A}$) and, hence, we can construct a single-valued sparse duality mapping. Finally, we showed in the third case, characterized by $p = +\infty$, that neither the set of dual conjugates nor the ones with the minimal rank are unique.

In Corollary 3.2, we highlight some consequences of Theorem 3.13 concerning the strict convexity of Schatten spaces and the continuity of the duality mapping.

Corollary 3.2. *The Banach space of m by n matrices equipped with the Schatten-p norm is strictly convex, if and only if $p \in (1, +\infty)$. In this case, the function $\mathrm{J}_{S_p} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is continuous.*

Proof. For $p \in (1, +\infty)$, we know from Theorem 3.13 that the duality mapping J_{S_p} is bijective. Moreover, it is known that all finite-dimensional Banach spaces are reflexive. Now, following [253], we deduce the strict convexity of the space of m by n matrices with Schatten-p norm.

For $p = 1$ and $p = +\infty$, we can readily verify that

$$\left\| \alpha \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \right\|_{S_1} = \left\| \begin{pmatrix} \alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-\alpha) \end{pmatrix} \right\|_{S_1} = 1, \tag{3.242}$$

$$\left\| \alpha \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \right\|_{S_\infty} = \left\| \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-\alpha) \end{pmatrix} \right\|_{S_\infty} = 1, \tag{3.243}$$

for all $\alpha \in (0, 1)$, which shows that the Schatten space is not strictly convex for $p = 1, +\infty$.

Finally, the Schatten-p norm is known to be Fréchet differentiable for $p \in (1, +\infty)$ [242]. Moreover, the duality mapping of any Banach space with Fréchet-differentiable norms is guaranteed to be continuous [254, 255]. Combining the two statements, we deduce the continuity of the duality mapping in this case. □

By contrast, the sparse duality mapping $\mathrm{J}_{S_1,\mathrm{rank}}(\cdot)$ is not continuous. This is best explained by providing a counterexample. Specifically, let us consider the sequence of 2 by 2 matrices

$$\mathbf{S}_k = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{k} \end{pmatrix}, \quad k \in \mathbb{N}. \tag{3.244}$$

It is clear that $\mathbf{S}_k \to \mathbf{S}_\infty = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. However, we have that

$$\forall k \in \mathbb{N} : \mathrm{J}_{S_1,\mathrm{rank}}(\mathbf{S}_k) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{while} \quad \mathrm{J}_{S_1,\mathrm{rank}}(\mathbf{S}_\infty) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.245}$$

which shows the discontinuity of $\mathrm{J}_{S_1,\mathrm{rank}}$ in the space of 2 by 2 matrices. This can be generalized to space of matrices with arbitrary dimensions $m, n \in \mathbb{N}$.

Although $\mathrm{J}_{S_1,\mathrm{rank}}$ is not continuous, we now show that it is Borel-measurable and, hence, that it can be approximated with arbitrary precision by a continuous mapping due to Lusin's theorem [94].

Proposition 3.6. *For any $m, n \in \mathbb{N}$, the sparse duality mapping $\mathrm{J}_{S_1,\mathrm{rank}}$ is a Borel-measurable matrix-valued function over the space of m by n matrices.*

Before going into the proof of Proposition 3.6, we present a preliminary result.

Lemma 3.4. *The set $\mathcal{R}_r \subseteq \mathbb{R}^{m\times n}$ of m by n matrices of rank r is Borel-measurable.*

Proof. First note that

$$\mathcal{R}_1 = \{\mathbf{u}\mathbf{v}^T : \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n\}. \tag{3.246}$$

The set $\mathcal{R}_1$ is the image of the continuous mapping $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m\times n} : (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u}\mathbf{v}^T$ and, hence, is Borel-measurable.

Now, denote by $\mathcal{R}_{\leq r} \subseteq \mathbb{R}^{m\times n}$, the set of matrices with rank no more than r. Using the identity

$$\mathcal{R}_{\leq r} = \mathcal{R}_1 + \cdots + \mathcal{R}_1, \quad \text{(r times)}, \tag{3.247}$$

we deduce that $\mathcal{R}_{\leq r}$ and, consequently, $\mathcal{R}_r = \mathcal{R}_{\leq r} \backslash \mathcal{R}_{\leq (r-1)}$ are also Borel-measurable sets. □

Proof of Proposition 3.6. Consider a Borel-measurable set $\mathcal{B} \subseteq \mathbb{R}^{m\times n}$. We show that $\mathcal{B}_{\mathrm{inv}} = \mathrm{J}^{-1}_{S_1,\mathrm{rank}}(\mathcal{B})$ is also Borel-measurable. By defining $\mathcal{B}_{\mathrm{inv},r} = \mathcal{B}_{\mathrm{inv}} \cap \mathcal{R}_r$, we can partition $\mathcal{B}_{\text{inv}}$ as

$$\mathcal{B}_{\text{inv}} = \bigcup_{r=1}^{\min(m,n)} \mathcal{B}_{\text{inv},r}. \tag{3.248}$$

Hence, it is sufficient to show that each partition $\mathcal{B}_{\text{inv},r}$ is Borel-measurable.

Define the set $\mathcal{P}_r \subseteq \mathcal{R}_r^2$ as

$$\mathcal{P}_r = \{(\mathbf{A}, \mathbf{B}) \in \mathcal{R}_r \times \mathcal{B} : \text{Tr}(\mathbf{A}^T\mathbf{B}) = \|\mathbf{A}\|_{S_1}\|\mathbf{B}\|_{S_\infty}, \quad \|\mathbf{A}\|_{S_1} = \|\mathbf{B}\|_{S_\infty}\}. \tag{3.249}$$

The set $\mathcal{P}_r$ introduces a relation over $\mathcal{R}_r$ whose domain is $\mathcal{B}_{\text{inv},r}$. In other words, we have that

$$\mathcal{B}_{\text{inv},r} = \{\mathbf{A} \in \mathcal{R}_r : \exists \mathbf{B} \in \mathcal{B}, (\mathbf{A}, \mathbf{B}) \in \mathcal{P}_r\}. \tag{3.250}$$

Since the trace and norm are continuous (and, consequently, Borel-measurable) functions and $\mathcal{R}_r \times \mathcal{B}$ is a Borel-measurable set (using Lemma 3.4), we deduce that the relation induced from $\mathcal{P}_r$ is Borel-measurable as well. Finally, we use [256, Proposition 2.1] to show that its domain is Borel-measurable. □

3.5.2 Hessian-Schatten Total Variation

We now introduce[9] our novel seminorm—the Hessian-Schatten total variation (HTV)—and we propose its use as a way to quantify the complexity of learning schemes. Our definition of the HTV is based on a second-order extension of the space of functions with bounded variation [257]. We show that the HTV seminorm satisfies the following desirable properties:

1. It assigns the zero value for linear regression, which is the simplest learning scheme.
2. It is invariant (up to a multiplicative factor) to simple transformations (such as linear isometries and scaling) over the input domain.
3. It is defined for both smooth and CPWL functions. Hence, it is applicable to a broad class of learning schemes, including ReLU neural networks and radial-basis functions.
4. It favors CPWL functions with a small number of linear regions, thus promoting a simpler (and, hence, more interpretable) representation of the data (Occam's razor principle).

We provide closed-form formulas for the HTV of both smooth and CPWL functions. For smooth functions, the HTV coincides with the Hessian-Schatten seminorm which is often used as a regularization term in linear inverse problems [219, 218]. For CPWL functions, the HTV is a convex relaxation of the number of linear regions. This is analogous to the classical ℓ_0 penalty in the field of compressed sensing, where it is often replaced by its convex proxy, the ℓ_1 norm, to ensure tractability [62, 66].

Mathematical Background

Throughout this work, we denote the input domain by $\Omega \subseteq \mathbb{R}^d$. We assume Ω to be an open ball of radius $R > 0$, with the convention that the case $R = +\infty$

[9] From our submitted work [111].

corresponds to $\Omega = \mathbb{R}^d$. It is left to reader to verify that the function spaces that were defined throughout this section can easily be adapted to this setting.

We denote by $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ the space of continuous matrix-valued functions $\mathbf{F} : \Omega \to \mathbb{R}^{d\times d}$ that vanish at infinity so that $\lim_{\|\boldsymbol{x}\|\to\infty} \|\mathbf{F}(\boldsymbol{x})\| = 0$ whenever the domain is unbounded. (Note that this definition does not depend on the choice of the norms, because they are all equivalent in finite-dimensional vector spaces.) Any matrix-valued function $\mathbf{F} : \Omega \to \mathbb{R}^{d\times d}$ has the unique representation

$$\mathbf{F} = [f_{i,j}] = \begin{pmatrix} f_{1,1} & \cdots & f_{1,d} \\ \vdots & \ddots & \vdots \\ f_{d,1} & \cdots & f_{d,d} \end{pmatrix}, \tag{3.251}$$

where each entry $f_{i,j} : \Omega \to \mathbb{R}$ is a scalar-valued function for $i, j = 1, \dots, d$. In this representation, the space $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ is the collection of matrix-valued functions of the form (3.251) with $f_{i,j} \in \mathcal{C}_0(\Omega)$.

Definition 3.7. *Let $q \in [1, +\infty]$. For any $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$, the L_∞-S_q mixed norm is defined as*

$$\|\mathbf{F}\|_{L_\infty, S_q} \triangleq \left\| \begin{pmatrix} \|f_{1,1}\|_{L_\infty} & \cdots & \|f_{1,d}\|_{L_\infty} \\ \vdots & \ddots & \vdots \\ \|f_{d,1}\|_{L_\infty} & \cdots & \|f_{d,d}\|_{L_\infty} \end{pmatrix} \right\|_{S_q}. \tag{3.252}$$

Remark 3.3. *In Definition 3.7, the S_q-norm appears as the outer norm. We remain faithful to this convention and always denote mixed norms in order of appearance, where the first is the inner-norm and the second the outer-norm.*

Following Theorem 2.3, we deduce that $\left(\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d}), \|\cdot\|_{L_\infty, S_q}\right)$ is a *bona fide* Banach space, whose dual is $(\mathcal{M}(\Omega, \mathbb{R}^{d\times d}), \|\cdot\|_{\mathcal{M}, S_p})$, where $\mathcal{M}(\Omega; \mathbb{R}^{d\times d})$ is the collection of matrix-valued Radon measures of the form

$$\mathbf{W} = [w_{i,j}] = \begin{pmatrix} w_{1,1} & \cdots & w_{1,d} \\ \vdots & \ddots & \vdots \\ w_{d,1} & \cdots & w_{d,d} \end{pmatrix}, \quad w_{i,j} \in \mathcal{M}(\Omega) \quad \forall i, j = 1, \dots, d, \tag{3.253}$$

and the mixed $\mathcal{M} - S_p$ norm is defined as

$$\|\mathbf{W}\|_{\mathcal{M},S_p} \triangleq \left\| \begin{pmatrix} \|w_{1,1}\|_{\mathcal{M}} & \cdots & \|w_{1,d}\|_{\mathcal{M}} \\ \vdots & \ddots & \vdots \\ \|w_{d,1}\|_{\mathcal{M}} & \cdots & \|w_{d,d}\|_{\mathcal{M}} \end{pmatrix} \right\|_{S_p}. \tag{3.254}$$

The duality product $\langle \cdot, \cdot \rangle : \mathcal{M}(\Omega; \mathbb{R}^{d\times d}) \times \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d}) \to \mathbb{R}$ is then defined as

$$\langle \mathbf{W}, \mathbf{F} \rangle \triangleq \sum_{i=1}^{d} \sum_{j=1}^{d} \langle w_{i,j}, f_{i,j} \rangle. \tag{3.255}$$

Likewise, we denote by $L_1(\Omega; \mathbb{R}^{d\times d})$, $\mathcal{D}(\Omega; \mathbb{R}^{d\times d})$, and $\mathcal{D}'(\Omega; \mathbb{R}^{d\times d})$, the matrix-valued generalizations of the spaces $L_1(\Omega)$, $\mathcal{D}(\Omega)$ and $\mathcal{D}'(\Omega)$, respectively.

Finally, we highlight that the operators $\frac{\partial^2 f}{\partial x_i \partial x_j} : \mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ are viewed as second-order weak partial derivatives. More precisely, for any $i, j = 1, \dots, d$ and any $w \in \mathcal{D}'(\Omega)$, the distribution $\frac{\partial^2 w}{\partial x_i \partial x_j}\{w\} \in \mathcal{D}'(\Omega)$ is defined as

$$\left\langle \frac{\partial^2}{\partial x_i \partial x_j} w, \varphi \right\rangle = \left\langle w, \frac{\partial^2}{\partial x_i \partial x_j} \varphi \right\rangle, \tag{3.256}$$

for all test functions $\varphi \in \mathcal{D}(\Omega)$. This leads to the following definition of the generalized Hessian operator over the space of distributions.

Definition 3.8. *The Hessian operator* $\mathrm{H} : \mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega; \mathbb{R}^{d\times d})$ *is defined as*

$$\mathrm{H}\{f\} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{pmatrix}. \tag{3.257}$$

Towards Defining The HTV

In order to properly define the HTV seminorm, we start by introducing a novel class of mixed norms over $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$.

Definition 3.9. *Let $q \in [1, +\infty]$. For any $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$, the $S_q - L_\infty$ mixed-norm is defined as*

$$\|\mathbf{F}\|_{S_q, L_\infty} = \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{S_q}. \tag{3.258}$$

As previously mentioned, the dual norm of $L_\infty - S_q$ mixed-norm is $\mathcal{M} - S_p$, which is defined over matrix-valued Radon measures. In Definition 3.9, we switched the order of application of the individual norms; however, the two norms induce the same topology over the space $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$.

Theorem 3.14. *Regarding the mixed norms defined in Definitions 3.7 and 3.9*

1. *The functional $\mathbf{F} \mapsto \|\mathbf{F}\|_{S_q, L_\infty}$ is a well-defined (finite) norm over $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$.*
2. *The $L_\infty - S_q$ and the $S_q - L_\infty$ mixed norms are equivalent, in the sense that there exists positive constants $A, B > 0$ such that, for all $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$, we have that*
$$A\|\mathbf{F}\|_{S_q, L_\infty} \leq \|\mathbf{F}\|_{L_\infty, S_q} \leq B\|\mathbf{F}\|_{S_q, L_\infty}. \tag{3.259}$$
3. *The normed space $\big(\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d}), \|\cdot\|_{S_q, L_\infty}\big)$ is a bona fide Banach space.*

Proof. It is known that all norms are equivalent in finite-dimensional vector spaces. Consequently, there exist positive constants $c_1, c_2 > 0$ such that

$$\forall \mathbf{A} = [a_{i,j}] \in \mathbb{R}^{d\times d}, \quad c_1\|\mathbf{A}\|_{\mathrm{sum}} \leq \|\mathbf{A}\|_{S_q} \leq c_2\|\mathbf{A}\|_{\mathrm{sum}}, \tag{3.260}$$

where $\|\mathbf{A}\|_{\mathrm{sum}} = \sum_{i=1}^d \sum_{j=1}^d |a_{i,j}|$. This immediately yields that

$$c_1 \sum_{i=1}^d \sum_{j=1}^d \|f_{i,j}\|_{L_\infty} \leq \|\mathbf{F}\|_{L_\infty, S_q} \leq c_2 \sum_{i=1}^d \sum_{j=1}^d \|f_{i,j}\|_{L_\infty}, \tag{3.261}$$

as well as that

$$c_1 \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\mathrm{sum}} \leq \|\mathbf{F}\|_{S_q, L_\infty} \leq c_2 \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\mathrm{sum}}, \tag{3.262}$$

for all $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$. On the one hand, we have that

$$\sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}} = \sup_{\boldsymbol{x}\in\Omega} \left(\sum_{i,j=1}^{d} |f_{i,j}(\boldsymbol{x})| \right) \leq \sum_{i,j=1}^{d} \sup_{\boldsymbol{x}\in\Omega} |f_{i,j}(\boldsymbol{x})| = \sum_{i,j=1}^{d} \|f_{i,j}\|_{L_\infty}. \tag{3.263}$$

Combining (3.261), (3.262) and (3.263), we then deduce that

$$\|\mathbf{F}\|_{S_q,L_\infty} \leq c_2 \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}} \leq c_2 \sum_{i,j=1}^{d} \|f_{i,j}\|_{L_\infty} \leq \frac{c_2}{c_1} \|\mathbf{F}\|_{L_\infty,S_q}. \tag{3.264}$$

On the other hand, using $\|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}} \geq |f_{i,j}(\boldsymbol{x})|$ for all $i,j = 1,\dots,d$, we obtain that

$$\|f_{i,j}\|_{L_\infty} = \sup_{\boldsymbol{x}\in\Omega} |f_{i,j}(\boldsymbol{x})| \leq \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}}, \quad \forall i,j = 1,\dots,d. \tag{3.265}$$

Summing over all $i,j = 1,\dots,d$ then gives that

$$\sum_{i=1}^{d} \sum_{j=1}^{d} \|f_{i,j}\|_{L_\infty} \leq d^2 \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}}. \tag{3.266}$$

Combining (3.261), (3.262), and (3.266), we obtain that

$$\|\mathbf{F}\|_{L_\infty,S_q} \leq c_2 \sum_{i=1}^{d} \sum_{j=1}^{d} \|f_{i,j}\|_{L_\infty} \leq c_2 d^2 \sup_{\boldsymbol{x}\in\Omega} \|\mathbf{F}(\boldsymbol{x})\|_{\text{sum}} \leq \frac{c_2}{c_1} d^2 \|\mathbf{F}\|_{S_q,L_\infty}. \tag{3.267}$$

Finally, the inequalities (3.264) and (3.267) yield (3.259) with $A = \frac{c_1}{c_2}$ and $B = \frac{c_2}{c_1} d^2$ (Item 2). Further, it guarantees that the functional $\mathbf{F} \mapsto \|\mathbf{F}\|_{S_q,L_\infty}$ is well-defined (finite) for all $\mathbf{F} \in \mathcal{C}_0(\Omega, \mathbb{R}^{d\times d})$. It is then easy to verify the remaining norm properties (positivity, homogeneity and the triangle inequality) of $\|\cdot\|_{S_q,L_\infty}$ (Item 1). As for Item 3, we note that the norm equivalence implies that both norms induce the same topology over $\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$. Hence, $\big(\mathcal{C}_0(\Omega; \mathbb{R}^{d\times d}), \|\cdot\|_{S_q,L_\infty}\big)$ is a *bona fide* Banach space. □

Using the outcomes of Theorem 3.14 and, in particular, Item 3, we are now ready to introduce the $S_p - \mathcal{M}$ mixed norm defined over the space of matrix-valued Radon measures.

Definition 3.10. *For any matrix-valued Radon measure* $\mathbf{W} \in \mathcal{M}(\Omega, \mathbb{R}^{d\times d})$*, the* $S_p - \mathcal{M}$ *mixed-norm is defined as*

$$\|\mathbf{W}\|_{S_p,\mathcal{M}} \triangleq \sup\left\{\langle \mathbf{W}, \mathbf{F}\rangle : \mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q,L_\infty} = 1\right\}. \tag{3.268}$$

Intuitively, the $S_p - \mathcal{M}$ norm of a matrix-valued function $\mathbf{F} : \Omega \to \mathbb{R}^{d\times d}$ is equal to the total-variation norm of the function $\boldsymbol{x} \mapsto \|\mathbf{F}(\boldsymbol{x})\|_{S_q}$. However, this intuition cannot directly lead to a general definition because the space $\mathcal{M}(\Omega; \mathbb{R}^{d\times d})$ contains elements that do not have a pointwise definition. We are therefore forced to define this norm by duality, as opposed to the $\mathcal{M} - S_p$ norm given in (3.254).

We also remark that, due to the dense embedding $\mathcal{D}(\Omega; \mathbb{R}^{d\times d}) \hookrightarrow \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$, one can alternatively express the $S_p - \mathcal{M}$ norm as

$$\|\mathbf{W}\|_{S_p,\mathcal{M}} = \sup\left\{\langle \mathbf{W}, \mathbf{F}\rangle : \mathbf{F} \in \mathcal{D}(\Omega; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q,L_\infty} = 1\right\}, \tag{3.269}$$

which is well-defined for all matrix-valued distributions. However, the only elements of $\mathcal{D}'(\Omega; \mathbb{R}^{d\times d})$ of finite $S_p - \mathcal{M}$ norm are precisely the matrix-valued finite Radon measures. In other words, $\mathcal{M}(\Omega, \mathbb{R}^{d\times d})$ is the largest subspace of $\mathcal{D}'(\Omega; \mathbb{R}^{d\times d})$ with finite $S_p - \mathcal{M}$ norm.

In what follows, we strengthen the intuition behind the $S_p - \mathcal{M}$ norm by computing it for two general classes of functions/distributions in $\mathcal{M}(\Omega, \mathbb{R}^{d\times d})$ that are particularly important in our framework: the absolutely integrable matrix-valued functions and the Dirac fence distributions.

Definition 3.11. *For any nonzero matrix* $\mathbf{A} \in \mathbb{R}^{d\times d}$*, any convex compact set* $C \subset \mathbb{R}^{d_1}$ *with* $d_1 < d$*, and any measurable transformation* $\mathbf{T} : \mathbb{R}^{d_1} \to \mathbb{R}^{d-d_1}$ *(not necessarily linear) such that* $C \times \mathbf{T}(C) \subseteq \Omega$*, we define the corresponding Dirac fence* $\mathbf{D} \in \mathcal{M}(\Omega; \mathbb{R}^{d\times d})$ *as*

$$\mathbf{D}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbf{A}\mathbb{1}_{\boldsymbol{x}_1\in C}\delta(\boldsymbol{x}_2 - \mathbf{T}\boldsymbol{x}_1\}), \quad \boldsymbol{x}_1 \in \mathbb{R}^{d_1}, \boldsymbol{x}_2 \in \mathbb{R}^{d-d_1}, (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \Omega. \tag{3.270}$$

Dirac fence distributions are natural generalizations of *the Dirac impulse* to nonlinear (and bounded) manifolds [258]. More precisely, for any test function $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ and any Dirac fence $\mathbf{D}$ of the form (3.270), we have that

$$\langle \mathbf{D}, \mathbf{F}\rangle = \int_C \mathrm{Tr}\left(\mathbf{A}^T\mathbf{F}(\boldsymbol{x}_1, \mathbf{T}\boldsymbol{x}_1)\right) \mathrm{d}\boldsymbol{x}_1 \in \mathbb{R}. \tag{3.271}$$

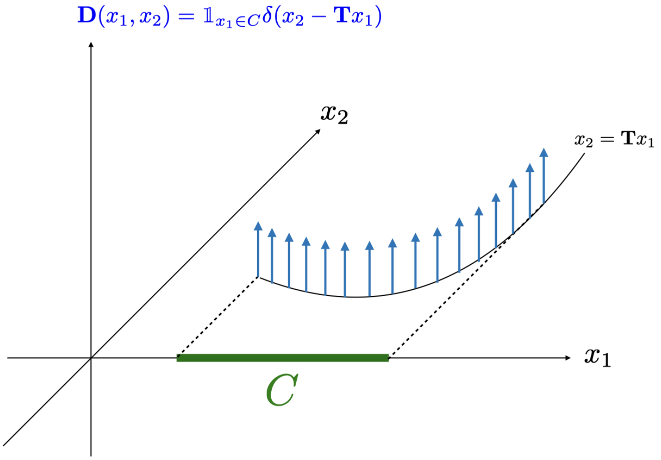

$\mathbf{D}(x_1, x_2) = \mathbb{1}_{x_1 \in C}\delta(x_2 - \mathbf{T}x_1)$
x_2
$x_2 = \mathbf{T}x_1$
x_1
C

Figure 3.18: Illustration of a Dirac fence with $d = 2$ and $d_1 = 1$.

Intuitively, this corresponds to considering a "continuum" of low-dimensional Dirac impulses on the d_1-dimensional compact manifold $C \times \mathbf{T}(C)$ that is embedded in Ω, as illustrated in Figure 3.18.

Theorem 3.15. *Let $p \in [1, +\infty)$.*

1. *For any matrix-valued function $\mathbf{W} \in L_1(\Omega, \mathbb{R}^{d\times d}) \subseteq \mathcal{M}(\Omega; \mathbb{R}^{d\times d})$, we have that*

$$\|\mathbf{W}\|_{S_p,\mathcal{M}} = \Big\| \|\mathbf{W}(\cdot)\|_{S_p} \Big\|_{L_1} = \int_\Omega \left(\sum_{i=1}^{d} |\sigma_i\left(\mathbf{W}(\boldsymbol{x})\right)|^p \right)^{\frac{1}{p}} \mathrm{d}\boldsymbol{x}. \tag{3.272}$$

2. *For any Dirac fence distribution $\mathbf{D}$ of the form* (3.270)*, we have that*

$$\|\mathbf{D}\|_{S_p,\mathcal{M}} = \|\mathbf{A}\|_{S_p}\mathrm{Leb}(C), \tag{3.273}$$

where $\mathrm{Leb}(C)$ *denotes the Lebesgue measure of $C \subseteq \mathbb{R}^{d_1}$.*

3. *Consider two Dirac fences* $\mathbf{D}_1$ *and* $\mathbf{D}_2$ *of the form*

$$\mathbf{D}_i(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbf{A}_i \mathbb{1}_{\boldsymbol{x}_1 \in C_i} \delta(\boldsymbol{x}_2 - \mathbf{T}_i \boldsymbol{x}_1\}), \quad i = 1, 2 \tag{3.274}$$

and assume that the "intersection" of the two fences is of measure zero, in the sense that $C_0 = \{\boldsymbol{x}_1 \in C_1 \cap C_2 : \mathbf{T}_1 \boldsymbol{x}_1 = \mathbf{T}_2 \boldsymbol{x}_1\}$ *is a subset of* $\mathbb{R}^{d_1}$ *whose Lebesgue measure is zero. Then, we have that*

$$\|\mathbf{D}_1 + \mathbf{D}_2\|_{S_p, \mathcal{M}} = \|\mathbf{D}_1\|_{S_p, \mathcal{M}} + \|\mathbf{D}_2\|_{S_p, \mathcal{M}}. \tag{3.275}$$

Proof. **Item 1:** We first show that the right-hand side of (3.272) is well-defined and admits a finite value. First, note that $\|\mathbf{W}(\cdot)\|_{S_p}$ is the composition of the measurable function $\mathbf{W} : \Omega \to \mathbb{R}^{d \times d}$ and the Schatten-p norm $\| \cdot \|_{S_p} : \mathbb{R}^{d \times d} \to \mathbb{R}$ that is continuous and, consequently, measurable. This implies that $\|\mathbf{W}(\cdot)\|_{S_p}$ is also a measurable function and, hence, its L_1 norm is well-defined. The last step is to show that the L_1-norm is finite. From the norm-equivalence property of finite-dimensional vector spaces, we deduce the existence of $b > 0$ such that, for any $\mathbf{A} = [a_{i,j}] \in \mathbb{R}^{d \times d}$, we have that

$$\|\mathbf{A}\|_{S_p} \leq b \|\mathbf{A}\|_{\mathrm{sum}}, \tag{3.276}$$

where $\|\mathbf{A}\|_{\mathrm{sum}} = \sum_{i=1}^d \sum_{j=1}^d |a_{i,j}|$. This implies that

$$\Big\| \|\mathbf{W}(\cdot)\|_{S_p} \Big\|_{L_1} = \int_\Omega \|\mathbf{W}(\boldsymbol{x})\|_{S_p} \mathrm{d}\boldsymbol{x} \leq b \int_\Omega \|\mathbf{W}(\boldsymbol{x})\|_{\mathrm{sum}} \mathrm{d}\boldsymbol{x} \overset{\text{(i)}}{=} b \sum_{i=1}^d \sum_{j=1}^d \|w_{i,j}\|_{L_1} < +\infty, \tag{3.277}$$

where we have used Fubini's theorem to deduce (i). Now, one readily verifies that

$$\begin{aligned} \langle \mathbf{W}, \mathbf{F} \rangle = \sum_{i,j=1}^d \langle w_{i,j}, f_{i,j} \rangle = \sum_{i,j=1}^d \int_\Omega w_{i,j}(\boldsymbol{x}) f_{i,j}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_\Omega \left(\sum_{i,j=1}^d w_{i,j}(\boldsymbol{x}) f_{i,j}(\boldsymbol{x}) \right) \mathrm{d}\boldsymbol{x} \\ \leq \int_\Omega \left| \sum_{i,j=1}^d w_{i,j}(\boldsymbol{x}) f_{i,j}(\boldsymbol{x}) \right| \mathrm{d}\boldsymbol{x} \overset{\text{(i)}}{\leq} \int_\Omega \|\mathbf{W}(\boldsymbol{x})\|_{S_p} \|\mathbf{F}(\boldsymbol{x})\|_{S_q} \mathrm{d}\boldsymbol{x} \overset{\text{(ii)}}{\leq} \Big\| \|\mathbf{W}(\cdot)\|_{S_p} \Big\|_{L_1} \|\mathbf{F}\|_{S_q}, \end{aligned} \tag{3.278}$$

where we have used the Hölder inequality for Schatten norms (see (3.186)) in (i) and the one for L_p norms in (ii). We conclude that

$$\|\mathbf{W}\|_{S_p,\mathcal{M}} \leq \left\| \, \|\mathbf{W}(\cdot)\|_{S_p} \right\|_{L_1}. \tag{3.279}$$

To show the equality, we need to prove that, for any $\epsilon > 0$, there exists an element $\mathbf{F}_\epsilon \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ with $\|\mathbf{F}_\epsilon\|_{S_q,L_\infty} = 1$ such that

$$\langle \mathbf{W}, \mathbf{F}_\epsilon \rangle \geq \left\| \, \|\mathbf{W}(\cdot)\|_{S_p} \right\|_{L_1} - \epsilon. \tag{3.280}$$

Consider the function $\mathbf{F} : \Omega \to \mathbb{R}^{d\times d}$ with

$$\mathbf{F}(\boldsymbol{x}) = \begin{cases} \frac{\mathrm{J}_{S_p,\mathrm{rank}}(\mathbf{W}(\boldsymbol{x}))}{\|\mathbf{W}(\boldsymbol{x})\|_{S_p}}, & \mathbf{W}(\boldsymbol{x}) \neq \mathbf{0} \\ 0, & \text{otherwise}, \end{cases} \tag{3.281}$$

where $\mathrm{J}_{S_p,\mathrm{rank}} : \mathbb{R}^{d\times d} \to \mathbb{R}$ is the sparse duality mapping (see, Definition 3.6). We first note that $\mathbf{F}$ is a measurable function. Indeed, from Proposition 3.6, we know that $\mathrm{J}_{S_p,\mathrm{rank}}$ is a measurable mapping over $\mathbb{R}^{d\times d}$. Hence, its composition with the measurable function $\mathbf{W}$ is also measurable. Moreover, norms are continuous (and, so, Borel-measurable) functionals. Therefore, we have that $\mathbf{F}(\boldsymbol{x}) = \mathbb{1}_{\mathbf{W}\neq\mathbf{0}} \frac{\mathrm{J}_{S_p,\mathrm{rank}}(\mathbf{W}(\boldsymbol{x}))}{\|\mathbf{W}(\boldsymbol{x})\|_{S_p}}$ is also Borel-measurable. Knowing the measurability of $\mathbf{F}$, we observe that

$$\int_\Omega \mathrm{Tr}(\mathbf{W}^T(\boldsymbol{x})\mathbf{F}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} = \int_\Omega \|\mathbf{W}(\boldsymbol{x})\|_{S_p}\mathrm{d}\boldsymbol{x} = \left\| \, \|\mathbf{W}(\cdot)\|_{S_p} \right\|_{L_1}. \tag{3.282}$$

We also note that $\|\mathbf{F}\|_{S_q,L_\infty} = 1$. The final step is to use Lusin's theorem (see [259, Theorem 7.10]) to find an ϵ-approximation $\mathbf{F}_\epsilon \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ of $\mathbf{F}$ on the unit $S_q - L_\infty$ ball so that

$$\left| \int_\Omega \mathrm{Tr}(\mathbf{W}^T(\boldsymbol{x})\mathbf{F}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} - \int_\Omega \mathrm{Tr}(\mathbf{W}^T(\boldsymbol{x})\mathbf{F}_\epsilon(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} \right| \leq \epsilon. \tag{3.283}$$

Now, combining (3.283) with (3.282), we deduce (3.280) which completes the proof.

Item 2: We first recall that the application of a distribution $\mathbf{D}$ of the form (3.270) to any element $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d\times d})$ can be computed as

$$\langle \mathbf{D}, \mathbf{F} \rangle = \int_C \mathrm{Tr}\left(\mathbf{A}^T\mathbf{F}(\boldsymbol{x}, \mathrm{T}\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}. \tag{3.284}$$

Using Hölder's inequality, for any $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ with $\|\mathbf{F}\|_{S_q, L_\infty} = 1$, we obtain that

$$\begin{aligned}\int_C \mathrm{Tr}\left(\mathbf{A}^T \mathbf{F}(\boldsymbol{x}, \mathrm{T}\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x} &\leq \int_C \|\mathbf{A}\|_{S_p} \|\mathbf{F}(\boldsymbol{x}, \mathrm{T}\boldsymbol{x})\|_{S_q} \mathrm{d}\boldsymbol{x} \\ &\leq \|\mathbf{A}\|_{S_p} \int_C 1 \mathrm{d}\boldsymbol{x} = \|\mathbf{A}\|_{S_1} \mathrm{Leb}(C),\end{aligned}$$

which implies that $\|\mathbf{D}\|_{S_p, \mathcal{M}} \leq \|\mathbf{A}\|_{S_p} \mathrm{Leb}(C)$. To verify the equality, we consider an element $\mathbf{F} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ whose restriction on C is the constant matrix $\mathbf{A}^* = \|\mathbf{A}\|_{S_p}^{-1} \mathrm{J}_{S_p, \mathrm{rank}}(\mathbf{A})$.

Item 3: Following the assumption that $\mathrm{Leb}(C_0) = 0$, for any $\epsilon > 0$, there exists a measurable set $E \subseteq \mathbb{R}^{d_1}$ with $\mathrm{Leb}(E) = \epsilon/2$ such that $C_0 \subseteq E$. From the construction, we deduce that the sets $C_1 \backslash E$ and $C_2 \backslash E$ are separable; hence, there exists a function $\mathbf{F}_\epsilon \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ with $\|\mathbf{F}_\epsilon\|_{S_q, L_\infty} = 1$ such that

$$\mathbf{F}_\epsilon(\boldsymbol{x}_1, \mathbf{T}_i \boldsymbol{x}_1) = \mathbf{A}_i^*, \quad \forall \boldsymbol{x}_1 \in C_i \backslash C_0, i = 1, 2, \tag{3.285}$$

where $\mathbf{A}_i^* = \|\mathbf{A}_i\|_{S_p}^{-1} \mathrm{J}_{S_p, \mathrm{rank}}(\mathbf{A}_i), i = 1, 2$. This implies that, for $i = 1, 2$, we have that

$$\begin{aligned}\langle \mathbf{D}_i, \mathbf{F}_\epsilon \rangle &= \int_{C_i} \mathrm{Tr}\left(\mathbf{A}_i^T \mathbf{F}_\epsilon(\boldsymbol{x}_1, \mathbf{T}_i \boldsymbol{x}_1)\right) \mathrm{d}\boldsymbol{x}_1 \\ &= \int_{C_0} \mathrm{Tr}\left(\mathbf{A}_i^T \mathbf{F}_\epsilon(\boldsymbol{x}_1, \mathbf{T}_i \boldsymbol{x}_1)\right) \mathrm{d}\boldsymbol{x}_1 + \int_{C_i \backslash C_0} \mathrm{Tr}\left(\mathbf{A}_i^T \mathbf{A}_i^*\right) \mathrm{d}\boldsymbol{x}_1 \\ &\geq -\mathrm{Leb}(C_0) \|\mathbf{A}_i\|_{S_p} + \mathrm{Leb}(C_i \backslash C_0) \|\mathbf{A}_i\|_{S_p} \\ &\geq \|\mathbf{A}_i\|_{S_p} (\mathrm{Leb}(C_i) - \epsilon).\end{aligned} \tag{3.286}$$

Hence, for any $\epsilon > 0$, we have that

$$\begin{aligned}\|\mathbf{D}_1 + \mathbf{D}_2\|_{S_p, \mathcal{M}} &\geq \langle \mathbf{D}_1 + \mathbf{D}_2, \mathbf{F}_\epsilon \rangle \\ &\geq \|\mathbf{A}_1\|_{S_p} \mathrm{Leb}(C_1) + \|\mathbf{A}_2\|_{S_p} \mathrm{Leb}(C_2) - \epsilon(\|\mathbf{A}_1\|_{S_p} + \|\mathbf{A}_2\|_{S_p}).\end{aligned}$$

By letting $\epsilon \to 0$, we deduce that $\|\mathbf{D}_1 + \mathbf{D}_2\|_{S_p, \mathcal{M}} \geq \|\mathbf{D}_1\|_{S_p, \mathcal{M}} + \|\mathbf{D}_2\|_{S_p, \mathcal{M}}$ which, together with the triangle inequality, yields the announced equality. □

We are now ready to define the HTV seminorm.

Definition 3.12. *Let $p \in [1, +\infty]$. The Hessian-Schatten total variation of any $f \in \mathcal{D}'(\Omega)$ is defined as*

$$\mathrm{HTV}_p(f) = \|\mathrm{H}\{f\}\|_{S_p,\mathcal{M}} = \sup\left\{\langle \mathrm{H}\{f\}, \mathbf{F}\rangle : \mathbf{F} \in \mathcal{D}(\Omega; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q,L_\infty} = 1\right\}, \tag{3.287}$$

where $q \in [1, +\infty]$ is the Hölder conjugate of p with $\frac{1}{p} + \frac{1}{q} = 1$.

We remark that the case $p = 2$ has been previously studied in the context of the space of functions with bounded Hessian [260, 261, 262, 263]. In our work, we complement their theoretical findings by extending the definition of the HTV to all Schatten norms with arbitrary value of $p \in [1, +\infty]$. We now prove some desirable properties of the HTV functional.

Theorem 3.16. *The HTV seminorm satisfies the following properties.*

1. **Null Space***: A distribution has a vanishing HTV if and only if it can be identified as an affine function. In other words, we have that*

$$\mathcal{N}_{\mathrm{HTV}_p}(\Omega) = \{f \in \mathcal{D}'(\Omega) : \mathrm{HTV}_p(f) = 0\} = \{\boldsymbol{x} \mapsto \boldsymbol{a}^T\boldsymbol{x} + b : \boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}\}. \tag{3.288}$$

2. **Invariance***: Let $\Omega = \mathbb{R}^d$. For any $f \in \mathcal{D}'(\mathbb{R}^d)$, we have that*

$$\mathrm{HTV}_p\left(f(\cdot - \boldsymbol{x}_0)\right) = \mathrm{HTV}_p\left(f\right), \qquad \forall \boldsymbol{x}_0 \in \mathbb{R}^d, \tag{3.289}$$

$$\mathrm{HTV}_p\left(f(\alpha\cdot)\right) = |\alpha|^{2-d}\mathrm{HTV}_p\left(f\right), \qquad \forall \alpha \in \mathbb{R}, \tag{3.290}$$

$$\mathrm{HTV}_p\left(f(\mathbf{U}\cdot)\right) = \mathrm{HTV}_p\left(f\right), \qquad \forall \mathbf{U} \in \mathbb{R}^{d\times d} : \textit{Orthonormal}. \tag{3.291}$$

Proof. **Item 1:** Starting from $\mathrm{H}\{f\} = \mathbf{0}$, we deduce that $\frac{\partial^2 f}{\partial x_i^2} = 0$ for $i = 1, \dots, d$. Following Proposition 6.1 in [48], we deduce that the null space of $\frac{\partial^2}{\partial x_1^2}$ can only contain (multivariate) polynomials. Using this, we infer that any p in the null space of $\frac{\partial^2}{\partial x_1^2}$ is of the form $p(\boldsymbol{x}) = a_1x_1 + q_1(\boldsymbol{x})$ for some $a_1 \in \mathbb{R}$ and some multivariate polynomial q_1 that does not depend on x_1. Finally, one verifies by induction that

$q_1(\boldsymbol{x}) = \sum_{i=2}^{d} a_i x_i + q_0(\boldsymbol{x})$, where q_0 is a multivariate polynomial that does not depend on any of its variables and so is constant, *i.e.* $q_0(\boldsymbol{x}) = b$ for some $b \in \mathbb{R}$. We conclude the proof by remarking that any affine mapping is indeed in the null space of H.

Item 2: By invoking that $\mathrm{H}\{f(\cdot - \boldsymbol{x}_0)\} = \mathrm{H}\{f\}(\cdot - \boldsymbol{x}_0)$, we immediately deduce that

$$\begin{aligned}\mathrm{HTV}_p\left(f(\cdot - \boldsymbol{x}_0)\right) &= \sup\left\{\langle \mathrm{H}\{f\}(\cdot - \boldsymbol{x}_0), \mathbf{F}\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \sup\left\{\langle \mathrm{H}\{f\}, \mathbf{F}(\cdot + \boldsymbol{x}_0)\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \mathrm{HTV}_p(f). \end{aligned} \tag{3.292}$$

Similarly, following the chain rule, we obtain that $\mathrm{H}\{f(\alpha\cdot)\} = \alpha^2 \mathrm{H}\{f\}(\alpha\cdot)$. This yields that

$$\begin{aligned}\mathrm{HTV}_p\left(f(\alpha\cdot)\right) &= \alpha^2 \sup\left\{\langle \mathrm{H}\{f\}(\alpha\cdot), \mathbf{F}\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \alpha^2 \sup\left\{\langle \mathrm{H}\{f\}, \alpha^{-d}\mathbf{F}(\alpha^{-1}\cdot)\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= |\alpha|^{2-d} \sup\left\{\langle \mathrm{H}\{f\}, \mathbf{F}(\cdot)\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= |\alpha|^{2-d} \mathrm{HTV}_p(f). \end{aligned} \tag{3.293}$$

As for the last invariance property, we use the formula for the Hessian of a rotated function

$$\mathrm{H}\{f(\mathbf{U}\cdot)\} = \mathbf{U}^T \mathrm{H}\{f\}(\mathbf{U}\cdot)\mathbf{U}. \tag{3.294}$$

This implies that

$$\begin{aligned}\mathrm{HTV}_p\left(f(\mathbf{U}\cdot)\right) &= \sup\left\{\langle \mathbf{U}^T \mathrm{H}\{f\}(\mathbf{U}\cdot)\mathbf{U}, \mathbf{F}\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \sup\left\{\langle \mathrm{H}\{f\}(\mathbf{U}\cdot), \mathbf{U}\mathbf{F}(\cdot)\mathbf{U}^T\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \sup\left\{\langle \mathrm{H}\{f\}, \mathbf{U}\mathbf{F}(\mathbf{U}^T\cdot)\mathbf{U}^T\rangle : \mathbf{F} \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^{d\times d}), \|\mathbf{F}\|_{S_q, L_\infty} = 1\right\} \\ &= \mathrm{HTV}_p(f), \end{aligned} \tag{3.295}$$

where the last equality follows from the invariance of Schatten norms under orthogonal transformations (as exploited, for example, in [219]). □

Closed-Form Expressions for the HTV of Special Functions

Although Definition 3.12 introduces a formal way to compute the HTV of a given element $f \in \mathcal{D}'(\Omega)$, it is still very abstract and not practical. This is the reason why we now provide closed-form expressions for the HTV of two general classes of functions.

Proposition 3.7 (Sobolev Compatibility). *Let $W_1^2(\Omega)$ be the Sobolev space of twice-differentiable functions $f : \Omega \to \mathbb{R}$ whose second-order partial derivatives are in $L_1(\Omega)$. Then, for any $p \in [1, +\infty]$ and any Sobolev function $f \in W_1^2(\Omega)$, we have that*

$$\mathrm{HTV}_p(f) = \|\mathrm{H}\{f\}\|_{S_p, L_1} = \int_\Omega \|\mathrm{H}\{f\}(\boldsymbol{x})\|_{S_p} \mathrm{d}\boldsymbol{x}. \tag{3.296}$$

Proof. This is a consequence of Theorem 3.15 since, for any $f \in W_1^2(\Omega)$, the matrix-valued function $\mathrm{H}\{f\} : \Omega \to \mathbb{R}^{d\times d} : \boldsymbol{x} \mapsto \mathrm{H}\{f\}(\boldsymbol{x})$ is measurable and is in $L_1(\Omega; \mathbb{R}^{d\times d})$. □

Interestingly, Proposition 3.7 demonstrates that our introduced seminorm is a generalization of the Hessian-Schatten regularization that has been used in inverse problems and image reconstruction [219, 218].

Theorem 3.17 (HTV of CPWL Mappings). *Let $f : \Omega \to \mathbb{R}$ be a CPWL function with linear regions $P_1, \dots, P_N$ and denote the gradient of f at the interior of P_n by $\boldsymbol{a}_n = \nabla f|_{P_n}$ for $n = 1, \dots, N$. Then, for any $p \in [1, +\infty]$, the corresponding HTV of f is given as*

$$\mathrm{HTV}_p(f) = \frac{1}{2} \sum_{n=1}^{N} \sum_{k \in \mathrm{adj}_n} \|\boldsymbol{a}_n - \boldsymbol{a}_k\|_2 H^{d-1}(P_n \cap P_k), \tag{3.297}$$

where adj_n is the set of indices $k \in \{1, \dots, N\}$ such that P_n and P_k are neighbors and H^{d-1} denotes the $(d-1)$-dimensional Hausdorff measure.

Before going to the proof, we first prove a useful lemma.

Lemma 3.5. *Let $f : \Omega \to \mathbb{R}$ be a CPWL function described in Theorem 3.17. Then, the gradient of f can be expressed as*

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \sum_{n=1}^{N} \boldsymbol{a}_n \mathbb{1}_{P_n}(\boldsymbol{x}), \tag{3.298}$$

for almost every $\boldsymbol{x} \in \Omega$.

Proof. The interior of P_n is denoted by U_n with $n = 1 \ldots, N$. We then note that $\Omega \backslash \left(\bigcup_{n=1}^{N} U_n \right)$ is a set of measure zero. Hence, it is sufficient to show that $\boldsymbol{\nabla} f(\boldsymbol{x}) = \boldsymbol{a}_n$ for any $\boldsymbol{x}_0 = (x_{0,1}, \ldots, x_{0,d}) \in U_n$. We define the functions $g_i : \mathbb{R} \to \mathbb{R}$ as

$$g_i(x) = f(x_{0,1}, \ldots, x_{0,i-1}, x, x_{0,i+1}, \ldots, x_{0,d}). \tag{3.299}$$

Following the definition of CPWL mappings, g_i is a linear spline (*i.e.*, a 1D continuous and piecewise-linear function). Hence, it is locally linear and can be expressed as $g_i(x) = a_{n,i}x + (\sum_{j \neq i} a_{n,j} x_{0,j} + b)$ in an open neighborhood of $x_{0,i}$. Moreover, it is clear that $a_{n,i} = g_i'(x_{0,i}) = \frac{\partial f}{\partial x_i}(\boldsymbol{x}_0)$. Hence,

$$\boldsymbol{\nabla} f(\boldsymbol{x}_0) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0), \ldots, \frac{\partial f}{\partial x_d}(\boldsymbol{x}_0) \right) = (a_{n,1}, \ldots, a_{n,d}) = \boldsymbol{a}_n. \tag{3.300}$$

□

Proof of Theorem 3.17. We start by introducing some notions that are required in the proof. For each $n = 1, \ldots, N$ and $k \in \mathrm{adj}_n$, we denote the intersection of P_n and P_k by $L_{n,k} = P_n \cap P_k$, which is itself a convex polytope with co-dimension $(d-1)$, in the sense that it lies on a hyperplane $H_{n,k} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{u}_{n,k}^T \boldsymbol{x} + \beta_{n,k} = 0\}$ for some normal vector $\boldsymbol{u}_{n,k} = (u_{n,k,i}) \in \mathbb{R}^d$ with $\|\boldsymbol{u}_{n,k}\|_2 = 1$ and some shift value $\beta_{n,k} \in \mathbb{R}$. We adopt the convention that $\boldsymbol{u}_{n,k}$ refers to the outward normal vector, so that $\boldsymbol{u}_{n,k}^T \boldsymbol{x} + \beta_{n,k} \leq 0$ for all $\boldsymbol{x} \in P_n$. We divide the proof in four steps:

Step 1: Transformation to the General Position. First, without any loss of generality, we assume that all entries of $\boldsymbol{u}_{n,k}$ for all $n = 1, \ldots, N$ and $k \in \mathrm{adj}_n$ are nonzero. Consider a unitary matrix $\mathbf{V} \in \mathbb{R}^{d \times d}$ such that $[\mathbf{V}\boldsymbol{u}_{n,k}]_i \neq 0$ for all

$n = 1, \dots, N$, $k \in \mathrm{adj}_n$, and $i = 1, \dots, d$. We remark that the function $g = f(\mathbf{V}\cdot)$ is CPWL with linear regions $\tilde{P}_n = \mathbf{V}^T P_n$ and affine parameters $\tilde{\boldsymbol{a}}_n = \mathbf{V}^T \boldsymbol{a}_n$ and $\tilde{b}_n = b_n$ for $n = 1, \dots, N$. Now, if (3.297) holds for g, then we can invoke the invariance properties of the HTV (see Theorem 3.16) to deduce that

$$\begin{aligned}\mathrm{HTV}_p(f) &= \mathrm{HTV}_p\,(g)\\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{k\in\mathrm{adj}_n} \|\tilde{\boldsymbol{a}}_n - \tilde{\boldsymbol{a}}_k\|_2 H^{d-1}(\tilde{P}_n \cap \tilde{P}_k)\\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{k\in\mathrm{adj}_n} \|\mathbf{V}^T(\boldsymbol{a}_n - \boldsymbol{a}_k)\|_2 H^{d-1}(\mathbf{V}^T(P_n \cap P_k))\\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{k\in\mathrm{adj}_n} \|\boldsymbol{a}_n - \boldsymbol{a}_k\|_2 H^{d-1}(P_n \cap P_k), \end{aligned} \tag{3.301}$$

where the last equality is due to the invariance of the Hausdorff measure and the ℓ_2 norm to orthonormal transformations.

Step 2: Calculation of the Hessian Distribution. From now on, we assume that all entries of $u_{n,k}$ are nonzero, with

$$u_{n,k,i} = 0, \qquad n = 0, \dots, N, \quad k \in \mathrm{adj}_n, \quad i = 1, \dots, d. \tag{3.302}$$

This allows us to view $H_{n,k}$ as the graph of the affine mapping $T_{n,k} : \mathbb{R}^{d-1} \to \mathbb{R}$, with

$$T_{n,k}(x_1, \dots, x_{d-1}) = \beta_{n,k} - \frac{\sum_{i=1}^{d-1} u_{n,k,i}x_i}{u_{n,k,d}}, \tag{3.303}$$

and to define $C_{n,k} = \{\boldsymbol{x} \in \mathbb{R}^{d-1} : (\boldsymbol{x}, T_{n,k}\boldsymbol{x}) \in L_{n,k}\} \subseteq \Omega$ as the preimage of $L_{n,k}$ over $T_{n,k}$. We also remark that, due to this affine projection, the $(d-1)$-dimensional Hausdorff measure of $L_{n,k}$ and the Lebesgue measure of $C_{n,k}$ are related by the coefficient $u_{n,k,d}$. Indeed, we have that $H^{d-1}(L_{n,k}) = \frac{\mathrm{Leb}(C_{n,k})}{|u_{n,k,d}|}$. Using these notions, we now compute the matrix-valued distribution $\mathrm{H}\{f\} \in \mathcal{M}(\Omega; \mathbb{R}^{d\times d})$. We first note that, for all $n = 0, \dots, N$ and $i = 1, \dots, d$, we have that

$$\frac{\partial \mathbb{1}_{P_n}}{\partial x_i}(\boldsymbol{x}) = \sum_{k\in\mathrm{adj}_n} -\mathrm{sgn}(u_{n,k,i})\delta\left(x_i + \frac{\sum_{j\neq i} u_{n,k,j}x_j + \beta_{n,k}}{u_{n,k,i}}\right)\mathbb{1}_{L_{n,k}}(\boldsymbol{x}). \tag{3.304}$$

Using the relation $\delta(\alpha\cdot) = |\alpha|^{-1}\delta(\cdot)$ for all $\alpha \in \mathbb{R}$, we obtain that

$$\frac{\partial \mathbb{1}_{P_n}}{\partial x_i}(\boldsymbol{x}) = \sum_{k \in \mathrm{adj}_n} \frac{-u_{n,k,i}}{|u_{n,k,d}|} \delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{L_{n,k}}(\boldsymbol{x}), \tag{3.305}$$

where $\boldsymbol{x}_1 = (x_1, \dots, x_{d-1}) \in \mathbb{R}^{d-1}$. Following the definition of $C_{n,k}$, we immediately get that

$$\delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{L_{n,k}}(\boldsymbol{x}) = \delta\left(x_d - T_{n,k}\boldsymbol{x}_2\right) \mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1), \tag{3.306}$$

which leads to

$$\frac{\partial \mathbb{1}_{P_n}}{\partial x_i}(\boldsymbol{x}) = \sum_{k \in \mathrm{adj}_n} \frac{-u_{n,k,i}}{|u_{n,k,d}|} \delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1). \tag{3.307}$$

Combining (3.307) with Lemma 3.5, we then deduce that

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \sum_{n=1}^{N} a_{n,j} \frac{\partial \mathbb{1}_{P_n}}{\partial x_i}(\boldsymbol{x}) = \sum_{n=1}^{N} a_{n,j} \sum_{k \in \mathrm{adj}_n} \frac{-u_{n,k,i}}{|u_{n,k,d}|} \delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1). \tag{3.308}$$

Now, since $L_{n,k} = P_n \cap P_k$ and $\boldsymbol{u}_{n,k} = (-\boldsymbol{u}_{k,n})$, we can rewrite the second-order partial derivatives as

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \frac{1}{2} \sum_{n=1}^{N} \sum_{k \in \mathrm{adj}_n} (a_{k,j} - a_{n,j}) \frac{u_{n,k,i}}{|u_{n,k,d}|} \delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1). \tag{3.309}$$

Putting it in matrix form, we conclude that the Hessian is a sum of disjoint Dirac fences, as in

$$\mathrm{H}\{f\}(\boldsymbol{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x})\right] = \frac{1}{2} \sum_{n=1}^{N} \sum_{k \in \mathrm{adj}_n} \left[(a_{k,j} - a_{n,j}) \frac{u_{n,k,i}}{|u_{n,k,d}|}\right] \delta\left(x_d - T_{n,k}\boldsymbol{x}_1\right) \mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1). \tag{3.310}$$

Step 3: Computation of the HTV. By invoking Item 3 of Theorem 3.15, we deduce that

$$\begin{aligned}\mathrm{HTV}_p(f) &= \|\mathrm{H}\{f\}\|_{S_p,\mathcal{M}}\\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{k\in\mathrm{adj}_n}\left\|\left[(a_{k,j}-a_{n,j})\frac{u_{n,k,i}}{|u_{n,k,d}|}\right]\delta\left(x_d-T_{n,k}\boldsymbol{x}_1\right)\mathbb{1}_{C_{n,k}}(\boldsymbol{x}_1)\right\|_{S_p,\mathcal{M}}\\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{k\in\mathrm{adj}_n}\left\|\left[(a_{k,j}-a_{n,j})\frac{u_{n,k,i}}{|u_{n,k,d}|}\right]\right\|_{S_p}\mathrm{Leb}(C_{n,k}),\end{aligned} \tag{3.311}$$

where the last equality results from Item 2 of Theorem 3.15.

Finally, we use the continuity of f to deduce that, for any pair of points $\boldsymbol{p}_1,\boldsymbol{p}_2\in H_{n,k}$, we have that

$$\boldsymbol{a}_n^T\boldsymbol{p}_i+b_n=\boldsymbol{a}_k^T\boldsymbol{p}_i+b_k,\qquad i=1,2. \tag{3.312}$$

Subtracting the above equalities for $i=1$ and $i=2$, we obtain that

$$\boldsymbol{a}_n^T(\boldsymbol{p}_1-\boldsymbol{p}_2)=\boldsymbol{a}_k^T(\boldsymbol{p}_1-\boldsymbol{p}_2). \tag{3.313}$$

However, $(\boldsymbol{p}_1-\boldsymbol{p}_2)$ is orthogonal to $\boldsymbol{u}_{n,k}$. Hence, the vector $(\boldsymbol{a}_k-\boldsymbol{a}_n)$ points in the direction of $\boldsymbol{u}_{n,k}$. This implies that the matrix

$$\left[(a_{k,j}-a_{n,j})\frac{u_{n,k,i}}{|u_{n,k,d}|}\right]=|u_{n,k,d}|^{-1}\boldsymbol{u}_{n,k}(\boldsymbol{a}_k-\boldsymbol{a}_n)^T=\frac{\|\boldsymbol{a}_k-\boldsymbol{a}_n\|_2}{|u_{n,k,d}|}\boldsymbol{u}_{n,k}\boldsymbol{u}_{n,k}^T \tag{3.314}$$

is rank-1 and symmetric. Hence, for any $p\in[1,+\infty]$, its Schatten-p norm is equal to the absolute value of its trace. The replacement of this in (3.311) and the use of $H^{d-1}(L_{n,k})=\frac{\mathrm{Leb}(C_{n,k})}{|u_{n,k,d}|}$ yields the announced expression (3.297). □

We conclude from (3.297) that the HTV seminorm accounts for the change of (directional) slope in all the junctions in the partitioning. Specifically, the HTV of a CPWL function is proportional to a weighted ℓ_1 penalty on the vector of slope changes, where the weights are proportional to the volume of the intersection region. This can be seen as a convex relaxation of the number of linear regions,. The latter has the disadvantage that is unable to differentiate between small and large changes of slope. Another noteworthy observation is the invariance of the HTV of CPWL functions to the value of $p \in [1, +\infty)$, which is unlike the case of Sobolev functions in Proposition 3.7. This is due to the extreme sparsity of the Hessian of CPWL functions. In fact, the Hessian matrix is zero everywhere except at the borders of linear regions. There, it is a Dirac fence weighted by a rank-1 matrix. The invariance then follows from the observation that the Schatten-p norms collapse to a single value in rank-1 matrices (*i.e.,* their only nonzero singular value).

3.5.3 Learning with HTV

To further illustrate the practical aspect of our framework, we present[10] a method to learn sparse CPWL functions with HTV regularization in 2D. We consider a CPWL search space spanned by linear-box-spline basis functions. These choices allow us to recast the infinite-dimensional learning problem as a finite one which can be efficiently solved using known optimization algorithms. Our framework presents itself as an alternative to training ReLU networks for the learning of CPWL functions, with the following advantages:

1. the enforcement of sparsity, in the sense that we follow Occam's razor principle by promoting solutions with the fewest CPWL regions;
2. the use of a rotation, scale and translation-invariant regularization;
3. the reliance on a single hyperparameter—the regularization weight λ. This is in contrast with the numerous hyperparameters found in neural networks such as the choice of architecture and its components, learning rate schemes, and batch size, among others;
4. an improved model interpretability since we provide a linear parametrization for the learned CPWL mapping.

Search Space

We remark that, although the HTV of CPWL functions has a simple closed-form expression, it requires the complete knowledge of the domain partition and the gradients in each polytope. To circumvent this, we construct a CPWL search space that is based on a uniform domain partition. This allows us to obtain a tractable formula for computing the HTV of any model in the space.

More precisely, we let this space be spanned by shifts of a CPWL basis function

[10] From our published work [109].

φ of the form

$$\varphi(x_1, x_2) = [1 - \max(0, a_1, a_2) + \min(0, a_1, a_2)]_+ , \tag{3.315}$$

where $a_1 = (x_1 - x_2/\sqrt{3})$, $a_2 = (-2x_2/\sqrt{3})$, and $[x]_+ \overset{\Delta}{=} \max(x, 0)$.

We note that φ is, in fact, a scaled hexagonal box spline [264] (see, Proposition 3.8 below). Box splines are multivariate extensions of B-splines [265]. In constrast to tensor products of 1D B-splines, they are non-separable, which makes them suitable for interpolation algorithms taylored to non-Cartesian (and often optimal) sampling lattices [266, 267, 264]. They also find applications in areas such as finite-element methods [268], computer-aided design [269], and edge detection [270].

In dimension $d = 2$, we denote by $k_{\boldsymbol{\Xi}} : \mathbb{R}^2 \mapsto \mathbb{R}$, the box spline associated to the matrix $\boldsymbol{\Xi} = \begin{bmatrix}\boldsymbol{\xi}_1 & \cdots & \boldsymbol{\xi}_n\end{bmatrix} \in \mathbb{R}^{2\times n}$. For $n = 2$ and an invertible matrix $\boldsymbol{\Xi} \in \mathbb{R}^{2\times 2}$, $k_{\boldsymbol{\Xi}}$ is an indicator function of the form

$$k_{\boldsymbol{\Xi}}(\mathbf{x}) = \frac{1}{|\det(\boldsymbol{\Xi})|} \mathbb{1}_{S_2}(\mathbf{x}), \tag{3.316}$$

where $S_2 = \left\{\boldsymbol{\Xi}\boldsymbol{\alpha} : \quad \boldsymbol{\alpha} \in [0, 1)^2\right\}$. For $n \geq 3$, the box spline is defined recursively as

$$k_{[\boldsymbol{\Xi} \quad \boldsymbol{\xi}]}(\mathbf{x}) = \int_0^1 k_{\boldsymbol{\Xi}}(\mathbf{x} - t\boldsymbol{\xi})\, \mathrm{d}t. \tag{3.317}$$

Box splines are nonnegative functions and have a unit integral over the entire space, with $\int_{\mathbb{R}^2} k_{\boldsymbol{\Xi}}(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$. Moreover, they are supported over the set $\left\{\boldsymbol{\Xi}\boldsymbol{\alpha} : \quad \boldsymbol{\alpha} \in [0, 1)^n\right\}$ and are symmetric with respect to the center of their support [269].

Proposition 3.8. *The basis function φ is a scaled box spline whose associated matrix $\boldsymbol{\Xi}$ is (see Figure 3.19)*

$$\boldsymbol{\Xi} = \begin{bmatrix}\boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \boldsymbol{\xi}_3\end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}. \tag{3.318}$$

Proof. Let $\alpha = 2/\sqrt{3}$ and $\boldsymbol{\Xi} \in \mathbb{R}^{2\times 3}$, the matrix given in (3.318). It follows from

Definition (3.317) that

$$k_{\boldsymbol{\Xi}}(\mathbf{x}) = \alpha \int_0^1 \mathbb{1}_{S_2}(\mathbf{x} - t\boldsymbol{\xi}_3), \in \mathrm{d}t = \alpha \operatorname{supp}\Big([0,1] \cap \{t : \ \mathbf{x} - t\boldsymbol{\xi}_3 \in S_2\}\Big), \quad (3.319)$$

where $\operatorname{supp}(B)$ is the length of the interval $B \in \mathbb{R}$ and $S_2 = \Big\{t_1\boldsymbol{\xi}_1 + t_2\boldsymbol{\xi}_2 : \ t_1, t_2 \in [0,1)\Big\}$. By expressing $\mathbf{x}$ in the basis $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\}$, with $\mathbf{x} = a_1\boldsymbol{\xi}_1 + a_2\boldsymbol{\xi}_2$, $a_1, a_2 \in \mathbb{R}$, and using the relation $(-\boldsymbol{\xi}_3) = \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2$, we deduce that

$$\begin{aligned} k_{h\boldsymbol{\Xi}}(\mathbf{x}) &= \alpha \operatorname{supp}\Big([0,1] \cap \{t : \ (a_1 + t)\boldsymbol{\xi}_1 + \\ &\quad (a_2 + t)\boldsymbol{\xi}_2 \in S_2\}\Big) \\ &= \alpha \operatorname{supp}\Big(\{t : \ 0 \leq t \leq 1, \ 0 \leq a_1 + t \leq 1, \\ &\quad 0 \leq a_2 + t \leq 1\}\Big) \\ &= \alpha \left[\min(1, 1 - a_1, 1 - a_2) - \max(0, -a_1, -a_2)\right]_+ \\ &= \alpha \left[1 - \max(0, a_1, a_2) + \min(0, a_1, a_2)\right]_+ . \end{aligned} \quad (3.320)$$

Finally, dividing both sides by α, we reach the desired result. □

Additionally, we construct a hexagonal lattice on which we shift these basis functions. This lattice is determined by the primitive vectors $\mathbf{r}_1 = \boldsymbol{\xi}_1$ and $\mathbf{r}_2 = (-\boldsymbol{\xi}_2)$. For ease of representation, we concatenate the primitive vectors into the lattice matrix $\mathbf{R} = \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2\end{bmatrix}$ [271]. Then, the search space with grid size $h \in \mathbb{R}_+$ is defined as

$$\mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2) = \operatorname{span}\left(\left\{\varphi\left(\frac{\cdot}{h} - \mathbf{R}\mathbf{k}\right)\right\}_{\mathbf{k} \in \mathbb{Z}^2}\right). \quad (3.321)$$

We observe that any model $f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ can be expressed as

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^2} c[\mathbf{k}]\varphi\left(\frac{\mathbf{x}}{h} - \mathbf{R}\mathbf{k}\right) \quad (3.322)$$

for some set of box-spline coefficients $\{c[\mathbf{k}]\}_{\mathbf{k} \in \mathbb{Z}^2}$.

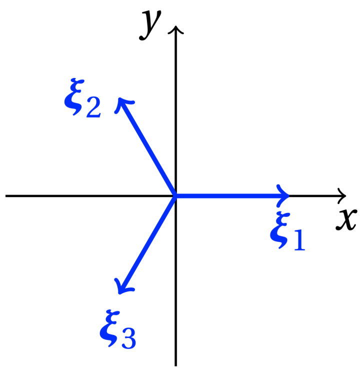


Figure 3.19: Hexagonal box-spline vectors.

Analogous to the space of cardinal linear splines [271, 33], our search space satisfies some desirable properties that are listed in Theorem 3.18.

Theorem 3.18. *The search space $\mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ satisfies the following properties.*

1. *It reproduces any affine mapping, in the sense that any function of the form $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ can be expressed as* (3.322)*.*
2. *The collection $\{\varphi(\cdot/h - \mathbf{Rk})\}_{\mathbf{k}\in\mathbb{Z}^2}$ forms a Riesz basis for $\mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$. This ensures a unique and stable link between each model function and its coefficients [35].*
3. *The approximation error of our search space decays with h^{-2} as $h \to 0$.*
4. *The atoms satisfy the interpolatory condition*

$$\forall \mathbf{k} \in \mathbb{Z}^2 : \quad \varphi(\mathbf{Rk}) = \begin{cases} 1, & \mathbf{k} = 0 \\ 0, & \textit{otherwise.} \end{cases} \tag{3.323}$$

Consequently, we have that $f(h\mathbf{Rk}) = c[\mathbf{k}]$ for any $f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ and any $\mathbf{k} \in \mathbb{Z}^2$.

5. *The basis element φ is refinable, in the sense that $\varphi(\cdot/2h)$ can be exactly represented in $\mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ with finitely many coefficients.*

Proof. We prove the properties for a unit grid size ($h = 1$), without any loss of generality. To do so, we rely on the Fourier-domain characterization of a generic box spline [272, Proposition 17]. Specifying it for the case of the hexagonal box spline whose vectors are given in (3.318), and using $\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 + \boldsymbol{\xi}_3 = 0$ and the relation $\varphi = \sqrt{3}/2\, k_{\boldsymbol{\Xi}}$, we get that

$$\widehat{\varphi}(\boldsymbol{\omega}) = \frac{\sqrt{3}}{2} \prod_{n=1}^{3} \mathrm{sinc}\left(\frac{\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle}{2}\right), \tag{3.324}$$

where $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$.

Item 1 (Reproduction of affine mappings): We begin by proving that $\{\varphi(\cdot - \mathbf{R}\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^2}$ satisfies the partition-of-unity property

$$\sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi(\mathbf{x} - \mathbf{R}\mathbf{k}) = 1, \qquad \forall \mathbf{x} \in \mathbb{R}^2. \tag{3.325}$$

This condition implies that the search space is able to reproduce any constant function.

Let $\tilde{\mathbf{R}}$ be the lattice matrix expressed as $\tilde{\mathbf{R}} = \begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 \end{bmatrix}$. One can readily verify that $\sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi(\mathbf{x} - \mathbf{R}\mathbf{k}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi(\mathbf{x} - \tilde{\mathbf{R}}\mathbf{k})$. From the Poisson-sum formula for lattices [271], the partition-of-unity property holds if and only if, for any $\mathbf{k} \in \mathbb{Z}^2$, we have that

$$\frac{1}{\left|\det(\tilde{\mathbf{R}})\right|} \widehat{\varphi}(2\pi \tilde{\mathbf{R}}^{-T}\mathbf{k}) = \begin{cases} 1, & \mathbf{k} = \mathbf{0} \\ 0, & \mathbf{k} \neq \mathbf{0}. \end{cases} \tag{3.326}$$

Evaluating the Fourier transform (3.324) at the selected locations and using $\left|\det(\tilde{\mathbf{R}})\right| = \frac{\sqrt{3}}{2}$, we infer that

$$\frac{1}{\left|\det(\tilde{\mathbf{R}})\right|} \widehat{\varphi}(2\pi \tilde{\mathbf{R}}^{-T}\mathbf{k}) = \prod_{n=1}^{3} \mathrm{sinc}\Big(\pi \langle \tilde{\mathbf{R}}^{-T}\mathbf{k}, \boldsymbol{\xi}_n \rangle\Big) = \prod_{n=1}^{3} \mathrm{sinc}\Big(\pi \langle \mathbf{k}, \tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_n \rangle\Big). \tag{3.327}$$

We then observe that $\tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_1 = (1,0)$, $\tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_2 = (0,1)$, and $\tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_3 = (-1,-1)$. This results in

$$\frac{1}{\left|\det(\tilde{\mathbf{R}})\right|}\widehat{\varphi}(2\pi\tilde{\mathbf{R}}^{-T}\mathbf{k}) = \mathrm{sinc}\Big(\pi(k_1+k_2)\Big)\prod_{n=1}^{2}\mathrm{sinc}(\pi k_n) = \begin{cases}1, & \mathbf{k}=\mathbf{0}\\ 0, & \mathbf{k}\neq\mathbf{0},\end{cases} \tag{3.328}$$

where, in the last equality, we have used that $\mathrm{sinc}(\pi k) = \delta[k]$ for any $k \in \mathbb{Z}$.

Now, we show that the search space can approximate any linear function. Following the Strang-Fix conditions [266, 273], we just need to prove that

$$\nabla\widehat{\varphi}(2\pi\tilde{\mathbf{R}}^{-T}\mathbf{k}) = \mathbf{0},\ \forall\,\mathbf{k} \in \mathbb{Z}^2\setminus\{\mathbf{0}\}. \tag{3.329}$$

Using the product rule for differentiation, we observe that

$$\nabla\widehat{\varphi}(\boldsymbol{\omega}) = \frac{\sqrt{3}}{2}\sum_{n=1}^{3}\nabla\mathrm{sinc}\left(\frac{\langle\boldsymbol{\omega},\boldsymbol{\xi}_n\rangle}{2}\right)\prod_{m\neq n}\mathrm{sinc}\left(\frac{\langle\boldsymbol{\omega},\boldsymbol{\xi}_m\rangle}{2}\right). \tag{3.330}$$

Evaluating this expression at $\boldsymbol{\omega} = 2\pi\tilde{\mathbf{R}}^{-T}\mathbf{k}$ and defining $\boldsymbol{\beta}_{n,\mathbf{k}} = \left(\sqrt{3}/2\right)\nabla\mathrm{sinc}(\pi\langle\tilde{\mathbf{R}}^{-T}\mathbf{k},\boldsymbol{\xi}_n\rangle)$ we obtain that

$$\nabla\widehat{\varphi}(2\pi\tilde{\mathbf{R}}^{-T}\mathbf{k}) = \sum_{n=1}^{3}\boldsymbol{\beta}_{n,\mathbf{k}}\prod_{m\neq n}\mathrm{sinc}\Big(\pi\langle\tilde{\mathbf{R}}^{-T}\mathbf{k},\boldsymbol{\xi}_m\rangle\Big). \tag{3.331}$$

Then, we use that $\mathrm{sinc}\Big(\pi\langle\tilde{\mathbf{R}}^{-T}\mathbf{k},\boldsymbol{\xi}_m\rangle\Big) = \mathrm{sinc}\Big(\pi\langle\mathbf{k},\tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_m\rangle\Big)$ to deduce that

$$\begin{aligned}\nabla\widehat{\varphi}(2\pi\tilde{\mathbf{R}}^{-T}\mathbf{k}) = {} & \boldsymbol{\beta}_{1,\mathbf{k}}\mathrm{sinc}(\pi k_2)\mathrm{sinc}(\pi(k_1+k_2))\\ & + \boldsymbol{\beta}_{2,\mathbf{k}}\mathrm{sinc}(\pi k_1)\mathrm{sinc}(\pi(k_1+k_2))\\ & + \boldsymbol{\beta}_{3,\mathbf{k}}\mathrm{sinc}(\pi k_1)\mathrm{sinc}(\pi k_2).\end{aligned} \tag{3.332}$$

Finally, since $\mathrm{sinc}(\pi k) = \delta[k]$ for any $k \in \mathbb{Z}$, all terms in (3.332) vanish for $\mathbf{k} \in \mathbb{Z}\setminus\{\mathbf{0}\}$.

Item 2 (Riesz basis): The collection $\{\varphi(\cdot - \mathbf{Rk})\}_{\mathbf{k}\in\mathbb{Z}^2}$ is a Riesz basis if there exist $\lambda_{\min} > 0$ and $\lambda_{\max} < +\infty$ such that, for any sequence $c \in \ell_2(\mathbb{Z})$, we have that

$$\lambda_{\min} \|c\|_2^2 \leq \left\| \sum_{\mathbf{k}\in\mathbb{Z}^2} c[\mathbf{k}]\varphi(\cdot - \mathbf{Rk}) \right\|_{L_2}^2 \leq \lambda_{\max} \|c\|_2^2 . \tag{3.333}$$

To show that (3.333) is valid for the collection of our shifted search-space atoms, we use Fourier-based conditions in the spirit of [271] and [35]. This leads to the bounds

$$\lambda_{\min} = \min_{[0,2\pi)^2} \frac{1}{|\det(\mathbf{R})|} \sum_{\mathbf{k}\in\mathbb{Z}^2} \left| \widehat{\varphi}(\mathbf{R}^{-T}\Big(\boldsymbol{\omega} + 2\pi\mathbf{k})\Big) \right|^2, \quad \lambda_{\max} = \max_{[0,2\pi)^2} \frac{1}{|\det(\mathbf{R})|} \sum_{\mathbf{k}\in\mathbb{Z}^2} \left| \widehat{\varphi}(\mathbf{R}^{-T}\Big(\boldsymbol{\omega} + 2\pi\mathbf{k})\Big) \right. \tag{3.334}$$

To obtain a more tractable expression for the summation on the right-hand side of (3.334), we set $\mathbf{x} = \mathbf{0}$ in the Poisson-sum formula for lattices [271] and deduce that

$$\sum_{\mathbf{k}\in\mathbb{Z}^2} f(\mathbf{Rk}) = \frac{1}{|\det(\mathbf{R})|} \sum_{\mathbf{k}\in\mathbb{Z}^2} \widehat{f}(2\pi\mathbf{R}^{-T}\mathbf{k}). \tag{3.335}$$

Then, we consider the function $f(\boldsymbol{\tau}) = c_{\varphi\varphi}(\boldsymbol{\tau})\mathrm{e}^{-\mathrm{j}\langle\mathbf{R}^{-T}\boldsymbol{\omega}_0,\boldsymbol{\tau}\rangle}$, where $c_{\varphi\varphi}(\boldsymbol{\tau}) = \langle\varphi(\cdot - \boldsymbol{\tau}), \varphi\rangle$, which results in

$$\sum_{\mathbf{k}\in\mathbb{Z}^2} \langle\varphi(\cdot - \mathbf{Rk}), \varphi\rangle \mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\mathbf{k}\rangle} = \frac{1}{|\det(\mathbf{R})|} \sum_{\mathbf{k}\in\mathbb{Z}^2} \left| \widehat{\varphi}\Big(\mathbf{R}^{-T}(\boldsymbol{\omega} + 2\pi\mathbf{k})\Big) \right|^2, \tag{3.336}$$

where we have used that $\widehat{c}_{\varphi\varphi}(\boldsymbol{\omega}) = |\widehat{\varphi}(\boldsymbol{\omega})|^2$ and taken advantage of the modulation property of the Fourier transform.

Due to the fact that φ is finitely supported, the summation on the left-hand side of (4.39) contains only 7 nonzero terms: 1 term corresponding to the energy of the atom and 6 others corresponding to the inner product with overlapping replicas. Therefore, (4.39) can be expanded as

$$\begin{aligned} \sum_{\mathbf{k}\in\mathbb{Z}^2} \langle\varphi(\cdot - \mathbf{Rk}), \varphi\rangle \mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\mathbf{k}\rangle} =& \langle\varphi(\cdot - \mathbf{r}_1), \varphi\rangle \mathrm{e}^{-\mathrm{j}\omega_1} + \langle\varphi(\cdot + \mathbf{r}_1), \varphi\rangle \mathrm{e}^{\mathrm{j}\omega_1} \qquad (3.337) \\ & + \langle\varphi(\cdot - \mathbf{r}_2), \varphi\rangle \mathrm{e}^{-\mathrm{j}\omega_2} + \langle\varphi(\cdot + \mathbf{r}_2), \varphi\rangle \mathrm{e}^{\mathrm{j}\omega_2} \qquad (3.338) \\ & + \langle\varphi(\cdot + \mathbf{r}_1 - \mathbf{r}_2), \varphi\rangle \mathrm{e}^{-\mathrm{j}(\omega_2 - \omega_1)} + \langle\varphi(\cdot - \mathbf{r}_1 + \mathbf{r}_2), \varphi\rangle \mathrm{e}^{\mathrm{j}(\omega_2 - \omega_1)} + \|\varphi\|_{L_2}^2 . \qquad (3.339) \end{aligned}$$

We remark that the pairs of conjugate exponentials in (3.339) do arise due to the symmetry in the location of the replicas. By simple computations, we deduce that $\|\varphi\|_{L_2}^2 = \frac{\sqrt{3}}{4}$ and $\langle \varphi_n, \varphi \rangle = \frac{\sqrt{3}}{12}$ for any of the replicas φ_n, $n = 1, \dots, 6$. (Due to symmetries, the inner products are all equal.) Combining (3.339) with (3.334) and (4.39), we conclude that

$$\lambda_{\min} = \frac{\sqrt{3}}{12} \min_{[0,2\pi)^2} (3 + \cos(\omega_1) + \cos(\omega_2) + \cos(-\omega_1 + \omega_2)) = \frac{\sqrt{3}}{8} > 0, \tag{3.340}$$

$$\lambda_{\max} = \frac{\sqrt{3}}{12} \max_{[0,2\pi)^2} (3 + \cos(\omega_1) + \cos(\omega_2) + \cos(-\omega_1 + \omega_2)) = \frac{\sqrt{3}}{2} < +\infty, \tag{3.341}$$

which completes the proof.

Item 3 (Order of approximation): From Items 1 and 2, we know that the collection of the search-space atoms $\{\varphi(\cdot - \mathbf{R}\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^2}$ forms a Riesz basis and reproduces first-degree polynomials. Hence, it satisfies the first-order Strang-Fix conditions [274, Theorem 2.2.]. It follows that

$$\left\| f - \mathrm{Proj}_{\mathcal{X}_{\mathbf{R}_\mathrm{h}}}\{f\} \right\|_{L_2} = \mathcal{O}(h^{-2}), \ h \to 0 \tag{3.342}$$

for any sufficiently smooth function $f : \mathbb{R}^2 \to \mathbb{R}$ [273].

Item 4 (Interpolatory atoms): Evaluating the partition of unity at $\mathbf{x} = \mathbf{R}\mathbf{k}'$, we have that

$$\sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi(\mathbf{R}(\mathbf{k}' - \mathbf{k})) = \varphi(\mathbf{0}) + \sum_{\mathbf{k}\neq\mathbf{k}'} \varphi(\mathbf{R}(\mathbf{k}' - \mathbf{k})) = 1. \tag{3.343}$$

Since $\varphi(\mathbf{0}) = 1$ and $\varphi(\mathbf{x}) \geq 0, \ \forall \mathbf{x} \in \mathbb{R}^2$, it follows that

$$\forall \mathbf{k} \in \mathbb{Z}^2 : \quad \varphi(\mathbf{R}\mathbf{k}) = \begin{cases} 1, & \mathbf{k} = 0 \\ 0, & \text{Otherwise.} \end{cases} \tag{3.344}$$

Item 5 (Refinable search space): We want to show that there exists a refinability filter $h \in \ell_2(\mathbb{R}^2)$ such that

$$\varphi\left(\frac{\mathbf{x}}{2}\right) = \sum_{\mathbf{k}\in\mathbb{Z}^2} h[\mathbf{k}]\varphi(\mathbf{x} - \tilde{\mathbf{R}}\mathbf{k}), \tag{3.345}$$

where $\tilde{\mathbf{R}} = \begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 \end{bmatrix}$. In the Fourier domain, this condition is equivalent to

$$2^2 \widehat{\varphi}(2\boldsymbol{\omega}) = \mathrm{H}(\mathrm{e}^{\mathrm{j}\tilde{\mathbf{R}}^T \boldsymbol{\omega}}) \widehat{\varphi}(\boldsymbol{\omega}). \tag{3.346}$$

Computing $2^2 \widehat{\varphi}(2\boldsymbol{\omega}) / \widehat{\varphi}(\boldsymbol{\omega})$, we deduce that

$$\mathrm{H}(\mathrm{e}^{\mathrm{j}\tilde{\mathbf{R}}^T \boldsymbol{\omega}}) = \frac{4\widehat{\varphi}(2\boldsymbol{\omega})}{\widehat{\varphi}(\boldsymbol{\omega})} = 4 \prod_{n=1}^{3} \frac{\operatorname{sinc}(\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle)}{\operatorname{sinc}\left(\frac{\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle}{2}\right)} = 2 \prod_{n=1}^{3} \frac{\sin(\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle)}{\sin\left(\frac{\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle}{2}\right)} = 4 \prod_{n=1}^{3} \cos\left(\frac{\langle \boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle}{2}\right), \tag{3.347}$$

where we have used the identity $\sin(x) = 2\cos(x/2)\sin(x/2)$. Observing that $\mathrm{H}(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) = \mathrm{H}(\mathrm{e}^{\mathrm{j}\tilde{\mathbf{R}}^T(\tilde{\mathbf{R}}^{-T}\boldsymbol{\omega})})$, we get that

$$\mathrm{H}(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) = 4 \prod_{n=1}^{3} \cos\left(\frac{\langle \boldsymbol{\omega}, \tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_n \rangle}{2}\right) = \frac{1}{2}\left(1 + \mathrm{e}^{-\mathrm{j}w_1}\right)\left(1 + \mathrm{e}^{-\mathrm{j}w_2}\right)\left(1 + \mathrm{e}^{\mathrm{j}(w_1 + w_2)}\right), \tag{3.348}$$

where $\langle \tilde{\mathbf{R}}^{-T}\boldsymbol{\omega}, \boldsymbol{\xi}_n \rangle = \langle \boldsymbol{\omega}, \tilde{\mathbf{R}}^{-1}\boldsymbol{\xi}_n \rangle$ and $\boldsymbol{\xi}_3 = (-\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$. Taking the inverse Fourier transform of (3.348), we write that

$$\begin{aligned} h[k_1, k_2] = \delta[k_1, k_2] + \frac{1}{2}\Big(&\delta[k_1 - 1, k_2] + \delta[k_1 + 1, k_2] \\ &+ \delta[k_1, k_2 - 1] + \delta[k_1, k_2 + 1] \\ &+ \delta[k_1 - 1, k_2 - 1] + \delta[k_1 + 1, k_2 + 1]\Big). \end{aligned} \tag{3.349}$$

Finally, replacing (3.349) in (3.345), and again using that $\boldsymbol{\xi}_3 = (-\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$, we obtain that

$$\begin{aligned} \varphi\left(\frac{\mathbf{x}}{2}\right) = \frac{1}{2} \sum_{\mathbf{k} \in \{0,1\}^3} &\varphi(\mathbf{x} - \boldsymbol{\Xi}\mathbf{k}) = \\ &+ \varphi(\mathbf{x}) + \frac{1}{2}\Big(\varphi(\mathbf{x} - \boldsymbol{\xi}_1) + \varphi(\mathbf{x} - \boldsymbol{\xi}_2) \\ &+ \varphi(\mathbf{x} - \boldsymbol{\xi}_3) + \varphi(\mathbf{x} - \boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) \\ &+ \varphi(\mathbf{x} - \boldsymbol{\xi}_1 - \boldsymbol{\xi}_3) + \varphi(\mathbf{x} - \boldsymbol{\xi}_2 - \boldsymbol{\xi}_3)\Big). \end{aligned} \tag{3.350}$$

□

Finally, we formalize the functional-learning problem in our search space as the minimization

$$\min_{f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)} \left(\sum_{m=1}^{M} E\Big(f(\mathbf{x}_m), y_m\Big) + \lambda \mathrm{HTV}(f) \right). \tag{3.351}$$

where $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a strictly convex loss function (*e.g.* $E(y,z) = (y-z)^2$ for the quadratic loss). Note that we removed the subscript p in denoting the HTV seminorm, due to the invariance of the latter to a specific choice of $p \in [1, +\infty]$ (see, Theorem 3.17).

Exact Discretization

We now detail the algorithm we have developed to solve problem (3.351). The first step is to express the value of a function $f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ at a given point $\mathbf{x} \in \mathbb{R}^2$ in terms of the spline coefficients $c[\cdot]$ associated to f. Due to the finite support of the atoms, we infer that, there are at most three basis functions that are active in the computation of $f(\mathbf{x})$. These active atoms are located at the vertices of the triangle to which $\mathbf{x}$ belongs. For each datapoint $\mathbf{x}_m$, let us denote its triangle by the index set $\{\mathbf{k}_{m,1}, \mathbf{k}_{m,2}, \mathbf{k}_{m,3}\}$. From this, we express $f(\mathbf{x}_m)$ as

$$f(\mathbf{x}_m) = \sum_{n=1}^{3} c[\mathbf{k}_{m,n}] \varphi\left(\frac{\mathbf{x}_m}{h} - \mathbf{R}\mathbf{k}_{m,n}\right) = \mathbf{h}_m^T(c[\mathbf{k}_{m,1}], c[\mathbf{k}_{m,2}], c[\mathbf{k}_{m,3}]), \tag{3.352}$$

where $h_{m,n} = \varphi(\mathbf{x}_m/h - \mathbf{R}\mathbf{k}_{m,n})$, $n = 1, 2, 3$.

The next step is to compute the HTV of any element in our search space.

Theorem 3.19. *For any $f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)$ of the form* (3.322)*, we have that*

$$\mathrm{HTV}(f) = \|d_1 * c\|_{1,1} + \|d_2 * c\|_{1,1} + \|d_3 * c\|_{1,1}, \tag{3.353}$$

where $\|\mathbf{A}\|_{1,1} = \|\mathrm{vec}(\mathbf{A})\|_1$ *is the sum of the absolute values of the entries of* $\mathbf{A}$*, and*

$$d_1 = \begin{bmatrix} a & -a & 0 \\ 0 & -a & a \end{bmatrix}, \, d_2 = \begin{bmatrix} a & 0 \\ -a & -a \\ 0 & a \end{bmatrix}, \, d_3 = \begin{bmatrix} -a & a \\ a & -a \end{bmatrix}, \tag{3.354}$$

with $a = \frac{2\sqrt{3}}{3}$*.*

Proof. Let Δ denote the set of triangles that form the domain partition of our search space. We have that

$$\mathrm{HTV}(f) = \frac{h}{2} \sum_{P \in \Delta} \sum_{\tilde{P} \in \mathrm{adj}(P)} \left\| \nabla f \Big|_P - \nabla f \Big|_{\tilde{P}} \right\|_2 . \tag{3.355}$$

Due to the specific form of our search space, we can rewrite (3.355) as a summation over the lattice vertices rather than the triangles and associate three junctions to each vertex. This leads to

$$\mathrm{HTV}(f) = h \sum_{\mathbf{n} \in \mathbb{Z}^2} \sum_{k=1}^{3} \|\mathbf{a}_{\mathbf{n}_k} - \mathbf{a}_{\mathbf{n}}\|_2 , \tag{3.356}$$

where $\mathbf{a}_{\mathbf{n}}$ is the gradient of the triangle $P_{\mathbf{n}}$ associated with the vertex $\mathbf{n}$. Similarly, the vector $\mathbf{a}_{\mathbf{n}_k}$ is the gradient of the neighboring triangle that shares a border with $P_{\mathbf{n}}$ in the direction of $\mathbf{r}_k$, where $\mathbf{r}_1 = (1, 0)$, $\mathbf{r}_2 = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, and $\mathbf{r}_3 = \mathbf{r}_2 - \mathbf{r}_1$. By changing the order of summation, we obtain that

$$\mathrm{HTV}(f) = h \sum_{\mathbf{n} \in \mathbb{Z}^2} \|\mathbf{a}_{\mathbf{n}_1} - \mathbf{a}_{\mathbf{n}}\|_2 + h \sum_{\mathbf{n} \in \mathbb{Z}^2} \|\mathbf{a}_{\mathbf{n}_2} - \mathbf{a}_{\mathbf{n}}\|_2 \quad + h \sum_{\mathbf{n} \in \mathbb{Z}^2} \|\mathbf{a}_{\mathbf{n}_3} - \mathbf{a}_{\mathbf{n}}\|_2 . \tag{3.357}$$

Each of the three terms of (3.357) can be computed via a filtering operation. Here, we just prove this for the last term in the summation and we deduce the other two using similar computations.

Using the notations $\mathbf{k}_1 = \mathbf{n} + \mathbf{r}_1$, $\mathbf{k}_2 = \mathbf{n} + \mathbf{r}_2$, and $\mathbf{k}_3 = \mathbf{n} + \mathbf{r}_1 + \mathbf{r}_2$, we write that

$$\begin{cases} \mathbf{a}_{\mathbf{n}_3}^T \mathbf{R}_h(\mathbf{k}_3 - \mathbf{k}_2) = c[\mathbf{k}_3] - c[\mathbf{k}_2] \\ \mathbf{a}_{\mathbf{n}_3}^T \mathbf{R}_h(\mathbf{k}_3 - \mathbf{k}_1) = c[\mathbf{k}_3] - c[\mathbf{k}_1] \\ \mathbf{a}_{\mathbf{n}}^T \mathbf{R}_h(\mathbf{k}_1 - \mathbf{n}) = c[\mathbf{k}_1] - c[\mathbf{n}] \\ \mathbf{a}_{\mathbf{n}}^T \mathbf{R}_h(\mathbf{k}_2 - \mathbf{n}) = c[\mathbf{k}_2] - c[\mathbf{n}] \end{cases} \Leftrightarrow \begin{cases} \mathbf{R}_h^T \mathbf{a}_{\mathbf{n}_3} = \begin{bmatrix} c[\mathbf{k}_3] - c[\mathbf{k}_2] \\ c[\mathbf{k}_3] - c[\mathbf{k}_1] \end{bmatrix} \\ \mathbf{R}_h^T \mathbf{a}_{\mathbf{n}} = \begin{bmatrix} c[\mathbf{k}_1] - c[\mathbf{n}] \\ c[\mathbf{k}_2] - c[\mathbf{n}]] \end{bmatrix}, \end{cases} \tag{3.358}$$

where $\mathbf{R}_h = h \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 \end{bmatrix}$ is the lattice matrix. Combining these equations, we obtain that

$$\mathbf{R}_h^T(\mathbf{a}_{\mathbf{n}_3} - \mathbf{a}_{\mathbf{n}}) = (c[\mathbf{n}] - c[\mathbf{k}_1] - c[\mathbf{k}_2] + c[\mathbf{k}_3])\mathbf{1}, \tag{3.359}$$

where $\mathbf{1} = (1, 1)$. The application of $\left(\mathbf{R}_h^{-1}\right)^T$ to both sides of (3.359) leads to

$$(\mathbf{a}_{\mathbf{n}_3} - \mathbf{a}_n) = (1, -1, -1, 1)^T \mathbf{z} \left(\mathbf{R}_h^{-1}\right)^T \mathbf{1}, \tag{3.360}$$

where $\mathbf{z} = (c[\mathbf{n}], c[\mathbf{k}_1], c[\mathbf{k}_2], c[\mathbf{k}_3])$. Using the homogeneity of the ℓ_2-norm, we verify that

$$\|\mathbf{a}_{\mathbf{n}_3} - \mathbf{a}_{\mathbf{n}}\|_2 = \left|(1, -1, -1, 1)^T \mathbf{z}\right| \left\| \left(\mathbf{R}_h^{-1}\right)^T \mathbf{1} \right\|_2 = \frac{2\sqrt{3}}{3h} \left|(1, -1, -1, 1)^T \mathbf{z}\right|. \tag{3.361}$$

By plugging in $\mathbf{k}_1 = \mathbf{n} + (1, 0)$, $\mathbf{k}_2 = \mathbf{n} + (0, 1)$ and $\mathbf{k}_3 = \mathbf{n} + (1, 1)$, we express the last term in (3.357) as

$$\begin{aligned} h \sum_{\mathbf{n} \in \mathbb{Z}^2} \|\mathbf{a}_{\mathbf{n}_3} - \mathbf{a}_{\mathbf{n}}\|_2 &= \frac{2\sqrt{3}}{3} \sum_{\mathbf{n} \in \mathbb{Z}^2} \Big| c[\mathbf{n}] - c[\mathbf{n} + (1, 0)] - \quad c[\mathbf{n} + (0, 1)] + c[\mathbf{n} + (1, 1)] \Big| \\ &= \|d_3 * c\|_{1,1}. \end{aligned} \tag{3.362}$$

□

Theorem 3.19 provides a simple algorithm to evaluate HTV(f) in terms of three convolutions. We also remark that, for any admissible CPWL model f, the output of the digital filters $d_n, n = 1, 2, 3$, are zero outside of a compact domain. This in effect allows us to consider an equivalent finite lattice to represent f in practice.

We now derive an exact finite-dimensional discretization of Problem (3.351). We consider a finite lattice of square size $(N \times N)$ in such a way that all training data are contained in it. The lattice coefficients are grouped into the vector $\mathbf{c} \in \mathbb{R}^{N^2}$ which is a (row-wise) vectorization of the 2D array $c[\mathbf{k}], \mathbf{k} \in \Omega$, where $\Omega \subset \mathbb{Z}^2$ is the set of lattice indices. Consequently, we define the regularization matrix $\mathbf{L} \in \mathbb{R}^{3N^2 \times N^2}$ as

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \mathbf{L}_3 \end{bmatrix}, \tag{3.363}$$

where $\mathbf{L}_n$ is a Toeplitz-like matrix associated to the 2D digital filter d_n such that, for $n = 1, 2, 3$, $\mathbf{L}_n\mathbf{c}$ is the vectorized version of $(d_n * c)[\mathbf{k}], \mathbf{k} \in \Omega$. Further, we define the forward matrix $\mathbf{H} \in \mathbb{R}^{M \times N^2}$ such that its mth row corresponds to the datapoint $\mathbf{x}_m$, with $f(\mathbf{x}_m) = [\mathbf{Hc}]_m$ for $m = 1, \dots, M$. Using these, we restate (3.351) as the finite-dimensional minimization

$$\arg\min_{\mathbf{c} \in \mathbb{R}^{N^2}} \sum_{m=1}^{M} E\left([\mathbf{Hc}]_m, y_m\right) + \lambda \left\|\mathbf{Lc}\right\|_1. \tag{3.364}$$

Ultimately, the finite-dimensional problem (3.364) has the composite structure of the generalized LASSO [275] which can be solved efficiently using known convex optimization solvers (*e.g.* ADMM or its variants [200, 276, 201]). We denote the corresponding solution by $\mathbf{c}_0$.

The discrete formulation (3.364) also highlights the sparsity-promoting effect of the HTV regularization, due to the presence of the ℓ_1 penalty in (3.364). Consequently, we expect to learn models with few linear regions. In order to find a sparser solution, we use the simplex algorithm [277, 278] to solve the minimization

$$\arg\min_{\mathbf{c} \in \mathbb{R}^{N^2}} \left\|\mathbf{Lc}\right\|_1, \ \ s.t. \ \mathbf{Hc} = \mathbf{Hc}_0. \tag{3.365}$$

This post-processing step is known to provide an extreme point of the solution set of (3.364) [74] which, in our case, often leads to a sparser CPWL mapping.

Numerical Illustration

We now demonstrate the advantages of our pipeline by comparing it to other existing learning methods. The Python code of all the experimental results is available on Github[11].

Minimum-Norm Interpolation

We demonstrate the sparsity-promoting effect of the HTV regularizer in a controlled environment in which we sample $M = 12$ points from a pyramid function f_{pyr} whose vertices are positioned on the lattice. This ensures that the target function can be represented exactly in our search space. To isolate the effect of the regularization, we use simplex solver to find

$$\arg\min_{f \in \mathcal{X}_{\mathrm{R}_h}(\mathbb{R}^2)} \mathrm{HTV}(f), s.t.\, f(\mathbf{x}_m) = f_{\mathrm{pyr}}(\mathbf{x}_m), m = 1, \dots, M \tag{3.366}$$

by recasting it as a discrete minimization problem of the form (3.365).

In Figure 3.20, we show the results of successive experiments where we use a lattice of size (20 × 20) (a total number of 421 parameters) with zero boundary conditions. We chose a colormap based on the triangle normals so that co-planarities can be identified. Due to randomness in the implementation of the algorithm, we obtained different solutions of (3.365). They all resulted in the same minimal $\mathrm{HTV}(f)$. We observe that the algorithm leads to sparse solutions in all cases, with few faces. Indeed, from a search space which can model functions with hundreds of faces (Figure 3.20a), we reached solutions with just 12 (3.20b), 7 (3.20c), and 6 (3.20b) faces, respectively.

Data-Fitting

In this experiment, we tackle a data-fitting problem and compare three approaches.

1. **Ours**, using HTV regularization and a CPWL search space (3.321).

[11] `https://github.com/joaquimcampos/HTV-Learn`

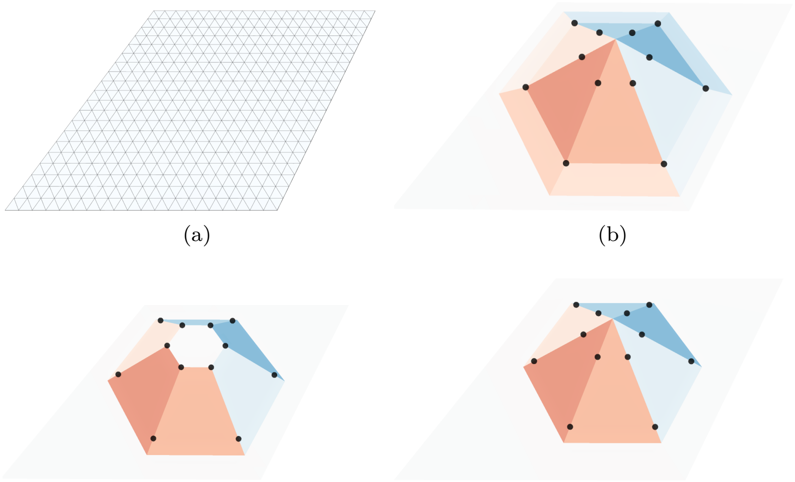


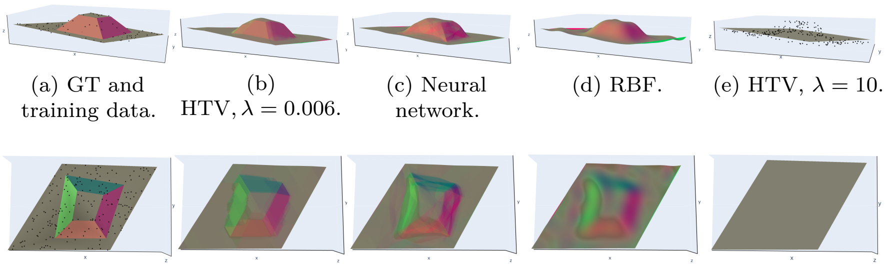
Figure 3.20: Solutions of the minimum-HTV interpolation problem.

2. **ReLU neural networks**, which also construct CPWL models.
3. **Radial basis functions** with Gaussian kernels—a classical approach in supervised learning [171, 174, 105].

The dataset consists of samples from a CPWL function with noisy labels. More precisely, the labels are of the form

$$y_m = h(\mathbf{x}_m) + \epsilon, \tag{3.367}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and h is the CPWL function shown in Figures 3.21a and 3.21f. We use 200 datapoints and set $\sigma = \frac{1}{20} \|f\|_{L_\infty}$. Note that the model cannot be represented exactly in our search space since the data points do not fall on the lattice; however, the error can be mitigated by a sufficient reduction of the stepsize of the grid.



(a) GT and training data. (b) HTV, $\lambda = 0.006$. (c) Neural network. (d) RBF. (e) HTV, $\lambda = 10$.

(f) GT and training data. (g) HTV, $\lambda = 0.006$. (h) Neural network. (i) RBF. (j) HTV, $\lambda = 10$.

Figure 3.21: A comparison between HTV regularization, ReLU neural network, and radial basis functions in the 2D data-fitting example

The setup is as follows: for the data-fidelity term in (3.364), we use the quadratic loss $E(y, z) = (y - z)^2$. For the ReLU network, we use a fully connected architecture with 4 hidden layers, each with 256 hidden neurons. The total number of parameters of the neural network is 198401. We train the neural network for 500 epochs using an Adam optimizer [279] with a batch size of 10 and weight decay. The initial learning rate is set to 10^{-3} and is decreased by 10 at epochs 375 and 425. For the HTV, we use a lattice size of size (64×64), giving a total of 4225 parameters. In all methods, we tune the corresponding hyperparameter on a validation set (regularization weight λ for the HTV and radial-basis function (RBF), kernel size γ for the RBF, and weight-decay parameter μ for the neural network) to have a fair comparison. To assess sparsity, we sample the learned neural network and RBF models in the position of the lattice vertices and vectorize these values (we denote the resulting vector by $\mathbf{c}$), as done for our method. Finally, for all methods, we compute the percentage of non-negligible "changes of slope" as $\frac{\|\mathbf{Lc}_{>\epsilon}\|_0}{3N^2} \cdot 100$, where $\epsilon = 10^{-4}$ and $3N^2$ is the number of rows of $\mathbf{L}$.

The results are shown in Table 3.3 and Figure 3.21, along with the groundthruth (GT). We observe that the HTV model performs significantly better than the radial-basis functions and on par with the neural network. Furthermore, as seen in the last column of the table and from the Figures, the HTV leads to a much

Model	Hyperparameters	Test MSE	Sparsity
HTV	$\lambda = 2 \times 10^{-3}$	$\mathbf{3.6 \times 10^{-5}}$	**27%**
HTV	$\lambda = 4 \times 10^{-3}$	3.8×10^{-5}	19%
HTV	$\lambda = 6 \times 10^{-3}$	4.4×10^{-5}	16%
HTV	$\lambda = 10$	3.1×10^{-3}	00%
ReLU	$\mu = 1 \times 10^{-6}$	$\mathbf{3.7 \times 10^{-5}}$	**63%**
RBF	$\lambda = 0.08, \gamma = 7$	5.5×10^{-5}	88%

Table 3.3: Test MSE and sparsity of each method in the data-fitting example.

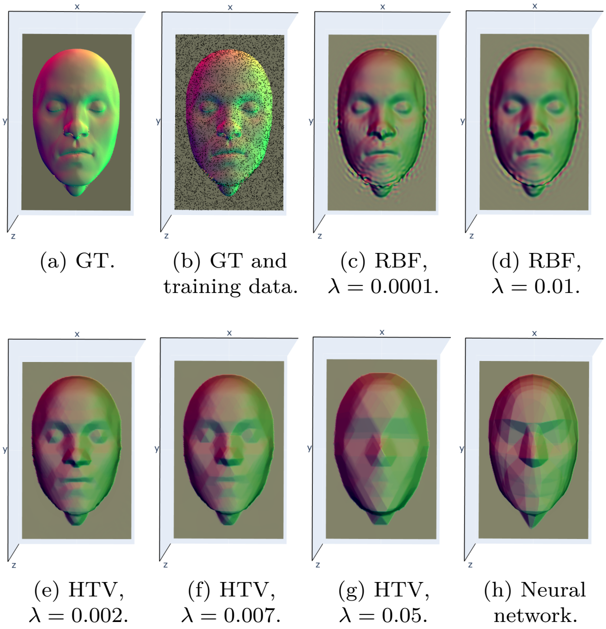
sparser result. Moreover, we observe that the level of sparsity for the HTV can be controlled with the regularization weight (higher leads to sparser results). In the extreme case $\lambda \to +\infty$, the model should converge to the least-squares linear approximation of the training data. The very high regularization weight $\lambda = 10$ allows us to verify this in practice. Indeed, the resulting model is linear and the data-fitting error is precisely the same as the one obtained with a least-squares fit.

Real Dataset

We now benchmark the three methods of the previous experiment on a (non-CPWL) facial dataset. This dataset is a 2D height map $f : \mathbb{R}^2 \to \mathbb{R}$ that we construct by cutting a 3D face model[12] (Figure 3.22a). We then sample 8000 data points for training (Figure 3.22b).

Relative to the previous experiment, the setup has the following differences: for the HTV, we use a lattice of size (194 × 194) (38025 parameters) and skip the simplex post-processing step; for the neural network, we incorporate one additional hidden layer (264193 parameters), increase the number of epochs to 2000 and the batch size to 100, and, lastly, decrease the initial learning rate at epochs 1750 and 1900.

[12] `https://www.turbosquid.com/3d-models/3d-male-head-model-1357522`



(a) GT.

(b) GT and training data.

(c) RBF, $\lambda = 0.0001$.

(d) RBF, $\lambda = 0.01$.

(e) HTV, $\lambda = 0.002$.

(f) HTV, $\lambda = 0.007$.

(g) HTV, $\lambda = 0.05$.

(h) Neural network.

Figure 3.22: A comparison between HTV regularization, ReLU neural network, and radial basis functions in the face dataset

The results are shown in Table 3.4 and Figure 3.22. The HTV achieved the lowest test mean-squared error (MSE) on par with the RBF which is expected to perform well due to the high density of datapoints and the absence of noise. Regarding the effect of the regularization, we again observe that, for the HTV, increasing it results in a model with fewer faces. In the case of the RBF, the solutions present ringing artifacts, especially in a low-regularization regime. Finally, we remark that the neural network constructs a coarse approximation of the data.

Finally, we introduce some gaps in the training data in order to make the fitting problem more challenging. The results are depicted in Figure 3.23. This example highlights that the HTV favors simple and intuitive models that are visually more

Model	Hyperparameters	Test MSE	Sparsity
HTV	$\lambda = 2 \times 10^{-3}$	$\mathbf{3.0 \times 10^{-6}}$	**10%**
HTV	$\lambda = 7 \times 10^{-3}$	4.8×10^{-6}	8%
HTV	$\lambda = 5 \times 10^{-2}$	1.9×10^{-5}	6%
ReLU	$\mu = 1 \times 10^{-6}$	$\mathbf{5.1 \times 10^{-6}}$	**12%**
RBF	$\lambda = 10^{-4}, \gamma = 50$	3.2×10^{-6}	31%
RBF	$\lambda = 10^{-2}, \gamma = 50$	3.4×10^{-6}	24%

Table 3.4: Test MSE and sparsity of each method in the face dataset.

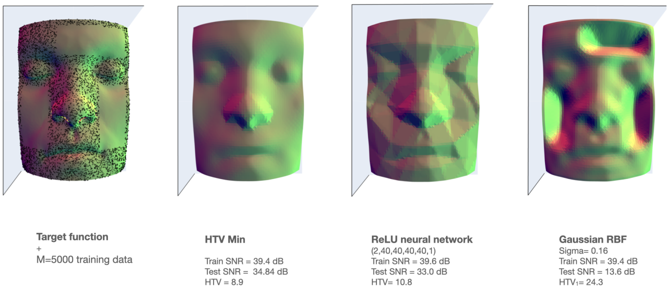


Figure 3.23: Learning of a 2D height map of a face from its nonuniform samples.

adequate.

3.5.4 Summary

In this section, we introduced a novel seminorm for learning continuous and piecewise linear multivariate functions. To that end, we first studied the duality mapping in finite-dimensional Schatten spaces. Based on a careful investigation of the cases where the Hölder inequality saturates, we provided an explicit form for this mapping when $p \in (1, +\infty)$. Furthermore, by adding a rank constraint, we proved that the mapping becomes single-valued for the special case $p = 1$. As for $p = +\infty$, we showed that the mapping yields a convex set whose elements are explicitly characterized.

Next, we rigorously defined the HTV seminorm and showed that it satisfies the desirable properties of a complexity measure for the study of learning schemes. Moreover, we computed the HTV of two general classes of functions. In each case, we derived simple formulas for the HTV that allowed us to interpret its underlying behavior.

Finally, we use the HTV seminorm as a regularization functional and proposed a method to learn two-dimensional sparse CPWL mappings. We formulated the problem in a search space consisting of shifts of CPWL box-splines in a lattice. By doing so, we were able to evaluate any model in the search space, as well as compute its HTV, from the values at the lattice points (model parameters). In particular, we showed that the latter can be computed with a three-filter convolutional structure; this allows us to discretize the problem exactly and to recast it in the form of the generalized LASSO. Finally, we demonstrated the sparsity-promoting effect of our framework via numerical examples where we compared its performance with ReLU neural networks and radial-basis functions.

Chapter 4

Linear Inverse Problems with Multicomponent Models

In this chapter[1], we demonstrate the applicability of our general representer theorem (see Chapter 2) in the context of solving linear inverse problems that admit a multicomponent prior on the signal of interest. To illustrate the concept, we focus on the problem of recovering continuous-domain 1D signals that can be written as the sum of two components. After providing a short overview on linear inverse problems (Section 4.1), we consider two different scenarios and treat each one separately. In the first scenario, both components are assumed to be sparse, albeit in different transform domains (Section 4.2). In the second scenario, we consider a composite "sparse-plus-smooth" model to address signals with components of different natures (Section 4.3). In both cases, we propose adequate variational formulations with corresponding representer theorems and optimal methods to discretize the problems exactly. Finally, we present a novel scheme for fitting curves to 2D point-clouds as an application of our proposed framework (Section 4.4). Our formulation adopts the hybrid setting of Section 4.2 in the periodic case and can be used in practice to

[1] This chapter is based on our published [3, 280, 281] works.

generate stylized fonts.

4.1 Overview on Linear Inverse Problem

In the traditional discrete formalism of linear inverse problems, the goal is to recover a signal $\mathbf{c}_0 \in \mathbb{R}^N$ based on some measurement vector $\mathbf{y} \in \mathbb{R}^M$. These measurements are typically acquired via a linear operator $\mathbf{H} \in \mathbb{R}^{M\times N}$ that models the physics of the acquisition system (forward model), so that $\mathbf{Hc}_0 \approx \mathbf{y}$. The recovery is often achieved by solving an optimization problem that aims at minimizing the discrepancy between the measurements $\mathbf{Hc}$ of the reconstructed signal $\mathbf{c}$ and the acquired data $\mathbf{y}$. This data fidelity is measured with a suitable convex loss functional $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$, the prototypical example being the quadratic error $E(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$. A regularization term is often added to the cost functional, which yields the optimization problem

$$\arg\min_{\mathbf{c}\in\mathbb{R}^N} \left\{ \underbrace{E(\mathbf{Hc}, \mathbf{y})}_{\text{Data fidelity}} + \underbrace{\lambda\mathcal{R}(\mathbf{Lc})}_{\text{Regularization}} \right\}, \tag{4.1}$$

where $\mathcal{R}$ is the regularization functional, $\mathbf{L}$ specifies a suitable transform domain, and $\lambda > 0$ is a tuning parameter that determines the strength of the regularization. The use of regularization can have multiple motivations:

1. to handle the ill-posedness of the inverse problem, which occurs when different signals yield identical measurements;
2. to favor certain types of reconstructed signal (*e.g.* sparse or smooth) based on our prior knowledge;
3. to improve the conditioning of the inverse problem and thus increase its numerical stability and robustness to noise.

Historically, the first instance of regularization dates back to Tikhonov [282] with a quadratic regularization functional $\mathcal{R} = \|\cdot\|_2^2$. Tikhonov regularization constrains the energy of $\mathbf{Lc}$ which, when $\mathbf{L}$ is a finite-difference matrix, leads to a smooth signal $\mathbf{c}$. Tikhonov regularization has the practical advantage of being mathematically tractable which leads to a closed-form solution. More recently, there has been growing interest in ℓ_1 regularization $\mathcal{R} = \|\cdot\|_1$, which has peaked in popularity for compressed sensing [62, 163, 66, 67]. With ℓ_1 regularization, the prior assumption is that the transform signal $\mathbf{Lc}_0$ is sparse, meaning that it has few nonzero coefficients. Indeed, the ℓ_1 norm can be seen as a convex relaxation of the ℓ_0 "norm", which counts the number of nonzero entries of a vector. The sparsity-promoting effect of ℓ_1 regularization is well understood and documented [214, 163, 163]. It is now generally considered to be superior to Tikhonov regularization for most applications [113]. Moreover, despite its non differentiability, numerous efficient proximal algorithms based on the proximity operator of the ℓ_1 norm have emerged to solve ℓ_1-regularized problems [176, 283, 200].

4.1.1 Continuous-Domain Problems

Until now, we have focused on the discrete setting, as it constitutes the vast majority of the inverse-problem literature for computational feasibility reasons. However, most real-world signals are inherently continuous. Therefore, when feasible, it is natural and desirable to formulate the inverse problem in the continuous domain. Similarly to the discrete setting, we are given measurements $\mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n} \in \mathbb{R}^M$, where $\boldsymbol{\nu} : s \to \boldsymbol{\nu}(s) \in \mathbb{R}^M$ is a continuous-domain linear measurement operator and $\mathbf{n} \in \mathbb{R}^M$ is some additive noise. The inverse problem can then be formulated through the minimization

$$\arg\min_{f \in \mathcal{F}} E(\boldsymbol{\nu}(f), \mathbf{y}) + \lambda \mathcal{R}\left(\mathrm{L}\{f\}\right). \tag{4.2}$$

In this case, the regularization operator L acts on signals that are defined over the continuum, the prototypical example being derivative operators of different orders, $\mathrm{L} = \mathrm{D}^n$, $n \in \mathbb{N} \setminus \{0\}$. The classical choice for the functional $\mathcal{R}$ is the squared L_2 norm, $\mathcal{R} = \|\cdot\|_{L_2}^2$, which corresponds to the *generalized Tikhonov* (gTikhonov) regularization and is known to promote smoothness in combination with the regularization operator L. More recently, the use of the total-variation norm, $\mathcal{R} = \|\cdot\|_{\mathcal{M}}$, has been also proposed [69, 73]. It can be viewed as the continuous counterpart of the discrete ℓ_1 norm and the regularization term is called generalized TV (gTV) due to the presence of the operator L. We refer to Section 3.2 for an application of gTV regularization in supervised learning.

4.1.2 Representer Theorem

A classical way of discretizing a continuous-domain problem is to reformulate it as a finite-dimensional one by relying on a *representer theorem* that gives a parametric form of the solution. Prominent examples include representer theorems for problems formulated over reproducing-kernel Hilbert spaces (RKHS) or semi-RKHS, which are foundational to the field of machine learning [8, 61]. As demonstrated in [74, Theorem 3], the minimization Problem (4.2) with gTikhonov regularization (*i.e.*, $\mathcal{R} = \|\cdot\|_{L_2}^2$) falls into this category. The representer theorem states that there is a unique solution of the form

$$s^*(x) = p(x) + \sum_{m=1}^{M} a_m h_m(x), \tag{4.3}$$

where the additional component p lies in the null space of L (*i.e.* $\mathrm{L}\{p\} = 0$), h_m is a (typically quite smooth) kernel function that is fully determined by the choice of ν_m and L, and $a_m \in \mathbb{R}$ are expansion coefficients. Therefore, to solve the continuous-domain problem, one needs only to optimize over the a_m coefficients and the null-space component p which lives in a finite-dimensional space. This leads to a standard finite-dimensional problem with Tikhonov regularization.

Concerning the minimization problems with gTV regularization (*i.e.*, $\mathcal{R} = \|\cdot\|_{\mathcal{M}}$), several representer theorems yield a parametric form of a sparse solution in different settings [284, 73, 72, 76]. The more specific case of our setting is tackled by [74, Theorem 4], which states that there is an L-spline solution of the form

$$s^*(x) = p(x) + \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(x - x_k), \tag{4.4}$$

where $a_k, x_k \in \mathbb{R}$, ρ_{L} is a Green's function of L (*i.e.* $\mathrm{L}\{\rho_{\mathrm{L}}\} = \delta$, where δ is the Dirac impulse), K is the number of atoms of s which is bounded by $K \leq (M - N_0)$, N_0

being the dimension of the null space of L, and p lies in the null space of L. For example, when $\mathrm{L} = \mathrm{D}^{N_0}$, the signal s is a piecewise polynomial of degree $(N_0 - 1)$ with smooth junctions at the knots x_k. These representer theorems have paved the way for various exact discretization methods. In the gTikhonov case, one can optimize over the a_m coefficients in (4.3) directly [74]. For the gTV case (4.4), grid-based techniques using a well-conditioned B-spline basis [166] as well as grid-free techniques [71] have been proposed.

4.2 Hybrid Models

In this section[2], we study one-dimensional continuous-domain inverse problems with multiple generalized total-variation regularization, which involves the joint use of several regularization operators. We first show that such inverse problems have *hybrid-spline* solutions with a total sparsity bounded by the number of measurements. We then show that such continuous-domain problems can be discretized in an exact way by using a union of B-spline dictionary bases matched to the regularization operators.

4.2.1 Context

We are interested in multicomponent signals $\mathbf{c} = \sum_{i=1}^{Q} \mathbf{c}_i$ such that each of the Q components $\mathbf{c}_i$ is sparse in a transform domain that is specified by the regularization matrix $\mathbf{L}_i$ $(i \in \{1, \dots, Q\})$. For simplicity of exposition, we set $Q = 2$. In the discrete setting, a natural way of formulating the recovery problem is by solving the minimization

$$(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2) = \underset{\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^N}{\arg\min} \left(\|\mathbf{H}(\mathbf{c}_1 + \mathbf{c}_2) - \mathbf{y}\|_2^2 + \lambda \Big(\|\mathbf{L}_1 \mathbf{c}_1\|_1 + \|\mathbf{L}_2 \mathbf{c}_2\|_1 \Big) \right), \tag{4.5}$$

which yields the reconstructed signal $\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_2$.

Although this setting introduces practical and theoretical difficulties (due to the high redundancy of the overcomplete dictionary), it is extremely useful in many applications when a single dictionary is insufficient to represent the richness of a signal. In particular, the problem of accurately reconstructing both components $\mathbf{c}_1$ and $\mathbf{c}_2$ is known as data separation [66, Chapter 11], and has been studied extensively both theoretically and practically. In fact, some of the first theoretical works concerning sparse vector recovery using ℓ_1-norm minimization involved a concatenated dictionary consisting of a mixture of sinusoids and spikes [285, 286]. The goal was to provide a condition under which ℓ_0 and ℓ_1 minimization yield the same solution. This sparked an abundance of research, which extended and improved

[2] This section is based on our published work [3].

these results for more general (non-orthonormal) dictionaries [86, 87, 287, 89]. An overview is given in [288]. Later, these results were extended to images to separate point-like and curve-like structures [289]. These works mostly tackle denoising problems characterized by $\mathbf{H} = \mathbf{I}_N$ and $M = N$. In the field of compressed sensing, in which we have $M \ll N$, [90] considers redundant dictionaries in general. On the practical side, data separation is intimately related to morphological component analysis (MCA), a method popularized by Starck *et al.* [88, 290, 291, 292, 293] with applications in inpainting removal or the separation of texture and natural parts of an image.

Several practicioners have used the above formulation for data-separation problems, most notably Starck *et al.* in the context of MCA [88, 290, 291, 292]. More recently, this formulation was applied to the task of separating cartoon and texture parts of an image in [294]. A similar approach is used in low-rank plus sparse decomposition methods [295].

Despite these empirical works, virtually no theoretical study of Problem (4.5) has been carried out. In [91], Candès *et al.* have named it the “split-analysis" problem. A theoretical study was later done by Lin *et al.* in [92], where they show that the data separation problem (*i.e.* the recovery of both components of the original signal) can be solved via Problem (4.5). This result requires that $\mathbf{H}$ satisfies the restricted isometry property adapted to a dictionary (D-RIP) and that $\mathbf{L}_1$ and $\mathbf{L}_2$ satisfy a mutual coherence condition.

While the literature on the topic is scarce in the discrete setting, to the best of our knowledge, it is nonexistent in the continuous domain. Since most real-world signals are continuously defined, the reconstruction of continuous-domain solutions is a desirable objective. Moreover, although handling discrete signals is obviously appealing from a computational perspective, it introduces discretization errors in the measurements. For instance, the discrete Fourier transform (DFT) is often used as surrogate for the continuous Fourier transform to model MRI measurements, which is by no means an exact discretization.

Main Contributions

In this work, we propose to use unions of dictionaries in a continuous-domain framework. Our goal is to reconstruct a multicomponent continuous-domain 1D signal $s = s_1 + s_2$, where s_1 and s_2 have different characteristics. We focus on continuous-domain inverse problems of the form

$$s^* = \arg\min_{f} \left(E(\boldsymbol{\nu}(f), \mathbf{y}) + \lambda \left((1-\alpha)\|\mathrm{L}_1\{f_1\}\|_{\mathcal{M}} + \alpha\|\mathrm{L}_2\{f_2\}\|_{\mathcal{M}} \right) \right), \tag{4.6}$$

where $E(\cdot,\cdot)$ is a convex loss function, $\lambda > 0$ is the regularization parameter and $\alpha \in (0,1)$ controls the weighing of the two regularization terms. We recall from Chapter 3 that the regularization norm $\|\cdot\|_{\mathcal{M}}$ generalizes the L_1 norm [73] and is the continuous counterpart of the ℓ_1 norm used in discrete problems. By adopting our general representer, presented in Chapter 2, we prove that, for differential operators L_i, Problem (4.6) leads to spline solutions $s^* = s_1^* + s_2^*$, where each component s_i^* is an L_i-spline. The reconstructed signal s^* is therefore a sum of different splines, which we coin as a *hybrid spline*. Moreover, the total sparsity of s^* in this union of spline dictionaries is no larger than the number M of measurements. Finally, we demonstrate how Problem (4.6) can be discretized in an exact way using B-splines, based on the methodology of [166].

4.2.2 Theory

Our aim is to recover a continuous-domain signal $s : \mathbb{R} \to \mathbb{R}$ given M noisy measurements modeled as $\mathbf{y} = \boldsymbol{\nu}(s) + \mathbf{n}$, where $\mathbf{n} \in \mathbb{R}^M$ is some additive noise. The noiseless measurements $\boldsymbol{\nu}(s)$ are acquired through M linear measurement functionals $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)$, with $\boldsymbol{\nu}(s) = (\langle \nu_1, s\rangle, \dots, \langle \nu_M, s\rangle)$. Here, $\langle \nu_m, s\rangle$ stands for the duality product, which is given by $\int_{\mathbb{R}} \nu_m(x)s(x)\mathrm{d}x$ when ν_m and s are ordinary functions. The ν_m functionals constitute the (assumed) forward model.

Next, we sum up all the relevant information and notations that concern the regularization operators L_i ($i \in \{1,2\}$).

1. For the sake of clarity, we restrict ourselves to the class of differential operators

$\mathrm{L}_i = \mathrm{D}^{N_i}$ for some integers $N_2 > N_1 \geq 1$. It therefore has a Green's function denoted by ρ_{L_i} which verifies $\mathrm{L}_i\{\rho_{\mathrm{L}_i}\} = \delta$ (see, Section 1.1).

2. The native space of L_i is denoted by $\mathcal{M}_{\mathrm{L}_i}(\mathbb{R})$ and verifies $\|\mathrm{L}_i\{f_i\}\|_{\mathcal{M}} < \infty$ for any $f_i \in \mathcal{M}_{\mathrm{L}_i}(\mathbb{R})$.

3. The null space of L_i is denoted by $\mathcal{N}_{\mathrm{L}_i}$ and contains polynomials of degree less than N_i, $\mathcal{N}_{\mathrm{L}_i} = \mathrm{span}\{p_n = (\cdot)^{n-1}\}_{n=1}^{N_i}$.

4. Following our convention $N_2 > N_1$, we note that the intersection of the null spaces is $\mathcal{N}_{\mathrm{L}_1} \cap \mathcal{N}_{\mathrm{L}_2} = \mathcal{N}_{\mathrm{L}_1}$. Consequently, we introduce the biorthogonal system $(\boldsymbol{\phi}_1, \mathbf{p}_1) = (\phi_n, p_n)_{n=1}^{N_1}$ for $\mathcal{N}_{\mathrm{L}_1}$ such that $\mathcal{N}_{\mathrm{L}_1} = \mathrm{span}(\mathbf{p}_1)$ and $\phi_n(p_m) = \delta[m-n]$ (Kronecker delta). Furthermore, we extend it to a biorthogonal system $(\boldsymbol{\phi}_2, \mathbf{p}_2) = (\phi_n, p_n)_{n=1}^{N_2}$ for $\mathcal{N}_{\mathrm{L}_2}$.

5. The restricted search space for L_1 is defined as

$$\mathcal{M}_{\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{R}) = \{f \in \mathcal{M}_{\mathrm{L}_1}(\mathbb{R}) : \boldsymbol{\phi}_1(f) = \mathbf{0}\}. \tag{4.7}$$

   The space $\mathcal{M}_{\mathrm{L}_i,\boldsymbol{\phi}_i}(\mathbb{R})$ is isometrically isomorphic to $\mathcal{M}(\mathbb{R})$. Indeed, from [73, Theorem 4] we deduce that there exist stable right-inverse operators $\mathrm{L}_{\tilde{\boldsymbol{\phi}}_i}^{-1} : \mathcal{M}(\mathbb{R}) \to \mathcal{M}_{\mathrm{L}_i,\boldsymbol{\phi}_i}(\mathbb{R})$ that act as isometries between these spaces.

6. Any $f \in \mathcal{M}_{\mathrm{L}_i}(\mathbb{R})$ has a unique representation as

$$f(x) = \mathrm{L}_{\tilde{\boldsymbol{\phi}}_i}^{-1}\{w_i\}(x) + \mathbf{c}_i^T \mathbf{p}_i(x), \tag{4.8}$$

   where $w_i \in \mathcal{M}(\mathbb{R})$, and $\mathbf{c}_i \in \mathbb{R}^{N_i}$ with $i \in \{1, 2\}$. Consequently, $f \in \mathcal{M}_{\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{R})$, if and only if $\mathbf{c}_1 = \mathbf{0}$.

We now have the necessary tools to present the main theoretical result of this work, on which our implementation is based.

Theorem 4.1 (Continuous-domain representer theorem)**.** *Let* $\mathrm{L}_i = \mathrm{D}^{N_i}$ *for* $i = 1, 2$ *with the convention* $N_2 > N_1 \geq 1$*. Let* $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)$ *be a linear measurement operator composed of the* M *linear functionals* $\nu_m : f \mapsto \nu_m(f) \in \mathbb{R}$ *which are weak*-continuous on both* $\mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$ *and* $\mathcal{M}_{\mathrm{L}_2}(\mathbb{R})$*. Assume that* $\mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\mathrm{L}_2} = \{0\}$*, where* $\mathcal{N}_{\boldsymbol{\nu}}$ *is the null space of* $\boldsymbol{\nu}$ *(well-posedness assumption). Then, for any strictly* *convex loss function* $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_{\geq 0}$*,* $\lambda > 0$*, and* $\alpha \in (0,1)$*, the linear inverse problem*

$$\mathcal{V} = \underset{\substack{f=f_1+f_2\\ f_1\in\mathcal{M}_{\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{R})\\ f_2\in\mathcal{M}_{\mathrm{L}_2}(\mathbb{R})}}{\arg\min} \left(E(\boldsymbol{\nu}(f),\mathbf{y}) + \lambda\left((1-\alpha)\|\mathrm{L}_1\{f_1\}\|_{\mathcal{M}} + \alpha\|\mathrm{L}_2\{f_2\}\|_{\mathcal{M}}\right)\right) \tag{4.9}$$

has a solution $s = s_1 + s_2$*, where the* s_i *are nonuniform splines associated to the operator* L_i *for* $i = 1, 2$ *so that the reconstructed signal* s *admits the form*

$$s(x) = \sum_{k=1}^{K_1} a_{1,k}\rho_{\mathrm{L}_1}(x - x_{1,k}) + \sum_{k=1}^{K_2} a_{2,k}\rho_{\mathrm{L}_2}(x - x_{2,k}) + q(x) \tag{4.10}$$

for $q \in \mathcal{N}_{\mathrm{L}_2}$*,* $a_{i,k}, x_{i,k} \in \mathbb{R}$*. Moreover, the sparsity indices* K_i *verify* $K_1 + K_2 \leq M - N_2$*.*

Proof. Following (4.8), there is a bijection between $\mathcal{V}$ and

$$\tilde{\mathcal{V}} = \underset{\substack{\mathbf{w}\in\mathcal{M}(\mathbb{R})^2\\ q\in\mathcal{N}_{\mathrm{L}_2}}}{\arg\min} \left(E(\tilde{\boldsymbol{\nu}}(\mathbf{w},q),\mathbf{y}) + \lambda\|\mathbf{w}\|_{\mathrm{hyb}}\right), \tag{4.11}$$

where $\mathbf{w} = (w_1, w_2)$, $\tilde{\boldsymbol{\nu}}(\mathbf{w}, q) = \boldsymbol{\nu}\left(\mathrm{L}_{\tilde{\boldsymbol{\phi}}_1}^{-1}\{w_1\} + \mathrm{L}_{\tilde{\boldsymbol{\phi}}_2}^{-1}\{w_2\} + q\right)$, and $\|\mathbf{w}\|_{\mathrm{hyb}} = (1-\alpha)\|w_1\|_{\mathcal{M}} + \alpha\|w_2\|_{\mathcal{M}}$. We first show two intermediate results:

Claim 1: $\tilde{\nu}_m : \mathcal{M}(\mathbb{R})^2 \times \mathcal{N}_{\mathrm{L}_2} \to \mathbb{R}$ is weak*-continuous for $m = 1, \ldots, M$. To verify this, we take a sequence $(\mathbf{w}_k, q_k)$ that converges to $(\mathbf{w}_{\lim}, q_{\lim})$ in the weak*-topology. This implies that

$$f_{1,k} = \mathrm{L}_{\tilde{\boldsymbol{\phi}}_1}^{-1}\{w_{1,k}\} \to f_{1,\lim} = \mathrm{L}_{\tilde{\boldsymbol{\phi}}_1}^{-1}\{w_{1,\lim}\} \tag{4.12}$$

and

$$f_{2,k} = \mathrm{L}_{\tilde{\boldsymbol{\phi}}_2}^{-1}\{w_{2,k}\} + q_k \to f_{2,\lim} = \mathrm{L}_{\tilde{\boldsymbol{\phi}}_2}^{-1}\{w_{2,\lim}\} + q_{\lim}, \tag{4.13}$$

where both convergences are in the weak*-topology as well. Now by invoking the weak*-continuity of ν_m over both $\mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$ and $\mathcal{M}_{\mathrm{L}_2}(\mathbb{R})$, we deduce that $\tilde{\nu}_m(\mathbf{w}_k, q_k) = \nu_m(f_{1,k} + f_{2,k}) \to \nu_m(f_{1,\lim} + f_{2,\lim}) = \tilde{\nu}_m(\mathbf{w}_{\lim}, q_{\lim})$.

Claim 2: The cross-correlation matrix $\mathbf{V} = [\langle \nu_m, p_n \rangle] \in \mathbb{R}^{M \times N_2}$ has full rank. This is a direct consequence of the well-posedness assumption $\mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\mathrm{L}_2} = \{0\}$. Consequently, we deduce that the mapping $\boldsymbol{\nu} : \mathcal{N}_{\mathrm{L}_2} \to \mathbb{R}^M$ is coercive.

Following Claims 1 and 2, we are able to invoke Theorem 2.5, which guarantees the existence of a minimizer $(\mathbf{w}^*, q)$ of (4.11), where $\mathbf{w}^*$ is a vector-valued atomic measure (*i.e.*, its components can be expressed as sums of Diracs). The final step is to use the decomposition (4.8) to deduce the announced characterization. □

The following observations can be made concerning Theorem 4.1:

- We use the restricted search space $\mathcal{M}_{\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{R})$ defined in (4.7) instead of the complete space $\mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$ in order to ensure that Problem (4.9) is well-posed [3]. This does not restrict the native space of the reconstructed signal s since $\mathcal{M}_{\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{R}) + \mathcal{M}_{\mathrm{L}_2}(\mathbb{R}) = \mathcal{M}_{\mathrm{L}_1}(\mathbb{R}) + \mathcal{M}_{\mathrm{L}_2}(\mathbb{R})$.
- Theorem 4.1 can readily be extended to Q operators $\mathrm{L}_1, \dots, \mathrm{L}_Q$. However, for $Q > 2$, the handling of the pairwise null space intersections would make the general formulation more tedious. For the sake of clarity, we therefore only consider the case $Q = 2$.
- A remarkable feature of Theorem 4.1 is that the bound on the sparsity of the solutions does not increase with the number Q of operators. This is particularly appealing from a theoretical point of view since, compared to the single-operator framework of [73], we essentially enrich our dictionary at no cost in terms of sparsity.

4.2.3 Exact Discretization

We now briefly explain our discretization method for Problem (4.9), which is based on the methodology of [166]. For more details, we refer to our supporting publication [3].

[3] An unbounded solution set would arise if we allowed ourselves to add contributions $(p, -p)$ to a solution, with arbitrary $p \in \mathcal{N}_{\mathrm{L}_1}$.

We first introduce the discretized search space $\mathcal{M}_{\mathrm{L},h}(\mathbb{R})$ associated to L as

$$\mathcal{M}_{\mathrm{L},h}(\mathbb{R}) \triangleq \left\{ p + \sum_{k\in\mathbb{Z}} a[k]\rho_{\mathrm{L}}(\cdot - kh) : a \in \ell_1(\mathbb{Z}), p \in \mathcal{N}_{\mathrm{L}} \right\} \subseteq \mathcal{M}_{\mathrm{L}}(\mathbb{R}), \tag{4.14}$$

where $h > 0$ is the grid size. Given the form of the solutions (4.10) of our continuous-domain inverse Problem (4.9), this choice of search space is a natural one. By picking h sufficiently small, the aforementioned search space contains functions that are arbitrarily close to (4.10). Moreover, we recall from Section 1.1 that the discretized search space can alternatively be represented using B-splines. Specifically, we have that

$$\mathcal{M}_{\mathrm{L},h}(\mathbb{R}) = \left\{ \sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{L},h}(\cdot - kh) : c \in \ell_{1,\mathrm{L}}(\mathbb{Z}) \right\}, \tag{4.15}$$

where $\beta_{\mathrm{L},h} = \beta_{\mathrm{L}}(\frac{\cdot}{h})$ is the scaled B-spline and

$$\ell_{1,\mathrm{L}}(\mathbb{Z}) = \left\{ (c[k])_{k\in\mathbb{Z}} : (d_{\mathrm{L}} * c) \in \ell_1(\mathbb{Z}) \right\}. \tag{4.16}$$

The use of the B-spline representation of $\mathcal{M}_{\mathrm{L},h}(\mathbb{R})$ leads to well-conditioned problems and, thus, to computationally effective algorithms. This is due to the advantageous properties of B-splines, namely, their finite support and the fact that they produce a Riesz basis [296, Theorem 1].

For the first component, we enforce the boundary conditions $\boldsymbol{\phi}_1(s_1) = 0$. To that end, we define the matching search space for B-spline coefficients as

$$\ell_{1,\mathrm{L}_1,\boldsymbol{\phi}_1}(\mathbb{Z}) = \left\{ c \in \ell_{1,\mathrm{L}_1}(\mathbb{Z}) : \boldsymbol{\phi}_{1,h}(c) = \mathbf{0} \right\}, \tag{4.17}$$

where $\boldsymbol{\phi}_{1,h} : \ell_{1,\mathrm{L}_1}(\mathbb{Z}) \to \mathbb{R}^{N_1}$ with

$$\boldsymbol{\phi}_{1,h}(c) = \boldsymbol{\phi}_1 \left(\sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{L},h}(\cdot - kh) \right). \tag{4.18}$$

Since $\beta_{\mathrm{D}^{N_{0,1}}}$ is supported in $[0, N_{0,1}]$, the boundary conditions $\boldsymbol{\phi}_{1,h}$ impose linear constraints on only a few coefficients of c.

We now show that for any element in the discretized search space, the cost function in (4.9) can be expressed as a functional of the B-spline coefficients. To see this, we first recall that for any $f \in \mathcal{M}_{\mathrm{L},h}(\mathbb{R})$, we have that

$$\|\mathrm{L}\{f\}\|_{\mathcal{M}} = \left\|\mathrm{L}\left\{p + \sum_{k\in\mathbb{Z}} a[k]\rho_{\mathrm{L}}(\cdot - kh)\right\}\right\|_{\mathcal{M}} = \left\|\sum_{k\in\mathbb{Z}} a[k]\delta(\cdot - kh)\right\|_{\mathcal{M}} = \|a\|_{\ell_1} = \|c{*}d_{\mathrm{L}}\|_{\ell_1}, \tag{4.19}$$

where the last equality follows from (1.13). For the measurement functional $\boldsymbol{\nu}$, we need to compute the cross-product terms $\mathbf{v}_{\mathrm{L}}[k] = (\langle \nu_m, \beta_{\mathrm{L},h}(\cdot - kh)\rangle)_{1\leq m\leq M} \in \mathbb{R}^M$. Using this quantity, we readily observe that for any $f \in \mathcal{M}_{\mathrm{L},h}(\mathbb{R})$, we have that

$$\boldsymbol{\nu}(f) = \sum_{k\in\mathbb{Z}} c[k]\mathbf{v}_{\mathrm{L}}[k] = \langle c, \mathbf{v}_{\mathrm{L}}\rangle. \tag{4.20}$$

All put together, we define the discretized problem as

$$\mathcal{V}_h = \underset{c_i\in\ell_{1,\mathrm{L}_i}(\mathbb{Z}), i=1,2}{\arg\min} \left(E\left(\langle c_1, \mathbf{v}_{\mathrm{L}_1}\rangle + \langle c_2, \mathbf{v}_{\mathrm{L}_2}\rangle, \mathbf{y}\right) + \lambda\left(\alpha\|c_1 * d_{\mathrm{L}_1}\|_{\ell_1} + (1-\alpha)\|c_2 * d_{\mathrm{L}_2}\|_{\ell_1}\right)\right),$$
$$\text{s.t.} \quad \boldsymbol{\phi}_{1,h}(c_1) = \mathbf{0}. \tag{4.21}$$

In most cases, the measurement functional is compactly supported and, hence, so is $\mathbf{v}_{\mathrm{L_i}}$. Moreover, by choosing compactly supported boundary conditions $\boldsymbol{\phi}_1$ (*e.g.*, sampling of the function and its higher-order derivatives), the condition $\boldsymbol{\phi}_{1,h}(c_1) = \mathbf{0}$ will also depend on a finite number of coefficients $c_1[k]$. Finally, due to the compact support of the filters $d_{\mathrm{L_i}}$, one expects to be able to extrapolate a compactly supported sequence in a way that the regularization penalty remains unchanged (see [166]). Hence, Problem (4.21) can be recast as a finite-dimensional minimization that is the reminiscent of the generalized LASSO and can be solved efficiently using iterative methods such as ADMM.

4.2.4 Numerical Illustration

In this part, we illustrate our framework with numerical examples. Our algorithms are implemented using GlobalBioIm [177], a Matlab library developed in our group.

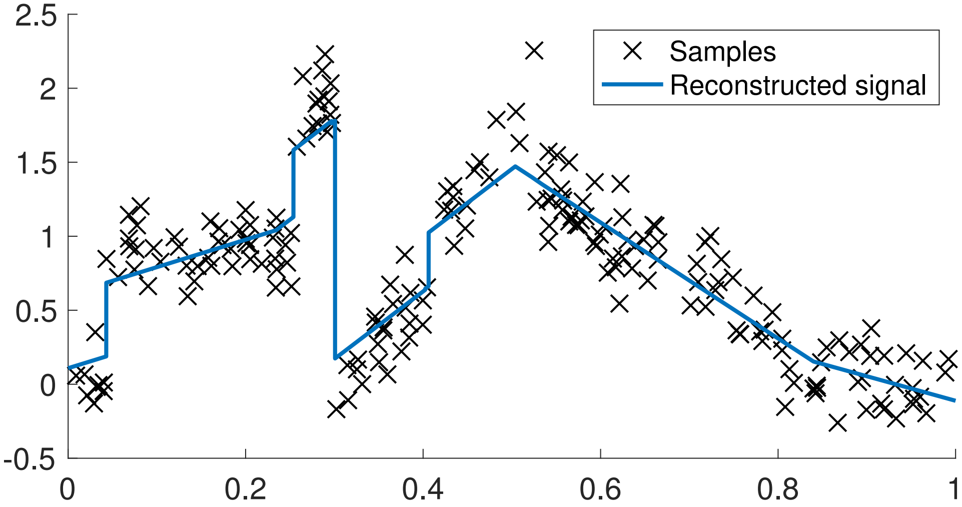


Figure 4.1: Curve fitting for $\mathrm{L}_1 = \mathrm{D}$, $\mathrm{L}_2 = \mathrm{D}^2$, $M = 200$, $\lambda = 1.3$, $\alpha = 0.05$.

Curve Fitting

Curve fitting is particularly well-suited for smoothing problems, which consist in fitting a continuous-domain function which is sparse in a certain dictionary basis from many noisy data points. This is usually done in a single-operator framework, where the dictionary is associated with a single brand of splines. In contrast with standard single-operator frameworks, our approach allows for the joint use of several families of basis functions, and can therefore represent a richer class of signals. An example of a curve-fitting reconstruction is given in Figure 4.1. The chosen regularization operators are $\mathrm{L}_1 = \mathrm{D}$ and $\mathrm{L}_2 = \mathrm{D}^2$; our dictionary thus consists of both piecewise-constant and piecewise-linear splines.

In this experiment, the data are $M = 200$ noisy samples of a hybrid spline

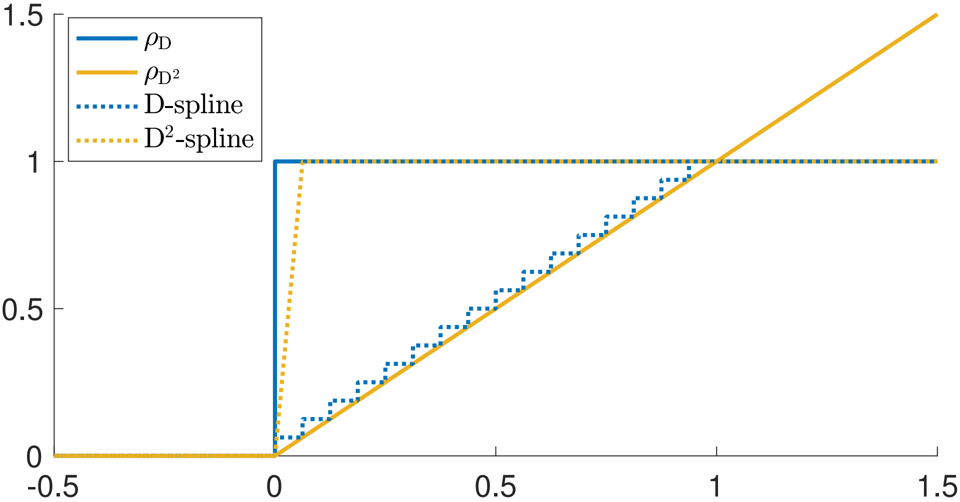


Figure 4.2: Approximation of ρ_{D} with a D^2-spline and vice-versa.

$s = s_1 + s_2$, where s_i is a sparse L_i-spline of the form

$$s_i(x) = \sum_{k=1}^{K_{s_i}} a_{k,i} \rho_{\mathrm{L}_i}(x - x_{k,i}) + \sum_{n=1}^{N_{0,i}} b_{n,i} p_{n,i}(x), \tag{4.22}$$

where $\{p_{n,i}\}_{n=1}^{N_{0,i}}$ form a basis of $\mathcal{N}_{\mathrm{L}_i}$. The knot locations $x_{i,k} \in I_T$ are chosen at random and the coefficients $a_{k,i}$ and $b_{n,i}$ are i.i.d. Gaussian random variables. We set $\lambda = 1.3$, $\alpha = 0.05$, $\epsilon = 10^{-3}$, and $h = 1/2^9$. The sparsity of the reconstructed signal is $K_1 = 7$ and $K_2 = 3$.

Observe that the reconstructed signal is very sparse and is satisfactory, in that it is close to what a human would reconstruct. Another promising feature of this experiment is that the balance between both dictionaries (*i.e.* D-splines and D^2-splines) is well-suited to the measurements. This was not obvious *a priori*, since at a very fine grid, a D-spline can be approximated very well by a D^2-spline with knots on the grid, and vice-versa. This is illustrated in Figure 4.2, where we voluntarily chose a coarse grid for visualization purposes. However, although these approximations (in dotted lines) might yield measurements similar to the corresponding Green's functions (solid lines), the regularization costs are discriminating. Therefore, when the weight parameter α is properly tuned, the reconstruction algorithm selects the most "natural" type of spline for a given path.

Compressed Sensing

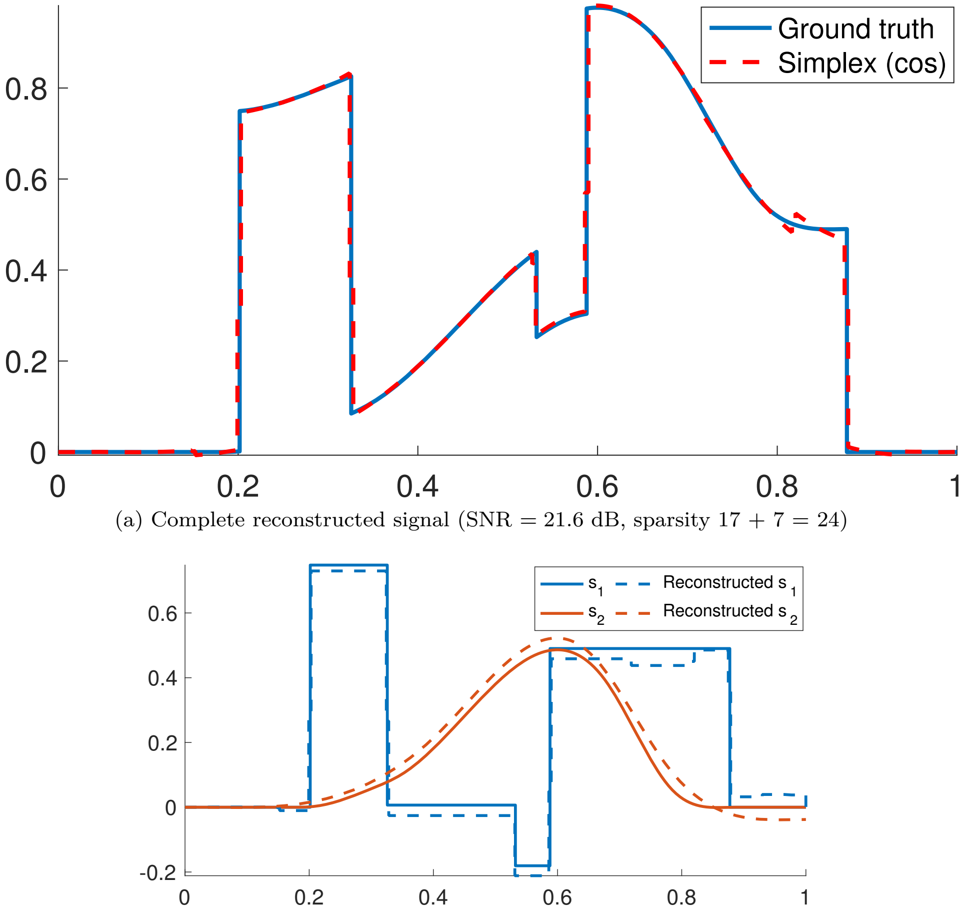
A second application of our framework is compressed-sensing-type problems, which attempt to the recover a sparse multicomponent signals given a small number of measurements [4]. We use the same type of test signals as in (4.22), namely hybrid splines with low sparsity $K_{s_1} + K_{s_2}$. The measurement functionals are assumed to take the form

$$\nu_m(s) = \int_0^T \cos(\omega_m x + \phi_m) s(x) \mathrm{d}x, \tag{4.23}$$

where $\omega_m \in \mathbb{R}$ are the sampling frequencies and $\phi_m \in [0, 2\pi)$ some given phase offsets. This amounts to sampling in the Fourier space — a rectangular window is applied in order to make the measurement functional compactly supported. We choose Fourier measurements because they are known to have good recovery properties according to the theory of compressed sensing [297]. However, the absence of a restricted isometry property (RIP)-type assumption prevents us from making any theoretical claims on the quality of the recovery.

An example run is shown in Figure 4.3 for a test signal with sparsity 15 ($K_{s_1} = 5$ and $K_{s_2} = 10$) and $M = 30$. Since the measurements are noiseless and we are interested in recovering the test signal as faithfully as possible, the data-fidelity term should be penalized much more than the regularization term. We therefore pick the regularization parameter $\lambda = 10^{-15} \ll 1$. The reconstructed signal in Figure 4.3a is remarkably close to the test signal, considering the difficulties of the problem. The separate components of the reconstructed signal compared to those of the test signal are provided in Figure 4.3b. We observe that the separation is not perfect: there is a small compensation effect between the two reconstructed components.

[4] Note that we are not interested in the data separation problem as in [92], but only in recovering the complete multicomponent signal



(a) Complete reconstructed signal (SNR = 21.6 dB, sparsity 17 + 7 = 24)

(b) Separate components

Figure 4.3: Reconstruction result with noiseless Fourier measurements for $\mathrm{L}_1 = \mathrm{D}$, $\mathrm{L}_2 = \mathrm{D}^4$, $M = 30$, $\lambda = 10^{-15}$, $\alpha = 5 \times 10^{-5}$. Final grid size: $h = 1/2^8$.

4.2.5 Summary

We have established a representer theorem that states that hybrid splines are solutions of continuous-domain inverse problems with multiple generalized total-variation regularization. The regularization operators L_1 and L_2 are taken to be multiple-order derivatives, which lead to piecewise-polynomial splines. This result implies that such problems can be solved exactly using a concatenated dictionary that consists of L_1 and L_2-splines. We proposed an exact B-spline-based discretization scheme to solve the continuous-domain problem in a suitable search space. Finally, we have illustrated our algorithm within some numerical examples. Our algorithm can be viewed as the continuous-domain counterpart of discrete data-separation problems, such as morphological component analysis.

4.3 Composite Models

In this section[5], we present a framework for the reconstruction of 1D composite signals assumed to be a mixture of two additive components, one sparse and the other smooth, given a finite number of linear measurements. We prove that these penalties induce reconstructed signals that indeed take the desired form of the sum of a sparse and a smooth component. We then discretize this problem which in order to to solve it numerically. Our discretization is exact in the sense that we are solving the continuous-domain problem over a restricted shift-invariant search space without any discretization error.

4.3.1 Context

In Section 4.2, we have demonstrated the limitation of single-component models for modelling real-world signals and we have presented a hybrid sparse-plus-sparse model. In this work, we investigate composite models of the form $s = s_1 + s_2$ where the first component s_1 is assumed to be sparse in some given domain and is treated with ℓ_1 regularization, while s_2 is assumed to be smooth and is treated with ℓ_2 regularization. In discrete settings, a natural way of reconstructing such signals is to solve the optimization problem

$$\min_{\mathbf{c}_1,\mathbf{c}_2\in\mathbb{R}^N} \left(E(\mathbf{H}(\mathbf{c}_1+\mathbf{c}_2),\mathbf{y}) + \lambda_1\|\mathbf{L}_1\mathbf{c}_1\|_1 + \lambda_2\|\mathbf{L}_2\mathbf{c}_2\|_2^2\right), \tag{4.24}$$

where $\mathbf{c}_1, \mathbf{c}_2$ are the two components of the signal $\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_2$, $\lambda_1, \lambda_2 > 0$ are tuning parameters, $\mathbf{L}_1 \in \mathbb{R}^{N\times N}$ is a sparsifying transform for $\mathbf{c}_1$, and $\mathbf{L}_2 \in \mathbb{R}^{N\times N}$ is a low-energy-promoting transform for $\mathbf{c}_2$. Amongst others, this modeling is considered in [298, 299, 300, 301, 302].

In this work, we adapt the discrete approach of (4.24) to 1D continuous-domain composite signals by solving an optimization problem of the form

$$\min_{s_1,s_2} \left(E(\boldsymbol{\nu}(s_1+s_2),\mathbf{y}) + \lambda_1\|\mathrm{L}_1\{s_1\}\|_{\mathcal{M}} + \lambda_2\|\mathrm{L}_2\{s_2\}\|_{L_2}^2\right), \tag{4.25}$$

[5] From our published work [280].

where s_1, s_2 are the two components of the signal $s = s_1 + s_2 : \mathbb{R} \to \mathbb{R}$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M) : s \mapsto \boldsymbol{\nu}(s) \in \mathbb{R}^M$ is the continuous-domain linear forward model, and L_1, L_2 are suitable continuously defined regularization operators. Typical choices are $\mathrm{L}_i = \mathrm{D}^{N_{0,i}}$ for $i \in \{1, 2\}$, where D is the derivative operator and $N_{0,i}$ the order of the derivative.

We remark that the model of Meyer [303] and its generalization by Vese and Osher [304, 305] follow the same idea as Problem (4.25), with the important difference that they use calculus-of-variation techniques to solve it. There is a connection as well with the Mumford-Shah functional [306], which is commonly used to segment an image in piecewise-smooth regions. The main difference lies in the fact that the optimization is not performed over the different components of the signal, but over the region boundaries. Another difference is that these models assume that one has full access to the noisy signal over a continuum, whereas (4.25) assumes that we only have access to some discrete measurements specified by the forward model $\boldsymbol{\nu}$.

We prove that there exists a solution to (4.25) of the form $s_1^* = s_1^* + s_2^*$ such that s_1^* is of the form (4.4) and s_2^* is of the form (4.3): a "sparse plus smooth" solution. Building on this representation, we propose an exact discretization scheme. Both components s_i for $i \in \{1, 2\}$ are expressed in a suitable Riesz basis as $s_i = \sum_k c_i[k]\varphi_{i,k}$, where $c_i[k]$ are the coefficients to be optimized. This leads to an infinite-dimensional optimization problem reminiscent of the infinite-dimensional compressed sensing framework of Adcock and Hansen [307] and the wavelet-based model of Daubechies *et al.* [308]. However, these frameworks differ from our original Problem (4.25) in that their native spaces admit a countable basis which leads to a more straighforward discretization process.

To solve this infinite-dimensional problem numerically, we cast it as a finite-dimensional problem under some mild assumptions. In our implementation, we choose basis functions $\varphi_{1,k} = \beta_{\mathrm{L}_1}(\cdot - k)$ and $\varphi_{2,k} = \beta_{\mathrm{L}_2^*\mathrm{L}_2}(\cdot - k)$, where β_{L} is the B-spline for the operator L. B-splines are popular choices of basis functions [10, 12, 6], in large part due to their minimal-support property. We show that optimizing over the spline coefficients leads to a discrete problem similar to (4.24) of the form

$$\min_{(\mathbf{c}_1,\mathbf{c}_2)\in\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}} \left(E(\mathbf{H}_1\mathbf{c}_1 + \mathbf{H}_2\mathbf{c}_2, \mathbf{y}) + \lambda_1\|\mathbf{L}_1\mathbf{c}_1\|_1 + \lambda_2\|\mathbf{L}_2\mathbf{c}_2\|_2^2\right), \tag{4.26}$$

where $\mathbf{H}_i \in \mathbb{R}^{M \times N_i}$ and $\mathbf{L}_i \in \mathbb{R}^{P_i \times N_i}$ for $i \in \{1, 2\}$. This discretization is exact in the sense that it is equivalent to the continuous problem (4.25) when each component s_i lies in the space generated by the basis functions $\{\varphi_{i,k}\}_{k \in \mathbb{Z}}$. This is a consequence of our informed choice of these basis functions $\varphi_{i,k}$. Moreover, the short support of the B-splines leads to well-conditioned $\mathbf{H}_i$ matrices and, thus, to a computationally feasible problem.

4.3.2 Theory

For the sake of clarity, we only consider the higher-order derivatives as our regularization operators, *i.e.* $\mathrm{L}_i = \mathrm{D}^{N_i}$ with $N_2 \geq N_1 \geq 1$[6] for $i = 1, 2$. For the sparse component, we use the same regularization (gTV) as we did for the individual components of the hybrid model, presented in Section 4.2. We recall that the restricted search space for the sparse component is defined as

$$\mathcal{M}_{\mathrm{L}_1, \boldsymbol{\phi}_1}(\mathbb{R}) = \{f \in \mathcal{M}_{\mathrm{L}_1}(\mathbb{R}) : \boldsymbol{\phi}_1(f) = \mathbf{0}\}, \tag{4.27}$$

where $(\boldsymbol{\phi}_1, \mathbf{p}_1) = (\phi_n, p_n)_{n=1}^{N_1}$ is a biorthogonal system for $\mathcal{N}_{\mathrm{L}_1}$. We also recall that any function $s_1 \in \mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$ has the unique decomposition

$$s_1 = \mathrm{L}_{1,\boldsymbol{\phi}_1}^{-1}\{w\} + \mathbf{c}_1^T \mathbf{p}_1, \tag{4.28}$$

where $w \in \mathcal{M}(\mathbb{R})$, $\mathbf{c}_1 = \boldsymbol{\phi}_1(s_1) \in \mathbb{R}^{N_1}$, and $\mathrm{L}_{1,\tilde{\boldsymbol{\phi}}_1}^{-1}$ is the pseudo-inverse operator of L_1 for the biorthogonal system $(\boldsymbol{\phi}_1, \mathbf{p}_1)$.

The regularization norm $\|\cdot\|_{L_2}$ for the smooth component s_2 in (4.25) is defined over the Hilbert space $L_2(\mathbb{R})$. The corresponding native space of the smooth component s_2 is the Hilbert space $\mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$ with the fundamental property that $f \in \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$ if and only if $\mathrm{L}_2\{f\} \in L_2(\mathbb{R})$. We highlight that the biorthogonal system for $\mathcal{N}_{\mathrm{L}_1}$ can be extended to a biorthogonal system $(\boldsymbol{\phi}_1, \mathbf{p}_1) = (\phi_n, p_n)_{n=1}^{N_2}$ for $\mathcal{N}_{\mathrm{L}_2}$. Similar to the sparse component, for any $s_2 \in \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$, there is a unique decomposition

$$s_2 = \mathrm{L}_{2,\boldsymbol{\phi}_2}^{-1}\{h\} + \mathbf{c}_2^T \mathbf{p}_2, \tag{4.29}$$

[6] Although the assumption $N_2 \geq N_1$ is not necessary, we put it here for the sake of simplicity.

where $\mathbf{c}_2 = \boldsymbol{\phi}_2(s_2) \in \mathbb{R}^{N_2}$ and $h \in L_2(\mathbb{R})$.

We now present in Theorem 4.1 our problem formulation to reconstruct sparse-plus-smooth composite signals. This representer Theorem gives a parametric form of a solution of our optimization problem.

Theorem 4.2 (Continuous-domain representer theorem)**.** *Let $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_{\geq 0}$ be a strictly convex functional. Let $\mathrm{L}_i = \mathrm{D}^{N_i}$ with $N_2 \geq N_1 \geq 1$ be regularization operators and let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_M)$ be a linear measurement operator composed of the M linear functionals $\nu_m : f \mapsto \nu_m(f) \in \mathbb{R}$ that are weak*-continuous over $\mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$ and over $\mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$. We assume that $\mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{N}_{\mathrm{L}_2} = \{0\}$, where $\mathcal{N}_{\boldsymbol{\nu}}$ is the null space of $\boldsymbol{\nu}$ (well-posedness assumption). Then, for any $\lambda_1, \lambda_2 > 0$, the optimization problem*

$$\mathcal{V} = \underset{\substack{s_1 \in \mathcal{M}_{\mathrm{L}_1, \boldsymbol{\phi}_0}(\mathbb{R}) \\ s_2 \in \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})}}{\arg\min} \left(E(\boldsymbol{\nu}(s_1 + s_2), \mathbf{y}) + \lambda_1 \|\mathrm{L}_1\{s_1\}\|_{\mathcal{M}} + \lambda_2 \|\mathrm{L}_2\{s_2\}\|_{L_2}^2 \right) \tag{4.30}$$

has a solution $(s_1^*, s_2^*) \in \mathcal{V}$ with the following components:

- *the component s_1^* is a nonuniform L_1-spline of the form*

$$s_1^*(x) = p_1(x) + \sum_{k=1}^{K_1} a_{1,k} \rho_{\mathrm{L}_1}(x - x_k) \tag{4.31}$$

  for some $K_1 \leq (M - N_1)$, where $p_1 \in \mathcal{N}_{\mathrm{L}_1}$, and $a_{1,k}, x_k \in \mathbb{R}$;

- *the component s_2^* is of the form*

$$s_2^*(x) = p_2(x) + \sum_{m=1}^{M} a_{2,m} h_m(x), \tag{4.32}$$

  where $h_m(x) = \nu_m * \mathcal{F}^{-1}\left\{\frac{1}{|\widehat{L}_2|^2}\right\}(x)$, $p_2 \in \mathcal{N}_{\mathrm{L}_2}$, $a_{2,k} \in \mathbb{R}$, and where $\sum_{m=1}^{M} a_{2,m} \langle q_2, \nu_m \rangle = 0$ for any $q_2 \in \mathcal{N}_{\mathrm{L}_2}$.

Moreover, for any pair of solutions $(s_1^*, s_2^*), (\tilde{s}_1^*, \tilde{s}_2^*) \in \mathcal{V}$, s_2^* and $\tilde{s}_2^*$ differ only up to an element of the null space $\mathcal{N}_{\mathrm{L}_2}$, so that $(s_2^* - \tilde{s}_2^*) \in \mathcal{N}_{\mathrm{L}_2}$.

Proof. Using decompositions (4.28) and (4.29), we deduce that there is bijection between $\mathcal{V}$ and the solution set

$$\tilde{\mathcal{V}} = \operatorname*{arg\,min}_{\substack{w\in\mathcal{M}(\mathbb{R})\\ h\in L_2(\mathbb{R})\\ \mathbf{c}\in\mathbb{R}^{N_2}}} \left(E(\tilde{\boldsymbol{\nu}}(w,h,\mathbf{c}),\mathbf{y}) + \lambda_1\|w\|_{\mathcal{M}} + \lambda_2\|h\|_{L_2}^2\right), \tag{4.33}$$

where $\tilde{\boldsymbol{\nu}} : \mathcal{M}(\mathbb{R}) \times L_2(\mathbb{R}) \times \mathbb{R}^{N_2} \to \mathbb{R}^M$ with $\tilde{\boldsymbol{\nu}}(w,h,\mathbf{c}) = \boldsymbol{\nu}(s_1+s_2)$. Similar to the proof of Theorem 4.1, we can show that $\tilde{\boldsymbol{\nu}}$ is weak*-continuous and the cross-product matrix $\mathbf{V} = [\langle\nu_m, p_n]\rangle] \in \mathbb{R}^{M\times N_2}$ has full rank. Consequently, we deduce from Theorem 2.5 that the solution set $\tilde{\mathcal{V}}$ is nonempty and contains an element $(\tilde{w},\tilde{h},\tilde{\mathbf{c}}) \in \tilde{\mathcal{V}}$. Following the one-to-one correspondence between $\tilde{\mathcal{V}}$ and $\mathcal{V}$, we deduce that $(\tilde{s}_1,\tilde{s}_2)$ is a solution of (4.30), where $\tilde{s}_1 = \mathrm{L}_{1,\phi_1}^{-1}\tilde{w}$ and $\tilde{s}_2 = \mathrm{L}_{2,\phi_2}^{-1}\tilde{h}+\tilde{\mathbf{c}}^T\mathbf{p}_2$. Using the representer theorem for the minimum-gTV interpolation problem (see [73]), we deduce that there exists a minimizer s_1^* of the form (4.31) for the minimization

$$\min_{s_1\in\mathcal{M}_{\mathrm{L}_1,\phi_0}(\mathbb{R})} \|\mathrm{L}_1\{s_1\}\|_{\mathcal{M}} \quad \text{s.t.} \quad \boldsymbol{\nu}(s_1) = \boldsymbol{\nu}(\tilde{s}_1). \tag{4.34}$$

One can also readily verify that $(s_1^*,\tilde{s}_2)$ is a minimizer of the original problem. Similarly, one can consider the minimization problem

$$\min_{s_2\in\mathcal{H}_{\mathrm{L}_2}(\mathbb{R})} \|\mathrm{L}_2\{s_2\}\|_{L_2} \quad \text{s.t.} \quad \boldsymbol{\nu}(s_2) = \boldsymbol{\nu}(\tilde{s}_2). \tag{4.35}$$

It is known from [74, Theorem 3] that (4.35) has a minimizer s_2^* of the form (4.32). Again, (s_1^*,s_2^*) is a solution of the original problem, which matches the form specified by Theorem 4.2. Finally, the uniqueness of the second component (up to a null-space element) directly follows from the strict convexity of the mapping $f \mapsto \|\mathrm{L}_2\{f\}\|_{L_2}^2$. □

4.3.3 Exact Discretization

In order to discretize Problem (4.30), we restrict the search spaces $\mathcal{M}_{\mathrm{L}_1,\phi_0}(\mathbb{R})$ and $\mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$. For the sparse component, we choose basis functions $\varphi_{1,k} = \beta_{\mathrm{L}_1,h}(\cdot - k)$

for all $k \in \mathbb{Z}$. With this choice, the restricted search space

$$V_1(\mathbb{R}) = \left\{ f = \sum_{k\in\mathbb{Z}} c_1[k]\varphi_{1,k} : c_1 \in V_1(\mathbb{Z}) \right\} \subset \mathcal{M}_{\mathrm{L}_1,\phi_0}(\mathbb{R}) \tag{4.36}$$

with the digital-filter space

$$V_1(\mathbb{Z}) = \left\{ (c_1[k])_{k\in\mathbb{Z}} : \ (d_{\mathrm{L}_1} * c_1) \in \ell_1(\mathbb{Z}) \text{ and } \sum_{k\in\mathbb{Z}} c_1[k]\boldsymbol{\phi}_1(\varphi_{1,k}) = \mathbf{0} \right\}, \tag{4.37}$$

is the largest possible reconstruction space [3, Equation (22)]. The choice of the basis functions $\varphi_{1,k}$ is guided by the following considerations:

- they generate the space of uniform L_1 splines. This conforms with Theorem 4.2, which states that the component s_1^* is an L_1-spline;
- they enable exact computations in the continuous domain. In particular, by invoking (1.13), we observe that $\|\mathrm{L}_1\{\sum_{k\in\mathbb{Z}} c_1[k]\varphi_{1,k}\}\|_{\mathcal{M}} = \|d_{\mathrm{L}_1} * c_1\|_{\ell_1}$;
- the Riesz-basis property of B-splines leads to a well-conditioned system matrix, which is paramount in numerical applications (see, Section 1.1.4).

Similarly, we define the restricted search space for the smooth component as

$$V_2(\mathbb{R}) = \left\{ f = \sum_{k\in\mathbb{Z}} c_2[k]\varphi_{2,k} : c_2 \in V_2(\mathbb{Z}) \right\} \subset \mathcal{H}_{\mathrm{L}_2}(\mathbb{R}), \tag{4.38}$$

where the basis functions $\varphi_{2,k}$ and the digital-filter space $V_2(\mathbb{Z})$ need to be chosen to have that $V_2(\mathbb{R}) \subseteq \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$.

At first glance, the most natural choice for $\varphi_{2,k}$ is to select the basis functions suggested by (4.32) in Theorem 4.2: h_m for $1 \leq m \leq M$ and a basis of $\mathcal{N}_{\mathrm{L}_2}$, which yield a finite number $M + N_2$ of basis functions. However, this approach runs into the following hitches:

- the basis functions h_m are typically increasing at infinity, which contradicts the Riesz-basis requirement and leads to severely ill-conditioned optimization tasks [74];
- depending on the measurements operator $\boldsymbol{\nu}$, h_m may lack a closed-form expression.

We therefore focus on these criteria, in a spirit similar to [309]. The $\varphi_{2,k}$ are chosen to be regular shifts of a generating function φ_2, with $\varphi_{2,k} = \varphi_2(\cdot - k)$ such that $\{\mathrm{L}_2\{\varphi_{2,k}\}\}_{k\in\mathbb{Z}}$ forms a Riesz basis. Contrary to $\varphi_{1,k}$, these requirements allow for many choices of $\varphi_{2,k}$. In order to perform exact discretization, one then only needs to compute the following autocorrelation filter.

Proposition 4.1 (Autocorrelation filter for the smooth component)**.** *Let φ_2 be a generating function such that $\{\varphi_{2,k} = \varphi_2(\cdot - k)\}_{k\in\mathbb{Z}}$ form a Riesz basis. Then, the following two items hold:*

- *the inner product $\langle \mathrm{L}_2\{\varphi_{2,k}\}, \mathrm{L}_2\{\varphi_{2,k'}\}\rangle_{L_2}$ only depends on the difference $(k - k')$. We can thus introduce the autocorrelation filter*

$$\rho[k] = \langle \mathrm{L}_2\{\varphi_{2,k}\}, \mathrm{L}_2\{\varphi_{2,0}\}\rangle_{L_2} = \langle \mathrm{L}_2\{\varphi_{2,k+k'}\}, \mathrm{L}_2\{\varphi_{2,k'}\}\rangle_{L_2} \tag{4.39}$$

  for any $k, k' \in \mathbb{Z}$;
- *the filter ρ is positive semidefinite, with $\sum_{k,k'\in\mathbb{Z}} c[k]c[k']\rho[k-k'] \geq 0$ for any finitely supported real digital filter c.*

Proof. The first item is proved with a simple change of variable in the integral that defines the inner product. The second item is derived by observing that, for any c_2, we have

$$\begin{aligned}\left\| \mathrm{L}_2 \left\{ \sum_{k\in\mathbb{Z}} c_2[k]\varphi_{2,k} \right\} \right\|_{L_2}^2 &= \sum_{k,k'\in\mathbb{Z}} \left(c_2[k]c_2[k'] \langle \mathrm{L}_2\{\varphi_{2,k}\}, \mathrm{L}_2\{\varphi_{2,k'}\}\rangle \right) \\ &= \sum_{k,k'\in\mathbb{Z}} c_2[k]c_2[k']\rho[k-k'] \geq 0. \end{aligned} \tag{4.40}$$

□

For our implementation, we make a specific choice of basis function φ_2 among the many choices for which the autocorrelation filter (4.39) can be computed analytically. We choose the scaled $\mathrm{L}_2^*\mathrm{L}_2$ B-spline basis $\varphi_2 = \beta_{\mathrm{L}_2^*\mathrm{L}_2,h}$ and $\varphi_{2,k} = \varphi_2(\cdot - k)$, where L_2^* denotes the adjoint operator of L_2. This choice has the following additional advantages:

- the generator φ_2 has a simple explicit expression that does not depend on the measurement operator $\boldsymbol{\nu}$;
- the autocorrelation filter ρ also has a simple expression, as will be shown in Proposition 4.2;
- in the special case of the sampling operator $\nu_m = \delta(\cdot - x_m)$, where the x_m are the sampling locations, this choice conforms with (4.32) in Theorem 4.2 since s_2^* is then an $\mathrm{L}_2^*\mathrm{L}_2$-spline. Note, however, that we do not exploit the knowledge that s_2^* has knots at the sampling locations x_m.

With this basis function $\varphi_2 = \beta_{\mathrm{L}_2^*\mathrm{L}_2}$, there is no straightforward choice for the digital-filter space $V_2(\mathbb{Z})$. Our practical choice is given in (4.46) which is motivated from our discretization method. To see this more clearly, we now prove a simpler form for the regularization term of the smooth component that is based on the factorization of the autocorrelation sequence.

Proposition 4.2 (Factorization of the autocorrelation filter). *When $\varphi_2 = \beta_{\mathrm{L}_2\mathrm{L}_2^*}$, the autocorrelation filter ρ defined in Proposition 4.1 is of the form*

$$\rho = d_{\mathrm{L}_2} * d_{\mathrm{L}_2}^{\vee} * b, \tag{4.41}$$

where $b[k] = \beta_{\mathrm{L}_2^*\mathrm{L}_2}(k)$ is the B-spline kernel of the operator $\mathrm{L}_2^*\mathrm{L}_2$, which is a positive-semidefinite filter of finite support. The filter ρ can thus be factorized as $\rho = g * g^{\vee}$ with

$$g = d_{\mathrm{L}_2} * b^{1/2}, \tag{4.42}$$

where the finite-support filter $b^{1/2}$ satisfies $b = b^{1/2} * (b^{1/2})^{\vee}$.

Proof. We have that

$$\begin{aligned}\rho[k] &= \langle \mathrm{L}_2\{\varphi_{2,k}\}, \mathrm{L}_2\{\varphi_{2,0}\}\rangle_{L_2} = \langle \mathrm{L}_2^*\mathrm{L}_2\{\varphi_{2,k}\}, \varphi_{2,0}\rangle_{\mathcal{H}'_{\mathrm{L}_2}\times\mathcal{H}_{\mathrm{L}_2}} \\ &= \langle \sum_{k'\in\mathbb{Z}} d_{\mathrm{L}_2^*\mathrm{L}_2}[k]\delta(\cdot-(k+k')), \varphi_{2,0}\rangle_{\mathcal{H}'_{\mathrm{L}_2}\times\mathcal{H}_{\mathrm{L}_2}} = \sum_{k'\in\mathbb{Z}} d_{\mathrm{L}_2^*\mathrm{L}_2}[k]b[k+k'] \\ &= (d_{\mathrm{L}_2^*\mathrm{L}_2} * b^\vee)[-k] = (d_{\mathrm{L}_2^*\mathrm{L}_2} * b)[k], \end{aligned} \tag{4.43}$$

where $\langle\cdot,\cdot\rangle_{\mathcal{H}'_{\mathrm{L}_2}\times\mathcal{H}_{\mathrm{L}_2}}$ denotes the duality product between $\mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$ and its dual $\mathcal{H}'_{\mathrm{L}_2}(\mathbb{R})$, and the last line results from the symmetry of ρ and b.

Next, we prove that b is positive-semidefinite. Indeed, for any finitely supported filter c, we have that

$$\sum_{k,k'\in\mathbb{Z}} c[k]c[k']b[k-k'] = \left\| \sum_{k\in\mathbb{Z}} c[k]\beta_{\mathrm{L}_2}(\cdot-k) \right\|_{L_2}^2 \geq 0, \tag{4.44}$$

where we have used the property

$$b[k] = (\beta_{\mathrm{L}_2} * \beta_{\mathrm{L}_2}^\vee)(k) = \langle \beta_{\mathrm{L}_2}, \beta_{\mathrm{L}_2}(\cdot-k)\rangle_{L_2}. \tag{4.45}$$

Finally, to prove the existence of $b^{1/2}$, we notice that b has the finite support due to the finite support of $\beta_{\mathrm{L}_2^*\mathrm{L}_2}$. Since b is also symmetric, its z-transform satisfies $B(z) = B(z^{-1})$; therefore, for any zero z_k of $B(z)$, z_k^{-1} is also a zero. Moreover, it is well known that $B(\pm1) \neq 0$, so that zeros must come in pairs $z_k \neq z_k^{-1}$. Hence, $B(z)$ can be written as $B(z) = \prod_k(1-z_kz)(1-z_kz^{-1})$. Hence, to take $b^{1/2}$ to be the inverse z-transform of $B^{1/2}(z) = \prod_{k=1}(1-z_kz^{-1})$ is a valid choice (we clearly have $b = b^{1/2} * (b^{1/2})^\vee$), and (4.42) is readily obtained. □

Using this factorization, we now define the discrete search space $V_2(\mathbb{Z})$ as

$$V_2(\mathbb{Z}) = \left\{ (c_2[k])_{k\in\mathbb{Z}} : \ (g * c_2) \in \ell_2(\mathbb{Z}) \right\}. \tag{4.46}$$

Following Proposition 4.2, we observe that for any $c \in V_2(\mathbb{Z})$, the function $s_2 = \sum_{k\in\mathbb{Z}} c_2[k]\varphi_{2,k}$ satisfies $\|\mathrm{L}_2\{s_2\}\|_{L_2}^2 = \|g*c_2\|_{\ell_2}^2 = \|b^{1/2}*(d_{\mathrm{L}_2}*c_2)\|_{\ell_2}^2 < +\infty$, which proves that $s_2 \in \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$.

These choices enable us to discretize Problem (4.30) in an exact way in the $V_i(\mathbb{R})$ spaces. Precisely, it can be readily seen that the discrete problem

$$\mathcal{V}_{\mathrm{d}} = \underset{(c_1,c_2)\in V_1(\mathbb{Z})\times V_2(\mathbb{Z})}{\arg\min} \left(E\left(\sum_{k_1\in\mathbb{Z}} c_1[k]\boldsymbol{\nu}(\varphi_{1,k}) + \sum_{k\in\mathbb{Z}} c_2[k]\boldsymbol{\nu}(\varphi_{2,k}), \mathbf{y} \right) + \lambda_1 \|d_{\mathrm{L}_1} * c_1\|_{\ell_1} + \lambda_2 \|g * c_2\|_{\ell_2}^2 \right) \tag{4.47}$$

is equivalent to the restricted continuous-domain problem

$$\mathcal{V}_{\mathrm{res}} = \underset{(s_1,s_2)\in V_1(\mathbb{R})\times V_2(\mathbb{R})}{\arg\min} \left(E(\boldsymbol{\nu}(s_1+s_2), \mathbf{y}) + \lambda_1 \|\mathrm{L}_1\{s_1\}\|_{\mathcal{M}} + \lambda_2 \|\mathrm{L}_2\{s_2\}\|_{L_2}^2 \right), \tag{4.48}$$

in the sense that there exists a bijective linear mapping $(c_1, c_2) \mapsto \left(\sum_{k\in\mathbb{Z}} c_1[k]\varphi_{1,k}, \sum_{k\in\mathbb{Z}} c_2[k]\varphi_{2,k}\right)$ from $\mathcal{V}_{\mathrm{d}}$ to $\mathcal{V}_{\mathrm{res}}$.

To solve Problem (4.47) numerically in an exact way, we note that the operators L_i for $i \in \{1, 2\}$ admit a B-spline with finite support, which implies that the filters d_{L_i} and g have finite support. We also assume that the measurement functionals ν_m are compactly supported. This is natural and is often fulfilled in practice, for instance in imaging with a finite field of view. The support length then roughly corresponds to the number of grid points in the interval of interest. These together with our choice of basis functions enable us to restrict Problem (4.47) to the interval of interest. The restriction to a finite number of active spline coefficients leads to finite-dimensional system and regularization matrices. This then can be solved efficiently using iterative methods. For the precise setting of boundary conditions and algorithmic details, we refer to our supporting paper [280].

4.3.4 Numerical Illustration

We now validate our reconstruction algorithm in a simulated setting. We generate a ground-truth signal $s^{\mathrm{GT}} = s_1^{\mathrm{GT}} + s_2^{\mathrm{GT}}$. The sparse component s_1^{GT} is chosen to be an L_1-spline with few jumps, for which gTV is an adequate choice of regularization, as demonstrated by (4.31) in our representer theorem. For the smooth component s_2^{GT}, we generate a realization of a solution s_2 of the stochastic differential equation

$\mathrm{L}_2 s_2 = w$, where w is a Gaussian white noise with standard deviation σ_2 by following the method of [310]. The operator L_2 then acts as a whitening operator for the stochastic process s_2. The reason for this choice is the connection between the minimum mean-square estimation of such stochastic processes and the solutions to variational problems with gTikhonov regularization $\|\mathrm{L}_2 s_2\|_{L_2}^2$ [8, 311, 57].

Our forward model is the Fourier-domain cosine sampling operator of the form $\nu_1(s) = \int_0^1 s(t)\mathrm{d}t$ (DC term) and

$$\nu_m(s) = \int_0^1 \cos(\omega_m t + \theta_m) s(t)\mathrm{d}t \tag{4.49}$$

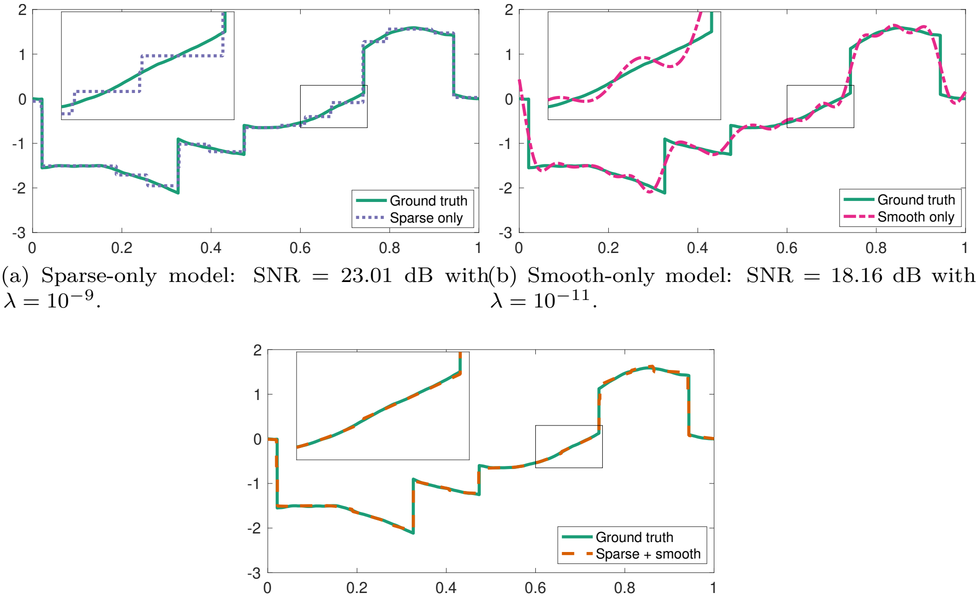
for $2 \leq m \leq M$, where the sampling pulsations ω_m are chosen at random within the interval $(0, \omega_{\max}]$, and the phases θ_m are chosen at random within the interval $[0, 2\pi)$. Notice that ν_m is a Fourier-domain measurement of the restriction of s to the interval of interest $[0, 1]$. Finally, we use the standard quadratic error $E(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$ for the data-fidelity term.

We validate our sparse-plus-smooth model against more standard non-composite models. More precisely, for $i \in \{1, 2\}$ we solve the regularized problems

$$\arg\min_{f \in \mathcal{X}_i} \left(E(\boldsymbol{\nu}(f), \mathbf{y}) + \lambda \mathcal{R}_i(f) \right) \tag{4.50}$$

with regularizers $\mathcal{R}_1(f) = \|\mathrm{L}_1\{f\}\|_{\mathcal{M}}$ (sparse model with native space $\mathcal{X}_1 = \mathcal{M}_{\mathrm{L}_1}(\mathbb{R})$) and $\mathcal{R}_2(f) = \|\mathrm{L}_2\{f\}\|_{L_2}$ (smooth model with native space $\mathcal{X}_2 = \mathcal{H}_{\mathrm{L}_2}(\mathbb{R})$). We discretize these problems using the reconstruction spaces $V_i(\mathbb{R})$ without restricting $V_1(\mathbb{R})$ with the boundary conditions $\boldsymbol{\phi}_1$. The sparse model thus amounts to an ℓ_1-regularized discrete problem which we solve using ADMM, while the smooth model has a closed-form solution that can be obtained by inverting a matrix.

For this comparison, we choose regularization operators $\mathrm{L}_1 = \mathrm{D}$ and $\mathrm{L}_2 = \mathrm{D}^2$ with $M = 50$ Fourier-domain measurements (cosine sampling with $\omega_{\max} = 100$). We generate the ground-truth signal with $K_1 = 5$ jumps whose i.i.d. Gaussian amplitudes have the variance $\sigma_1^2 = 1$ for s_1^{GT}. For the smooth component s_2^{GT}, we generate a realization of a Gaussian white noise w with the variance $\sigma_2^2 = 100$, such that $\mathrm{L}_2\{s_2^{\mathrm{GT}}\} = w$. The measurements are corrupted by some i.i.d. Gaussian white noise $\mathbf{n} \in \mathbb{R}^M$ so that $\mathbf{y} = \boldsymbol{\nu}(s_{\mathrm{GT}}) + \mathbf{n}$. We set the signal-to-noise ratio (SNR)



(a) Sparse-only model: SNR = 23.01 dB with $\lambda = 10^{-9}$.

(b) Smooth-only model: SNR = 18.16 dB with $\lambda = 10^{-11}$.

(c) Sparse-plus-smooth model: SNR = 27.02 dB with $\lambda_1 = 8 \cdot 10^{-7}$ and $\lambda_2 = 5 \cdot 10^{-10}$.

Figure 4.4: Comparison between our sparse-plus-smooth model and non-composite models with regularization operators $\mathrm{L}_1 = \mathrm{D}$, $\mathrm{L}_2 = \mathrm{D}^2$, and $M = 50$ Fourier-domain measurements.

between $\boldsymbol{\nu}(s_{\mathrm{GT}})$ and $\mathbf{n}$ to be 50 dB. The regularization parameters are selected through a grid search with $h = 1/2^9$ to maximize the SNR of the reconstructed signal s with respect to the ground truth. The SNR is defined as

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\int_0^T (s^{\mathrm{GT}}(t))^2 \mathrm{d}t}{\int_0^T (s(t) - s^{\mathrm{GT}}(t))^2 \mathrm{d}t} \right). \tag{4.51}$$

The results of this comparison in terms of SNR are shown in Table 4.1 for varying grid sizes h. For all methods, the SNR values increase when the grid size decreases,

h	Sparse plus smooth	Sparse only	Smooth only
2^6	18.02	17.72	18.16
2^7	21.87	21.08	18.16
2^8	23.46	22.07	18.16
2^9	27.02	23.01	18.16
2^{10}	25.70	23.04	18.16

Table 4.1: SNR values (in dB) of the reconstructed signal with respect to the ground truth with varying grid size h.

which is to be expected since the grids are embedded. The only exception is the sparse-plus-smooth reconstruction for the finest grid size, which is likely due to numerical issues arising from the increased dimension ($N \approx 2^{11}$) of the optimization problem. The effect of the grid size on the quality of the reconstruction varies between the models: it is almost non-existent for the smooth-only model, whereas it is most significant for our sparse-plus-smooth model. Over all grid sizes, due to the fact that our sparse-plus-smooth signal model matches the ground truth, our reconstructed signal yields a higher SNR (27.02 dB) than the sparse-only (23.04 dB) and smooth-only (18.16 dB) models.

The reconstruction results for the grid size $h = 1/2^9$ are shown in Figure 4.4. Our sparse-plus-smooth reconstruction is qualitatively much more satisfactory. As can be observed in the zoomed-in section, the sparse-only model is subject to a staircasing phenomenon in the smooth regions of the ground-truth signal, a well-known shortcoming of total-variation regularization. Conversely, our reconstruction is remarkably accurate in the smooth regions. Finally, the smooth-only model fails both visually and in terms of SNR, due to its inability to represent sharp jumps.

4.3.5 Summary

We have introduced a continuous-domain framework for the reconstruction of multicomponent signals. It assumes two additive components, the first one being sparse and the other one being smooth. The reconstruction is performed by solving a regularized inverse problem, using a finite number of measurements of the signal. The form of a solution to this problem is given by our representer theorem. This form justifies the choice of the search space in which we discretize the problem. Our discretization is exact, in the sense that it amounts to solving a continuous-domain optimization problem restricted to our search space. The discretized problem is then solved using our ADMM-based algorithm, which we validate on simulated data.

4.4 Curve Fitting

To conclude this chapter, we present[7] a practical application of our multicomponent model for the reconstruction of continuous-domain univariate signals. To that end, we formulate as an inverse problem the construction of sparse parametric continuous curve models that fit a sequence of contour points. Our prior is incorporated as a regularization term that encourages rotation invariance and sparsity. We prove that an optimal solution to the inverse problem is a closed curve with spline components. We then show how to efficiently solve the task using B-splines as basis functions. We extend our problem formulation to curves made of two distinct components with complementary smoothness properties and solve it using hybrid splines. We illustrate the performance of our model on contours having different smoothness properties. Our experimental results show that we can faithfully reconstruct any general contour using few parameters, even in the presence of imprecisions in the measurements.

4.4.1 Context

Contour tracing is a common yet rich subject in the image processing and computer graphics community. It produces a contour that accurately separates two regions of a given image. This task is, however, not without difficulties. Firstly, the edges suffer from discretization effects and intrinsic image noise. Secondly, the smoothness of the contour may need to be nonuniform, since contours are often made of smooth parts joined by sharp discontinuities.

Our goal is to extract a continuous stylized sparse parametric curve that explains a given set of ordered edge points given by possibly inaccurate two-dimensional coordinates, which is particularly relevant for raster-to-vector conversion [312]. Our search for sparsity intimately follows Occam's razor principle of simplicity. Indeed, it heightens our probability of approaching the true curve, as many real-world signals are sparse. This is the principle on which compressed sensing hinges [62, 66].

[7] This section is based on our submitted work [281].

Two main approaches come to mind when thinking of contour tracing. The first one consists in joint edge detection and curve fitting. Parametric active contours are popular examples of this approach as these methods provide efficient tools for the extraction of a contour from an image. The contour consists in continuous curves that evolve through the optimization of an energy functional and iteratively approximate an image edge [313, 314]. A plethora of parametric snake models can be found in the literature [315, 316, 317, 318, 319]. Of particular relevance to this work is a snake model implementation that uses basis functions and that allows for tangent control, a useful property when the smoothness of the contours is nonuniform [29].

The second approach to contour tracing is discrete contour extraction and subsequent curve fitting. In the first approach, the entire image was used to iteratively update the contour, whereas the second approach interpolates a continuous parametric curve from a list of coordinates. A popular way to tackle this is through a regularized minimization problem, the regularization enforcing prior knowledge about the curve [320, 321]. The method presented in this work follows the paradigm of the second approach.

To attain our goal, we solve a bipartite optimization problem. On one hand, we want that the candidate curve fits the existing contour points exactly. This is achieved through a data-fidelity term. On the other hand, as an infinity of curves could satisfy this fit, we have to enforce prior knowledge into our model. This prior knowledge is introduced as a regularization cost coupled with a regularization operator, the result aiming at the enforcement of desired properties. First, it is likely that the true curve has few variations, which implies that the curve has a sparse representation. Second, it is frequent that variations happen over both the horizontal and the vertical axes simultaneously. Moreover, the recovered curve should be possibly denoised. Finally, the optimization cost should not depend on a rotation of the system of coordinates. We show that these specifications lead us to a regularization cost that consists of a mixed (TV-ℓ_2) norm. A similar setting was already explored in [322] to recover signals using a sparsity-promoting regularizer, but the signals were discrete. In this work, we explore the continuous setting, as we aim at the recovery of a continuous 2D curve. Similarly to [166], we explore generalized total-variation (TV) regularization for continuous-domain signal reconstruction using B-splines as basis functions for an exact discretization.

Finally, we choose to represent the curve with hybrid splines, which give us the tools to represent curves with nonuniform smoothness. While [3] addressed signal reconstruction using hybrid splines, this manuscript extends the setting for the handling of curves in 2D, which calls for a new regularizer.

Our main contribution is threefold. Firstly, we introduce a continuous rotation-invariant TV (RI-TV) norm as a regularization for the recovery of curves. It effectively reconstructs sparse parametric curves from given contour points while being robust to noise. Secondly, we prove a representer theorem according to which there exists a curve with spline components that is a global minimizer of our optimization problem. Building upon this, we propose a curve construction using B-splines, which allows us to discretize the continuous-domain problem exactly with numerical efficiency. Finally, we present the combination of such RI-TV norm with a hybrid framework to generate stylized curves with nonuniform smoothness properties.

4.4.2 Theory

Our goal is to recover a 2D parametric curve $\mathbf{r}(t) = (x(t), y(t))$ that best fits a given ordered list of points $\mathbf{p}[m] = (\mathrm{p}_x[m],\ \mathrm{p}_y[m]),\ m = 0, \dots, M-1$. Contours being closed curves, we consider the coordinate functions $x(t)$ and $y(t)$ to be periodic in t. Since we have M data locations, it is convenient to deal with M-periodic functions. We consequently set $t \in \mathbb{T}_M = [0, M]$.

Concurrently, we want to control the sparsity of the fitted curve. This can be achieved by limiting the number of the singularities in the higher-order derivatives of its components. This effectively means that $\mathbf{r}(t)$ admits a sparse representation. To that end, we introduce two new elements: a differential operator L and the RI-TV regularization functional.

The mathematical foundation on which this work is built is based on Schwartz' theory of distributions [169] over the torus. This is in contrast to previous parts of this thesis, where Euclidean spaces were used. To that end, we review previously defined notions once more to highlight similarities and differences between the periodic and nonperiodic settings. To that end, let us denote the Schwartz' space of

M-periodic smooth functions by $\mathcal{S}(\mathbb{T}_M)$. Its topological dual, $\mathcal{S}'(\mathbb{T}_M)$ is the space of tempered distributions over the torus. Moreover, the space of periodic finite Radon measures is denoted by $\mathcal{M}(\mathbb{T}_M)$. It is a Banach space equipped with the total-variation norm

$$\|w\|_{\mathrm{TV}} \stackrel{\Delta}{=} \sup_{\substack{\varphi\in\mathcal{S}(\mathbb{T}_M)\\ \|\varphi\|_\infty=1}} \langle w,\varphi\rangle. \tag{4.52}$$

The first element we introduce in our problem formulation is $\mathrm{L}=\mathrm{D}^{\alpha+1}$, the derivative operator whose order $(\alpha+1)$, with $\alpha\in\mathbb{N}\setminus\{0\}$, determines the smoothness of the components of the constructed curve. The native space associated to the pair $(\mathrm{L},\mathcal{M}(\mathbb{T}_M))$ is defined as $\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)=\{f\in\mathcal{S}'(\mathbb{T}_M):\|\mathrm{L}\{f\}\|_{\mathrm{TV}}<+\infty\}$. It has been shown that $\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$ is isometrically isomorphic to $\mathcal{M}_0(\mathbb{T}_M)\times\mathbb{R}$, where $\mathcal{M}_0(\mathbb{T}_M)=\{w\in\mathcal{M}(\mathbb{T}_M):\langle w,1\rangle=0\}$ is the space of Radon measures with zero mean [323]. The explicit form of such an isometry between spaces (and its inverse) is given by

$$\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)\to\mathcal{M}_0(\mathbb{T}_M)\times\mathbb{R}: f\mapsto(\mathrm{L}\{f\},\langle f,1\rangle), \tag{4.53}$$

$$\mathcal{M}_0(\mathbb{T}_M)\times\mathbb{R}\to\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M):(w,a)\mapsto\mathrm{L}^\dagger\{w\}+a, \tag{4.54}$$

where $\mathrm{L}^\dagger$ is the pseudoinverse of L. Finally, we note that the Green's function of $\mathrm{L}=\mathrm{D}^{(\alpha+1)}$, defined as $g_{\mathrm{L}}=\mathrm{L}^\dagger\{Ш\}$, is a continuous periodic function for all integers $\alpha\geq 1$ [323].

Next, we define the periodic L-splines with respect to the operator L. A periodic L-spline is a function $s:\mathbb{T}_M\to\mathbb{R}$ that verifies that

$$\mathrm{L}\{s\}(t)=\sum_{k=0}^{K-1}a[k]Ш_M(t-t_k), \tag{4.55}$$

where $Ш_M(t)=\sum_{k\in\mathbb{Z}}\delta(t-Mk)\in\mathcal{S}'(\mathbb{T}_M)$ is the M-periodic Dirac comb, $K\in\mathbb{N}\setminus\{0\}$ is the number of knots, $a[k]\in\mathbb{R}$ is the amplitude of the kth jump, and $t_k\in\mathbb{R}$ are distinct knot locations.

The second element is $\mathcal{R}$, a sparsity-promoting regularization functional with key characteristics. Firstly, for 2D curves, the minimization of $\mathcal{R}(\mathrm{L}\{\mathbf{r}\})$, where

$\mathrm{L}\{\mathbf{r}\} = (\mathrm{L}\{x\}, \mathrm{L}\{y\})$, should enforce sparsity jointly for the two components of $\mathbf{r}$. Indeed, we want $\mathbf{r}$ to have few variations, and they often should occur along both components simultaneously. Secondly, $\mathcal{R}$ should be invariant to a rotation of the system of coordinates. Similarly, $\mathcal{R}$ should be equivariant to isotropic scaling, meaning that there exists a function A such that $\mathcal{R}(\mathrm{L}\{a\mathbf{r}\}) = \mathrm{A}(a)\ \mathcal{R}(\mathrm{L}\{\mathbf{r}\})$ for any $a \neq 0$. We now introduce the RI-TV norm, which consists in a mixed continuous (TV-ℓ_2) norm and satisfies our specifications.

Definition 4.1. *Let $p \in [1, +\infty]$. The (TV-ℓ_p) norm of any vector-valued tempered distribution $\mathbf{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \in \mathcal{S}'(\mathbb{T})^2$ is defined as*

$$\|\mathbf{w}\|_{\mathrm{TV}-\ell_p} \triangleq \sup_{\substack{\boldsymbol{\varphi}=(\varphi_1,\varphi_2)\in\mathcal{S}(\mathbb{T}_M)^2 \\ \|\boldsymbol{\varphi}\|_{q,\infty}=1}} \left(\langle w_1, \varphi_1\rangle + \langle w_2, \varphi_2\rangle\right), \tag{4.56}$$

where $q \in [1, \infty]$ is the Hölder conjugate of p with $\frac{1}{p} + \frac{1}{q} = 1$ and $\|\cdot\|_{q,\infty}$ is the $(\ell_q - L_\infty)$ mixed norm, defined for any $\boldsymbol{\varphi} \in \mathcal{S}(\mathbb{T}_M)^2$ as

$$\|\boldsymbol{\varphi}\|_{q,\infty} \triangleq \sup_{t\in\mathbb{T}_M} \|\boldsymbol{\varphi}(t)\|_q. \tag{4.57}$$

A mixed norm similar to (4.56) was previously introduced in [324] in the context of the recovery of Dirac distributions. In Theorem 4.3, we compute the (TV-ℓ_p) norm for two general classes of functions or distributions.

Theorem 4.3. *1. For any curve $\mathbf{f} = (f_1, f_2)$ with absolutely integrable components $f_i \in L_1(\mathbb{T}_M)$, $i = 1, 2$, we have that*

$$\|[f_1 \quad f_2]\|_{\mathrm{TV}-\ell_p} = \int_0^M \|\mathbf{f}(t)\|_p \mathrm{d}t. \tag{4.58}$$

2. Let $\mathbf{w} = (w_1, w_2)$ be a vector-valued distribution of the form $\mathbf{w} = \sum_{k=1}^K \mathbf{a}[k]\text{Ш}_M(\cdot - t_k)$ with $\mathbf{a}[k] \in \mathbb{R}^2, k = 0, \ldots, K-1$. Then, we have that

$$\|[w_1 \quad w_2]\|_{\mathrm{TV}-\ell_p} = \sum_{k=0}^{K-1} \|\mathbf{a}[k]\|_p. \tag{4.59}$$

Proof. **Item** 1: Let $\boldsymbol{\varphi} = (\varphi_1, \varphi_2) \in \mathcal{S}(\mathbb{T}_M)^2$ be an arbitrary smooth curve with $\|\boldsymbol{\varphi}\|_{q,\infty} = 1$. On the one hand, the Hölder inequality for vectors implies that, for any $t \in \mathbb{T}_M$,

$$|f_1(t)\varphi_1(t) + f_2(t)\varphi_2(t)| \leq \|\mathbf{f}(t)\|_p \|\boldsymbol{\varphi}(t)\|_q \leq \|\mathbf{f}(t)\|_p, \tag{4.60}$$

where the last inequality is due to $\|\boldsymbol{\varphi}\|_{q,\infty} = 1$. On the other hand, the inclusion $f_i \in L_1(\mathbb{T}_M)$ allows us to express the duality product $\langle f_i, \varphi_i\rangle$ as a simple integral of the form

$$\langle f_i, \varphi_i\rangle = \int_0^M f_i(t)\varphi_i(t)\mathrm{d}t, \quad i = 1, 2. \tag{4.61}$$

Combining (4.61) with (4.60), we obtain that

$$\langle f_1, \varphi_1\rangle + \langle f_2, \varphi_2\rangle = \int_0^M (f_1(t)\varphi_1(t) + f_2(t)\varphi_2(t))\,\mathrm{d}t \leq \int_0^M \|\mathbf{f}(t)\|_p \mathrm{d}t. \tag{4.62}$$

Taking the supremum over all $\boldsymbol{\varphi} \in \mathcal{S}(\mathbb{T}_M)^2$ with $\|\boldsymbol{\varphi}\|_{q,\infty} = 1$ then yields that $\|[f_1 \;\; f_2]\|_{\mathrm{TV}-\ell_p} \leq \int_0^M \|\mathbf{f}(t)\|_p \mathrm{d}t$. To prove the equality, we first define the functions

$$g_i : \mathbb{T}_M \to \mathbb{R} : t \mapsto \mathbb{1}_{\mathbf{f}(t)\neq\mathbf{0}} \frac{\mathrm{sgn}(f_i(t))|f_i(t)|^{(p-1)}}{\|\mathbf{f}(t)\|_p^{(p-1)}}, \tag{4.63}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A. We note that $g_i, i = 1, 2$ are Borel-measurable with $\|g_i\|_{L_\infty} \leq 1$ for $i = 1, 2$. Further, one readily verifies that

$$\int_0^M (f_1(t)g_1(t) + f_2(t)g_2(t))\,\mathrm{d}t = \int_0^M \|\mathbf{f}(t)\|_p \mathrm{d}t. \tag{4.64}$$

By invoking a variant of Lusin's theorem (see [259, Theorem 7.10]) on the space $\mathcal{C}(\mathbb{T}_M)$ of M-periodic continuous functions, we then consider the ϵ-approximations $g_{i,\epsilon} \in \mathcal{C}(\mathbb{T}_M)$ of g_i such that $\|g_{i,\epsilon}\|_{L_\infty} \leq \|g_i\|_{L_\infty} \leq 1$, and $\int_E |f_i(t)|\mathrm{d}t \leq \epsilon/8, i = 1, 2$, where $E = \{t \in \mathbb{T}_M : g_{i,\epsilon}(t) \neq g_i(t)\}$. This in effect implies that

$$\int_0^M |f_i(t)| \cdot |g_{i,\epsilon}(t) - g_i(t)|\,\mathrm{d}t = \int_E |f_i(t)| \cdot |g_{i,\epsilon}(t) - g_i(t)|\,\mathrm{d}t \leq \|\mathbb{1}_E f_i\|_{L_1} \|g_{i,\epsilon} - g_i\|_{L_\infty} \leq \frac{\epsilon}{4}. \tag{4.65}$$

We then invoke the denseness of $\mathcal{S}(\mathbb{T}_M)$ in $\mathcal{C}(\mathbb{T}_M)$ to deduce the existence of $\varphi_{i,\epsilon} \in \mathcal{S}(\mathbb{T}_M)$ with $\|g_{i,\epsilon} - \varphi_{i,\epsilon}\|_{L_\infty} \leq \frac{\epsilon}{4\|f_i\|_{L_1}}$. This gives us the upper-bound

$$\int_0^M |f_i(t)| \cdot |\varphi_{i,\epsilon}(t) - g_{i,\epsilon}(t)| \,\mathrm{d}t \leq \|f_i\|_{L_1} \|\varphi_{i,\epsilon} - g_{i,\epsilon}\|_{L_\infty} \leq \frac{\epsilon}{4}. \tag{4.66}$$

Next, we use the triangle inequality to obtain the lower-bound

$$\begin{aligned}\langle f_i, \varphi_{i,\epsilon}\rangle &\geq \int_0^M f_i(t) g_i(t) \mathrm{d}t - \int_0^M |f_i(t)| \cdot |g_i(t) - g_{i,\epsilon}(t)| \mathrm{d}t - \int_0^M |f_i(t)| \cdot |g_{i,\epsilon}(t) - \varphi_{i,\epsilon}(t)| \mathrm{d}t \\ &\geq \int_0^M f_i(t) g_i(t) \mathrm{d}t - \frac{\epsilon}{2}, \quad i = 1, 2, \end{aligned} \tag{4.67}$$

where the last inequality follows the combination of (4.65) and (4.66). Finally, we use (4.64) to conclude that

$$\|[f_1 \;\; f_2]\|_{\mathrm{TV}-\ell_p} \geq \frac{\langle f_1, \varphi_{1,\epsilon}\rangle + \langle f_2, \varphi_{2,\epsilon}\rangle}{\|(\varphi_{1,\epsilon}, \varphi_{2,\epsilon})\|_{q,\infty}} \geq \frac{\int_0^M \|\mathbf{f}(t)\|_p \mathrm{d}t - \epsilon}{1 + O(\epsilon)}. \tag{4.68}$$

We complete the proof by letting $\epsilon \to 0$.

Item 2: Similarly to the previous part, for any smooth curve $\boldsymbol{\varphi} = (\varphi_1, \varphi_2) \in \mathcal{S}(\mathbb{T}_M)^2$ with $\|\boldsymbol{\varphi}\|_{q,\infty} = 1$, we have that

$$\langle w_1, \varphi_1\rangle + \langle w_2, \varphi_2\rangle = \sum_{k=0}^{K-1} (a_1[k]\varphi_1(t_k) + a_2[k]\varphi_2(t_k)) \leq \sum_{k=0}^{K-1} \|\mathbf{a}[k]\|_p \|\boldsymbol{\varphi}(t_k)\|_q \leq \sum_{k=0}^{K-1} \|\mathbf{a}[k]\|_p \tag{4.69}$$

Taking the supremum over $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ with $\|\boldsymbol{\varphi}\|_{q,\infty} = 1$ then yields that $\|\mathbf{w}\|_{\mathrm{TV}-\ell_p} \leq \sum_{k=0}^{K-1} \|\mathbf{a}[k]\|_p$. To prove the equality, we first define a set of vectors $\boldsymbol{\varphi}_k \in \mathbb{R}^2$ such that $\|\boldsymbol{\varphi}_k\|_\infty = 1$ and $\mathbf{a}[k]^T \boldsymbol{\varphi}_k = \|\mathbf{a}[k]\|_p$ for $k = 0, \dots, K-1$. We then consider a smooth curve $\boldsymbol{\varphi}^* \in \mathcal{S}(\mathbb{T}_M)^2$ with $\|\boldsymbol{\varphi}^*\|_{q,\infty} = 1$ such that $\boldsymbol{\varphi}^*(t_k) = \boldsymbol{\varphi}_k$. Using this, we then verify that

$$\|\mathbf{w}\|_{\mathrm{TV}-\ell_p} \geq \langle w_1, \varphi_1^* \rangle + \langle w_2, \varphi_2^* \rangle = \sum_{k\in\mathbb{Z}} \|\mathbf{a}[k]\|_p. \tag{4.70}$$

□

The outer TV norm promotes sparsity, as it is the continuous counterpart of the ℓ_1 norm [166]. The inner ℓ_p norm in Item 1 induces a coupling of the f_1 and f_2 components. Indeed, it first aggregates the f_1 and f_2 curve components, which the outer TV norm then jointly sparsifies. This is true of any ℓ_p norm for $p \neq 1$. For $p = 1$, the components are no longer coupled due to the separability of the norm. For $p = 2$ and any curve $\mathbf{f} = (f_1, f_2)$, we set

$$\mathcal{R}(\mathbf{f}) = \|[f_1 \quad f_2]\|_{\mathrm{TV}-\ell_2}\,. \tag{4.71}$$

Proposition 4.3. *The (TV-ℓ_2) norm, noted $\mathcal{R}$, is invariant to rotation, in the sense that $\mathcal{R}(\mathbf{R}_\theta \mathbf{f}) = \mathcal{R}(\mathbf{f})$, where $\mathbf{R}_\theta$ is a rotation matrix. Furthermore, the (TV-ℓ_2) norm is the only (TV-ℓ_p) norm that is rotation invariant.*

Proof. By substitution of $\mathbf{R}_\theta \mathbf{f}$ in (4.56), we have that

$$\begin{aligned}\mathcal{R}(\mathbf{R}_\theta \mathbf{f}) &= \sup_{\substack{\boldsymbol{\varphi}\in\mathcal{S}(\mathbb{T}_M)^2\\ \|\boldsymbol{\varphi}\|_{2,\infty}=1}} \left(\langle \cos(\theta) f_1 - \sin(\theta) f_2, \varphi_1\rangle + \langle \sin(\theta) f_1 + \cos(\theta) f_2, \varphi_2\rangle\right)\\ &= \sup_{\substack{\boldsymbol{\varphi}\in\mathcal{S}(\mathbb{T}_M)^2\\ \|\boldsymbol{\varphi}\|_{2,\infty}=1}} \left(\langle f_1, \cos(\theta)\varphi_1 + \sin(\theta)\varphi_2\rangle + \ \langle f_2, -\sin(\theta)\varphi_1 + \cos(\theta)\varphi_2\rangle\right).\end{aligned} \tag{4.72}$$

We perform the change of variable $\boldsymbol{\psi} = \mathbf{R}_{-\theta}\boldsymbol{\varphi}$. We readily conclude that, since $\mathbf{R}_{-\theta}$ is bijective over $\mathcal{S}(\mathbb{T}_M)^2$, for any $\boldsymbol{\varphi} \in \mathcal{S}(\mathbb{T}_M)^2$, we have that $\boldsymbol{\psi} = \mathbf{R}_{-\theta}\boldsymbol{\varphi} \in$

$\mathcal{S}(\mathbb{T}_M)^2$. Additionally, and in accordance with (4.57), we have that

$$\|\boldsymbol{\psi}\|_{2,\infty} = \sup_{t\in\mathbb{T}_M} \|\boldsymbol{\psi}(t)\|_2 = \sup_{t\in\mathbb{T}_M} \|\mathbf{R}_{-\theta}\boldsymbol{\varphi}(t)\|_2 = \sup_{t\in\mathbb{T}_M} \|\boldsymbol{\varphi}(t)\|_2, \tag{4.73}$$

since $\mathbf{R}_{-\theta}$ is an isometry. Hence, it does not change the ℓ_2 norm of a vector. Consequently, we have that

$$\mathcal{R}(\mathbf{R}_\theta\mathbf{f}) = \sup_{\substack{\boldsymbol{\psi}\in\mathcal{S}(\mathbb{T}_M)^2 \\ \|\boldsymbol{\psi}\|_{2,\infty}=1}} (\langle f_1, \psi_1\rangle + \langle f_2, \psi_2\rangle) = \mathcal{R}(\mathbf{f}). \tag{4.74}$$

Moreover, according to Item 1 of Theorem 4.3, for any curve $\mathbf{f} = (f_1, f_2)$ with absolutely integrable components $f_i \in L_1(\mathbb{T}_M),\ i = 1, 2$, the $\mathrm{TV} - \ell_p$ norm is

$$\|\mathbf{f}\|_{\mathrm{TV}-\ell_p} = \int_0^M (|f_1(t)|^p + |f_2(t)|^p)^{\frac{1}{p}}\ \mathrm{d}t. \tag{4.75}$$

We take $f_1(t) = 1,\ f_2(t) = 0$, and $\theta = \frac{\pi}{4}$. This gives us

$$\|\mathbf{f}\|_{\mathrm{TV}-\ell_p} = \int_0^M (|1|^p + |0|^p)^{\frac{1}{p}}\ \mathrm{d}t = M. \tag{4.76}$$

When applying the planar rotation $\mathbf{R}_\theta$ to the curve $\mathbf{f}$, we have that

$$\begin{aligned}\|\mathbf{R}_\theta\mathbf{f}\|_{\mathrm{TV}-\ell_p} &= \int_0^M (|f_1(t)\cos(\theta) - f_2(t)\sin(\theta)|^p + |f_1(t)\sin(\theta) + f_2(t)\cos(\theta)|^p)^{\frac{1}{p}}\ \mathrm{d}t \\ &= \int_0^M (|\cos(\theta)|^p + |\sin(\theta)|^p)^{\frac{1}{p}}\ \mathrm{d}t \qquad (4.77) \\ &= \int_0^M \left(2\left|\frac{\sqrt{2}}{2}\right|^p\right)^{\frac{1}{p}}\ \mathrm{d}t = 2^{\frac{1}{p}-\frac{1}{2}}M. \qquad (4.78)\end{aligned}$$

We conclude that $\|\mathbf{f}\|_{\mathrm{TV}-\ell_p} = \|\mathbf{R}_\theta\mathbf{f}\|_{\mathrm{TV}-\ell_p}$ if and only if $p = 2$, which proves that the $\mathrm{TV} - \ell_p$ norm is not rotation invariant for $p \neq 2$. □

Our proposed minimization problem consists of two terms. The first one—the data-fidelity term—ensures that the candidate curve $\mathbf{r}(t)$ is close to the points $\mathbf{p}[m]$. The second one, called regularization, introduces our *a priori* desiderata for the reconstructed curve. The importance of these two terms is weighted by a parameter $\lambda > 0$. The solution set of the minimization problem is

$$\mathcal{V} = \underset{\mathbf{r}\in\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)}{\arg\min} \left(\sum_{m=0}^{M-1} \| \mathbf{r}(t)|_{t=m} - \mathbf{p}[m]\|_2^2 + \lambda\mathcal{R}(\mathrm{L}\{\mathbf{r}\}) \right), \tag{4.79}$$

where the search space $\mathcal{X}_{\mathrm{L}}$ is defined as

$$\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M) = \{\mathbf{r} \in \mathcal{S}'(\mathbb{T}_M)^2 : \mathcal{R}(\mathrm{L}\{\mathbf{r}\}) < +\infty\}. \tag{4.80}$$

In the following proposition, we characterize the topological structure of the search space.

Proposition 4.4. *The search space $\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$ can be expressed as*

$$\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M) = \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M) \times \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M). \tag{4.81}$$

Moreover, the mapping

$$T_{\mathrm{L}} : \mathcal{X}_{\mathrm{L}}(\mathbb{T}_M) \to \mathcal{M}_0(\mathbb{T}_M)^2 \times \mathbb{R}^2 \tag{4.82}$$

$$T_{\mathrm{L}}(\mathbf{r}) = (\mathrm{L}\{r_1\}, \mathrm{L}\{r_2\}, \langle r_1, 1\rangle, \langle r_2, 1\rangle) \tag{4.83}$$

is an isomorphism between $\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$ and $\mathcal{M}_0(\mathbb{T}_M)^2 \times \mathbb{R}^2$ whose inverse is

$$T_{\mathrm{L}}^{-1} : \mathcal{M}_0(\mathbb{T}_M)^2 \times \mathbb{R}^2 \to \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M) \tag{4.84}$$

$$T_{\mathrm{L}}^{-1}(\mathbf{w}, \mathbf{a}) = \left(\mathrm{L}^\dagger\{w_1\} + a_1, \mathrm{L}^\dagger\{w_2\} + a_2\right). \tag{4.85}$$

Proof. Let $\mathbf{r} = (r_1, r_2) \in \mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$. We have that

$$
\begin{aligned}
\mathcal{R}\left(\mathrm{L}\{\mathbf{r}\}\right) &= \sup_{\substack{\boldsymbol{\varphi}\in\mathcal{S}(\mathbb{T}_M)^2\\ \|\boldsymbol{\varphi}\|_{2,\infty}=1}} \left(\langle \mathrm{L}\{r_1\},\varphi_1\rangle + \langle \mathrm{L}\{r_2\},\varphi_2\rangle\right) && (4.86)\\
&\geq \sup_{\substack{\varphi_1\in\mathcal{S}(\mathbb{T}_M)\\ \|(\varphi_1,0)\|_{2,\infty}=1}} \langle \mathrm{L}\{r_1\},\varphi_1\rangle = \sup_{\substack{\varphi_1\in\mathcal{S}(\mathbb{T}_M)\\ \|\varphi_1\|_{\infty}=1}} \langle \mathrm{L}\{r_1\},\varphi_1\rangle && (4.87)\\
&= \|\mathrm{L}\{r_1\}\|_{\mathrm{TV}}, && (4.88)
\end{aligned}
$$

from which we deduce that $r_1 \in \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$. Similarly, we get that $r_2 \in \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$ and, hence, we have that $\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M) \subseteq (\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M))^2$. For the reverse inclusion, let $r_1, r_2 \in \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$. Using the inequalities $\|\boldsymbol{\varphi}\|_{2,\infty} \geq \|\varphi_i\|_\infty$ for $i = 1, 2$, we deduce that

$$
|\langle \mathrm{L}\{r_i\},\varphi_i\rangle| \leq \|\mathrm{L}\{r_i\}\|_{\mathrm{TV}}\|\varphi_i\|_\infty \leq \|\mathrm{L}\{r_i\}\|_{\mathrm{TV}}\|\boldsymbol{\varphi}\|_{2,\infty}. \qquad (4.89)
$$

Hence, we have that

$$
\langle \mathrm{L}\{r_1\},\varphi_1\rangle + \langle \mathrm{L}\{r_2\},\varphi_2\rangle \leq \left(\|\mathrm{L}\{r_1\}\|_{\mathrm{TV}} + \|\mathrm{L}\{r_2\}\|_{\mathrm{TV}}\right)\|\boldsymbol{\varphi}\|_{2,\infty}, \qquad (4.90)
$$

which implies that

$$
\mathcal{R}\left(\mathrm{L}\{\mathbf{r}\}\right) \leq \|\mathrm{L}\{r_1\}\|_{\mathrm{TV}} + \|\mathrm{L}\{r_2\}\|_{\mathrm{TV}} < +\infty. \qquad (4.91)
$$

Hence, we have the inclusion $\mathbf{r} \in \mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$.

Following (4.91) and (4.88), we deduce that the norm topology of $\mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$ is equivalent to the product topology induced from $\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M) \times \mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$. This, together with the fact that $\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$ is isometrically isomorphic to $\mathcal{M}_0(\mathbb{T}_M) \times \mathbb{R}$, implies that T_{L} is an isomorphism. Its inverse is readily deduced from (4.54). □

The data-fidelity term in (4.79) penalizes the Euclidean distance between the sample $\mathbf{r}(t)|_{t=m}$ of the curve and the point $\mathbf{p}[m]$ for every $m = 0, \dots, M-1$. The fact that $\mathbf{r}$ is sampled uniformly along the parameter axis encourages the reconstructed curve to be parametrized by its curvilinear abscissa, promoting the arc length of $\mathbf{r}(t)$ to be a linear function of the parameter t. The underlying assumption behind this statement is that the points $\mathbf{p}[m]$ are spread approximately uniformly along the curve. This is an important assumption, since the regularization term in (4.79) involves the derivatives of $\mathbf{r}(t)$ and thus heavily depends on the choice of parametrization. In that respect, the curvilinear abscissa is a desirable choice. In practice, it often results in rough curves being penalized heavily by our regularization, which other parametrizations may fail to achieve [316].

Our representer theorem specifies the form of the solution of (4.79).

Theorem 4.4. *The solution set* (4.79) *is nonempty, convex, and weak*-compact. Moreover, any extreme point* $\mathbf{r}^*$ *of* $\mathcal{V}$ *is a periodic* L*-spline curve* $\mathbf{r}^*$ *with at most* $K \leq 2M + 2$ *knots. Indeed, we have that*

$$\mathrm{L}\{\mathbf{r}^*\} = \sum_{k=0}^{K-1} \mathbf{a}_k \text{Ш}_M(\cdot - t_k) \tag{4.92}$$

for some distinct knot locations $t_k \in \mathbb{T}_M$ *and amplitude vectors* $\mathbf{a}_k \in \mathbb{R}^2$.

Our strategy to characterize the solution set $\mathcal{V}$ defined in (4.79) consists in invoking the main result of Boyer *et al.* [76], which requires the knowledge of the form of extreme points of the unit ball of the regularization functional. To that end, we prove that the extreme points of the RI-TV unit ball are vector-valued Dirac combs.

Proposition 4.5. *An element* $\mathbf{w}^* \in \mathcal{M}(\mathbb{T}_M)^2$ *is an extreme point of the RI-TV unit ball* $B = \{\mathbf{w} \in \mathcal{M}(\mathbb{T}_M)^2 : \mathcal{R}(\mathbf{w}) \leq 1\}$ *if and only if it is a vector-valued Dirac comb of the form* $\mathbf{w}^* = \mathbf{a}\text{Ш}_M(\cdot - t_0)$ *for some* $t_0 \in \mathbb{T}_M$ *and* $\mathbf{a} \in \mathbb{R}^2$ *with* $\|\mathbf{a}\|_2 = 1$.

Proof. Assume by contradiction that there exists an extreme point $\mathbf{w}^*$ of B that is not a Dirac comb. This implies that there exists an interval $I \subseteq \mathbb{T}_M$ such that $\mathbf{w}_1 = \mathbf{w}^* \mathbb{1}_I$ and $\mathbf{w}_2 = \mathbf{w}^* \mathbb{1}_{I^c}$ are both nonzero Radon measures that satisfy $\mathbf{w}^* = \mathbf{w}_1 + \mathbf{w}_2$.

Due to their disjoint support, we have that $\mathcal{R}(\mathbf{w}^*) = \mathcal{R}(\mathbf{w}_1) + \mathcal{R}(\mathbf{w}_2)$. Let us now define the measures $\mathbf{w}_+ = (1+\epsilon)\mathbf{w}_1 + (1-\delta)\mathbf{w}_2$ and $\mathbf{w}_- = (1-\epsilon)\mathbf{w}_1 + (1+\delta)\mathbf{w}_2$, where $\epsilon, \delta > 0$ are small constants such that $\epsilon\mathcal{R}(\mathbf{w}_1) = \delta\mathcal{R}(\mathbf{w}_2)$. By observing that $\mathcal{R}(\mathbf{w}_+) = \mathcal{R}(\mathbf{w}_-) = 1$ and $\mathbf{w}^* = \frac{\mathbf{w}_+ + \mathbf{w}_-}{2}$, we conclude that $\mathbf{w}^*$ is not an extreme point of B, which yields a contradiction. Hence, the extreme points of B can only be vector-valued Dirac combs.

To prove the reverse inclusion, let $\mathbf{w}^* = \mathbf{a}\text{Ш}_M(\cdot - t_0)$ with $\|\mathbf{a}\|_2 = 1$. We now prove that $\mathbf{w}^*$ is an extreme point of B. Assume that there exist $\mathbf{w}_1, \mathbf{w}_2 \in B$ such that $\mathbf{w}^* = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2)$. Let us define the measure $\mathbf{w}_0 = \mathbf{w}_1 \mathbb{1}_{t \neq t_0} \in \mathcal{M}(\mathbb{T}_M)^2$ so that $\mathbf{w}_1 = \mathbf{w}_0 + \mathbf{a}_1\text{Ш}_M(\cdot - t_0)$ for some $\mathbf{a}_1 \in \mathbb{R}^2$. We then must have $\mathbf{w}_2 = (-\mathbf{w}_0) + \mathbf{a}_2\text{Ш}_M(\cdot - t_0)$ with $\mathbf{a} = \frac{1}{2}(\mathbf{a}_1 + \mathbf{a}_2)$. The construction implies that

$$1 = \mathcal{R}(\mathbf{w}_i) = \mathcal{R}(\mathbf{w}_0) + \|\mathbf{a}_i\|_2 \geq \|\mathbf{a}_i\|_2, i = 1, 2. \tag{4.93}$$

This, together with the triangle inequality, yields

$$2 = \|2\mathbf{a}\|_2 \leq \|\mathbf{a}_1\|_2 + \|\mathbf{a}_2\|_2 \leq 1 + 1 = 2. \tag{4.94}$$

Hence, all inequalities must be saturated. In particular, we must have that $\mathbf{w}_0 = \mathbf{0}$ and $\|\mathbf{a}\|_2 = \frac{1}{2}(\|\mathbf{a}_1\|_2 + \|\mathbf{a}_2\|_2)$. Finally, we invoke the strict convexity of the ℓ_2 norm to conclude that $\mathbf{a} = \mathbf{a}_1 = \mathbf{a}_2$ and, thus, that $\mathbf{w}_1 = \mathbf{w}_2$, which in turn implies that $\mathbf{w}^*$ is an extreme point of B. □

Proof of Theorem 4.4. Let us define the cost functional $E : \mathcal{M}(\mathbb{T}_M)^2 \times \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ as

$$E(\mathbf{w}, \mathbf{a}) = \sum_{m=0}^{M-1} \|\boldsymbol{\nu}_m(\mathbf{w}) + \mathbf{a} - \mathbf{p}[m]\|_2^2 + \sum_{i=1}^{2} \chi_{\langle w_i, 1\rangle = 0}, \tag{4.95}$$

where χ_A denotes the characteristic function of the set A, and $\boldsymbol{\nu}_m = (\nu_{m,1}, \nu_{m,2})$ with

$$\nu_{m,i}(\mathbf{w}) = \left(\mathrm{L}^{\dagger}\{w_i\}(t)\right)\big|_{t=m} = \langle \mathrm{L}^{\dagger}\{w_i\}, Ш(\cdot - m)\rangle = \langle w_i, \mathrm{L}^{\dagger *}\{Ш(\cdot - m)\}\rangle = \langle w_i, g_{\mathrm{L}}(m - \cdot)\rangle \tag{4.96}$$

for $i = 1, 2$. We note that $\boldsymbol{\nu}_m$ is weak*-continuous in the topology of $\mathcal{M}_{\mathrm{L}}(\mathbb{T}_M)$ due to the inclusion $g_{\mathrm{L}} \in \mathcal{C}(\mathbb{T}_M)$.

Then, we formulate a minimization problem that admits the solution set

$$\tilde{\mathcal{V}} = \underset{\substack{\mathbf{w}\in\mathcal{M}(\mathbb{T}_M)^2\\ \mathbf{a}\in\mathbb{R}^2}}{\arg\min}\ \mathcal{J}(\mathbf{w}, \mathbf{a}), \tag{4.97}$$

where $\mathcal{J}(\mathbf{w}, \mathbf{a}) = E(\mathbf{w}, \mathbf{a}) + \lambda\mathcal{R}(\mathbf{w})$. We now characterize $\tilde{\mathcal{V}}$ by invoking Proposition 4.5 and Theorem 2.5 which, together, imply that $\tilde{\mathcal{V}}$ is a nonempty, convex and weak*-compact set such that for any of its extreme points $(\mathbf{w}^*, \mathbf{a}_0)$, we have that $\mathbf{w}^* = \sum_{k=0}^{K-1} \mathbf{a}_k Ш(\cdot - t_k)$ with $K \leq 2M + 2$ (the total number of linear measurements in (4.97)) for some $\mathbf{a}_k \in \mathbb{R}^2$ and $t_k \in \mathbb{T}_M$. The final step is to observe that the isomorphism T_{L} defined in Proposition 4.4 allows us to write

$$E(T_{\mathrm{L}}(\mathbf{r})) = \sum_{m=0}^{M-1} \|\, \mathbf{r}(t)|_{t=m} - \mathbf{p}[m]\|_2^2 \tag{4.98}$$

for any $\mathbf{r} \in \mathcal{X}_{\mathrm{L}}(\mathbb{T}_M)$, from which we conclude that $\tilde{\mathcal{V}} = T_{\mathrm{L}}(\mathcal{V})$, which yields the announced characterization. □

Theorem 4.4 states that the solution set (4.79) contains periodic L-splines. Even though our work uses a mixed (TV-ℓ_2) norm as regularization, this result is reminiscent of [284, 73, 323], which proves that inverse problems with TV regularization have spline solutions. However, not all curves can be faithfully represented with a single type of spline. We propose to cater to this by modeling our closed function as a sum of two components $\mathbf{r}(t) = \mathbf{r}_1(t) + \mathbf{r}_2(t)$. Similarly to the non-hybrid setting, we have M points $\mathbf{p}[m] = (\mathrm{p}_x[m],\ \mathrm{p}_y[m])$, $m = 0, \dots, M-1$. Hence, we again have that $\mathbf{r}$ is M-periodic with $t \in \mathbb{T}_M$. Following the formulation for one-dimensional signals in Section 4.2 and extending it to two dimensions, we consider continuous problems of the form

$$\underset{\substack{\mathbf{r}_i \in \mathcal{X}_{\mathrm{L}_i}(\mathbb{T}_M) \\ \mathbf{r}_1(0)=\mathbf{0}}}{\arg\min} \sum_{m=0}^{M-1} \left\| \mathbf{r}_1(t)|_{t=m} + \mathbf{r}_2(t)|_{t=m} - \mathbf{p}[m] \right\|_2^2 + \lambda_1 \left\| \mathrm{L}_1\{\mathbf{r}_1\} \right\|_{\mathrm{TV}-\ell_2} + \lambda_2 \left\| \mathrm{L}_2\{\mathbf{r}_2\} \right\|_{\mathrm{TV}-\ell_2}, \tag{4.99}$$

where $\lambda_1,\ \lambda_2 > 0$ are the two regularization parameters weighting the two regularization terms, and L_1 and L_2 are derivative operators of different orders. The constraint $\mathbf{r}_1(0) = \mathbf{0}$ is necessary to handle the ill-posedness of the problem. Indeed, without this constraint, for any solution $(\mathbf{r}_1(t), \mathbf{r}_2(t))$ of Problem (4.99), the pair $(\mathbf{r}_1 + \mathbf{v}, \mathbf{r}_2 - \mathbf{v})$, where $\mathbf{v}$ is an arbitrary constant vector, would clearly also be a solution. This implies that the solution set would be unbounded, which can be problematic for numerical implementations. The constraint $\mathbf{r}_1(0) = \mathbf{0}$ resolves this issue without any restriction on the reconstructed curve, since any constant offset can be included in the $\mathbf{r}_2$ component. See Section 4.2 for more details concerning this question and implementation details.

Theorem 4.5. *There exists a global minimizer* $\mathbf{r}^*$ *of* (4.99) *that can be decomposed as* $\mathbf{r}^* = \mathbf{r}_1^* + \mathbf{r}_2^*$*, where* $\mathbf{r}_i^*$ *is a periodic* L_i*-spline (see,* (4.55)*) with* K_i *knots,* $i = 1, 2$*. Moreover, we have the bound* $K_1 + K_2 \leq 2M + 2$ *for the total number of knots of* $\mathbf{r}^*$*.*

Proof. By invoking Proposition 4.4, we deduce that there is a bijection between $\mathcal{V}_{\mathrm{hyb}}$ and the solution set

$$\tilde{\mathcal{V}}_{\mathrm{hyb}} = \underset{\substack{\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{M}_0(\mathbb{T})^2 \\ \mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2}}{\arg\min}\ E(\mathbf{w}_1, \mathbf{a}_1, \mathbf{w}_2, \mathbf{a}_2) + \lambda_1 \mathcal{R}(\mathbf{w}_1) + \lambda_2 \mathcal{R}(\mathbf{w}_2), \tag{4.100}$$

where the data fidelity cost $E : \left(\mathcal{M}(\mathbb{T})^2 \times \mathbb{R}^2\right)^2 \to \mathbb{R}_{\geq 0}$ satisfies

$$E(T_{\mathrm{L}_1}(\mathbf{r}_1), T_{\mathrm{L}_1}(\mathbf{r}_2)) = \sum_{m=0}^{M-1} \left\| \mathbf{r}_1(t)|_{t=m} + \mathbf{r}_2(t)|_{t=m} - \mathbf{p}[m] \right\|_2^2. \tag{4.101}$$

This implies that there is a bijection between $\mathcal{V}_{\mathrm{hyb}}$ and $\tilde{\mathcal{V}}_{\mathrm{hyb}}$. The last step is to note that for any extreme point $(\mathbf{w}_1^*, \mathbf{w}_2^*)$ of the unit ball $\{(\mathbf{w}_1, \mathbf{w}_2) \in \mathcal{M}(\mathbb{T})^4 : \lambda_1 \mathcal{R}(\mathbf{w}_1) + \lambda_2 \mathcal{R}(\mathbf{w}_2) \leq 1\}$, we have that $\mathbf{w}_1^* = \mathbf{0}$ or $\mathbf{w}_2^* = \mathbf{0}$. This together with Theorem 2.5 concludes the proof. □

4.4.3 Exact Discretization

In this section, we discuss our method of discretizing the original problem. For the sake of simplicity, we only focus on the single setting; the hybrid formalism is the direct extension. As suggested by Theorem 4.4, we take the stance of recasting the continuous-domain problem in (4.79) as a finite-dimensional optimization problem by restricting the search space to periodic L-splines with knots on a uniform grid. This allows us to effectively reduce the complexity of our algorithmic resolution. To do so, we describe our closed curves $\mathbf{r}(t)$ as linear combinations of N shifts of the M-periodized basis function $\varphi_M(t) = \sum_{k\in\mathbb{Z}} \varphi(\frac{t-Mk}{h})$, where $h = \frac{M}{N}$ is the grid stepsize. Following Theorem 4.4, we choose φ_M to be the M-periodization and h-dilation of the B-spline generator $\varphi = \beta_{\mathrm{D}^{\alpha+1}}$. These basis functions are weighted by two vectors of coefficients $\mathbf{c}_x = (\mathrm{c}_x[n])_{n=0}^{N-1}$ and $\mathbf{c}_y = (\mathrm{c}_y[n])_{n=0}^{N-1}$. Finally, the weighted functions are shifted by multiples of the grid size h in order to describe

$$\mathbf{r}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} \mathrm{c}_x[n]\ \varphi_M(t-nh) \\ \sum_{n=0}^{N-1} \mathrm{c}_y[n]\ \varphi_M(t-nh) \end{bmatrix}. \tag{4.102}$$

Our choice (4.102) of curve parametrization allows us to optimize solely on the coefficients $\mathbf{c}_x, \mathbf{c}_y \in \mathbb{R}^N$ of two curve components. We implement a system matrix that interpolates the coefficients using φ_M and samples them at the measurement locations. We introduce $\mathbf{H} \in \mathbb{R}^{M\times N}$ with

$$[\mathbf{H}]_{m,n} = \varphi_M\left(m - nh\right). \tag{4.103}$$

The regularization operator L becomes a circulant regularization matrix $\mathbf{L} \in \mathbb{R}^{N\times N}$ composed of shifted versions of the αth-order derivative operator coefficients $d_{\mathrm{D}^{\alpha+1}}$. The regularization matrix $\mathbf{L}$ is therefore constructed as

$$[\mathbf{L}]_{m,n} = \frac{1}{h^\alpha} d_{\mathrm{D}^{\alpha+1}}[(m-n)_{\bmod N}]. \tag{4.104}$$

Our mixed-norm regularization involves, in the discrete setting, an $\ell_1 - \ell_2$ norm given by

$$\|[\mathbf{f}_1 \;\; \mathbf{f}_2]\|_{\ell_1-\ell_2} \triangleq \sum_{n=0}^{N-1} \sqrt{(\mathbf{f}_1[n])^2 + (\mathbf{f}_2[n])^2}, \tag{4.105}$$

for $\mathbf{f}_1, \mathbf{f}_2 \in \mathbb{R}^N$. Indeed, we have that (see Theorem 4.3):

$$\|[\mathrm{L}\{x\} \; \mathrm{L}\{y\}]\|_{\mathrm{TV}-\ell_2} = \|\mathbf{L}\,[\mathbf{c}_x \;\, \mathbf{c}_y]\|_{\ell_1-\ell_2}\,. \tag{4.106}$$

Our discrete optimization problem therefore aims at finding $\mathbf{c}_x$ and $\mathbf{c}_y$ such that

$$\arg\min_{\mathbf{c}_x,\mathbf{c}_y\in\mathbb{R}^N} \left\| \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{c}_x \\ \mathbf{c}_y \end{bmatrix} - \begin{bmatrix} \mathbf{p}_x \\ \mathbf{p}_y \end{bmatrix} \right\|_2^2 + \lambda \, \|\mathbf{L}\,[\mathbf{c}_x \;\, \mathbf{c}_y]\|_{\ell_1-\ell_2}\,. \tag{4.107}$$

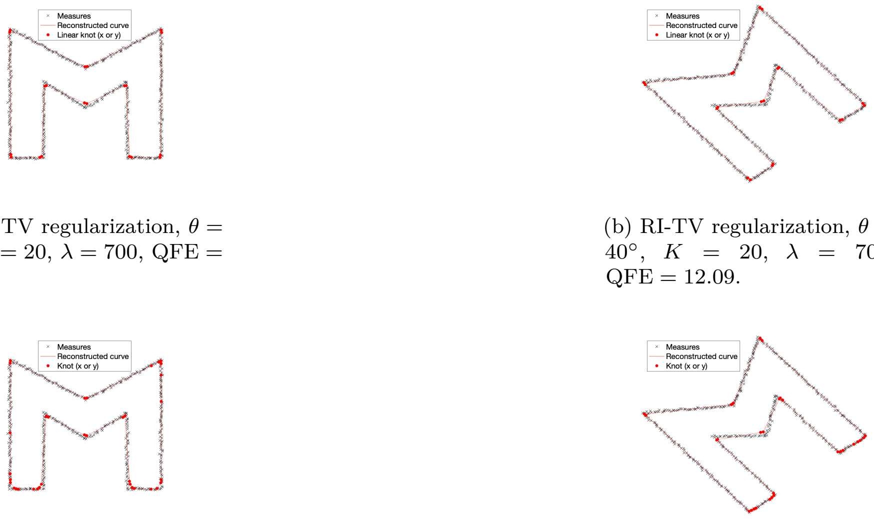
To solve Problem (4.107), we use the alternating-direction method of multipliers (ADMM) solver [200] as implemented in the GlobalBioIm Matlab library [177] dedicated to the solution of inverse problems.

4.4.4 Numerical Illustration

We evaluate the distance between the constructed curves and the contour points through the quadratic fitting error (QFE) defined as

$$\mathrm{QFE} = \frac{1}{M} \sum_{m=0}^{M-1} \| \mathbf{r}(t)|_{t=m} - \mathbf{p}[m]\|_2^2\,. \tag{4.108}$$

It is noteworthy that the QFE can be used at the same time in the single-spline setting and in the hybrid setting. Indeed, by replacing the hybrid curve $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ in (4.108), we obtain a QFE that is consistent with the data-fidelity term in (4.99).



(a) RI-TV regularization, $\theta = 0°$, $K = 20$, $\lambda = 700$, QFE $= 12.09$.

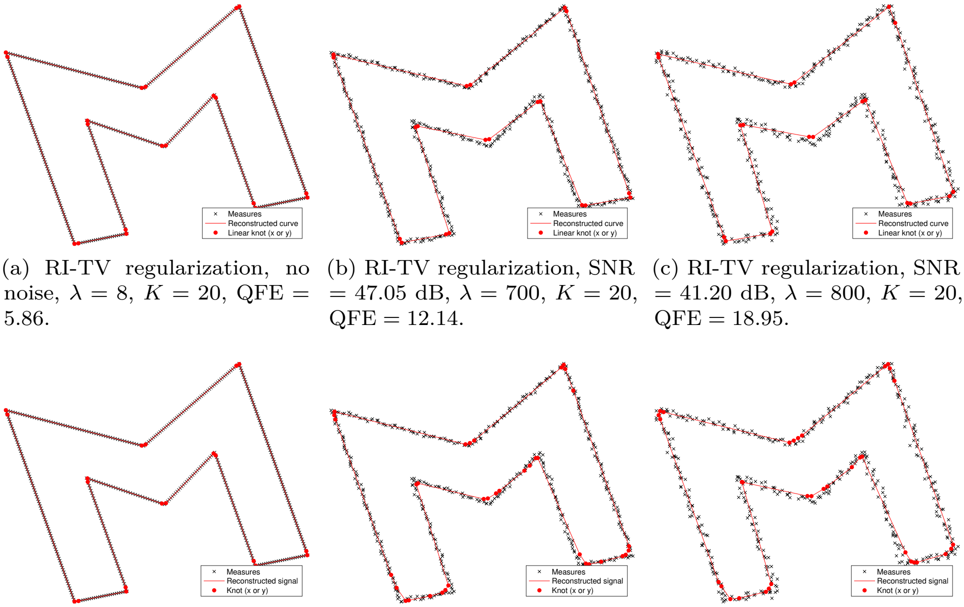
(b) RI-TV regularization, $\theta = 40°$, $K = 20$, $\lambda = 700$, QFE $= 12.09$.

(c) (TV-ℓ_1) regularization, $\theta = 0°$, $K = 37$, $\lambda = 482.13$, QFE $= 12.09$.

(d) (TV-ℓ_1) regularization, $\theta = 40°$, $K = 29$, $\lambda = 500.93$, QFE $= 12.09$.

Figure 4.5: Solutions as a function of the rotation angle θ for RI-TV regularization and (TV-ℓ_1) regularization for a same contour. $M = 488$, grid stepsize $h = 1.9062$, $\varphi = \beta^1$.

To verify that our regularization norm is truly rotation-invariant, we apply a planar rotation of angle θ to our data before we reconstruct the curve with the regularization operator $\mathrm{L} = \mathrm{D}^2$. We have added to the data a Gaussian perturbation with a signal-to-noise ratio (SNR) of 47.28 dB. We compare the curves reconstructed with our regularization to the curves resulting from the (TV-ℓ_1) regularization of Definition 4.1. Indeed, ℓ_1 regularization is widely used in the signal-processing community as a sparsifying prior. To do so, we choose λ in the non-rotated (TV-ℓ_1) regularized curve (Figure 4.5c) so that the QFE matches the QFE from the non-rotated RI-TV regularized curve (Figure 4.5a). When rotating the measurements,



(a) RI-TV regularization, no noise, $\lambda = 8$, $K = 20$, QFE = 5.86.

(b) RI-TV regularization, SNR = 47.05 dB, $\lambda = 700$, $K = 20$, QFE = 12.14.

(c) RI-TV regularization, SNR = 41.20 dB, $\lambda = 800$, $K = 20$, QFE = 18.95.

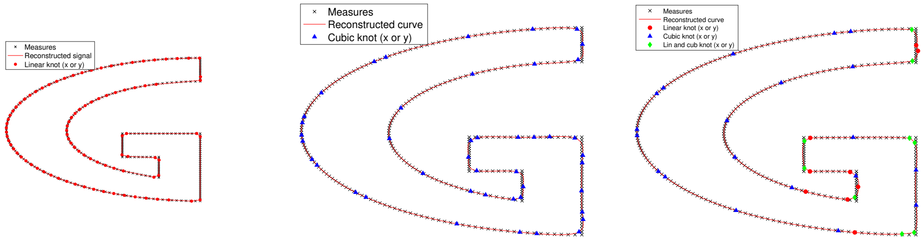
(d) (TV-ℓ_1) regularization, no noise, $\lambda = 10$, $K = 20$, QFE = 5.86.

(e) (TV-ℓ_1) regularization, SNR = 47.05 dB, $\lambda = 459.45$, $K = 36$, QFE = 12.14.

(f) (TV-ℓ_1) regularization, SNR = 41.20 dB, $\lambda = 531.35$, $K = 35$, QFE = 18.95.

Figure 4.6: Resilience to noise for RI-TV and (TV-ℓ_1) regularizations.

we adjust λ again so that the QFE of the (TV-ℓ_1)-regularized curve on the rotated points matches the one of the curve constructed with RI-TV regularization with rotated points. We see in Figure 4.5 that the RI-TV-regularized problem provides the same solution regardless of θ. Indeed, the knot locations do not differ between Figures 4.5a and 4.5b, nor does the number K of knots. This is not the case for the purely (TV-ℓ_1)-regularized problem. Not only are the knot locations different when a rotation is applied to the measurements, but the number K of knots varies with θ as well as the QFE of the curve. Additionally, one needs to adapt λ to obtain the same QFE between the constructed curves on rotated and non-rotated measurements.



(a) Spline degree: 1, $\lambda = 31.87$, $K = 89$.

(b) Spline degree: 3, $\lambda = 24.72$, $K = 44$.

(c) Spline degrees: 1 and 3, $\lambda_1 = 80$, $\lambda_2 = 90$, $K = 37$

Figure 4.7: Noiseless curve reconstruction with a single spline, a hybrid setting, and RI-TV regularization. All three reconstructions have a constant QFE = 8.88.

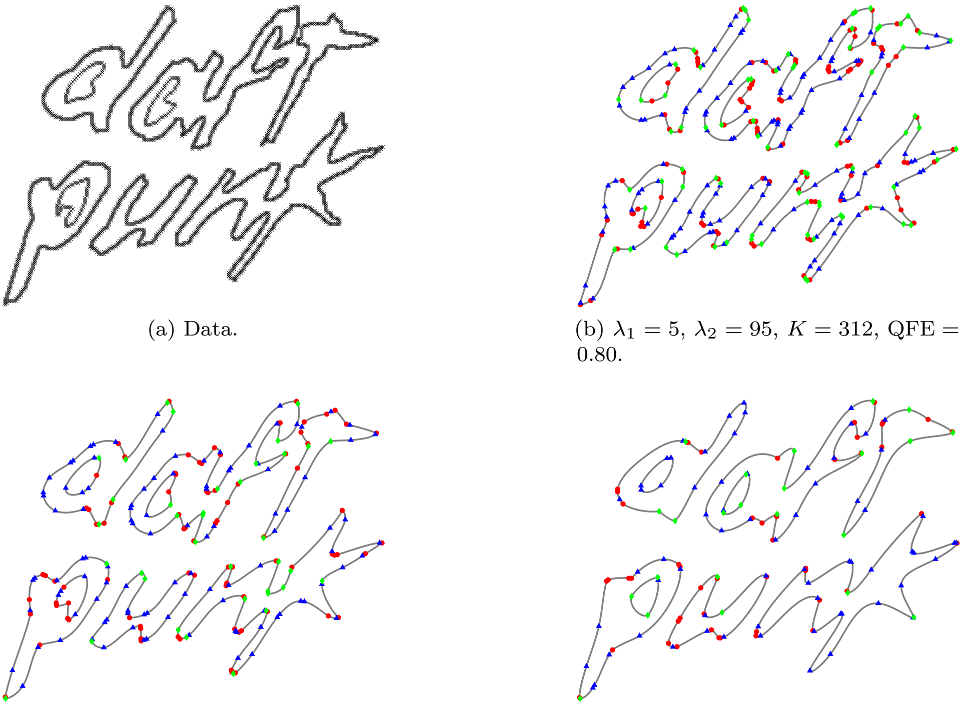
A beneficial feature derived from the enforcement of joint sparsity in the two curve components is resilience of our reconstructions to imprecisions in the contour points. Indeed, when we expect our data to be imprecise, we can choose to increase the regularization parameter λ at the cost of data fidelity, as the curve cannot rely as much on the data. Particularly, when the regularizer is TV-based, an increase in λ tends to smoothen sharp variations. This is visible in Figure 4.6, where several curves have been reconstructed using linear B-splines β^1. Figures 4.6a, 4.6b, and 4.6c are reconstructions of increasingly inaccurate measurements using RI-TV. Figures 4.6d, 4.6e, and 4.6f depict reconstructions resulting from a sparsifying regularization without coupling (TV-ℓ_1), using a λ tuned so that the QFE matches the QFE of the curves regularized by RI-TV. When TV regularization is used, and as the contour becomes more inaccurate, the number K of knots drastically increases and the angles are deformed. On the contrary, for the reconstructions in Figures 4.6a, 4.6b, even as the noise and λ increase, the number K of knots remains unchanged and the angles are fairly well preserved.

The single-component framework only allows for the use of one kind of B-spline per curve. However, when the contour under consideration is composed of smooth sections and kinks, no single type of B-spline can provide a faithful and sparse reconstruction. An example of curve fitting that depicts this problem is given in Figure 4.7. We reconstructed the contour using first β^1 and β^3 as basis functions, giving piecewise-linear and piecewise-cubic curves in Figures 4.7a and 4.7b, respectively. Figure 4.7c contains a reconstruction under the hybrid setting $\mathrm{L}_1 = \mathrm{D}^2$ and $\mathrm{L}_2 = \mathrm{D}^4$, thus producing a curve that has both a linear and a cubic component. While all three reconstructions yield the same QFE with respect to the data, the hybrid curve has by far the smallest number of knots. Moreover, upon visual inspection, the hybrid curve in Figure 4.7c portrays the most faithful reconstruction, as it does round neither the angles nor the straight lines, nor does it straighten the smooth sections.

We can observe the effect of the parameters λ_1 and λ_2 on the constructed curve when the hybrid reconstruction setting is applied to real contour points for a constant ratio of knots $\frac{K_1}{K_2} = 0.86$. In Figure 4.8, as λ_1 and λ_2 increase, the total number K of knots decreases and the curve becomes more stylized. In addition, for all values of λ_1 and λ_2, our algorithm preserves the kinks of the contour while mimicking its smooth segments.

4.4.5 Summary

We have introduced a framework to reconstruct sparse continuous curves from a list of possibly inaccurate contour points using an RI-TV regularization norm. We have proved that an optimal solution to our minimization problem is a curve that uses splines as basis functions, and we have leveraged this result to provide an exact discretization of the continuous-domain framework using B-splines. Furthermore, we have extended our formulation to reconstruct curves with components of distinct smoothness properties using hybrid splines. We have experimental confirmation of the rotation invariance of our regularizer. In addition, our experimental results demonstrate that our formulation yields sparse reconstructions that are close to the data points even when their noise increases, unlike other regularization methods. Finally, our hybrid-curve experiments demonstrate that we are able to faithfully reconstruct any contour with a low number of knots.



(a) Data.

(b) $\lambda_1 = 5$, $\lambda_2 = 95$, $K = 312$, QFE $=$ 0.80.

(c) $\lambda_1 = 20$, $\lambda_2 = 980$, $K = 229$, QFE $=$ 1.11.

(d) $\lambda_1 = 100$, $\lambda_2 = 9900$, $K = 139$, QFE $= 2.82$.

Figure 4.8: Effect of λ_1 and λ_2 on the reconstructed hybrid curve for $M = 2714$ under RI-TV regularization. The reconstructed curve is represented by the solid line. The round markers and the triangular markers indicate the location of the linear and cubic knots, respectively. The diamond-shaped markers indicate the superimposition of a linear knot and a cubic knot. The data are extracted from the official Daft Punk logo [8].

Chapter 5

A Stochastic View on Splines

In the last part of this thesis[1], we provide a stochastic view on sparsity and splines, the two main themes of this thesis. Precisely, we study the family of sparse stochastic processes that are known to be the limit point of "random splines". By relying on the latter density result, we propose a method based on B-splines to generate trajectories of these processes (Section 5.2). We then characterize their compressibility by providing a precise identification of the Besov regularity of their corresponding innovation models (Section 5.3). While our characterization is sharp for most popular classes of stochastic processes, it only provides a lower-bound for the family of random splines. Hence, we treat this case separately in Section 5.4 by introducing a different machinery for studying their compressibility rate.

[1] This chapter is based on our published works [310, 325, 326].

5.1 Overview on Sparse Stochastic Processes

5.1.1 Context

The statistical modelling of data plays a central role in numerous research domains, such as signal processing [327] and pattern recognition [328]. In that regard, Gaussian models have been the first and by far the most considered ones, thanks to their desirable mathematical properties and relatively simple characterization. For instance, the Karhunen-Loève transform (KLT) identifies the optimal basis for representing data with Gaussian priors [329] and Kalman filters are optimal denoisers of Gaussian signals [330], both in the mean-square sense. These facts, among others, have made Gaussian statistical priors very convenient in practice. They also reveal the fundamental relationship between Fourier-based signal representations and Gaussian models.

However, it has been a long standing observation that Gaussian models fail to capture several key statistical properties of most naturally-occurring signals [331, 332]. Indeed, the latter frequently have heavy-tailed marginals [333, 48, 334, 335] or richer structure of dependencies than Gaussian ones [336, 337]. Real-world signals are highly structured and often admit concise representations, typically on wavelet bases that appear to be genuinely versatile [338, 339]. This has led to the current paradigm in modern data science where *sparsity* plays one of the central roles in statistical learning [340, 341] and signal modelling [65, 342, 48]. Classical Gaussian priors cannot model sparsity as they tend to produce poorly compressible signals [343, 344]. Many recent efforts in signal processing have been directed towards the development of deterministic frameworks that are better tailored to the reconstruction or synthesis of sparse signals, such as traditional compressed sensing [62, 163, 67] and its infinite-dimensional extensions [345, 307, 66, 73].

The development of wavelet methods, based on the pioneering works of I. Daubechies, Y. Meyer, and S. Mallat in the late 80's [346, 347, 348], has shed new lights on signal representation. Repeated numerical observations confirmed that wavelet-based compression techniques such as JPEG-2000 [349] outperform classical Fourier-based standards (*e.g.*, JPEG) for natural images. This is despite the fact that the discrete Fourier transform (DFT) and its real-valued counterpart, the discrete cosine transform (DCT) [350], are asymptotically equivalent to KLT and, hence, are optimal for representing signals with Gaussian prior [351].

Sparsity being an essential component of modern signal processing [338, 62, 339], the authors of [48] proposed a wider stochastic framework that encompasses both Gaussian and sparsity-compatible models. Within this framework, a continuous-domain signal is a realization of a stochastic process s that can be whitened by some linear, shift-invariant operator L. The key here is that the resulting white noise, or innovation, is not necessarily Gaussian. Put formally, signals are solutions of

$$\mathrm{L}s = w, \tag{5.1}$$

where w is a well defined innovation process called a Lévy white noise [352]. The term *Lévy* here comes from the fact that w is an object that can be interpreted as the derivative of a Lévy process in the sense of distributions [170, 353]. Whenever w is non-Gaussian, the realizations of s can be shown to be sparse. Accordingly, they have been named sparse stochastic processes [48]. Specific instances of such processes have been used to model natural signals such as images [354, 355], RF echoes in ultrasound [356], and network traffic in communication systems [357, 358, 359].

5.1.2 Generalized Random Processes

The theory of generalized random processes was initiated independently by K. Itô [360] and I.M. Gel'fand [361] in the 50's and corresponds to the probabilistic counterpart of the theory of generalized functions of L. Schwartz. It was later brought to light by Gel'fand himself together with N.Y. Vilenkin in [362, Chapter III]. In this framework, a generalized random process is characterized by its effects against test functions. This allows to consider random processes that are not necessarily defined pointwise, as is the case for the Lévy white noise. The theory of generalized random processes, besides being very general, appears to be very flexible for the construction and analysis of random processes. It is a powerful alternative to more classic approaches, as argumented in [363, 364]. The theory of generalized random processes is used as the ground for generalized CARMA processes [365] and fields [366, 367], for conformal field theory in statistical physics [368], for studying the solutions of stochastic differential PDEs [369, 370, 371], and as random models in signal processing [372, 373, 334, 48].

We recall that $\mathcal{S}(\mathbb{R}^d)$ is the space of rapidly decaying smooth functions from $\mathbb{R}^d$ to $\mathbb{R}$ which is endowed with its natural Fréchet nuclear topology [374]. Its topological dual is the space of tempered generalized functions $\mathcal{S}'(\mathbb{R}^d)$. It is endowed with the strong topology and $\mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ denotes the Borelian σ-field for this topology. Note that $\mathcal{S}'(\mathbb{R}^d)$ can be endowed with other natural σ-fields: the one associated to the weak-* topology or the cylindrical σ-field generated by the cylinders

$$\{u \in \mathcal{S}'(\mathbb{R}^d),\ (\langle u, \varphi_1\rangle, \dots, \langle u, \varphi_N\rangle) \in B\} \tag{5.2}$$

for $N \geq 1$, $\varphi_n \in \mathcal{S}(\mathbb{R}^d)$, and B a Borelian subset of $\mathbb{R}^N$. However, these different σ-fields are known to coincide in this case [375, Proposition 3.8 and Corollary 3.9][2]. See also Itô's [370] and Fernique's monographs [364] for general discussions on the measurable structures of function spaces. Throughout this chapter, we fix a complete probability space $(\Omega, \mathcal{F}, \mathscr{P})$.

Definition 5.1. *A measurable function s from $(\Omega, \mathcal{F})$ to $(\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d)))$ is called a* generalized random process*. Its* probability law *is the probability measure on $\mathcal{S}'(\mathbb{R}^d)$ defined for $B \in \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ by*

$$\mathscr{P}_s(B) = \mathscr{P}(\{\omega \in \Omega,\ s(\omega) \in B\}). \tag{5.3}$$

The characteristic functional *of s is the functional $\widehat{\mathscr{P}_s} : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ such that*

$$\widehat{\mathscr{P}_s}(\varphi) = \int_{\mathcal{S}'(\mathbb{R})} \mathrm{e}^{\mathrm{j}\langle u, \varphi\rangle} \mathrm{d}\mathscr{P}_s(u). \tag{5.4}$$

It turns out that the characteristic functional is continuous, positive-definite over $\mathcal{S}(\mathbb{R}^d)$, and normalized such that $\widehat{\mathscr{P}_s}(0) = 1$. The converse of this result is also true: if $\widehat{\mathscr{P}}$ is a continuous and positive-definite functional over $\mathcal{S}(\mathbb{R}^d)$ such that $\widehat{\mathscr{P}}(0) = 1$, then it is the characteristic functional of a generalized random process in $\mathcal{S}'(\mathbb{R})$. This is known as the Bochner-Minlos theorem. It was initially proved

[2] This is true in general for the dual of a nuclear Fréchet space. Note that this is not obvious and is typically not true for other spaces of generalized functions, such as $\mathcal{D}'(\mathbb{R}^d)$ [370].

in [376] and uses the nuclearity of $\mathcal{S}'(\mathbb{R})$. See [377, Theorem 2.3] for an elegant proof based on the Hermite expansion of tempered generalized functions [168]. It means in particular that one can define generalized random processes via the specification of their characteristic functional. Following Gel'fand and Vilenkin, we use this principle to introduce Lévy white noises.

We consider functionals of the form $\widehat{\mathscr{P}}(\varphi) = \exp\left(\int_{\mathbb{R}^d} \Psi(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right)$. It is known that $\widehat{\mathscr{P}}$ is a characteristic functional over the space $\mathcal{D}(\mathbb{R})$ of compactly supported smooth functions if and only if the function $\Psi : \mathbb{R} \to \mathbb{C}$ is continuous, conditionally positive-definite, with $\Psi(0) = 0$ [362, Section III-4, Theorems 3 and 4]. A function Ψ that satisfies these conditions is called a *Lévy exponent* and can be decomposed according to the Lévy-Khintchine theorem [170, Theorem 8.1] as

$$\Psi(\xi) = \mathrm{j}\mu\xi - \frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{j}\xi t} - 1 - \mathrm{j}\xi t \mathbb{1}_{|t|\leq 1})\mathrm{d}\nu(t), \tag{5.5}$$

where $\mu \in \mathbb{R}$, $\sigma^2 \geq 0$, and ν is a *Lévy measure*, which means a positive measure on $\mathbb{R}$ such that $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} \inf(1, t^2)\mathrm{d}\nu(t) < \infty$. The triplet (μ, σ^2, ν) is unique and called the *Lévy triplet* of Ψ.

In our case, we are only interested in the definition of Lévy white noises over $\mathcal{S}'(\mathbb{R}^d)$. This requires an adaptation of the construction of Gel'fand and Vilenkin. We say that the characteristic exponent Ψ satisfies the *ϵ-condition* if there exists some $\epsilon > 0$ such that $\int_{\mathbb{R}} \inf(|t|^{\epsilon}, t^2)\mathrm{d}\nu(t) < \infty$, with ν the Lévy measure of Ψ. Then, the functional $\widehat{\mathscr{P}}(\varphi) = \exp\left(\int_{\mathbb{R}} \Psi(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right)$ is a characteristic functional over $\mathcal{S}(\mathbb{R})$ if and only if Ψ is a characteristic exponent that satisfies the ϵ-condition. The sufficiency is proved in [378, Theorem 3] and the necessity in [379, Theorem 3.13] and, also, in [380].

Definition 5.2. *A* Lévy white noise *in $\mathcal{S}'(\mathbb{R}^d)$ (or simply a Lévy white noise) is a generalized random process w with characteristic functional of the form*

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}} \Psi(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right) \tag{5.6}$$

for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where Ψ is a characteristic exponent that satisfies the ϵ-condition.

The Lévy triplet of w is denoted by (μ, σ^2, ν). Then, we say that w is a Gaussian white noise *if $\nu = 0$, a* compound Poisson white noise *if $\mu = \sigma^2 = 0$ and $\nu = \lambda P$, with $\lambda > 0$ and P a probability measure on $\mathbb{R}$ such that $P(\{0\}) = 0$, and a* Lévy white noise with finite moments *if $\mathbb{E}[|\langle w, \varphi \rangle|^p] < \infty$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $p > 0$.*

The ϵ-condition is extremely mild. Lévy white noises in $\mathcal{S}'(\mathbb{R}^d)$ include stable white noises, symmetric-gamma white noises, and compound Poisson white noises whose jumps probability measure P admits a finite moment ($\int_{\mathbb{R}^d} |t|^\epsilon P(\mathrm{d}t) < \infty$ for some $\epsilon > 0$) [381, Section 2.1.3]. Lévy white noises are stationary and independent at every point, meaning that $\langle w, \varphi_1 \rangle$ and $\langle w, \varphi_2 \rangle$ are independent as soon as φ_1 and $\varphi_2 \in \mathcal{S}(\mathbb{R}^d)$ have disjoint supports [362, Section III-4, Theorem 6].

One can extend the space of test functions a given Lévy white noise can be applied to. This is done by approximating a test function φ with functions in $\mathcal{S}(\mathbb{R}^d)$ and showing that the underlying sequence of random variables converges in probability to a random variable that we denote by $\langle w, \varphi \rangle$. This principle is developed with more generality in [382] by connecting the theory of generalized random processes to independently scattered random measures in the sense of B.S. Rajput and J. Rosinski [383]; see also [384]. In particular, as soon as $\varphi \in L_2(\mathbb{R}^d)$ is compactly supported, the random variable $\langle w, \varphi \rangle$ is well-defined [382, Proposition 5.10].

Remark 5.1. *The random variable $\langle w, \varphi \rangle$ can be interpreted as a stochastic integral with respect to a Lévy sheets $s : \mathbb{R}^d \to \mathbb{R}$ such that $\mathrm{D}_1 \ldots \mathrm{D}_d\{s\} = w$, where D_i is the partial derivative along direction $1 \leq i \leq d$. We recall that Lévy sheets are multivariate generalizations of the Lévy processes [379, 385, 384]. In that case, we have the formal relation $\langle w, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) \mathrm{d}s(\boldsymbol{x})$, whose precise meaning has been investigated in [379, 382].*

To summarise, the three most important operational properties of Lévy white noises for our purpose are:

1. **Stationarity:** For any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\boldsymbol{\tau} \in \mathbb{R}^d$, the random variables $\langle \varphi, w \rangle$ and $\langle \varphi(\cdot - \boldsymbol{\tau}), w \rangle$ are identically distributed.

2. **Independence:** For any $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$ with disjoint supports, the random variables $\langle \varphi_1, w \rangle$ and $\langle \varphi_2, w \rangle$ are independent.

3. **Characterization of the probability law:** For any Lévy white noises w in $\mathcal{S}'(\mathbb{R}^d)$ and for any test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the characteristic function of the random variable $X_\varphi = \langle \varphi, w \rangle$ can be specified as

$$\widehat{\mathcal{P}}_{X_\varphi}(\xi) = \mathbb{E}[\mathrm{e}^{\mathrm{j}\xi\langle \varphi, w \rangle}] = \exp\left(\int_{\mathbb{R}} \Psi(\xi\varphi(\mathbf{x}))\mathrm{d}\mathbf{x} \right). \tag{5.7}$$

5.1.3 Infinite Divisible Random Variables

Let us remark that in dimension $d = 1$, the Lévy exponent can alternatively be expressed as

$$\Psi(\xi) = \log\left(\widehat{\mathcal{P}}_{X_{\mathrm{rect}}}(\xi) \right), \tag{5.8}$$

where $X_{\mathrm{rect}} = \langle \mathrm{rect}_{[0,1]}, w \rangle$ is the observation of w through the rectangular window

$$\mathrm{rect}_{[0,1]}(x) = \begin{cases} 1, & 0 < x \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

The distribution of X_{rect} gives us the Lévy exponent from which we can determine all the statistics of w. In particular, the following Proposition from [48] connects the second-order statistics of w to those of X_{rect}.

Proposition 5.1 ([48], Theorem 4.15)**.** *Let w be a Lévy white noise such that $X_{\mathrm{rect}} = \langle \mathrm{rect}_{[0,1]}, w \rangle$ has zero mean and a finite variance $\sigma_w^2 = \mathbb{E}[X_{\mathrm{rect}}^2]$. Then,*

$$\forall \varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}), \quad \mathbb{E}[\langle \varphi_1, w \rangle \langle \varphi_2, w \rangle] = \sigma_w^2 \langle \varphi_1, \varphi_2 \rangle. \tag{5.10}$$

It turns out that X_{rect} is an infinitely divisible random variable in the sense of Definition 5.3 [344].

Definition 5.3. *A real-valued random variable X is said to be infinitely divisible if, for any natural number $M \in \mathbb{N}$, there exist M independent and identically distributed random variables $X_1, ..., X_M$ such that*

$$X = X_1 + \cdots + X_M. \tag{5.11}$$

Table 5.1: Infinitely divisible distributions and their Lévy exponents

Distribution	**Lévy exponent**
Gaussian (μ, σ)	$\mathrm{j}\mu\xi - \sigma^2\xi^2/2$
Symmetric α-stable $(\alpha, c), \alpha \in (0, 2]$	$-\lvert c\xi\rvert^\alpha$
Gamma(α, β)	$-\beta \log(1 - \mathrm{j}\xi/\alpha)$
Laplace (μ, b)	$\mathrm{j}\mu\xi - \log(1 + b^2\xi^2)$

To check the infinite divisibility of X_{rect}, one can note that, for any $M \in \mathbb{N}$, we have that

$$X_{\mathrm{rect}} = \langle \mathrm{rect}_{[0,1]}, w \rangle = \langle \sum_{m=0}^{M-1} \mathrm{rect}_{[\frac{m}{M}, \frac{m+1}{M}]}, w \rangle = \sum_{m=0}^{M-1} \langle \mathrm{rect}_{[\frac{m}{M}, \frac{m+1}{M}]}, w \rangle. \tag{5.12}$$

The terms in the sum (5.12) are independent and identically distributed random variables as a consequence of the independence and stationarity properties of white noises, which certifies that $\langle \mathrm{rect}_{[0,1]}, w \rangle$ is infinitely divisible.

The converse is also true: for any infinitely divisible random variable X with Lévy exponent $\Psi(\xi) = \log(\mathbb{E}[\mathrm{e}^{\mathrm{j}\xi X}])$ that satisfies the ϵ-condition, there exists a well defined Lévy white noise w whose statistics are determined by (5.7) [344, 378, 379]. This shows that there is a one-to-one correspondence between infinitely divisible distributions and *Lévy* white noises through $\langle \mathrm{rect}_{[0,1]}, w \rangle$.

The Gaussian, gamma, and α-stable distributions are classical examples of infinitely divisible distributions [170]. We can plug in their Lévy exponents in (5.7) to define their corresponding Lévy white noises. We repeat in Table 5.1 some infinitely divisible distributions of interest, along with their Lévy exponents[344].

5.1.4 Sparse Stochastic Processes

The sparse-stochastic-process framework of Unser *et al.* [48] is a comprehensive theory of generalized Lévy Processes. These are stochastic processes that can be whitened by some admissible linear, shift-invariant operator. More precisely, the generalized random process $s \in \mathcal{S}'(\mathbb{R})$ is a generalized Lévy process if there exists a linear, shift-invariant operator $\mathrm{L} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ such that $w = \mathrm{L}s$ is a Lévy white noise. Equivalently, one may view generalized Lévy processes as the solution of the stochastic differential equation

$$\mathrm{L}s = w. \tag{5.13}$$

It has been shown that, under mild technical assumptions on L and w, a solution s of (5.13) exists and constitutes a properly defined generalized stochastic process over $\mathcal{S}'(\mathbb{R})$ [378].

When L is an operator with a trivial null space, such as $\mathrm{L} = (\mathrm{D} - \alpha \mathrm{j})$ with $\Re(\alpha) \neq 0$, we can write that

$$s = \mathrm{L}^{-1} w, \tag{5.14}$$

where L^{-1} is the inverse of L. However, when the null space is nontrivial, for instance when L corresponds to an unstable ordinary differential equation, the specification of the boundary conditions become necessary to uniquely identify the solution. The boundary conditions take the form

$$\phi_\ell(s) = c_\ell, \quad \ell = 1, \dots, N_0, \tag{5.15}$$

where $\phi_\ell : s \mapsto \phi(s) \in \mathbb{R}$ are appropriate linear functionals, $c_\ell \in \mathbb{R}$, and N_0 is the dimension of the null space of L. For instance, one can impose that the process s takes fixed values at reference locations $t_1 < \ldots < t_{N_0}$; that is, $\phi_\ell(s) = s(t_\ell) = c_\ell$ for $\ell = 1, \dots, N_0$. Such boundary conditions appear in the classical definition of Lévy processes (including Brownian motion), where we have that $\phi(s) = s(0) = 0$ (Chapter 7 of [48]). We formally write

$$s = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w, \tag{5.16}$$

where $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is the right inverse of L. It incorporates the boundary conditions (5.15) (Chapter 5.4 of [48]).

5.1.5 Compound-Poisson Processes

A case of special interest is when the Lévy exponent corresponds to a compound-Poisson distribution. A compound-Poisson random variable X, with rate λ and amplitude law ν, is defined as

$$X = \sum_{k=1}^{K} A_k, \tag{5.17}$$

where the number K is a Poisson random variable with parameter λ and $(A_k)_{k=1}^{K}$ is an i.i.d. sequence drawn according to ν. We refer to the corresponding Lévy white noise w as a compound-Poisson innovation. It is known to be equal in law to

$$w = \sum_{k\in\mathbb{Z}} A_k \delta(\cdot - \tau_k), \tag{5.18}$$

where the sequence $(\tau_k)_{k\in\mathbb{Z}}$ of locations of Diracs is a stationary Poisson point-process (independent of A_ks) with rate λ (see [386] for a formal definition of point processes). The key property regarding the Dirac locations is that the number N of τ_k in any interval $[a,b]$ with $a<b$ is a Poisson random variable with parameter $\lambda(b-a)$. Furthermore, condition to the event $N=n$, the locations of jumps that are in $[a,b]$ are drawn independently from a uniform law over $[a,b]$ [386, Section 2.1]. This implies that, if we denote by $\mathbf{x}=(x_1,\dots,x_N)$, the ordered set of Dirac locations that are in $[a,b]$, then condition to the event $\{N=n\}$ for any $n\geq 1$, the probability density function (PDF) of $\mathbf{x}$ is

$$p_{\mathbf{x}}(\boldsymbol{u}|N=n) = \frac{n!}{(b-a)^n}\mathbb{1}_{a\leq u_1\leq\ldots\leq u_n\leq b} \tag{5.19}$$

for any vector $\boldsymbol{u}=(u_1,\dots,u_n)\in\mathbb{R}^n$. It is worth noting that the probability density of $\mathbf{x}$, once condition to $N=n\geq 1$, does not depend on λ anymore.

On any finite interval, compound-Poisson innovations have a finite representation. They can be stored on a computer with the quantization of real numbers as sole source of information loss. They are therefore well adapted to simulation purposes.

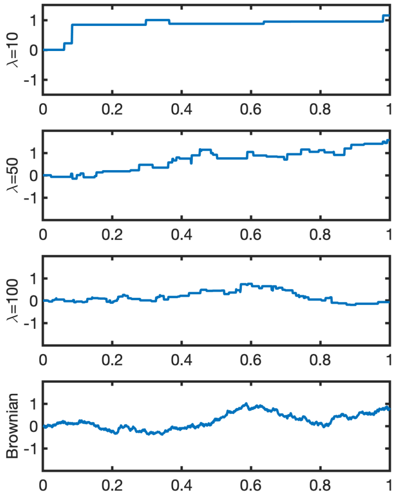


Figure 5.1: Trajectories of compound Poisson processes with Gaussian jumps (different values of λ) and a Brownian motion. All processes are normalized to have unit variance.

When w is a compound-Poisson innovation of the form (5.18), the process $s = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w$ is called a generalized Poisson process. In particular, compound Poisson processes with the whitening operator $\mathrm{L} = \mathrm{D}$) are piecewise constant processes (see Figure 5.1).

5.2 Generating Trajectories of Sparse Stochastic Processes

In this section[3], we provide an algorithm to generate trajectories of sparse stochastic processes that are solutions of linear ordinary differential equations driven by Lévy white noises. Our method is based on a theoretical observation which states that these processes are limits in law of generalized compound-Poisson processes. We derive an off-the-grid algorithm that generates arbitrarily close approximations of the target process. Our method relies on a B-spline representation of generalized compound-Poisson processes. We illustrate numerically the validity of our approach.

5.2.1 Context

The goal of this work is to generate realizations of the stochastic process s, defined in (5.1) given its whitening operator L and a statistical characterization of its innovation process w. The computer generation of these signals can be of great interest to practitioners who wish to evaluate their reconstruction algorithms. We are thinking of works such as [387, 388, 389, 390], where optimal estimators for interpolating and denoising such processes have been derived.

A possible approach to generate realizations of s would be to notice that, if L is a differential operator such as $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}$ or a polynomial in D, then (5.1) defines a stochastic differential equation (SDE) [391]. This becomes more apparent when notating w with the alternative notation $\mathrm{d}Z_t$, where $(Z_t)_{t\in\mathbb{R}^+}$ is a Lévy process (Chapter 7.4 in [48]). For example, $(\mathrm{D} - \alpha\mathrm{I})s = w$ can be rewritten as $\mathrm{d}S_t = \alpha S_t \mathrm{d}t + \mathrm{d}Z_t$. A suitable SDE solver, such as the one studied in [392], can then be used to generate an approximation of the signal. In particular, a common method is to solve the linear system of stochastic difference equations that is obtained by considering the discrete counter-part of the operator L (*e.g.* using finite differences instead of the derivative), and by replacing the innovation process w with a discrete white noise (see, for example, [393]).

[3] This section is based on our published work [310].

It turns out that generic SDE solvers do not exploit the linearity of L. Here, the analytic treatment of (5.1) can be pushed further to obtain an explicit solution. Brockwell shows in [394] that s corresponds to the integral of a deterministic function with respect to a Lévy process. The integral can then be approximated by substituting it with a Riemann sum defined on a partition of the integration interval [395, Theorem 21].

These approaches, although valid, have drawbacks when it comes to the generation of synthetic signals for the evaluation of algorithms. First, they directly depend on the existence of a grid on which the approximation of the continuous process is sampled. This can lead to complication in the context of the multi-resolution algorithms that manipulate grid-free descriptions of signals. Second, the generated approximations are not solutions of an SDE in the form of (5.1). In other words, the approximations are not mathematical objects of the same nature as s.

In what follows, we propose a method that addresses both issues. It is based on a theoretical result by Fageot *et al.* [393] that states that any sparse stochastic process s that is the solution of (5.13) can be specified as the limit in law of a sequence $\{s_n\}_{n\in\mathbb{N}}$ of generalized Poisson processes [393]. The corresponding driving processes $w_n = \mathrm{L}s_n$ are compound-Poisson innovations of the form

$$w_n = \sum_{k\in\mathbb{Z}} A_{k,n}\delta(\cdot - \tau_{k,n}) \tag{5.20}$$

with rates $\lambda_n = n$ and with i.i.d. amplitudes $A_{k,n}$ that are infinitely divisible random variables with Lévy exponent $f_n = \frac{1}{n}f$, where f is the Lévy exponent of w. These simpler processes have the advantage of having a grid-free numerical representation despite having a continuously defined domain. They fall within the category of (random) signals with a finite rate of innovation [396, 397]. They also have the desirable property of being whitened by the same operator L as the approximated signal. This implies that they all have the same correlation structure as the target signal (see Proposition 5.1).

Our method takes a sufficiently large value for n and generates a realization of the process s_n on a chosen interval. To do so, we consider an intermediary process called the generalized increment process. Interestingly, this process can be represented as a weighted sum of shifted B-splines and can be sampled very efficiently [296, 398]. The desired stochastic process s_n is then obtained from the latter by recursive filtering.

5.2.2 Mathematical Background: Rational Operators

We restrict ourselves to rational operators in $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{dt}}$, written $\mathrm{L} = P(\mathrm{D})Q(\mathrm{D})^{-1}$, where P and Q are polynomials such that $\deg(P) > \deg(Q)$. The latter assumption is crucial to have the minimum required regularity (point-wise definition) for the solution s of (5.1). The case $\mathrm{L} = \mathrm{D}$ is a typical choice that appears, for example, in the modeling of Brownian motion.

Rational operators are defined through their frequency response

$$\widehat{\mathrm{L}}(\omega) = \frac{P(\mathrm{j}\omega)}{Q(\mathrm{j}\omega)}. \tag{5.21}$$

They provide a succinct representation of the equation $P(\mathrm{D})s = Q(\mathrm{D})w$ that we can simply rewrite as $\mathrm{L}s = w$. The Green's function of a differential operator L is a tempered distribution $\rho_{\mathrm{L}} \in \mathcal{S}'(\mathbb{R})$ that satisfies

$$\mathrm{L}\rho_{\mathrm{L}} = \delta. \tag{5.22}$$

It can be viewed as the impulse response of the inverse of L. The canonical Green's function is

$$\rho_{\mathrm{L}} = \mathcal{F}^{-1}\left\{\frac{1}{\widehat{\mathrm{L}}(\omega)}\right\}, \tag{5.23}$$

where $\widehat{\mathrm{L}}$ is the frequency response of L (Chapter 5.2 of [48]). This definition can be made to stay valid even when $\widehat{\mathrm{L}}$ vanishes at some points, as long as $\frac{1}{\widehat{\mathrm{L}}(\omega)}$ is in $\mathcal{S}'(\mathbb{R})$.

Here, we describe a method to compute Green's functions of rational operators. We begin with the intermediate computation of the Green's function of $\mathrm{L} = (\mathrm{D} - \alpha\mathrm{j})^k$. We have that

$$\rho_{\alpha,k}(t) = \mathcal{F}^{-1}\left\{\frac{1}{(\mathrm{j}\omega - \alpha)^k}\right\}(t) = \begin{cases} \mathbb{1}_+(t)\frac{t^{k-1}}{(k-1)!}\mathrm{e}^{\alpha t}, & \Re(\alpha) \le 0 \\ -\mathbb{1}_+(-t)\frac{t^{k-1}}{(k-1)!}\mathrm{e}^{\alpha t}, & \text{otherwise} \end{cases} \tag{5.24}$$

is a Green's function of L.

Now, recall that rational operators are of the form $\mathrm{L} = P(\mathrm{D})Q(\mathrm{D})^{-1}$, where P and Q are polynomials. Taking $\{\alpha_1, ..., \alpha_m\}$ to be the roots of P with multiplicity $\{\gamma_1, ..., \gamma_m\}$, the inverse of the frequency response is given by

$$\frac{1}{\widehat{\mathrm{L}}(\omega)} = \frac{Q(\mathrm{j}\omega)}{\prod_{i=1}^{m}(\mathrm{j}\omega - \alpha_i)^{\gamma_i}}. \tag{5.25}$$

This inverse is known to admit a partial-fraction decomposition of the form

$$\frac{1}{\widehat{\mathrm{L}}(\omega)} = \sum_{i=1}^{m}\sum_{k=1}^{\gamma_i}\frac{c_{ik}}{(\mathrm{j}\omega - \alpha_i)^k} \tag{5.26}$$

for some constants $c_{ik} \in \mathbb{C}$. The corresponding Green's function is then given by:

$$\rho_{\mathrm{L}}(t) = \mathcal{F}^{-1}\left\{\frac{1}{\widehat{\mathrm{L}}(\omega)}\right\}(t) = \sum_{i=1}^{m}\sum_{k=1}^{\gamma_i} c_{ik}\mathcal{F}^{-1}\left\{\frac{1}{(\mathrm{j}\omega - \alpha_i)^k}\right\}(t) = \sum_{i=1}^{m}\sum_{k=1}^{\gamma_i} c_{ik}\rho_{\alpha_i,k}(t) \tag{5.27}$$

5.2.3 Method

We now introduce our method for generating (approximate) trajectories of a sparse stochastic process s that is whitened by a rational operator L and whose innovation noise is w. When necessary, we assume general boundary conditions of the form $\phi_\ell(s) = 0$ for $\ell = 1, ..., N_0$, where N_0 is the dimension of the null space of L.

As mentioned earlier, the process s is the limit of generalized compound-Poisson processes s_n driven by $w_n = \mathrm{L}s_n$, a compound-Poisson innovation of the form (5.20). The process s_n can therefore be written

$$s_n = \sum_{k\in\mathbb{Z}} A_{k,n}\rho_{\mathrm{L}}(\cdot - \tau_{k,n}) + p_{0,n}, \tag{5.28}$$

where ρ_{L} is a Green's function of L and $p_{0,n}$ is an element of the null space of L determined by boundary conditions (it vanishes when L is invertible). Indeed, we have that

$$\mathrm{L}\left\{\sum_{k\in\mathbb{Z}} A_{k,n}\rho_{\mathrm{L}}(\cdot-\tau_{k,n})+p_{0,n}\right\}=\sum_{k\in\mathbb{Z}} A_{k,n}\mathrm{L}\{\rho_{\mathrm{L}}(\cdot-\tau_{k,n})\}=\sum_{k\in\mathbb{Z}} A_{k,n}\delta(\cdot-\tau_{k,n})=w_n.$$

For large values of n, the process s_n is assumed to be a good approximation of s. So, our goal is to generate samples of s_n on any uniform grid over any interval $[0,T]$. More precisely, once an interval $[0,T]$ is specified and a regular grid with step size h is provided, our aim is to obtain the vector $\mathbf{s}_n$ whose components are $[\mathbf{s}_n]_i = s_n(ih)$, for $i = 0, ..., \left(\lceil \frac{T}{h} \rceil - 1\right)$.

We begin by obtaining a realization of the driving innovation w_n. It consists of a sequence of impulse locations $(\tau_{k,n})$ and a corresponding sequence of amplitudes $(A_{k,n})$.

The sequence $(\tau_{k,n})$ is a point Poisson process. Its realization on the interval $[0,T]$ is simulated in two steps. First, a Poisson random variable K with parameter $\lambda = nT$ is generated. Then, K impulse locations $(\tau_{k,n})_{k\in\{1,...,K\}}$ are sampled uniformly on $[0,T]$.

The next step is to simulate the K corresponding amplitudes $(A_{k,n})_{k\in\{1,...,K\}}$. The characteristic function of the amplitudes variable A is

$$\xi \mapsto \exp\left(\frac{1}{n}f(\xi)\right). \tag{5.29}$$

We refer to it as the nth root of the law of $\langle \mathrm{rect}_{[0,1]}, w\rangle$. Our assumption in this work is that there exists, for any $n \in \mathbb{N}$, a known method[4] to generate infinitely divisible variables with Lévy exponent $\frac{1}{n}f(\xi)$. For common parametric distributions such as α-stable, Laplace, and gamma distributions, such sampling methods[400] are well known and implemented in scientific computing libraries[5]. Simulating from their nth root is a simple matter of rescaling their parameters, as summarized in Table 5.2. By applying the correct rescaling, we simulate K independent amplitudes and thus obtain the sequence $(A_{k,n})_{k\in\{1,...,K\}}$.

[4] Workarounds exists for when a sampling method for the nth root $\frac{1}{n}f(\xi)$ is unavailable. For instance, one can opt for an approximate sampling scheme such as in [399].

[5] E. Jones, et al., "SciPy: Open source scientific tools for Python," 2001.

Table 5.2: The nth root of infinitely divisible distributions

Distribution	n**th Root**
Gaussian (μ, σ)	Gaussian $(\frac{\mu}{n}, \frac{\sigma}{\sqrt{n}})$
α-Stable (α, β, μ, c)	If $\alpha \neq 1$, $(\alpha, \beta, \frac{\mu}{n}, \frac{c}{n^{\frac{1}{\alpha}}})$,
	If $\alpha = 1$, $(\alpha, \beta, \frac{\mu}{n} - \frac{2}{\pi} c\beta \frac{\log(n)}{n}, \frac{c}{n})$
Gamma(α, β)	Gamma$(\frac{\alpha}{n}, \beta)$
Compound-Poisson of intensity λ	Compound-Poisson of intensity $\frac{\lambda}{n}$
Laplace (μ, b)	$X_n = \frac{\mu}{n} + b(G_1^{(n)} - G_2^{(n)})$
	with $G_1^{(n)}, G_1^{(n)} \sim$ Gamma$(\frac{1}{n}, 1)$

With the impulse locations $(\tau_{k,n})_{k \in \{1,\dots,K\}}$ and amplitudes $(A_{k,n})_{k \in \{1,\dots,K\}}$ in hand, we can compute samples of

$$s_n(\cdot) = \sum_{k=1}^{K} A_{k,n} \rho_{\mathrm{L}}(\cdot - \tau_{k,n}) + p_{0,n} \tag{5.30}$$

on a grid.

A direct approach to generate $\mathbf{s}_n$ is to use the expansion (5.30) and represent the process as a sum of shifted Green's functions. However in this case, the determination of $s_n(t)$ at any point $t \in [0, T]$ may require nontrivial computation of each and every term in (5.30). This stems from the fact that Green's functions are infinitely supported in general. There are therefore potential drawbacks to expansions in the basis of shifted Green's functions like (5.30). To overcome these issues, we propose instead an alternative method based on B-splines.

Recall that L is a rational operator of the form $P(\mathrm{D})Q(\mathrm{D})^{-1}$, where we take $\{\alpha_1, \dots, \alpha_{\deg(P)}\}$ to be the roots of P, with possible repetitions. Its discrete counterpart $\mathrm{L}_{\mathrm{d}}^{h}$ is defined as

$$\mathrm{L}_{\mathrm{d}}^{h}\{f\} = \sum_{m=0}^{\deg(P)} r[m] f(\cdot - mh), \tag{5.31}$$

where the sequence r is determined through its Fourier transform

$$R(\mathrm{e}^{\mathrm{j}\omega}) = \sum_{m=0}^{\deg(P)} r[m]\mathrm{e}^{-\mathrm{j}\omega m} = \prod_{m=1}^{\deg(P)} (1 - \mathrm{e}^{\alpha_m h}\mathrm{e}^{-\mathrm{j}\omega h}). \tag{5.32}$$

It is a finite impulse-response filter (FIR). Its null space contains the null space of L [296]. The function $\beta_{\mathrm{L}}^{h} := \mathrm{L}_{\mathrm{d}}^{h}\{\rho_{\mathrm{L}}\}$ is called the B-spline corresponding to L [7]. The B-spline has the fundamental property of being the shortest possible function within the space of cardinal L-splines (its support is included in $[0, \deg(P) \times h]$) [401, 18]. This will turn out to be crucial for the numerical efficiency of our method. Moreover, they reproduce both the Green's function and elements in the null space of their corresponding operator L [48, Section 6.4.].

The application of $\mathrm{L}_{\mathrm{d}}^{h}$ to s_n yields

$$u_n(t) = \mathrm{L}_{\mathrm{d}}^{h}\{s_n\}(t) = \sum_{m=0}^{\deg(P)} r[m] s_n(t - mh). \tag{5.33}$$

The process u_n in (5.33) is called the generalized increment process. Interestingly, it can be written as a sum of compactly supported terms, like

$$u_n(t) = \mathrm{L}_{\mathrm{d}}^{h}\{s_n\}(t) = \sum_{k=1}^{K} A_{k,n}\mathrm{L}_{\mathrm{d}}^{h}\{\rho_{\mathrm{L}}(\cdot - \tau_{k,n})\}(t) + \mathrm{L}_{\mathrm{d}}^{h}\{p_{0,n}\}(t) = \sum_{k=1}^{K} A_{k,n}\beta_{\mathrm{L}}^{h}(t - \tau_{k,n}) + 0. \tag{5.34}$$

The process u_n, along with boundary conditions, is our alternate representation of s_n. Now, let $\mathbf{u}_n$ be the vector whose components are $[\mathbf{u}_n]_i = u(ih)$, for $i = 1, ..., \left(\lceil \frac{T}{h} \rceil - 1\right)$. This vector can be computed more efficiently than $\mathbf{s}_n$ since the process u_n admits a representation with compactly supported terms. Moreover, $\mathbf{u}_n$ is linearly related to the vector $\mathbf{s}_n$ via a discrete system of difference equations. Indeed, we have that

$$[\mathbf{u}_n]_i = \sum_{m=0}^{\deg(P)} r[m][\mathbf{s}_n]_{i-m}, \tag{5.35}$$

for $\deg(P) \leq i \leq \left(\lceil \frac{T}{h} \rceil - 1\right)$. For $0 < i < \deg(P)$, we have that

$$[\mathbf{u}_n]_i = \sum_{m=0}^{\deg(P)} r[m]\mathbf{s}_n((i-m)h), \tag{5.36}$$

where the values $\mathbf{s}_n(-mh)$ for $m = 0, ..., (\deg(P) - 1)$ provide the boundary values. These relations are established by writing (5.33) with $t = ih$. The boundary values are determined by the null-space term $p_{0,n}$, which is itself determined by the boundary conditions.

Thus, once we have evaluated $\mathbf{u}_n$, we can obtain $\mathbf{s}_n$ by solving (5.35), which is accomplished by applying a recursive reverse filter to $\mathbf{u}_n$. This is performed by rewriting (5.35) as

$$[\mathbf{s}_n]_i = \frac{1}{r[0]}\left([\mathbf{u}_n]_i - \sum_{m=1}^{\deg(P)} r[m][\mathbf{s}_n]_{i-m}\right). \tag{5.37}$$

By substitution of the boundary values when necessary (*i.e.*, taking $\mathbf{s}_n((i-m)h)$ instead of $[\mathbf{s}_n]_{i-m}$ when $(i-m) \leq 0$), (5.37) allows one to recursively compute the components of $\mathbf{s}_n$.

We now describe an efficient procedure to compute the generalized increment process. The components of $\mathbf{u}_n$ are given by

$$[\mathbf{u}_n]_i = \sum_{k=0}^{K} A_{k,n}\beta_{\mathrm{L}}^{h}(ih - \tau_{k,n}). \tag{5.38}$$

The naive approach here would be to iterate through each grid point i independently and compute $[\mathbf{u}_n]_i$. Doing so would require one to read the entire sequence of impulse locations (τ_k) for each i. This cannot be avoided since there is no information on the sequence (τ_k), aside from its inclusion in $[0, T]$. We simply would not know *which* B-spline terms are inactive, so we would have to iterate through them all. A more efficient approach is to iterate through the list of impulses instead of the grid points.

The idea is as follows: First, initialize the vector $\mathbf{u}_n$ to zeros. Then, read the list of impulse locations one by one. For each impulse at τ_k, find the grid points that lie within the support of the B-spline at τ_k. Then, increment the value of $\mathbf{u}_n$ on those grid points by the contribution of the considered B-spline. In one pass over the list of impulses, this method computes the values $[\mathbf{u}_n]_i = \mathbf{u}_n(ih)$.

This intermediate computation of the generalized-increment process provides a considerable gain in terms of efficiency. Instead of having a number of operations that scales with $\lceil \frac{T}{h} \rceil \times K$ for the Green's function representation, we have one that scales with $\deg(P) \times \big(\lceil \frac{T}{h} \rceil + K\big)$.

Here is a summary of the procedure that generates trajectories of $\mathrm{L}s_n = w_n$. The pseudocode of our method is also provided in Algorithm 1.

1. First, fix the infinitely divisible distribution[6] that corresponds to w and define the operator L by identifying the polynomials P and Q.
2. Pick a sufficiently large value for n. Intuitively, n should be large enough to ensure the occurrence of several jumps in each bin. In other words, we expect n to be of the same order as h^{-1}. This has been validated with our numerical experiments as well, where we show that it provides a good approximation of the underlying statistics of the process (see, Figures 5.4, 5.5, and 5.6).
3. Pick a simulation interval $[0, T]$ and generate w_n. Determine an explicit form for ρ_{L}. At this point, the grid-free approximation s_n (expressed as in (5.30)) is available and can be stored.
4. Fix a grid on $[0, T]$ by choosing a step size h. Then determine the vector $\mathbf{s}_n$ with component $[\mathbf{s}_n]_i = s_n(ih)$. Compute the FIR filter $\mathrm{L}_{\mathrm{d}}^h$ and obtain $\beta_{\mathrm{L}}^h = \mathrm{L}_{\mathrm{d}}^h\{\rho_{\mathrm{L}}\}$. Then, compute the generalized increment vector $\mathbf{u}_n$.
5. To obtain $\mathbf{s}_n$, apply the reverse filter to $\mathbf{u}_n$ following (5.37). Take the values $\mathbf{s}_n(-mh)$ for $m = 0, ..., (\deg(P) - 1)$ to be zero for most cases except when L has a nontrivial null space, in which case it is derived from boundary conditions.

[6] The choice here is restricted to parametric families we can rescale and simulate.

```
Input  : Coefficients of $P$ and $Q$, approximation level $n$, interval size $T$,
         step size $h$
Output : Vector $\mathbf{s}_n$
Compute $\rho_{\mathrm{L}}$ and the FIR filter $r[m]$
Compute $\beta_L^h = \mathrm{L}_d^h\{\rho_{\mathrm{L}}\}$
Generate $[(\tau_1, A_1), ..., (\tau_K, A_K)]$
Initialize $\mathbf{u}_n$ with zeros as an array of size $\lceil \frac{T}{h} \rceil$
foreach $(\tau_k, A_k)$ do
    Find closest grid point $i_{\mathrm{grid}} = \lfloor \frac{\tau_i}{h} \rfloor$
    foreach $i$ in $\{i_{\mathrm{grid}}, \dots, i_{\mathrm{grid}} + \deg(P)\}$ do
        $[\mathbf{u}_n]_i \leftarrow [\mathbf{u}_n]_i + A_k \times \beta_L^h(ih - \tau_k)$
    end
end
Recursively apply a reverse filter to $\mathbf{u}_n$ following (5.37)
```

Algorithm 1: Procedure to obtain $\mathbf{s}_n$.

To show a merit of our proposed method, we prove that the generated approximations preserve the correlation structure of the target process. We first note that for any white Lévy noise w, we have that

$$w_n \xrightarrow{\mathcal{L}} w, \tag{5.39}$$

where the sequence of compound-Poisson innovations $(w_n)_{n\in\mathbb{N}}$ is defined in (5.20). We refer to this approximating sequence in Proposition 5.2.

Proposition 5.2. *Let w be a Lévy white noise such that $X_{\mathrm{rect}} = \langle \mathrm{rect}_{[0,1]}, w\rangle$ has zero mean and the finite variance $\sigma_w^2 = \mathbb{E}[X_{\mathrm{rect}}^2]$. Let $n \in \mathbb{N}$ and let w_n be a compound-Poisson innovation that approximates w as defined in* (5.20)*. Denoting $X_{\mathrm{rect},n} = \langle \mathrm{rect}_{[0,1]}, w_n\rangle$, we have that*

$$\mathbb{E}[X_{\mathrm{rect},n}] = \mathbb{E}[X_{\mathrm{rect}}] = 0 \tag{5.40}$$

and

$$\sigma_{w_n}^2 = \mathbb{E}[X_{\mathrm{rect},n}^2] = \mathbb{E}[X_{\mathrm{rect}}^2] = \sigma_w^2. \tag{5.41}$$

Proof. Since w_n is a compound-Poisson innovation, $X_{\mathrm{rect,\ n}}$ is a compound-Poisson random variable. It can be written

$$X_{\mathrm{rect,n}} = \sum_{i=1}^{N} A_i \tag{5.42}$$

where N is a Poisson random variable with rate $\lambda = n$ and the A_i are independent identically distributed infinitely divisible random variables with Lévy exponent $\frac{1}{n} f$ independent from N. We have by independence of the (A_i), that

$$X_{\mathrm{rect,n}} = \sum_{i=1}^{n} A_i =_d X_{\mathrm{rect}} \tag{5.43}$$

because the characteristic function of $\sum_{i=1}^{n} A_i$ is $(\mathrm{e}^{\frac{1}{n}})^n = \mathrm{e}^f$. This directly implies that $X_{\mathrm{rect,n}}$ and X_{rect} have the same moments. □

Now, if $s_n = \mathrm{L}^{-1} w_n$ is a generalized Poisson process that approximates $s = \mathrm{L}^{-1} w$, then

$$\begin{aligned}
\mathbb{E}[\langle \varphi_1, s_n \rangle \langle \varphi_2, s_n \rangle] &= \mathbb{E}[\langle \varphi_1, \mathrm{L}^{-1} w_n \rangle \langle \varphi_2, \mathrm{L}^{-1} w_n \rangle] \\
&= \mathbb{E}[\langle \mathrm{L}^{-1*} \varphi_1, w_n \rangle \langle \mathrm{L}^{-1*} \varphi_2, w_n \rangle] \\
&= \sigma_{w_n}^2 \langle \mathrm{L}^{-1*} \varphi_1, \mathrm{L}^{-1*} \varphi_2 \rangle \\
&= \sigma_{w}^2 \langle \mathrm{L}^{-1*} \varphi_1, \mathrm{L}^{-1*} \varphi_2 \rangle \\
&= \mathbb{E}[\langle \varphi_1, s \rangle \langle \varphi_2, s \rangle].
\end{aligned} \tag{5.44}$$

From (5.44), we concluded that, more than just approximated, the correlation structure is preserved *exactly* in our method.

5.2.4 Numerical Illustration

In this section, we validate our approach by conducting several numerical experiments. Let us also mention that a Python library that implements our algorithm can be

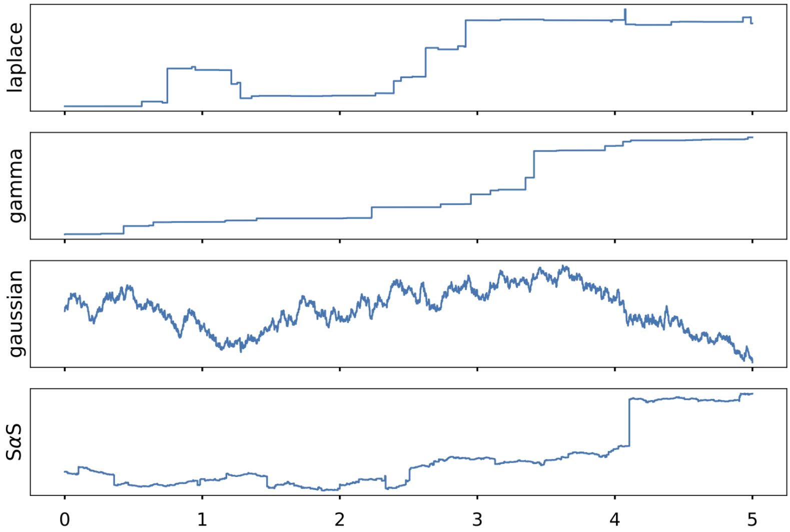


Figure 5.2: Trajectories of Lévy processes (L = D) with different innovations. From top to bottom: Laplace(0, 1), gamma(1, 1), Gaussian(0,1), and symmetric-α-stable with $\alpha = 1.23$.

found online[7]. Moreover, an accompanying web interface is also designed and is available [8].

Among all processes we can generate, those that are solutions to $\mathrm{D}s = w$ are called Lévy processes when the boundary condition is $s(0) = 0$. We showcase in Figure 5.2 different Lévy processes that correspond to several infinitely divisible distributions. For all four simulations, we took $n = 1{,}000$ and $h = 0.001$. As we demonstrate later, a reasonable choice for these parameter is to set nh to be a small integer (here, $nh = 1$). The visual appearance of the trajectories matches our expectations: The trajectory driven by a Gaussian innovation has the appearance of Brownian motion; the gamma Lévy process is nondecreasing.

[7] https://github.com/Biomedical-Imaging-Group/Generating-Sparse-Processes

[8] https://saturdaygenfo.pythonanywhere.com

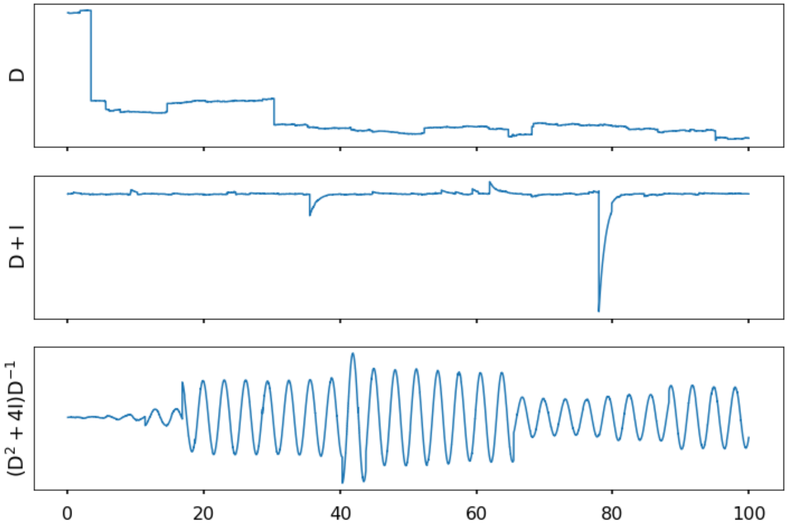


Figure 5.3: Trajectories of the solution s of $\mathrm{L}s = w$ for different operators L. In all cases, we considered a symmetric-α-stable white noise w with $\alpha = 1.23$.

Our framework allows for any rational operator of the form $P(\mathrm{D})Q(\mathrm{D})^{-1}$, so long as $\deg(P) > \deg(Q)$. In Figure 5.3, we generate trajectories of s that are solution of $\mathrm{L}s = w$, where w is a symmetric-α-stable innovation with $\alpha = 1.23$. Here we took $n = 200$ and $h = 0.001$. We see that, for various choices of L, the characteristics of the signal are markedly different, which exhibits the breadth of the modeling framework proposed in [48].

In Figure 5.4, we illustrate how an increase in n improves the approximation. In addition, we have depicted the convergence of moments in Figure 5.5. While the two figures emphasize the effect of n, they are insufficient to provide a quantitative way to choose n.

Here, we propose a measure that is based on the statistics of the generalized increment process. Since the process u_n is maximally decoupled, we can estimate

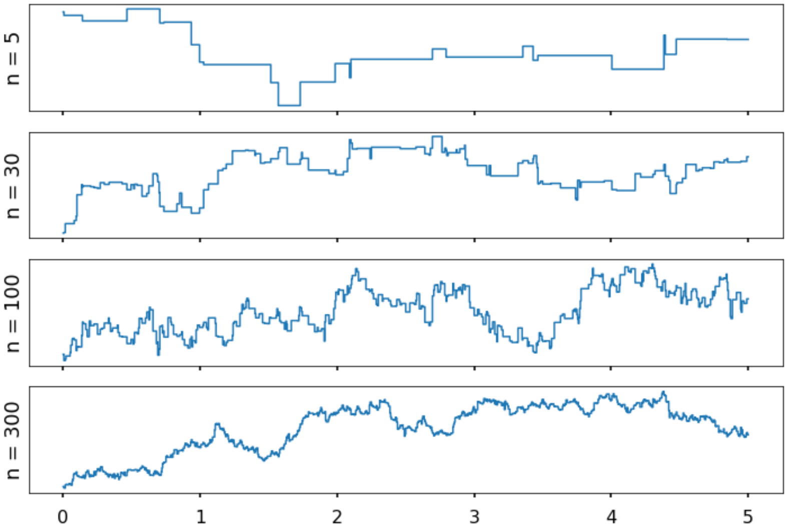


Figure 5.4: Approximations of Brownian motion (solution to $\mathrm{D}s = w$, with w a Gaussian white Lévy noise) as n increases.

the distribution of $U_n = \langle \beta_{\mathrm{L}}^{h\vee}, w_n \rangle$ from the samples $\{[\mathbf{u}_n]_i\}_i$ of the generalized increment process on the grid and obtain the empirical cumulative distribution function (CDF) $\tilde{F}_n(\cdot)$ of U_n. We then compare this empirical function to the reference CDF $F(\cdot)$ of $U = \langle \beta_{\mathrm{L}}^{h\vee}, w \rangle$. For the comparison, we use the Kolmogorov-Smirnov (KS) divergence [402] defined as

$$\mathcal{KS}(\tilde{F}_n, F) = \max_{x \in \mathbb{R}} |\tilde{F}_n(x) - F(x)|. \tag{5.45}$$

We then select n such that the KS-divergence is smaller than a certain threshold (*e.g.*, smaller than 0.1). The choice of the threshold is conditioned by the desired numerical precision: The lower the threshold, the more faithful the trajectories, but the higher the computational cost of the algorithm.

Intuitively, we expect that it is necessary to have several jumps in each bin in order to properly approximate the statistics of the process. The average number of

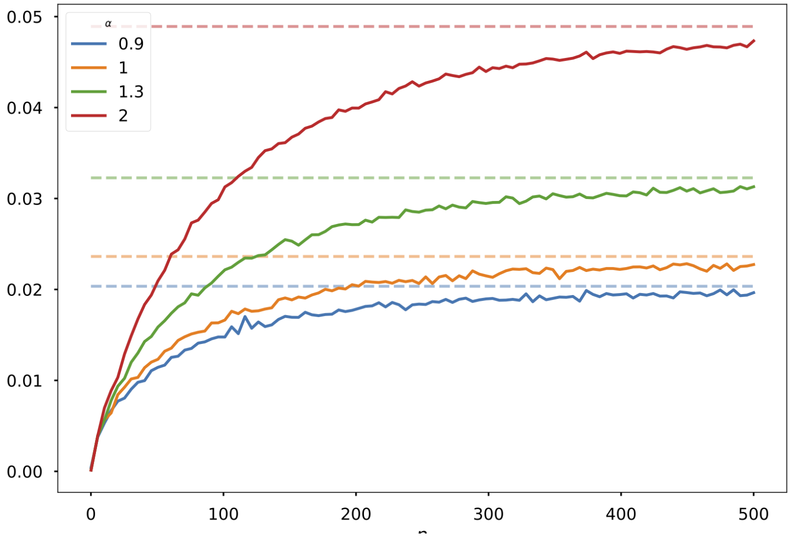


Figure 5.5: Convergence of $\mathbb{E}[|\langle \mathrm{rect}_{[0,h]}, s_n\rangle|^p]$ to $\mathbb{E}[|\langle \mathrm{rect}_{[0,h]}, s\rangle|^p]$ for $p = 0.4$, $h = 0.01$, and several symmetric-α-stable Lévy white noises w. The expectations are estimated with 10,000 trajectories for each n.

jumps in each bin of length h is $N_{\mathrm{jumps}} = nh$, so we expect n to be in the order of h^{-1}.

In Figure 5.6, we have validated this intuition by plotting the KS-divergence for different values of N_{jumps} in various settings. In all cases, as N_{jumps} increases, the KS-divergence decreases to a baseline error value, due to the finite-sample estimation of the underlying distribution.

Recall that a main motivation for our algorithm was to make it compatible with multi-grid methods. In our approach, the approximation s_n lives off the grid. It is only after the specification of the step size h that s_n is sampled on a grid. The generation of the random variables to determine s_n and the sampling on a grid are completely decoupled. This means that the same approximation s_n can be

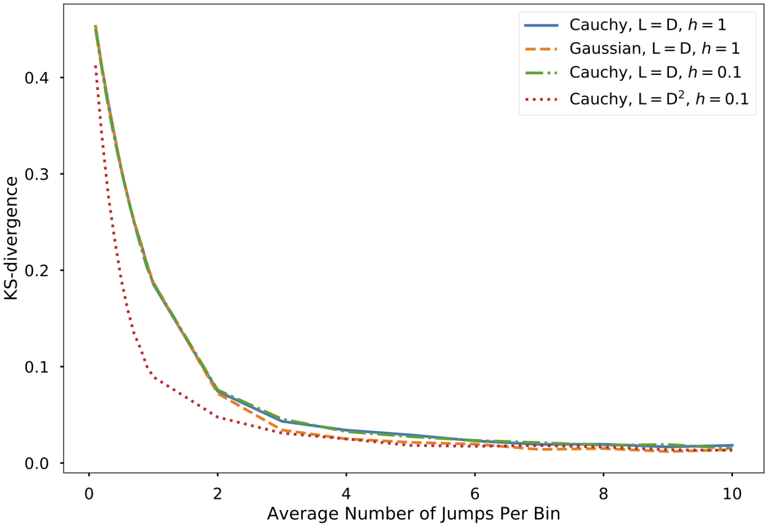


Figure 5.6: Kolmogorov-Smirnov (KS) divergence versus the average number of jumps per bin ($N_{\text{jumps}} = nh$).

viewed through different grids, which we illustrate in Figure 5.7. The solution to $(\mathrm{D}+1)^2 s = w$, where w is a Gaussian white Lévy noise, is first approximated by s_{1000}. Then, it is viewed on different regular grids on $[0, 1]$.

A crucial component of our approach is the computation of the generalized increment $\mathbf{u}$ in order to obtain the values of s_n on a grid. This provides a gain in numerical efficiency that can be felt even on moderately sized simulations. As can be seen in Figure 5.8, using a Green's function representation requires significantly more time than using an intermediate B-spline representation.

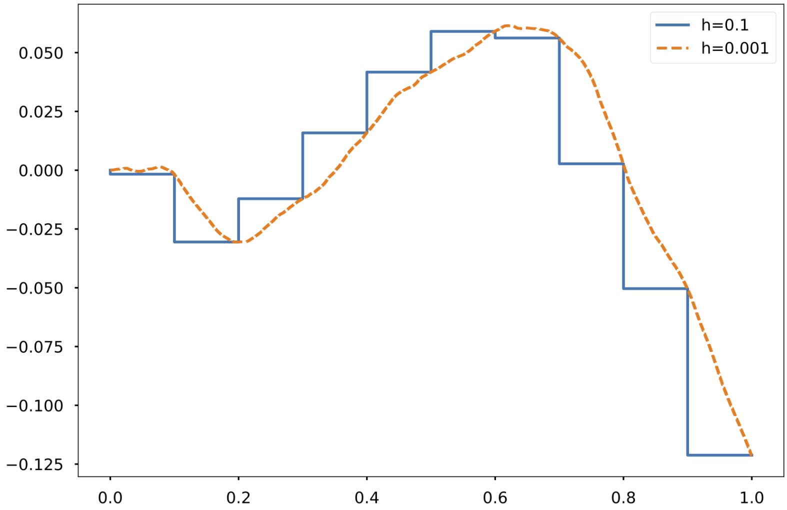


Figure 5.7: Single grid-free approximation sampled on grids that differ by their step size.

5.2.5 Summary

We have described a novel approach for generating sparse stochastic processes. Our method leverages the properties of B-splines to guarantee good numerical efficiency. We have illustrated numerically the merits of our proposed algorithm.

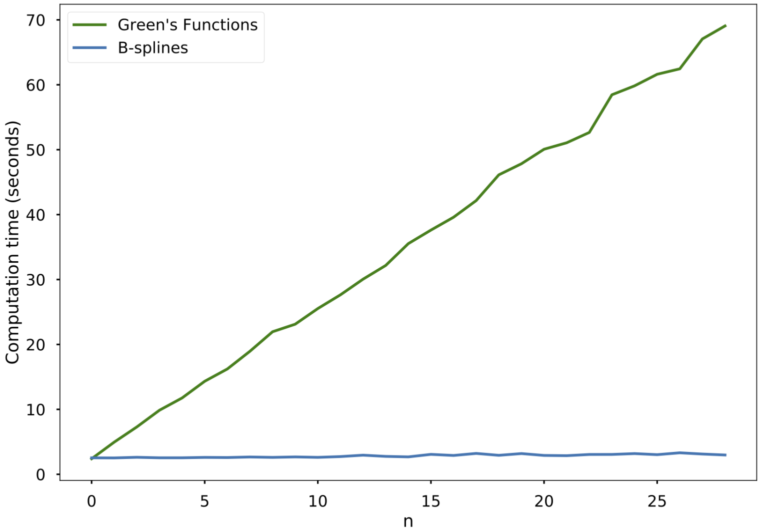


Figure 5.8: Average computation time for a trajectory of the solution of $(\mathrm{D}-0.5\mathrm{j})s = w$, where w a Gaussian white noise. The simulation interval is $[0, 1]$ with step size $h = 0.001$.

5.3 Besov Regularity of Lévy White Noises

In this section[9], we study the Besov regularity of Lévy white noises. More precisely, we identify the local smoothness and the asymptotic growth rate of the Lévy white noise. We do so by characterizing the weighted Besov spaces in which it is located. We extend known results in two ways. First, we obtain new bounds for the local smoothness via the Blumenthal-Getoor indices of the Lévy white noise. We also deduce the critical local smoothness when the two indices coincide, which is true for symmetric-α-stable, compound Poisson, and symmetric-gamma white noises to name a few. Second, we express the critical asymptotic growth rate in terms of the moment properties of the Lévy white noise. Previous analyses only provided lower

[9] This section is based on our published work [325].

bounds for both the local smoothness and the asymptotic growth rate. Showing the sharpness of these bounds requires us to determine in which Besov spaces a given Lévy white noise is (almost surely) *not*. Our methods are based on the wavelet-domain characterization of Besov spaces and precise moment estimates for the wavelet coefficients of the Lévy white noise.

5.3.1 Introduction and Summary of the Main Results

We study the Besov regularity of Lévy white noises. We are especially interested in identifying the critical local smoothness and the critical asymptotic growth rate of those random processes for any integrability parameter $p \in (0, \infty]$. In a nutshell, our contributions are as follows.

1. *Wavelet Methods for Lévy White Noises.* First appearing in the eighties, especially in the works of Y. Meyer [348], I. Daubechies [346], and S. Mallat [347], wavelet techniques have become primary tools in functional analysis [403]. As such, they are a natural choice to study random processes, as is done, for instance, with fractional Brownian motion [404], SαS processes [405], and with solutions of singular stochastic partial differential equations [406, 407]. In this work, we demonstrate that wavelet methods are also adapted to the analysis of the Lévy white noise. In particular, all our results are derived using the wavelet characterization of weighted Besov spaces.

2. *New Moment Estimates for Lévy White Noise.* The wavelet method allows us to obtain lower and upper bounds for the moments of a Lévy white noise as a function of the wavelet scale. Our moment estimates, which are new contributions to the rich literature on the moments of Lévy and Lévy-type processes [408, 409, 410, 411, 412], are fundamental for our study of the Besov regularity of the Lévy white noise.

3. *Besov Regularity of Lévy White Noises.* Regularity properties are usually stated in terms of the inclusion of the process in some weighted Besov spaces (positive result). In order to show that such a characterization is sharp, it is of interest to identify the smoothness spaces in which the process is *not included*

(negative result). To the best of our knowledge, very little is known in this direction. A precise answer to this question requires a more evolved analysis as compared to positive results. We achieve this goal thanks to the use of wavelets.

It is worth noting that our analysis requires the identification of a new index associated to a Lévy white noise, characterized by moment properties. By relying on this index, our negative results suggest moreover that some of the previous state-of-the-art inclusions are not sharp. We are then able to improve some of these results, in particular for the growth properties of the Lévy white noise.

4. *Critical Local Smoothness and Asymptotic Rate.* The combination of positive and negative results allows us to determine the critical Besov parameters of a Lévy white noise, both for the local smoothness and the asymptotic behavior. The results are summarized in Theorem 5.1. Two consequences are the characterization of the critical Sobolev and Hölder-Zigmund regularities of the Lévy white noise in Corollary 5.1.

Local Smoothness and Asymptotic Rate of Tempered Generalized Functions

Besov spaces are denoted by $B^{\tau}_{p,q}(\mathbb{R}^d)$, with $\tau \in \mathbb{R}$ the *smoothness*, $p \in (0, \infty]$ the *integrability parameter*, and $q \in (0, \infty]$ a secondary parameter. In this work, we focus on the case $p = q$ and we use the simplified notation $B^{\tau}_{p,p}(\mathbb{R}^d) = B^{\tau}_{p}(\mathbb{R}^d)$ for those spaces which are also referred to as Slobodeckij spaces after [413]. See [414] or [415, Section 2.2.1] for more details. We say that f is in the weighted Besov space $B^{\tau}_{p}(\mathbb{R}^d; \rho)$ with *weight exponent* $\rho \in \mathbb{R}$ if $\langle \cdot \rangle^{\rho} \times f$ is in the classic Besov space $B^{\tau}_{p}(\mathbb{R}^d)$, with the notation $\langle \boldsymbol{x} \rangle = (1 + \|\boldsymbol{x}\|^2)^{1/2}$. We precisely define weighted Besov spaces in Section 5.3.2 in terms of wavelet expansions. For the time being, it is sufficient to remember that the space of tempered generalized functions satisfies [416, Proposition 1]

$$\mathcal{S}'(\mathbb{R}^d) = \bigcup_{\tau, \rho \in \mathbb{R}} B^{\tau}_{p}(\mathbb{R}^d; \rho) \tag{5.46}$$

for any fixed $0 < p \leq \infty$. Ideally, we aim at identifying in which weighted Besov space a given $f \in \mathcal{S}'(\mathbb{R}^d)$ is. The relation (5.46) implies that, for any p, there exists some $\tau, \rho \in \mathbb{R}$ for which this is true. For a fixed p, Besov spaces are continuously embedded in the sense that, for $\tau, \tau_0, \tau_1 \in \mathbb{R}$ and $\rho, \rho_0, \rho_1 \in \mathbb{R}$ such that $\tau_0 \geq \tau_1$ and $\rho_0 \geq \rho_1$, we have

$$B_p^{\tau_0}(\mathbb{R}^d; \rho) \subseteq B_p^{\tau_1}(\mathbb{R}^d; \rho) \quad \text{and} \quad B_p^{\tau}(\mathbb{R}^d; \rho_0) \subseteq B_p^{\tau}(\mathbb{R}^d; \rho_1). \tag{5.47}$$

To characterize the properties of $f \in \mathcal{S}'(\mathbb{R}^d)$, the key is to determine the two critical exponents $\tau_p(f) \in (-\infty, \infty]$ and $\rho_p(f) \in (-\infty, \infty]$ such that

- if $\tau < \tau_p(f)$ and $\rho < \rho_p(f)$, then $f \in B_p^{\tau}(\mathbb{R}^d; \rho)$; while
- if $\tau > \tau_p(f)$ or $\rho > \rho_p(f)$, then $f \notin B_p^{\tau}(\mathbb{R}^d; \rho)$.

The case $\tau_p(f) = \infty$ corresponds to infinitely smooth functions, and $\rho_p(f) = \infty$ means that f is rapidly decaying. The quantity $\tau_p(f)$ measures the *local smoothness* and $\rho_p(f)$ the *asymptotic rate* of f for the integrability parameter p. When $\rho_p(f) \leq 0$ (which will be the case for the Lévy white noise), we talk about the *asymptotic growth rate* of f.

Local Smoothness and Asymptotic Growth Rate of Lévy White Noises

Our main contributions concern the inclusion of a Lévy white noise w in weighted Besov spaces. It includes positive (w is almost surely *in* a given Besov space) and negative (w is almost surely *not in* a given Besov space) results. In order to characterize the local smoothness $\tau_p(w)$ and the asymptotic growth rate $\rho_p(w)$, let Ψ denotes the Lévy exponent of w. We associate to w, its *Blumenthal-Getoor indices*, defined as

$$\beta_\infty = \inf \left\{ p > 0 \;\middle|\; \lim_{|\xi| \to \infty} \frac{|\Psi(\xi)|}{|\xi|^p} = 0 \right\}, \tag{5.48}$$

$$\underline{\beta}_\infty = \inf \left\{ p > 0 \;\middle|\; \liminf_{|\xi| \to \infty} \frac{|\Psi(\xi)|}{|\xi|^p} = 0 \right\}. \tag{5.49}$$

The distinction is that β_∞ considers the limit, while $\underline{\beta}_\infty$ deals with the inferior limit. In general, one has that $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$. The Blumenthal-Getoor indices are linked to the local behavior of Lévy processes and Lévy white noises (see Section 5.3.2 for more details). In addition, we introduce the *moment index* of the Lévy white noise w as

$$p_{\max} = \sup \left\{ p > 0 \mid \mathbb{E}[|\langle w, \mathbb{1}_{[0,1]^d} \rangle|^p] < \infty \right\}, \tag{5.50}$$

which is closely related—but in general not identical—to the Pruitt index (see Section 5.3.2). As we shall see, $p_{\max} \in (0, \infty]$ fully characterizes the asymptotic growth rate of w. The class of Lévy white noises is rich, and includes Gaussian and compound Poisson white noises. We summarize the results of this work in Theorem 5.1. We use the convention that $1/p = 0$ when $p = \infty$.

Theorem 5.1. *Consider a Lévy white noise w with Blumenthal-Getoor indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$ and moment index $0 < p_{\max} \leq \infty$. We fix $0 < p \leq \infty$.*

- *If w is a Gaussian white noise, then, almost surely,*

$$\tau_p(w) = -\frac{d}{2} \text{ and } \rho_p(w) = -\frac{d}{p}. \tag{5.51}$$

- *If w is a compound Poisson white noise, then, almost surely,*

$$\tau_p(w) = \frac{d}{p} - d \text{ and } \rho_p(w) = -\frac{d}{\min(p, p_{\max})}. \tag{5.52}$$

- *If w is a Lévy white noise and not a Gaussian white noise, then, almost surely,*

$$\frac{d}{\max(p, \beta_\infty)} - d \leq \tau_p(w) \leq \frac{d}{\max(p, \underline{\beta}_\infty)} - d. \tag{5.53}$$

  In particular, if $p \geq \beta_\infty$, then $\tau_p(w) = d/p - d$.

- *If w is a Lévy white noise and not a Gaussian white noise, then, almost surely, if $p \in (0, 2)$, p is an even integer, or $p = \infty$,*

$$\rho_p(w) = -\frac{d}{\min(p, p_{\max})}, \tag{5.54}$$

and for any $0 < p \leq \infty$,

$$\rho_p(w) \geq -\frac{d}{\min(p, p_{\max})}. \tag{5.55}$$

In a nutshell, Theorem 5.1 provides:

1. A full characterization of the local smoothness and the asymptotic growth rate for Gaussian and compound Poisson white noises;
2. A characterization of the asymptotic growth rate for any Lévy white noise for integrability parameter $p \leq 2$, p an even integer, or $p = \infty$; and
3. A full characterization of the local smoothness of a Lévy white noise for which $\underline{\beta}_\infty = \beta_\infty$; that is, for any $p \in (0, \infty]$,

$$\tau_p(w) = \frac{d}{\max(p, \beta_\infty)} - d. \tag{5.56}$$

We discuss the remaining cases—the local smoothness for $\underline{\beta}_\infty < \beta_\infty$ and the asymptotic growth rate for $p > 2, p \notin 2\mathbb{N}$—in Section 5.3.9. Two direct consequences are the identification of the Sobolev ($p = 2$) and Hölder-Zigmund ($p = \infty$) regularity of Lévy white noises.

Corollary 5.1. *Let w be a Lévy white noise in $\mathcal{S}'(\mathbb{R}^d)$ with Blumenthal-Getoor indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$ and moment index $0 < p_{\max} \leq \infty$. Then, the Sobolev local smoothness and asymptotic growth rate ($p = 2$) are*

$$\tau_2(w) = -\frac{d}{2} \quad \textit{and} \quad \rho_2(w) = -\frac{d}{\min(2, p_{\max})}. \tag{5.57}$$

Moreover, the Hölder-Zigmund local smoothness and asymptotic growth rate ($p = \infty$) are

$$\tau_\infty(w) = -\frac{d}{2} \quad \textit{and} \quad \rho_\infty(w) = 0 \quad \textit{if } w \textit{ is Gaussian, and} \tag{5.58}$$

$$\tau_\infty(w) = -d \quad \textit{and} \quad \rho_\infty(w) = -\frac{d}{p_{\max}} \quad \textit{otherwise.} \tag{5.59}$$

Proof. The case $p = 2$ is directly deduced from Theorem 5.1 and the relation $\underline{\beta}_\infty \leq \beta_\infty \leq 2$. For $p = \infty$, we use again Theorem 5.1 with $p = \infty$. The Gaussian case follows from (5.51). For the non-Gaussian case, (5.53) gives $\tau_\infty(w)$, while (5.54) gives $\rho_\infty(w)$. □

Local Smoothness of Lévy Processes

It is worth noting that the Sobolev regularity is the same—$\tau_2(w) = -d/2$—for any Lévy white noise. We also observe that the Hölder-Zigmund regularity of any non-Gaussian Lévy white noise is $(-d)$ (which is also the regularity of a Dirac impulse), the Gaussian case being different and reaching a smoothness of $(-d/2)$. In the one-dimensional setting ($d = 1$), this is reminiscent to the fact that the Brownian motion is the only continuous random process with independent and stationary increments, the other Lévy processes being only *càdlàg* (French acronym for functions that are right continuous with left limits at every points) [417]. Using Theorem 5.1, we deduce the local smoothness of Lévy processes in Corollary 5.2.

Corollary 5.2. *Let $s : \mathbb{R} \to \mathbb{R}$ be a one-dimensional Lévy process with Blumenthal-Getoor indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$. Then, we have almost surely that, for any $0 < p \leq \infty$,*

$$\tau_p(s) = \frac{1}{2} \quad \textit{if } s \textit{ is the Wiener process, and} \tag{5.60}$$

$$\tau_p(s) = \frac{1}{p} \quad \textit{if } s \textit{ is a compound Poisson process.} \tag{5.61}$$

In the general case, we have almost surely that, for any $0 < p \leq \infty$,

$$\frac{1}{\max(p, \beta_\infty)} \leq \tau_p(s) \leq \frac{1}{\max(p, \underline{\beta}_\infty)}. \tag{5.62}$$

Proof. A one-dimensional Lévy white noise w is the weak derivative of the corresponding Lévy process s with identical characteristic exponent. This well-known fact has been rigorously shown in the sense of generalized random processes in [379, Definition 3.4 and Proposition 3.17]. A direct consequence is that $\tau_p(s) = \tau_p(w) + 1$, where $w = s'$. Then, Corollary 5.2 is a reformulation of the local smoothness results of Theorem 5.1 with $d = 1$. □

Related Works on Lévy Processes and Lévy White Noises

In this section, for comparison purposes, we reinterpret all the results in terms of the critical smoothness and asymptotic growth rate of the considered random processes.

Lévy Processes. Most of the attention has been so far devoted to classic Lévy processes. The Wiener process was studied in [418, 419, 420, 421, 422, 423], while [419] also contains results on the Besov regularity of fractional Brownian motions and SαS processes. By exploiting the self-similarity of the stable processes, Ciesielski *et al.* obtained the following results for the Gaussian [419, Theorem IV.3] and stable non-Gaussian [419, Theorem VI.1] scenarios:

$$\frac{1}{2} \leq \tau_p(s_{\mathrm{Gauss}}) \leq \frac{1}{\min(2,p)}, \tag{5.63}$$

$$\frac{1}{\max(p,\alpha)} \leq \tau_p(s_\alpha) \leq \frac{1}{\min(p,\alpha)}, \tag{5.64}$$

for any $p \geq 1$, where s_{Gauss} is the Brownian motion and s_α is the SαS process with parameter $1 < \alpha < 2$.

The complete family of Lévy processes—and, more generally, of Lévy-type processes—has been considered by R. Schilling in a series of papers [424, 425, 426] synthesized in [427, Chapter V] and by V. Herren [428]. To summarize, Schilling has shown that, for a Lévy process s with indices $0 \leq \beta_\infty \leq 2$ and $0 < p_{\max} \leq \infty$,

$$\frac{1}{\max(p,\beta_\infty)} \leq \tau_p(s) \leq \frac{1}{p}, \tag{5.65}$$

$$-\frac{1}{p} - \frac{1}{\min(p_{\max},2)} \leq \rho_p(s). \tag{5.66}$$

We observe that (5.63), (5.64), and (5.65) are consistent with Corollary 5.2. Moreover, our results provide an improvement by showing that the lower bounds of (5.63) and (5.64) are actually sharp. Finally, we significantly improve the upper bound of (5.65) for general Lévy processes.

In contrast to the smoothness, the growth rate (5.66) of the Lévy process s does not seem to be related to the one of its derivative the Lévy white noise $w = s'$ by a simple relation. In particular, the rate of s is expressed in terms of the Pruitt index $\beta_0 = \min(p_{\max}, 2)$, conversely to $p_{\max}$ for w (see Section 5.3.2). This needs to be confirmed by a precise estimation of $\rho_p(s)$ for which only a lower bound is known.

Lévy White Noises. M. Veraar extensively studied the local Besov regularity of the d-dimensional Gaussian white noise. As a corollary of [429, Theorem 3.4], one then deduces that $\tau_p(w_{\text{Gauss}}) = -d/2$. This work is based on the specific properties of the Fourier series expansion of the random process under the Gaussianity assumption, and cannot be directly adapted to Lévy white noises.

In our own works, we have investigated the question for general Lévy white noises in dimension d in the periodic [430] and global settings [431]. We obtained the lower bounds

$$\begin{aligned} \frac{d}{\max(p, \beta_\infty)} - d \leq& \tau_p(w), \qquad \text{and} && (5.67) \\ -\frac{d}{\min(p, p_{\max}, 2)} \leq& \rho_p(w). && (5.68) \end{aligned}$$

These estimates are improved by Theorem 5.1, which provides an upper bound for $\tau_p(w)$ and shows that (5.67) is sharp when $\beta_\infty = \underline{\beta}_\infty$. It is also worth noticing that the lower bound of (5.68) is sharp if and only if $p_{\max} \leq 2$.

Sketch of Proof and the Role of Wavelet Methods

Our techniques are based on the wavelet characterization of Besov spaces, as presented by H. Triebel in [403]. We shall see that wavelets are especially relevant to the analysis of Lévy white noises.

We briefly present the strategy of the proof of Theorem 5.1 when the ambiant dimension is $d = 1$. The general case $d \geq 1$ is analogous and will be comprehensively addressed in the rest of the section. Let (ψ_M, ψ_F) be the (mother, father) Daubechies wavelets of a fixed order (the choice of the order has no influence on the results as soon as it is large enough). For $j \in \mathbb{N}$ and $k \in \mathbb{Z}$, we define the rescaled and shifted functions $\psi_{F,k} = \psi_F(\cdot - k)$ and $\psi_{j,M,k} = 2^{j/2}\psi_M(2^j \cdot - k)$. Then, the family

$(\psi_{F,k})_{k\in\mathbb{Z}} \cup (\psi_{j,M,k})_{j\in\mathbb{N},k\in\mathbb{Z}}$ forms an orthonormal basis of $L_2(\mathbb{R})$ [432]. For a given one-dimensional Lévy white noise w, one considers the family of random variables

$$((\langle w, \psi_{F,k}\rangle))_{k\in\mathbb{Z}} \cup ((\langle w, \psi_{j,M,k}\rangle))_{j\in\mathbb{N},k\in\mathbb{Z}}. \tag{5.69}$$

We then have that $w = \sum_{k\in\mathbb{Z}}\langle w, \psi_{F,k}\rangle\psi_{F,k} + \sum_{j\in\mathbb{N}}\sum_{k\in\mathbb{Z}}\langle w, \psi_{j,M,k}\rangle\psi_{j,M,k}$, where the convergence is almost sure in $\mathcal{S}'(\mathbb{R})$.

Then, for $0 < p < \infty$ (the case $p = \infty$ will be deduced by embedding and is not discussed in this section) and $\tau, \rho \in \mathbb{R}$, the random variable

$$\|w\|_{B_p^\tau(\mathbb{R};\rho)} = \left(\sum_{k\in\mathbb{Z}}\langle k\rangle^{\rho p}|\langle w, \psi_{F,k}\rangle|^p + \sum_{j\in\mathbb{N}} 2^{j(\tau p-1+\frac{p}{2})}\sum_{k\in\mathbb{Z}}\langle 2^{-j}k\rangle^{\rho p}|\langle w, \psi_{j,M,k}\rangle|^p\right)^{1/p} \tag{5.70}$$

is well-defined and takes values in $[0,\infty]$. Here, $\|w\|_{B_p^\tau(\mathbb{R};\rho)}$ is the Besov (quasi-)norm of the Lévy white noise (see Section 5.3.2). This means that w is a.s. (almost surely) in $B_p^\tau(\mathbb{R};\rho)$ if and only if $\|w\|_{B_p^\tau(\mathbb{R};\rho)} < \infty$ a.s., and a.s. not in $B_p^\tau(\mathbb{R};\rho)$ if and only if $\|w\|_{B_p^\tau(\mathbb{R};\rho)} = \infty$ a.s.

We then fix $0 < p < \infty$. We assume that we have guessed the values $\tau_p(w)$ and $\rho_p(w)$ introduced in Section 5.3.1. Here are the main steps leading to the proof that these values are effectively the critical ones.

- For $\tau < \tau_p(w)$ and $\rho < \rho_p(w)$, we show that $\|w\|_{B_p^\tau(\mathbb{R};\rho)} < \infty$ a.s. For $p < p_{\max}$ (see (5.50)), we establish the stronger result $\mathbb{E}\left[\|w\|^p_{B_p^\tau(\mathbb{R};\rho)}\right] < \infty$. This requires moment estimates for the wavelet coefficients of a Lévy white noise, which gives a precise estimation of the behavior of $\mathbb{E}[|\langle w, \psi_{j,M,k}\rangle|^p]$ as j goes to infinity. When $p > p_{\max}$, the random variables $\langle w, \psi_{j,M,k}\rangle$ have an infinite pth moment and the present method is not applicable. In that case, we actually deduce the result using embedding relations between Besov spaces. It turns out that this approach is sufficient to obtain sharp results.
- For $\tau > \tau_p(w)$, we show that $\|w\|_{B_p^\tau(\mathbb{R};\rho)} = \infty$ a.s. To do so, we only consider

the mother wavelet and truncate the sum over k to yield the lower bound

$$\|w\|^p_{B^\tau_p(\mathbb{R};\rho)} \geq C \sum_{j\in\mathbb{N}} 2^{j(\tau p - 1 + \frac{p}{2})} \sum_{0\leq k<2^j} |\langle w, \psi_{j,M,k}\rangle|^p \tag{5.71}$$

for some constant C such that $\langle 2^{-j}k\rangle^{\rho p} \geq C$ for every $j \in \mathbb{N}$ and $0 \leq k < 2^j$. We then need to show that the wavelet coefficients $\langle w, \psi_{j,M,k}\rangle$ cannot be too small altogether using Borel-Cantelli-type arguments. Typically, this requires us to control the evolution of quantities such as $\mathbb{P}(|\langle w, \psi_{j,M,k}\rangle| > x)$ with respect to j and is again based on moment estimates.

- For $\rho > \rho_p(w)$, we show again that $\|w\|_{B^\tau_p(\mathbb{R};\rho)} = \infty$ a.s. This time, we only consider the father wavelet in (5.70) and use the lower bound

$$\|w\|^p_{B^\tau_p(\mathbb{R};\rho)} \geq \sum_{k\in\mathbb{Z}} \langle k\rangle^{\rho p} |\langle w, \psi_{F,k}\rangle|^p. \tag{5.72}$$

A Borel-Cantelli-type argument is again used to show that the $|\langle w, \psi_{F,k}\rangle|$ cannot be too small altogether, and that the Besov norm is a.s. infinite.

5.3.2 Mathematical Background

The Lévy-Itô Decomposition of Lévy White Noises

The Lévy-Itô decomposition is a fundamental result of the theory of Lévy processes. It reveals that a Lévy process $s = (s(t))_{t\in\mathbb{R}}$ can be decomposed as $s = s_1 + s_2 + s_3$, where s_1 is a Wiener process, s_2 is a compound Poisson process, and s_3 is a square integrable pure jump martingale, which corresponds to the small jumps of s [433, Theorem 2.4.16], [170, Theorems 19.2 and 19.3]. The extension of the Lévy-Itô decomposition to the multivariate setting requires to define Lévy fields, for which different constructions are possible [385, 434, 435]. This includes Lévy sheets, that we already mentioned and for which the Lévy-Itô decomposition has been extended for Lévy sheets in [436, Theorem 4.6]. Using the Lévy-Itô decomposition of Lévy sheets, we are able to provide an identical result for the Lévy white noise. This is based on the connection between Lévy sheets and Lévy white noises, which is one of the main contribution of [379]. Indeed, the Lévy white noise w satisfies the relation

$$\mathrm{D}_1 \dots \mathrm{D}_d\{s\} = w \tag{5.73}$$

for some Lévy sheet s in $\mathcal{S}'(\mathbb{R}^d)$ [379, Defintion 3.4 and Proposition 3.17]. In dimension $d = 1$, we recover that the (weak) derivative of the Lévy process is a Lévy white noise.

Proposition 5.3. *A Lévy white noise w can be decomposed as*

$$w = w_1 + w_2 + w_3 \tag{5.74}$$

with w_1 a Gaussian white noise, w_2 a compound Poisson white noise, and w_3 a Lévy white noise with finite moments, the three being independent.

Proof. According to (5.73), $w = \mathrm{D}_1 \dots \mathrm{D}_d\{s\}$ for some Lévy sheet $s : \mathbb{R}^d \to \mathbb{R}$. Then, according to [436, Theorem 4.6], s can be decomposed as $s = s_1 + s_2 + s_3$ where s_1 is a Brownian sheet, s_2 is a compound Poisson sheet, and s_3 is Lévy sheet which is a square integrable pure jump martingale (s_3 corresponds to the small jumps of s). Moreover, the three random fields s_1, s_2, s_3 are independent from each other. Then, the jumps of s_3 are bounded by construction, implying that it has finite moments [433, Theorem 2.4.7]. Finally, we have that

$$w = \mathrm{D}_1 \dots \mathrm{D}_d\{s\} = \mathrm{D}_1 \dots \mathrm{D}_d\{s_1\} + \mathrm{D}_1 \dots \mathrm{D}_d\{s_2\} + \mathrm{D}_1 \dots \mathrm{D}_d\{s_3\} := w_1 + w_2 + w_3, \tag{5.75}$$

where w_1 is a Gaussian white noise, w_2 is a compound Poisson white noise, and w_3 is a Lévy white noise with finite moments. This last point is indeed ensured by the fact that the Lévy measure ν_3 associated to s_3 and therefore w_3 has a compact support. Hence, we have that $\int_{\mathbb{R}} |t|^p \mathrm{d}\nu_3(t) < \infty$ for any $p > 0$. This implies that $\mathbb{E}[|\langle w_3, \varphi\rangle|^p] < \infty$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $p > 0$ according to [170, Theorem 25.3] (see also Proposition 5.4 thereafter). Note moreover that w_1, w_2, and w_3 are independent, because the corresponding Lévy sheets are. □

Indices of Lévy White Noises

We introduce various indices associated to Lévy white noises. First of all, we exclude Lévy white noises with dominant drift via the following classic notion that appears for instance in [408, 425].

Definition 5.4. *We say that a Lévy white noise w with characteristic exponent Ψ satisfies the* sector condition *if there exists $M > 0$ such that*

$$\forall \xi \in \mathbb{R}, \quad |\Im\{\Psi(\xi)\}| \leq M \, |\Re\{\Psi(\xi)\}| \,. \tag{5.76}$$

This condition ensures that no drift is dominating the Lévy white noise. For instance, the deterministic Lévy white noise $w = \mu \neq 0$ a.s., which corresponds to the Lévy triplet $(\mu, 0, 0)$, is such that $\Psi(\xi) = \mathrm{j}\mu\xi$ and does not satisfy the sector condition. This is also the case for $w = \mu + w_\alpha$ where w_α is a SαS process with $\alpha \in (0, 2]$. It is worth noting that the characteristic exponent of a symmetric Lévy white noise is real, and therefore satisfies the sector condition. In the rest of this section, we will always assume that the sector condition is satisfied without further mention.

In Theorem 5.1, the smoothness and growth rate of Lévy white noises is characterized in terms of the indices (5.48), (5.49), and (5.50). We give here some additional insight about these quantities. The index β_∞ was introduced by R. Blumenthal and R. Getoor [437] to characterize the behavior of Lévy processes at the origin. This quantity appears to be related to many local properties of random processes driven by Lévy white noises, including the Hausdorff dimension of the image set [427], the spectrum of singularities [434, 438], the Besov regularity [427, 431, 424, 426] and more generally sample path properties [439, 440, 441], the local self-similarity [442], or the local compressiblity [443]. Finally, the index $\underline{\beta}_\infty$ plays a crucial role in the specification of negative results, such as the identification of the Besov spaces in which the Lévy white noises are not. It satisfies moreover the relation $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$.

In [444], W. Pruitt proposed the index

$$\beta_0 = \sup \left\{ p > 0 \;\middle|\; \lim_{|\xi| \to 0} \frac{|\Psi(\xi)|}{|\xi|^p} = 0 \right\} \tag{5.77}$$

as the asymptotic counterpart of β_∞. This quantity appears in the asymptotic growth rate of the supremum of Lévy(-type) processes [425] and the asymptotic self-similarity of random processes driven by Lévy white noises [442]. The Pruitt index differs from the index $p_{\max}$ that appears in Theorem 5.1. Actually, the two quantities are linked by the relation $\beta_0 = \min(p_{\max}, 2)$. This is shown by linking β_0 to the Lévy measure [427] and knowing that $\beta_0 \leq 2$ (see the appendix of [408] for a short and elegant proof). This means that $\beta_0 < p_{\max}$ when the Lévy white noise has some finite pth moments fo $p > 2$, and one cannot recover $p_{\max}$ from β_0 in this case. It is therefore necessary to introduce the index $p_{\max}$ in addition to the Pruitt index in our analysis. Note moreover that the moment index fully characterizes the moment properties of the Lévy white noise in the following sense.

Proposition 5.4. *Let $0 < p < \infty$ and w be a Lévy white noise with moment index $p_{\max} \in (0, \infty]$. We also fix a compactly supported and bounded test function $\varphi \neq 0$. If $p < p_{\max}$, then*

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] < \infty, \tag{5.78}$$

while if $p > p_{\max}$, then

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] = \infty. \tag{5.79}$$

Proposition 5.4 can be deduced from more general results presented in [383] and [382], where the set of test functions φ such that $\mathbb{E}[|\langle w, \varphi\rangle|^p] < \infty$ is fully characterized. For us, it is enough to know that the result is true for compactly supported bounded test functions, which includes Daubechies wavelets. We provide a proof thereafter for the sake of completeness, since this result is not exactly stated as such in the literature and known results require to introduce tools that are unnecessary for this work. The first part (5.78) allows one to consider the moments of $\langle w, \varphi\rangle$ for any $p < p_{\max}$. The second part (5.79) shows that some moments are infinite and will appear to be useful later on.

Proof of Proposition 5.4. The proof relies on the link between the moments of w and the moments of its Lévy measure ν. According to [170, Theorem 25.3], a random variable X with Lévy measure ν is such that

$$\mathbb{E}[|X|^p] < \infty \iff \int_{|t|>1} |t|^p \mathrm{d}\nu(t) < \infty. \tag{5.80}$$

Applying this to $X = \langle w, \mathbb{1}_{[0,1)^d} \rangle$ (whose Lévy measure is indeed ν)) and using the definition of $p_{\max}$ in (5.50), we deduce that

$$\int_{|t|>1} |t|^p \mathrm{d}\nu(t) < \infty \quad \text{if } p < p_{\max}, \text{ and } \int_{|t|>1} |t|^p \mathrm{d}\nu(t) = \infty \text{ if } p > p_{\max}. \tag{5.81}$$

According to [382, Proposition 3.14], we have that $\mathbb{E}[|\langle w, \varphi \rangle|^p] < \infty$ if and only if

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}} |t\varphi(\boldsymbol{x})|^p \mathbb{1}_{|t\varphi(\boldsymbol{x})|>1} \mathrm{d}\nu(t) \mathrm{d}\boldsymbol{x} < \infty. \tag{5.82}$$

Let K be the compact support of φ. Moreover, the test function being non identically zero, there exists $m > 0$ such that $\mathrm{Leb}(\{|\varphi| \geq m\}) > 0$ where Leb is the Lebesgue measure. Set $p < p_{\max}$, then for every $t \in \mathbb{R}$ and every $\boldsymbol{x} \in K$, we have that $\mathbb{1}_{|t\varphi(\boldsymbol{x})|>1} \leq \mathbb{1}_{|t| \|\varphi\|_\infty > 1}$. Therefore, according to the left side of (5.81),

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}} |t\varphi(\boldsymbol{x})|^p \mathbb{1}_{|t\varphi(\boldsymbol{x})|>1} \mathrm{d}\nu(t) \mathrm{d}\boldsymbol{x} &\leq \int_K \int_{\mathbb{R}} |t|^p \|\varphi\|_\infty^p \mathbb{1}_{|t| \|\varphi\|_\infty > 1} \mathrm{d}\nu(t) \mathrm{d}\boldsymbol{x} \\ &= \mathrm{Leb}(K) \|\varphi\|_\infty^p \int_{|t|>1/\|\varphi\|_\infty} |t|^p \mathrm{d}\nu(t) < \infty, \end{aligned} \tag{5.83}$$

proving (5.78). Moreover, if $p > p_{\max}$, then, using the right side of (5.81) and the inequality $\mathbb{1}_{|t\varphi(\boldsymbol{x})|>1} \geq \mathbb{1}_{|t|m>1}$ for every $\boldsymbol{x} \in \{|\varphi| \geq m\}$, we have

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}} |t\varphi(\boldsymbol{x})|^p \mathbb{1}_{|t\varphi(\boldsymbol{x})|>1} \mathrm{d}\nu(t) \mathrm{d}\boldsymbol{x} &\geq \int_{|\varphi| \geq m} \int_{\mathbb{R}} |t|^p m^p \mathbb{1}_{|t|m>1} \mathrm{d}\nu(t) \mathrm{d}\boldsymbol{x} \\ &= \mathrm{Leb}(\{|\varphi| \geq m\}) m^p \int_{|t|>1/m} |t|^p \mathrm{d}\nu(t) = \infty, \end{aligned} \tag{5.84}$$

and (5.79) is proved. □

We now summarize how the indices of the Lévy-Itô decomposition of a Lévy white noise behave in Proposition 5.5.

Proposition 5.5. *Let w be a Lévy white noise w and let $w = w_1 + w_2 + w_3$ be its Lévy-Itô decomposition according to* (5.74) *in Proposition 5.3, where w_1 is Gaussian, w_2 is compound Poisson, and w_3 have finite moments.*

(i) If w_1, w_2, and w_3 are nonzero, then

$$\underline{\beta}_\infty(w_1) = \beta_\infty(w_1) = 2, \quad \underline{\beta}_\infty(w_2) = \beta_\infty(w_2) = 0, \quad \textit{and} \quad p_{\max}(w_1) = p_{\max}(w_3) = \infty. \tag{5.85}$$

(ii) If $w = w_2 + w_3$ has no Gaussian part ($w_1 = 0$) with w_3 nonzero, then

$$\underline{\beta}_\infty(w) = \underline{\beta}_\infty(w_3), \quad \beta_\infty(w) = \beta_\infty(w_3), \quad \textit{and} \quad p_{\max}(w) = p_{\max}(w_2). \tag{5.86}$$

(iii) If $w = w_1 + w_2 + w_3$ with w_1 and w_3 non zero, then

$$\underline{\beta}_\infty(w) = \beta_\infty(w) = 2 \quad \textit{and} \quad p_{\max}(w) = p_{\max}(w_2). \tag{5.87}$$

Proof. Let Ψ be the characteristic exponent of w, and (μ, σ^2, ν) be its Lévy triplet. We assume that $\mu = 0$, what has no impact on the indices. The Lévy-Itô decomposition corresponds to the following sum for the characteristic exponent:

$$\Psi(\xi) = \underbrace{-\frac{\sigma^2\xi^2}{2}}_{\Psi_1(\xi)} + \underbrace{\int_{\mathbb{R}} (\mathrm{e}^{\mathrm{j}\xi t} - 1)\mathbb{1}_{|t|>1}\mathrm{d}\nu(t)}_{\Psi_2(\xi)} + \underbrace{\int_{\mathbb{R}} (\mathrm{e}^{\mathrm{j}\xi t} - 1 - \mathrm{j}\xi t)\mathbb{1}_{|t|\leq 1}\mathrm{d}\nu(t)}_{\Psi_3(\xi)}, \tag{5.88}$$

where, Ψ_1, Ψ_2, and Ψ_3 are the characteristic exponents of w_1, w_2, and w_3 respectively, with respective triplets $(0, \sigma^2, 0)$, $(0, 0, \mathbb{1}_{|\cdot|>1}\nu)$, and $(0, 0, \mathbb{1}_{|\cdot|\leq 1}\nu)$.

(i) The characteristic exponent of w_1 is $\Psi_1(\xi) = -\sigma^2\xi^2/2$, hence $\underline{\beta}_\infty(w_1) = \beta_\infty(w_1) = 2$ follows directly from the definition of the indices in (5.48) and (5.49). Moreover, Ψ_2 is bounded due to $|\Psi_2(\xi)| \leq 2\int_{|t|>1} \mathrm{d}\nu(t) < \infty$, therefore $\underline{\beta}_\infty(w_2) = \beta_\infty(w_2) = 0$. Finally, we have seen in the proof of Proposition 5.3 that the moments of w_3 are finite, hence $p_{\max}(w_3) = \infty$. It is moreover clear that the moments of the Gaussian white noise are finite, hence $p_{\max}(w_1) = \infty$ and the relations (5.85) are proved.

(ii) Assume that $w_1 = 0$. The characteristic exponent Ψ_2 is bounded by some constant $C > 0$. Hence, since $\Psi(\xi) = \Psi_2(\xi) + \Psi_3(\xi)$, we deduce that

$$|\Psi_3(\xi)| - C \leq |\Psi(\xi)| \leq |\Psi_3(\xi)| + C \tag{5.89}$$

for every $\xi \in \mathbb{R}$. The left inequality (5.89) implies that $\underline{\beta}_\infty(w) \geq \underline{\beta}_\infty(w_3)$ and $\beta_\infty(w) \geq \beta_\infty(w_3)$. The right inequality gives the other inequalities for the Blumenthal–Getoor indices and therefore $\underline{\beta}_\infty(w) = \underline{\beta}_\infty(w_3)$ and $\beta_\infty(w) = \beta_\infty(w_3)$.

For the moment index, we recall that $|a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$ (by convexity of $x \mapsto x^p$ on $\mathbb{R}^+$) for every $a, b \in \mathbb{R}$ and $p \geq 1$ and that $|a+b|^p \leq (|a|^p + |b|^p)$ when $0 < p < 1$ (since $x \mapsto |x|^p$ is subadditive). We set $c_p = 2^{p-1}$ if $p \geq 1$ and $c_p = 1$ if $0 < p < 1$. Then, if X and Y are two random variables such that $\mathbb{E}[|Y|^p] < \infty$, then we have that

$$c_p^{-1}\mathbb{E}[|X|^p] - \mathbb{E}[|Y|^p] \leq \mathbb{E}[|X+Y|^p] \leq c_p(\mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p]). \tag{5.90}$$

Applying (5.90) to $X = \langle w_2, \mathbb{1}_{[0,1]^d}\rangle$ and $Y = \langle w_3, \mathbb{1}_{[0,1]^d}\rangle$, the later having finite pth moments for any $p > 0$, we deduce that

$$\mathbb{E}\left[\left|\langle w, \mathbb{1}_{[0,1]^d}\rangle\right|^p\right] = \mathbb{E}[|X+Y|^p] < \infty \iff \mathbb{E}\left[\left|\langle w_2, \mathbb{1}_{[0,1]^d}\rangle\right|^p\right] = \mathbb{E}[|X|^p] < \infty. \tag{5.91}$$

Hence, w and w_2 have the same moment index.

(iii) Assume that $w_1 \neq 0$. Using that $\Psi_1(\xi) = -\sigma^2\xi^2$ and that Ψ_2 and Ψ_3 (like every characteristic exponent, see for instance [381, Proposition 2.4]), is asymptotically dominated by $\xi \mapsto \xi^2$, we deduce that $C_1|\Psi_1(\xi)| \leq |\Psi(\xi)| \leq C_2|\Psi_1(\xi)|$ for some constants $C_1, C_2 > 0$ and every $\xi \in \mathbb{R}$ such that $|\xi| \geq 1$. Therefore, w and w_1 have the same Blumenthal-Getoor indices $\underline{\beta}_\infty = \beta_\infty = 2$. We have shown the equalities on Blumenthal-Getoor indices in (5.86) and (5.87). The proof for the moment index is identical to the case $w_1 = 1$, this time with $Y = \langle w_1 + w_3, \mathbb{1}_{[0,1]^d}\rangle$.

□

Weighted Besov Spaces

As we have seen in Section 5.3.1, Besov spaces are natural candidates for characterizing the regularity of Lévy processes and Lévy white noises. We define the family of weighted Besov spaces based on wavelet methods, as exposed in [403]. Besov spaces have a long history in functional analysis [415]. They were successfully revisited by the introduction of wavelet methods following the works of Y. Meyer [348] and applied to the analysis of stochastic processes, including the Brownian motion [418, 419, 421], the fractional Brownian motion [445, 404], sparse random processes [334, 443, 405, 48], and general solutions of SPDEs [446, 447].

Essentially, weighted Besov spaces are subspaces of $\mathcal{S}'(\mathbb{R}^d)$ that are characterized by weighted sequence norms of the wavelet coefficients. Following H. Triebel, we use the compactly supported wavelets discovered by I. Daubechies [346], which we introduce first. The scale and shift parameters of the wavelets are respectively denoted by $j \in \mathbb{N}$ and $\boldsymbol{k} \in \mathbb{Z}^d$. The symbols M and F refer to the *gender* of the wavelet (M for the mother wavelets and F for the father wavelet). Consider two functions ψ_M and $\psi_F \in L_2(\mathbb{R})$. We set $\mathcal{G}^0 = \{M, F\}^d$ and for $j \geq 1$, $\mathcal{G}^j = \{M, F\}^d \backslash \{(F, \ldots, F)\}$. Therefore, the cardinal of $\mathcal{G}^0$ is $\mathrm{Card}(\mathcal{G}^0) = 2^d$, while $\mathrm{Card}(\mathcal{G}^j) = 2^d - 1$ for $j \geq 1$. For $\boldsymbol{G} = (G_1, \ldots, G_d) \in \mathcal{G}^0$, called a gender, we set, for every $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\psi_{\boldsymbol{G}}(\boldsymbol{x}) = \prod_{i=1}^d \psi_{G_i}(x_i)$. For $j \in \mathbb{N}$, $\boldsymbol{G} \in \mathcal{G}^j$, and $\boldsymbol{k} \in \mathbb{Z}^d$, we define

$$\psi_{j,\boldsymbol{G},\boldsymbol{k}} := 2^{jd/2} \psi_{\boldsymbol{G}}(2^j \cdot - \boldsymbol{k}). \tag{5.92}$$

We shall also use the notations $\boldsymbol{F} = (F, \ldots, F)$ and $\boldsymbol{M} = (M, \ldots, M)$ for the purely father and purely mother genders. It is known that, for any $r_0 \geq 1$, there exists two functions $\psi_M, \psi_F \in L_2(\mathbb{R})$, called *Daubechies wavelets*, that are compactly supported, with at least r_0 continuous derivatives and vanishing moments up to order at least $(r_0 - 1)$, and such that the family[10] $\{\psi_{j,\boldsymbol{G},\boldsymbol{k}}\}_{(j,\boldsymbol{G},\boldsymbol{k}) \in \mathbb{N} \times \mathcal{G}^j \times \mathbb{Z}^d}$ is an orthonormal basis of $L_2(\mathbb{R}^d)$ [403, Section 1.2.1].

We now introduce the family of weighted Besov spaces $B_p^\tau(\mathbb{R}^d; \rho)$. Traditionally, Besov spaces also depend on the additional parameter $q \in (0, \infty]$ (see for instance

[10] There is a slight abuse of notation when we write $(j, \boldsymbol{G}, \boldsymbol{k}) \in \mathbb{N} \times \mathcal{G}^j \times \mathbb{Z}^d$, since j appears as the first element of the triplet $(j, \boldsymbol{G}, \boldsymbol{k})$ and specifies the location of the gender $\boldsymbol{G} \in \mathcal{G}^j$. We keep this notation for its convenience.

[403, Definition 1.22]). We shall only consider the case $q = p$ in this work, so that we do not refer to this parameter.

We introduce weighted Besov spaces in Definition 5.5 relying on the wavelet decomposition of (generalized) functions. This construction is equivalent to the more usual Fourier-based definitions, as proved in [403, Theorem 1.26]. We use the notation $(x)_+ = \max(x, 0)$.

Definition 5.5. *Let $\tau, \rho \in \mathbb{R}$ and $0 < p \leq \infty$. Fix an integer $r_0 > \max(\tau, d(1/p - 1)_+ - \tau)$ and consider a family of Daubechies wavelets $\{\psi_{j,\boldsymbol{G},\boldsymbol{k}}\}_{(j,\boldsymbol{G},\boldsymbol{k}) \in \mathbb{N} \times \mathcal{G}^j \times \mathbb{Z}^d}$, where ψ_M and ψ_F have at least r_0 continuous derivatives and ψ_M has vanishing moments up to order at least $(r_0 - 1)$. The* weighted Besov space $B_p^\tau(\mathbb{R}^d; \rho)$ *is the collection of tempered generalized functions $f \in \mathcal{S}'(\mathbb{R}^d)$ that can be written as*

$$f = \sum_{j \in \mathbb{N}} \sum_{\boldsymbol{G} \in \mathcal{G}^j} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} c_{j,\boldsymbol{G},\boldsymbol{k}} \psi_{j,\boldsymbol{G},\boldsymbol{k}}, \tag{5.93}$$

where the $c_{j,\boldsymbol{G},\boldsymbol{k}}$ satisfy

$$\sum_{j \in \mathbb{N}} 2^{j(\tau p - d + \frac{dp}{2})} \sum_{\boldsymbol{G} \in \mathcal{G}^j} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j} \boldsymbol{k} \rangle^{\rho p} |c_{j,\boldsymbol{G},\boldsymbol{k}}|^p < \infty \tag{5.94}$$

and where the convergence (5.93) *holds on $\mathcal{S}'(\mathbb{R}^d)$. The usual adaptation is made for $p = \infty$; that is,*

$$\sup_{j \in \mathbb{N}} 2^{j(\tau + \frac{d}{2})} \sup_{\boldsymbol{G} \in \mathcal{G}^j} \sup_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j} \boldsymbol{k} \rangle^{\rho} |c_{j,\boldsymbol{G},\boldsymbol{k}}| < \infty. \tag{5.95}$$

The integer r_0 in Definition 5.5 is chosen such that the mother wavelet has enough vanishing moments and the mother and father wavelets are regular enough to be applied to a function of $B_p^\tau(\mathbb{R}^d; \rho)$. We refer the reader to [403, Section 1.2.1] and references therein for more details about the role of the smoothness and the vanishing moments of Daubechies wavelets. When the convergence (5.93) occurs, the duality product $\langle f, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle$ is well defined and we have $c_{j,\boldsymbol{G},\boldsymbol{k}} = \langle f, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle$. Moreover, for $p < \infty$, the quantity

$$\|f\|_{B_p^\tau(\mathbb{R}^d;\rho)} := \left(\sum_{j \in \mathbb{N}} 2^{j(\tau p - d + \frac{dp}{2})} \sum_{\boldsymbol{G} \in \mathcal{G}^j} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j} \boldsymbol{k} \rangle^{\rho p} |\langle f, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle|^p \right)^{1/p} \tag{5.96}$$

is finite for any $f \in B_p^\tau(\mathbb{R}^d;\rho)$ and specifies a norm (a quasi-norm, respectively) on the space $B_p^\tau(\mathbb{R}^d;\rho)$, with $p \geq 1$ ($p < 1$, respectively). The space $B_p^\tau(\mathbb{R}^d;\rho)$ is a Banach (a quasi-Banach, respectively) for this norm (quasi-norm, respectively) [403, Theorem 1.26]. For $p = \infty$, (5.96) becomes

$$\|f\|_{B_\infty^\tau(\mathbb{R}^d;\rho)} := \sup_{j\in\mathbb{N}} 2^{j(\tau+\frac{d}{2})} \sup_{\boldsymbol{G}\in\mathcal{G}^j} \sup_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^\rho |\langle f, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle|, \tag{5.97}$$

and $B_\infty^\tau(\mathbb{R}^d;\rho)$ is a Banach space for this norm.

Proposition 5.6 (Embeddings between weighted Besov spaces)**.** *Let* $0 < p_0 \leq p_1 \leq \infty$ *and* $\tau_0, \tau_1, \rho_0, \rho_1 \in \mathbb{R}$.

- *We have the embedding* $B_{p_0}^{\tau_0}(\mathbb{R}^d;\rho_0) \subseteq B_{p_1}^{\tau_1}(\mathbb{R}^d;\rho_1)$ *as soon as*

$$\tau_0 - \tau_1 \geq \frac{d}{p_0} - \frac{d}{p_1} \quad \text{and} \quad \rho_0 \geq \rho_1. \tag{5.98}$$

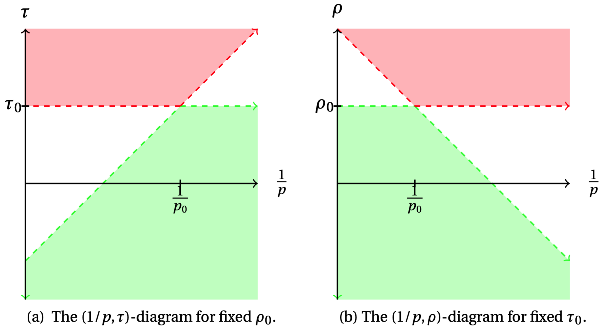
- *We have the embedding* $B_{p_1}^{\tau_1}(\mathbb{R}^d;\rho_1) \subseteq B_{p_0}^{\tau_0}(\mathbb{R}^d;\rho_0)$ *as soon as*

$$\rho_1 - \rho_0 > \frac{d}{p_0} - \frac{d}{p_1} \quad \text{and} \quad \tau_1 > \tau_0. \tag{5.99}$$

The embedding for the conditions (5.98) was proved by D.E. Edmunds and H. Triebel [448, Equation (9), Section 4.2.3] for general weights. The embedding for the conditions (5.99) was obtained in [431, Section 2.2.2]. Note that (5.47) is deduced from (5.98) by taking $p_0 = p_1 = p$. The embedding relations are summarized in the two Triebel diagrams[11] of Figure 5.9.

As a simple example, we obtain the Besov localization of the Dirac distribution. This result is of course well-known (an alternative proof can be found for instance in [449]) but we provide a new proof for two reasons: (1) it illustrates how to use the wavelet-based characterization of Besov spaces and (2) the result will be used to obtain sharp results for compound Poisson white noises.

[11] The representation of the smoothness properties in diagrams with axis $(1/p, \tau)$ is inherited from the work of H. Triebel. It is very convenient because the smoothness has often a simple formulation in terms of $1/p$, as appear typically in our Theorem 5.1. This is also valid for the asymptotic rate ρ.



(a) The $(1/p, \tau)$-diagram for fixed ρ_0.

(b) The $(1/p, \rho)$-diagram for fixed τ_0.

Figure 5.9: Representation of the embeddings between Besov spaces: If $f \in B^{\tau_0}_{p_0}(\mathbb{R}^d; \rho_0)$, then f is in every Besov space that is in the lower shaded green regions. Conversely, if $f \notin B^{\tau_0}_{p_0}(\mathbb{R}^d; \rho_0)$, then f is in none of the Besov spaces of the upper shaded red regions.

Proposition 5.7. *Let $0 < p < \infty$, $\tau \in \mathbb{R}$, and $\rho \in \mathbb{R}$. Then, the Dirac impulse δ is in $B^{\tau}_{p}(\mathbb{R}^d; \rho)$ if and only if $\tau < \frac{d}{p} - d$. Moreover, $\delta \in B^{\tau}_{\infty}(\mathbb{R}^d; \rho)$ if and only if $\tau \leq -d$.*

We remark that the weight $\rho \in \mathbb{R}$ plays no role in Proposition 5.7. This is a simple consequence of the fact that δ is compactly supported, and therefore insensitive to the weight, as will appear in the proof.

Proof of Proposition 5.7. We first treat the case $p < \infty$. The wavelet coefficients of δ are $c_{j,\boldsymbol{G},\boldsymbol{k}} = 2^{jd/2}\psi_{\boldsymbol{G}}(-\boldsymbol{k})$, hence the Besov (quasi-)norm of the Dirac impulse is given by

$$\|\delta\|^p_{B^{\tau}_{p}(\mathbb{R}^d;\rho)} = \sum_{j\in\mathbb{N}} 2^{j(\tau p - d + dp)} \sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} |\psi_{\boldsymbol{G}}(-\boldsymbol{k})|^p. \tag{5.100}$$

We first introduce some notations. The 2^d wavelets $\psi_{\boldsymbol{G}}$ with gender $\boldsymbol{G}$ describing $\mathcal{G}^0$ are bounded, hence the constant $b = \max_{\boldsymbol{G}\in\mathcal{G}^0} \|\psi_{\boldsymbol{G}}\|_\infty$ is finite. We denote by

K the set of multi-integers $\boldsymbol{k} \in \mathbb{Z}^d$ such that $\psi_{\boldsymbol{G}}(-\boldsymbol{k}) \neq 0$ for some gender $\boldsymbol{G} \in \mathcal{G}^0$. The set K is finite because the 2^d wavelets are compactly supported and we set $n_0 = \mathrm{Card}(K)$. Moreover, the set K is non empty; otherwise, (5.100) would imply that $\|\delta\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} = 0$, hence $\delta = 0$, which is absurd. We fix some element $\boldsymbol{k}_0 \in K$ and $\boldsymbol{G}_0$ a gender such that $\psi_{\boldsymbol{G}_0}(-\boldsymbol{k}_0) \neq 0$. We set $a = |\psi_{\boldsymbol{G}_0}(-\boldsymbol{k}_0)| > 0$. Then, there exists a constant $M > 0$ such that $\|2^{-j}\boldsymbol{k}\| \leq M$ for any $j \in \mathbb{N}$ and $(-\boldsymbol{k}) \in K$. In particular, for such $\boldsymbol{k}$ and j, we have that $1 \leq \langle 2^{-j}\boldsymbol{k}\rangle = (1 + \|2^{-j}\boldsymbol{k}\|^2)^{1/2} \leq \langle M\rangle$, and therefore,

$$0 < \min(1, \langle M\rangle^{\rho p}) \leq \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} \leq \max(1, \langle M\rangle^{\rho p}) < \infty. \tag{5.101}$$

Fix $j \geq 0$. Then, we have the lower bound

$$\sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} |\psi_{\boldsymbol{G}}(-\boldsymbol{k})|^p \geq \langle 2^{-j}\boldsymbol{k}_0\rangle^{\rho p} a^p \geq A := \min(1, \langle M\rangle^{\rho p}) a^p. \tag{5.102}$$

We recall that $\mathrm{Card}(\mathcal{G}^j) \leq 2^d$. Then, we also have the upper bound

$$\sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} |\psi_{\boldsymbol{G}}(-\boldsymbol{k})|^p \leq \sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in K} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} b^p \leq B := 2^d n_0 \max(1, \langle M\rangle^{\rho p}) b^p. \tag{5.103}$$

Combining (5.102) and (5.103), we therefore deduce that

$$A \sum_{j\in\mathbb{N}} 2^{j(\tau p - d + dp)} \leq \|\delta\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} \leq B \sum_{j\in\mathbb{N}} 2^{j(\tau p - d + dp)}. \tag{5.104}$$

The sum converges for $(\tau p - d + dp) < 0$ and diverges otherwise, implying the result.

We now adapt the argument to the case $p = \infty$ for which the Besov norm is given by

$$\|\delta\|_{B^\tau_\infty(\mathbb{R}^d;\rho)} = \sup_{j\in\mathbb{N}} 2^{j(\tau+d)} \sup_{\boldsymbol{G}\in\mathcal{G}^j} \sup_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho} |\psi_{\boldsymbol{G}}(-\boldsymbol{k})|. \tag{5.105}$$

We have that, for any $j \in \mathbb{N}$,

$$A' := a \min(1, \langle M\rangle^{\rho}) \leq \sup_{\boldsymbol{G}\in\mathcal{G}^0} \sup_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho} |\psi_{\boldsymbol{G}}(-\boldsymbol{k})| \leq B' := b \max(1, \langle M\rangle^{\rho}). \tag{5.106}$$

Therefore, we deduce that

$$A' \sup_{j \in \mathbb{N}} 2^{j(\tau+d)} \leq \|\delta\|_{B^{\tau}_{\infty}(\mathbb{R}^d;\rho)} \leq B' \sup_{j \in \mathbb{N}} 2^{j(\tau+d)}, \tag{5.107}$$

and $\|\delta\|_{B^{\tau}_{\infty}(\mathbb{R}^d;\rho)} < \infty$ if and only if $\tau + d \leq 0$. □

5.3.3 Gaussian White Noise

Our goal in this section is to prove the Gaussian part of Theorem 5.1. Without loss of generality, we focus on the Gaussian white noise with zero mean and unit variance.

The Gaussian case is much simpler than the general one since the wavelet coefficients of the Gaussian white noise are independent and identically distributed. We present it separately for three reasons: (i) it can be considered as an instructive toy problem that already contains some of the technicalities that will appear for the general case; (ii) it cannot be deduced from the other sections, where the results are based on a careful study of the Lévy measure; and (iii) the localization of the Gaussian white noise in *weighted* Besov spaces has not been addressed in the literature, to the best of our knowledge. We first state three simple lemma that will be useful throughout this work.

Lemma 5.1. *Let $(N_k)_{k\geq 1}$ be a sequence such that $N_k \to \infty$ when $k \to \infty$. Assume that we have i.i.d. random variables Y_k^i with $k \geq 1$ and $1 \leq i \leq N_k$ such that $Y_k^i \geq 0$ and $\mathscr{P}(Y_k^i > 0) > 0$. Let $(Z_k)_{k\geq 1}$ be the family of random variables such that $Z_k = \frac{1}{N_k}\sum_{i=1}^{N_k} Y_k^i$. Then, $\sum_{k\geq 1} Z_k = \infty$ a.s.*

Proof. Let $\mu = \mathbb{E}[Y_k^i] \in (0,\infty]$. If $\mu = \infty$, we set $\tilde{Y}_k^i = \min(Y_k^i, 1)$. Then, the $\tilde{Y}_k^i$ are i.i.d., non-negative with $\mathscr{P}(\tilde{Y}_k^i > 0) > 0$, and such that $\tilde{\mu} = \mathbb{E}[\tilde{Y}_k^i] \leq 1 < \infty$. Moreover, we have that

$$\sum_{k\geq 1} Z_k = \sum_{k\geq 1} \frac{1}{N_k} \sum_{i=1}^{N_k} Y_k^i \geq \sum_{k\geq 1} \frac{1}{N_k} \sum_{i=1}^{N_k} \tilde{Y}_k^i = \sum_{k\geq 1} \tilde{Z}_k, \tag{5.108}$$

where we set $\tilde{Z}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} \tilde{Y}_k^i$. Due to (5.108), it is then sufficient to demonstrate Lemma 5.1 for $\mu < \infty$, which we do now.

Let $(W_k)_{k\geq 1}$ be a family of i.i.d. random variables whose common law is the one of the Y_k^i and define $\overline{W}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} W_i$. The weak law of large numbers implies that $\mathscr{P}(|\overline{W}_k - \mu| \geq x)$ vanishes when $k \to \infty$ for any $x > 0$. Taking $x = \mu/2$, we readily deduce that $\mathscr{P}(\overline{W}_k \geq \mu/2)$ goes to 1 when $k \to \infty$. Moreover, we have the equality $Z_k \overset{(\mathcal{L})}{=} \overline{W}_k$, therefore, we also have that

$$\mathscr{P}(Z_k \geq \mu/2) \underset{k\to\infty}{\longrightarrow} 1. \tag{5.109}$$

This implies in particular that $\sum_{k\geq 1} \mathscr{P}(Z_k \geq \mu/2) = \infty$. The events $\{Z_k \geq \mu/2\}$ are moreover independent due to the independence of the Y_k^i. Using the Borel-Cantelli lemma, we deduce that $Z_k \geq \mu/2$ for infinitely many k a.s. An obvious consequence is then that $\sum_{k\geq 1} Z_k = \infty$ a.s.. □

As a consequence of Lemma 5.1, we deduce Lemma 5.2.

Lemma 5.2. *Assume that $(X_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{Z}^d}$, is a sequence of i.i.d. random variables such that $\mathscr{P}(|X_{\boldsymbol{k}}| > 0) > 0$. Then,*

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{|X_{\boldsymbol{k}}|}{\langle \boldsymbol{k}\rangle^d} = \infty \ a.s. \tag{5.110}$$

Proof. First of all, the result for any dimension d is easily deduced from the one-dimensional case. Moreover, $|k|$ and $\langle k\rangle$ are equivalent asymptotically, hence it is equivalent to show that $\sum_{k\geq 1} \frac{|X_k|}{k} = \infty$ for X_k i.i.d. with $\mathscr{P}(|X_k| > 0) > 0$. Setting $Z_k = \frac{1}{2^{k-1}} \sum_{\ell=2^{k-1}}^{2^k-1} |X_\ell|$ for $k \geq 1$, we deduce (5.110) by applying Lemma 5.1 with $N_k = 2^{k-1}$ and $Y_k^i = X_{2^{k-1}+i-1}$ and observing that $\sum_{k\geq 1}\frac{|X_k|}{k} \geq \frac{1}{2}\sum_{k\geq 1} Z_k = \infty$. □

Finally, we state the last lemma that deals with supremum of i.i.d. sequences of random variables.

Lemma 5.3. *Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables such that $\mathscr{P}(|X_k| \geq M) > 0$ for every $M \geq 0$. Then, we have that, almost surely,*

$$\sup_{k\geq 1} |X_k| = \infty. \tag{5.111}$$

Proof. Let $M > 0$. The assumption $\mathscr{P}(|X_k| \geq M) > 0$ and the fact that the events $\{|X_k| \geq M\}$ are independent implies, thanks to the Borel-Cantelli lemma, that there exists almost surely (infinitely many) $k \geq 1$ such that $|X_k| \geq M$. Hence, $\sup_{k\geq 1} |X_k| \geq M$ almost surely. This being true for every $M > 0$, we deduce (5.111). □

We characterize the Besov regularity of the Gaussian white noise in Proposition 5.8.

Proposition 5.8. *Fix $0 < p \leq \infty$ and $\tau, \rho \in \mathbb{R}$. The Gaussian white noise w is*

- *almost surely in $B_p^\tau(\mathbb{R}^d; \rho)$ if $\tau < -d/2$ and $\rho < -d/p$, and*
- *almost surely not in $B_p^\tau(\mathbb{R}^d; \rho)$ if $\tau \geq -d/2$ or $\rho \geq -d/p$.*

A direct consequence of Proposition 5.8 is Corollary 5.3, where we identify the local smoothness and the asymptotic growth rate of the Gaussian white noise, that are defined for (deterministic) generalized functions in Section 5.3.1.

Corollary 5.3. *Let w be a Gaussian white noise and $0 < p \leq \infty$. Then, we have almost surely that*

$$\tau_p(w) = -\frac{d}{2} \quad \textit{and} \quad \rho_p(w) = -\frac{d}{p}. \tag{5.112}$$

Remark. The determination of the local smoothness and the asymptotic growth rate is insensitive to the fact that the generalized function f is or is not in the critical space $B_p^{\tau_p(f)}(\mathbb{R}^d; \rho_p(f))$. For the Gaussian white noise, Proposition 5.8 implies that (almost surely) $w \notin B_p^{\tau_p(w)}(\mathbb{R}^d; \rho_p(w))$ for every $0 < p \leq \infty$. In that sense, Proposition 5.8 contains more information than Corollary 5.3, since we cannot deduce the critical cases treated in the proposition from the result of the corollary.

Proof of Proposition 5.8. Recall that we restrict, without loss of generality, to Gaussian white noise with unit variance $\sigma^2 = 1$. Then, $\langle w, \varphi_1 \rangle$ and $\langle w, \varphi_2 \rangle$ are independent if and only if $\langle \varphi_1, \varphi_2 \rangle = 0$. Moreover, $\langle w, \varphi \rangle$ is a Gaussian random variable with variance $\|\varphi\|_2^2$ [362, Section 2.5]. The family of functions $\{\psi_{j,\boldsymbol{G},\boldsymbol{k}}\}_{(j,\boldsymbol{G},\boldsymbol{k}) \in \mathbb{N} \times \mathcal{G}^j \times \mathbb{Z}^d}$ being orthonormal, the random variables $\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle$ are therefore i.i.d. with law $\mathcal{N}(0,1)$.

Case $p < \infty$, $\tau < -d/2$, and $\rho < -d/p$. For $p > 0$, we denote by C_p the pth moment of a Gaussian random variable with zero mean and unit variance. In particular, we have that $\mathbb{E}\left[|\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle|^p\right] = C_p$ for any $j \in \mathbb{N}, \boldsymbol{G} \in \mathcal{G}^j, \boldsymbol{k} \in \mathbb{Z}^d$, and therefore

$$\begin{aligned}
\mathbb{E}\left[\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)}\right] &= \sum_{j \in \mathbb{N}} 2^{j(\tau p - d + \frac{dp}{2})} \sum_{\boldsymbol{G} \in \mathcal{G}^j} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p} \mathbb{E}\left[|\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}} \rangle|^p\right] \\
&= C_p \sum_{j \in \mathbb{N}} 2^{j(\tau p - d + \frac{dp}{2})} \mathrm{Card}(\mathcal{G}^j) \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p} \\
&\leq 2^d C_p \sum_{j \in \mathbb{N}} 2^{j(\tau p - d + \frac{dp}{2})} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p}. \qquad (5.113)
\end{aligned}$$

The last inequality is due to $\mathrm{Card}(\mathcal{G}^j) \leq 2^d$. Since $\rho p < -d$ and $\langle 2^{-j}\boldsymbol{k} \rangle \underset{\|\boldsymbol{k}\| \to \infty}{\sim} 2^{-j}\|\boldsymbol{k}\|$, we have that $\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p} < \infty$. Moreover, we recognize a Riemann sum and have the convergence

$$2^{-jd} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p} \underset{j \to \infty}{\longrightarrow} \int_{\mathbb{R}^d} \langle \boldsymbol{x} \rangle^{\rho p} \mathrm{d}\boldsymbol{x} < \infty. \qquad (5.114)$$

In particular, the series $\sum_j 2^{j(\tau p + \frac{dp}{2})} \left(2^{-jd} \sum_{\boldsymbol{k}} \langle 2^{-j}\boldsymbol{k} \rangle^{\rho p}\right)$ converges if and only if the series $\sum_j 2^{j(\tau p + \frac{dp}{2})}$ does; in other words, if and only if $\tau < -d/2$. Finally, if $\tau < -d/2$ and $\rho < -d/p$, we have shown that $\mathbb{E}[\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)}] < \infty$ and therefore $w \in B^\tau_p(\mathbb{R}^d;\rho)$ almost surely.

Case $p < \infty$ and $\tau \geq -d/2$. Then, we have $2^{j(\tau p - d + dp/2)} \geq 2^{-jd}$. We aim at establishing a lower bound for the Besov norm of w and we restrict to the purely mother wavelet with gender $\boldsymbol{G} = \boldsymbol{M} = (M, \dots, M) \in \mathcal{G}^j$ for any $j \in \mathbb{N}$. For $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$ such that $0 \leq k_i < 2^j$ for every $i = 1, \dots, d$, we have that

$$\langle 2^{-j} \boldsymbol{k} \rangle^{\rho p} \geq C := \min_{\|\boldsymbol{x}\|_\infty \leq 1} \langle \boldsymbol{x} \rangle^{\rho p}. \tag{5.115}$$

For $\|\boldsymbol{x}\|_\infty \leq 1$, we have that $1 \leq \langle \boldsymbol{x} \rangle = (1 + \|\boldsymbol{x}\|^2)^{1/2} \leq (1+d)^{1/2}$, hence $C \geq \min(1, (1+d)^{\rho p/2}) > 0$. Then, we have

$$\begin{aligned} \|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} &\geq \sum_{j \in \mathbb{N}} 2^{-jd} \sum_{0 \leq k_1, \dots, k_d < 2^j} \langle 2^{-j} \boldsymbol{k} \rangle^{\rho p} |\langle w, \psi_{j, \boldsymbol{M}, \boldsymbol{k}} \rangle|^p \\ &\geq C \sum_{j \in \mathbb{N}} 2^{-jd} \sum_{0 \leq k_1, \dots, k_d < 2^j} |\langle w, \psi_{j, \boldsymbol{M}, \boldsymbol{k}} \rangle|^p. \end{aligned} \tag{5.116}$$

The random variables $\langle w, \psi_{j, \boldsymbol{M}, \boldsymbol{k}} \rangle$ are i.i.d. We can therefore apply Lemma 5.1 with blocks of size 2^{jd}, which goes to infinity when $j \to \infty$ to conclude that $\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} = \infty$ a.s.

Case $p < \infty$ and $\rho \geq -d/p$. We retain only the father wavelet $\phi = \psi_{\boldsymbol{F}}$ where $\boldsymbol{F} = (F, \dots, F) \in \mathcal{G}^0$ and the scale $j = 0$ and exploit the relation $\rho p \geq -d$ to deduce the lower bound

$$\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} \geq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \langle \boldsymbol{k} \rangle^{\rho p} |\langle w, \phi_{\boldsymbol{k}} \rangle|^p \geq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \frac{|\langle w, \phi_{\boldsymbol{k}} \rangle|^p}{\langle \boldsymbol{k} \rangle^d}, \tag{5.117}$$

with the notation $\phi_{\boldsymbol{k}} = \phi(\cdot - \boldsymbol{k})$. Finally, the random variables $\langle w, \phi_{\boldsymbol{k}} \rangle$ being i.i.d., Lemma 5.2 applies and $\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} = \infty$ almost surely.

Case $p = \infty$, $\tau < -d/2$, and $\rho < 0$. This case is deduced using the embeddings between Besov spaces. First of all, for $\epsilon \leq \min(-d/2 - \tau, -\rho)$, we have that

$$B^{-d/2-\epsilon}_\infty(\mathbb{R}^d; -\epsilon) \subseteq B^{-d/2-\epsilon}_\infty(\mathbb{R}^d; \rho) \subseteq B^\tau_\infty(\mathbb{R}^d; \rho) \tag{5.118}$$

using (5.47) first for the weight and then for the smoothness parameters. It therefore suffices to show the existence of $0 < \epsilon \leq \min(-d/2 - \tau, -\rho)$ such that

$w \in B_{\infty}^{-d/2-\epsilon}(\mathbb{R}^d; -\epsilon)$. Fix such an ϵ. Then, for every $p < \infty$, we already proved that $w \in B_p^{-d/2-\epsilon/2}(\mathbb{R}^d; -d/p - \epsilon/2)$ a.s. Applying this to $p = 2d/\epsilon$, we then remark that

$$B_p^{-d/2-\epsilon/2}(\mathbb{R}^d; -d/\rho - \epsilon/2) = B_{2d/\epsilon}^{-d/2-\epsilon/2}(\mathbb{R}^d; -\epsilon). \tag{5.119}$$

Moreover, applying Proposition 5.6 with $p_0 = 2d/\epsilon < p_1 = \infty$, $\tau_0 = -d/2 - \epsilon/2$, $\tau_1 = -d/2 - \epsilon$, $\rho_0 = -d/p_0 - \epsilon/2 = -\epsilon = \rho_1$, we easily verify that (5.98) is satisfied and therefore, using also (5.118), we have a.s. that

$$w \in B_{2d/\epsilon}^{-d/2-\epsilon/2}(\mathbb{R}^d; -\epsilon) \subset B_{\infty}^{-d/2-\epsilon}(\mathbb{R}^d; -\epsilon) \subseteq B_{\infty}^{\tau}(\mathbb{R}^d; \rho), \tag{5.120}$$

concluding this case.

Case $p = \infty$ and $\tau \geq -d/2$. Note that the case $\tau > -d/2$ can be deduced from the results for $p < \infty$ by embedding, but one cannot deduce the case $\tau = -d/2$. By keeping only the purely mother wavelet $\psi_{\boldsymbol{M}}$ with $\boldsymbol{M} = (M, \dots, M)$ and the shift parameter $\boldsymbol{k} = \boldsymbol{0}$, the Besov norm (5.97) applied to the Gaussian white noise is

$$\|w\|_{B_{\infty}^{\tau}(\mathbb{R}^d;\rho)} \geq \sup_{j \in \mathbb{N}} |\langle w, \psi_{j,\boldsymbol{M},\boldsymbol{0}} \rangle|. \tag{5.121}$$

The Gaussian random variables $\langle w, \psi_{j,\boldsymbol{M},\boldsymbol{0}} \rangle$ are independent and verifies the conditions of Lemma 5.3 (since $\mathscr{P}(|\langle w, \psi_{j,\boldsymbol{M},\boldsymbol{0}} \rangle| \geq M) > 0$ for every $M \geq 0$). This implies that $\sup_{j \in \mathbb{N}} |\langle w, \psi_{j,\boldsymbol{M},\boldsymbol{0}} \rangle| = \infty$ a.s., and therefore $w \notin B_{\infty}^{\tau}(\mathbb{R}^d; \rho)$ a.s. due to (5.121).

Case $p = \infty$ and $\rho \geq 0$. Again, the case $\rho > 0$ can be deduced from the results for $p < \infty$ by embedding, but the case $\rho = 0$ cannot. Using that $\langle \boldsymbol{k} \rangle^{\rho} \geq 1$ for $\rho \geq 0$ and keeping only the father wavelet $\phi = \psi_{\boldsymbol{F}}$ with $\boldsymbol{F} = (F, \dots, F) \in \mathcal{G}^0$ and the scale $j = 0$, the Besov norm (5.97) of w is lower bounded by

$$\|w\|_{B_{\infty}^{\tau}(\mathbb{R}^d;\rho)} \geq \sup_{\boldsymbol{k} \in \mathbb{Z}} |\langle w, \phi_{\boldsymbol{k}} \rangle|. \tag{5.122}$$

Again, Lemma 5.3 applies to the Gaussian random variables $\langle w, \phi_{\boldsymbol{k}} \rangle$ and $\max_{\boldsymbol{k} \in \mathbb{Z}} |\langle w, \phi_{\boldsymbol{k}} \rangle| = \infty$ a.s., implying that $\|w\|_{B_{\infty}^{\tau}(\mathbb{R}^d;\rho)} = \infty$ a.s. due to (5.122). □

The proof of Proposition 5.8 for the case $\rho \geq -d/p$ uses an argument that can be easily adapted to any Lévy white noise. We hence state this result in full generality.

Proposition 5.9. *Fix $0 < p \leq \infty$ and $\tau, \rho \in \mathbb{R}$. If w is a non-constant Lévy white noise, then $w \notin B_p^\tau(\mathbb{R}^d;\rho)$ as soon as $\rho \geq -d/p$.*

Proof. The proof is very similar to the one of Proposition 5.8 for $\rho \geq -d/p$ and $p < \infty$, and for $\rho \geq 0$ and $p = \infty$, except that we only consider father wavelet $\phi = \psi_{\boldsymbol{F}}$ and its shifts $\phi_{\boldsymbol{k}} = \psi_{0,\boldsymbol{F},\boldsymbol{k}}$ with $\boldsymbol{k} = k_0\mathbb{Z}^d$, where $k_0 \in \mathbb{N}\backslash\{0\}$ is chosen such that the $\phi_{\boldsymbol{k}}$ have disjoint supports. In particular, this implies the random variables $\langle w, \phi_{\boldsymbol{k}}\rangle$, are independent for $\boldsymbol{k} \in k_0\mathbb{Z}^d$ (the support of the test functions being disjoint) and independent (the Lévy white noise being stationary). As a consequence, (5.117) becomes

$$\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} \geq \sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d} \langle \boldsymbol{k}\rangle^{\rho p} |\langle w, \phi_{\boldsymbol{k}}\rangle|^p \geq \sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d} \frac{|\langle w, \phi_{\boldsymbol{k}}\rangle|^p}{\langle \boldsymbol{k}\rangle^d} \tag{5.123}$$

and Lemma 5.2 applies again, implying that $\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} = \infty$ a.s. Similarly, (5.122) becomes

$$\|w\|_{B_\infty^\tau(\mathbb{R}^d;\rho)} \geq \sup_{\boldsymbol{k}\in k_0\mathbb{Z}} |\langle w, \phi_{\boldsymbol{k}}\rangle|. \tag{5.124}$$

Then, it suffices to observe that $\mathscr{P}(|\langle w, \varphi\rangle| \geq M) > 0$ for any Lévy white noise w and any test function $\varphi \neq 0$. Indeed, the probability measure of an infinitely divisible is not compactly supported, except for constant Lévy white noise $w = \mu \in \mathbb{R}$, what we have excluded [170, Corollary 24.4]. Applying Lemma 5.3, we deduce finally that $w \notin B_\infty^\tau(\mathbb{R}^d;\rho)$ a.s. □

5.3.4 Compound Poisson White Noise

Compound Poisson white noises are almost surely made of countably many Dirac impulses, what will be crucial in their analysis. Our positive results are based on a careful estimation of the moments of the compound Poisson white noise presented in Proposition 5.10.

Proposition 5.10. *Let w be a compound Poisson white noise with moment index $p_{\max} \in (0,\infty]$ and $0 < p < p_{\max}$. Then, there exists a constant C such that*

$$\mathbb{E}[|\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle|^p] \leq C 2^{jpd/2 - jd} \tag{5.125}$$

for every $j \in \mathbb{N}$, $\boldsymbol{G} \in \mathcal{G}^j$, and $\boldsymbol{k} \in \mathbb{Z}^d$.

Proof. We recall that the Lebesgue measure is denoted by Leb. Let $\lambda > 0$ and P be respectively the sparsity parameter and the law of the jumps of w. Then, we have that

$$w \overset{(\mathcal{L})}{=} \sum_{k \in \mathbb{N}} a_k \delta(\cdot - \boldsymbol{x}_k), \tag{5.126}$$

where the a_k are i.i.d. with law P, and the $\boldsymbol{x}_k$, independent from the a_k, are randomly located such that $\mathrm{Card}\{k \in \mathbb{N},\ \boldsymbol{x}_k \in B\}$ is a Poisson random variable with parameter $\lambda \mathrm{Leb}(B)$ for any Borel set $B \subset \mathbb{R}^d$ with finite Lebesgue measure. For a demonstration that the right term in (5.126) specifies a compound Poisson white noise in the sense of a generalized random process with the adequate characteristic functional, we refer the reader to [450, Theorem 1].

Let $\psi \in L_2(\mathbb{R}^d)\backslash\{0\}$ be a compactly supported function and I_ψ be the closed convex hull of its support. In particular, $0 < \mathrm{Leb}(I_\psi) < \infty$. We set $N(\psi) = \mathrm{Card}\{k \in \mathbb{N},\ \boldsymbol{x}_k \in I_\psi\}$, which is a Poisson random variable with parameter $\lambda \mathrm{Leb}(I_\psi)$. We denote by a'_n and $\boldsymbol{x}'_n$, $n = 1, \dots, N(\psi)$, the weights and Dirac locations of the compound Poisson white noise w on I_ψ. That is, $\langle w, \psi\rangle = \sum_{n=1}^{N(\psi)} a'_n \psi(\boldsymbol{x}'_n)$. By conditioning on $N(\psi)$, we then have that

$$
\begin{aligned}
\mathbb{E}[|\langle w,\psi\rangle|^p] &= \sum_{N=1}^{\infty} \mathscr{P}(N(\psi)=N)\mathbb{E}\left[|\langle w,\psi\rangle|^p \,|N(\psi)=N\right] \\
&= \sum_{N=1}^{\infty} \mathscr{P}(N(\psi)=N)\mathbb{E}\left[\left|\sum_{n=1}^{N} a'_n \psi(\boldsymbol{x}'_n)\right|^p |N(\psi)=N\right] \\
&\overset{(i)}{\leq} \sum_{N=1}^{\infty} \mathscr{P}(N(\psi)=N)\mathbb{E}\left[N^{\max(0,p-1)} \sum_{n=1}^{N} |a'_n \psi(\boldsymbol{x}'_n)|^p \,|N(\psi)=N\right] \\
&\leq \|\psi\|_\infty^p \sum_{N=1}^{\infty} \left(N^{\max(0,p-1)} \mathscr{P}(N(\psi)=N) \left(\sum_{n=1}^{N} \mathbb{E}\left[|a'_n|^p \,|N(\psi)=N\right] \right) \right) \\
&\overset{(ii)}{=} \|\psi\|_\infty^p \sum_{N=1}^{\infty} \left(N^{\max(0,p-1)} \mathscr{P}(N(\psi)=N) \left(\sum_{n=1}^{N} \mathbb{E}\left[|a'_n|^p\right] \right) \right) \\
&\overset{(iii)}{=} \|\psi\|_\infty^p \mathbb{E}\left[|a'_1|^p\right] \sum_{N=1}^{\infty} N^{\max(1,p)} \mathscr{P}(N(\psi)=N), \qquad (5.127)
\end{aligned}
$$

where (i) uses the relation $\left|\sum_{n=1}^N y_n\right|^p \leq N^{\max(0,p-1)} \sum_{n=1}^N |y_n|^p$, valid for any $p>0$ and $y_n \in \mathbb{R}$ [393, Eq.(50)], (ii) is due to $\mathbb{E}\left[|a'_n|^p \,|N(\psi)=N\right] = \mathbb{E}\left[|a'_n|^p\right]$, a'_n and $N(\psi)$ being independent, and (iii) comes from $\sum_{n=1}^N \mathbb{E}\left[|a'_n|^p\right] = N\mathbb{E}\left[|a'_1|^p\right]$, the a'_n sharing the same law.

Our goal is now to apply (5.127) to $\psi = \psi_{j,\boldsymbol{G},\boldsymbol{k}}$. For fixed $j \geq 1$ and $\boldsymbol{k} \in \mathbb{Z}^d$, the Lebesgue measure of the convex hull $I_{\psi_{j,\boldsymbol{G},\boldsymbol{k}}}$ of the support of the $\psi_{j,\boldsymbol{G},\boldsymbol{k}}$ is $\mathrm{Leb}(I_{\psi_{j,\boldsymbol{G},\boldsymbol{k}}}) = 2^{-jd}\mathrm{Leb}(I_{\psi_{\boldsymbol{G}}})$. Therefore, $N(\psi_{j,\boldsymbol{G},\boldsymbol{k}}) = \mathrm{Card}\{k \in \mathbb{N},\ \boldsymbol{x}_k \in \psi_{j,\boldsymbol{G},\boldsymbol{k}}\}$ is a Poisson random variable with parameter $2^{-jd}\lambda\mathrm{Leb}(I_{\psi_{\boldsymbol{G}}})$. As a consequence, we have

$$\sum_{N=1}^{\infty} N^{\max(1,p)} \mathscr{P}(N(\psi_{j,\boldsymbol{G},\boldsymbol{k}}) = N) = \sum_{N=1}^{\infty} N^{\max(1,p)} \mathrm{e}^{-2^{jd}\lambda \mathrm{Leb}(I_{\psi_{\boldsymbol{G}}})} \frac{(\lambda \mathrm{Leb}(I_{\psi_{\boldsymbol{G}}}))^N 2^{-jdN}}{N!}$$
$$\leq \left(\sum_{N=1}^{\infty} N^{\max(1,p)} \frac{(\lambda \mathrm{Leb}(I_{\psi_{\boldsymbol{G}}}))^N}{N!} \right) 2^{-jd}, \tag{5.128}$$

where we used that $2^{-jdN} \leq 2^{-jd}$ and $\mathrm{e}^{-2^{jd}\lambda \mathrm{Leb}(I_{\psi_{\boldsymbol{G}}})} \leq 1$. We have moreover the relation

$$\|\psi_{j,\boldsymbol{G},\boldsymbol{k}}\|_{\infty}^{p} = 2^{jpd/2} \|\psi_{\boldsymbol{G}}\|_{\infty}^{p}. \tag{5.129}$$

Applying inequalities (5.128) and (5.129) in (5.127) with $\psi = \psi_{j,\boldsymbol{G},\boldsymbol{k}}$, we finally deduce (5.125) (for the finite constant $C = \|\psi_{\boldsymbol{G}}\|_{\infty}^{p} \mathbb{E}\left[|a_1'|^p\right] \sum_{N\geq 1} N^{\max(1,p)} (\lambda \mathrm{Leb}(I_{\psi_{\boldsymbol{G}}}))^N / N!$).
□

Proposition 5.11. *Fix $0 < p \leq \infty$ and $\tau, \rho \in \mathbb{R}$. Let w be a compound Poisson white noise with index $p_{\max} \in (0, \infty]$. If $0 < p < \infty$, then, w is*

- *almost surely in $B_p^{\tau}(\mathbb{R}^d; \rho)$ if $\tau < d/p - d$ and $\rho < -d/\min(p, p_{\max})$, and*
- *almost surely not in $B_p^{\tau}(\mathbb{R}^d; \rho)$ if $\tau \geq d/p - d$ or $\rho > -d/\min(p, p_{\max})$, or $\rho \geq -d/p$.*

If $p = \infty$, then w is

- *almost surely in $B_{\infty}^{\tau}(\mathbb{R}^d; \rho)$ if $\tau < -d$ and $\rho < -d/p_{\max}$, and*
- *almost surely not in $B_{\infty}^{\tau}(\mathbb{R}^d; \rho)$ if $\tau > -d$ or $\rho > -d/p_{\max}$, or $\rho \geq 0$.*

In Proposition 5.11, we have split the results in two scenarii depending on what is know when ρ and/or τ are equal to the critical values. The only remaining cases that are not covered by Proposition 5.11 is when $p_{\max} < \infty$ and $\rho = -d/p_{\max}$ and when $p = \infty$ and $\tau = -d$. As for the Gaussian white noise in Corollary 5.3, a direct consequence of Proposition 5.11 is the identification of the local smoothness and the asymptotic growth rate of a compound Poisson white noise.

Corollary 5.4. *Let w be a compound Poisson white noise with moment index $0 < p_{\max} \leq \infty$ and $0 < p \leq \infty$. Then, we have almost surely that*

$$\tau_p(w) = \frac{d}{p} - d \quad \text{and} \quad \rho_p(w) = -\frac{d}{\min(p, p_{\max})}. \tag{5.130}$$

Proof. The result is easily deduced from the definition of $\tau_p(w)$ and $\rho_p(w)$ in Section 5.3.1. Note that the two first cases treated in Proposition 5.11 gives together that $\tau_p(w) = \frac{d}{p} - d$ and $\rho_p(w) = -\frac{d}{\min(p,p_{\max})}$ for $p < \infty$ while the two last ones gives $\tau_\infty(w) = -d$ and $\rho_\infty(w) = -\frac{d}{p_{\max}}$. Finally, (5.130) condenses all the results. □

Proof of Proposition 5.11. **Case $p < p_{\max}$, $\tau < d/p - d$, and $\rho < -d/p$.** Under these assumptions, we apply Proposition 5.10 to deduce that

$$\begin{aligned} \mathbb{E}\left[\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)}\right] &= \sum_{j\in\mathbb{N}} 2^{j(\tau p - d + dp/2)} \sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} \mathbb{E}[|\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle|^p] \\ &\leq C 2^d \sum_{j\in\mathbb{N}} 2^{j(\tau p - d + dp)} \frac{1}{2^{jd}} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p}, \end{aligned} \tag{5.131}$$

where C is the constant appearing in (5.125), and using that $\text{Card}(\mathcal{G}^j) \leq 2^d$. Then,

$$\frac{1}{2^{jd}} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{\rho p} \underset{j\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} \langle \boldsymbol{x}\rangle^{\rho p} \mathrm{d}\boldsymbol{x} < \infty$$

because $\rho < -d/p$. The sum in (5.131) is therefore finite if and only if $\sum_j 2^{j(\tau p - d + dp)} < \infty$, which happens here due to our assumption $(\tau p - d + dp) < 0$. This shows that $w \in B^\tau_p(\mathbb{R}^d; \rho)$ almost surely.

Case $p_{\max} \leq p$, $\tau < d/p - d$, and $\rho < -d/p_{\max}$. We prove that $w \in B^\tau_p(\mathbb{R}^d; \rho)$ a.s. using the embeddings between Besov spaces and the study of the case $p < p_{\max}$ before. From the conditions on τ and ρ, one can find $p_0 \in (0, p_{\max})$, $\tau_0 \in \mathbb{R}$, and

$\rho_0 \in \mathbb{R}$ such that

$$\rho < \rho_0 < -\frac{d}{p_0} < -\frac{d}{p_{\max}}, \tag{5.132}$$

$$\tau + \frac{d}{p_0} - \frac{d}{p} < \tau_0 < \frac{d}{p_0} - d. \tag{5.133}$$

Then, in particular, $p_0 < p$, $\tau_0 - \tau > d/p_0 - d/p$, and $\rho_0 > \rho$, so that $B_{p_0}^{\tau_0}(\mathbb{R}^d; \rho_0) \subset B_p^{\tau}(\mathbb{R}^d; \rho)$ (according to (5.98)). Moreover, $p_0 < p_{\max}$, $\tau_0 < d/p_0 - d$, and $\rho_0 < -d/p_0$. We are therefore back to the first case, for which we have already shown that $w \in B_{p_0}^{\tau_0}(\mathbb{R}^d; \rho_0)$ a.s. In conclusion, $w \in B_p^{\tau}(\mathbb{R}^d; \rho)$ a.s.

Combining these first two cases, we obtain that $w \in B_p^{\tau}(\mathbb{R}^d; \rho)$ if $\tau < d/p - d$ and $\rho < -d/\min(p, p_{\max})$ for every $p \in (0, \infty]$.

Case $p < \infty$ and $\tau \geq d/p - d$. We use the representation (5.126) of the compound Poisson white noise. Assume that w is in $B_p^{\tau}(\mathbb{R}^d; \rho)$ for some $\rho \in \mathbb{R}$. Then, the product of w by any compactly supported smooth test function φ is well-defined and also in $B_p^{\tau}(\mathbb{R}^d; \rho)$. Choosing a (random) test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$ such that $\varphi(\boldsymbol{x}_0) = 1$ and $\varphi(\boldsymbol{x}_n) = 0$ for $n \neq 0$, we get

$$\varphi w = \varphi \cdot a_0 \delta(\cdot - \boldsymbol{x}_0) = a_0 \delta(\cdot - \boldsymbol{x}_0) \in B_p^{\tau}(\mathbb{R}^d; \rho), \tag{5.134}$$

where $a_0 \neq 0$ a.s. This is absurd due to Proposition 5.7, proving that $w \notin B_p^{\tau}(\mathbb{R}^d; \rho)$ for all $\rho \in \mathbb{R}$.

Case $p = \infty$ and $\tau > -d$. The same argument than for the case $p < \infty$ and $\tau \geq d/p - d$ applies, using this time that $w \notin B_{\infty}^{\tau}(\mathbb{R}^d; \rho)$ for any $\rho > 0$, again due to Proposition 5.7.

Case $\rho \geq -d/p$. This case has been treated in full generality in Proposition 5.9.

Case $p > p_{\max}$ and $\rho > -d/p_{\max}$. This means in particular that $p_{\max} < \infty$.

We treat the case $\rho < 0$, the extension for $\rho \geq 0$ clearly follows from the embedding relations between Besov spaces. We set $q := -d/\rho > p_{\max}$. In particular, according to Proposition 5.4, we have that $\mathbb{E}[|\langle w, \varphi \rangle|^q] = \infty$ for any compactly supported and bounded function $\varphi \neq 0$. Proceeding as for (5.123), we have that

$$\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} \geq \sum_{\boldsymbol{k} \in k_0 \mathbb{Z}^d} \frac{|\langle w, \phi_{\boldsymbol{k}} \rangle|^p}{\langle \boldsymbol{k} \rangle^{dp/q}}, \tag{5.135}$$

where $k_0 \geq 1$ is chosen such that the functions $\phi_{\boldsymbol{k}} = \phi(\cdot - \boldsymbol{k})$ have disjoint supports. Then, the random variables $X_{\boldsymbol{k}} = \langle w, \phi_{\boldsymbol{k}} \rangle$ are i.i.d. The independence implies that the events $A_{\boldsymbol{k}} = \{X_{\boldsymbol{k}} \geq \langle \boldsymbol{k} \rangle^{d/q}\}$ are independent themselves. Then, the $X_{\boldsymbol{k}}$ having the same law, we have

$$\sum_{\boldsymbol{k} \in k_0 \mathbb{Z}^d} \mathscr{P}(A_{\boldsymbol{k}}) = \sum_{\boldsymbol{k} \in k_0 \mathbb{Z}^d} \mathscr{P}(|X_{\boldsymbol{k}}| \geq \langle \boldsymbol{k} \rangle^{d/q}) = \sum_{\boldsymbol{k} \in k_0 \mathbb{Z}^d} \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq \langle \boldsymbol{k} \rangle^{d}) \geq \sum_{m \geq 1} \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq mk_0 \tag{5.136}$$

Moreover, we have $\mathbb{E}[|X_{\boldsymbol{0}}|^q] = \int_0^\infty \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq x) \mathrm{d}x = \sum_{m \geq 1} \int_{mk_0}^{(m+1)k_0} \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq x) \mathrm{d}x$. Exploiting that $\mathscr{P}(|X_{\boldsymbol{0}}|^q \geq x)$ is decreasing in x, we moreover have that $\int_{mk_0}^{(m+1)k_0} \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq x) \mathrm{d}x \leq k_0 \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq mk_0)$, and therefore,

$$\mathbb{E}[|X_{\boldsymbol{0}}|^q] \leq \sum_{m \geq 1} \mathscr{P}(|X_{\boldsymbol{0}}|^q \geq mk_0). \tag{5.137}$$

The choice of q implies moreover that $\mathbb{E}[|X_{\boldsymbol{0}}|^q] = \mathbb{E}[|\langle w, \phi \rangle|^q] = \infty$ due to Proposition 5.4. Hence, from (5.136) and (5.137), we deduce that $\sum_{\boldsymbol{k} \in k_0 \mathbb{Z}^d} \mathscr{P}(A_{\boldsymbol{k}}) \geq \mathbb{E}[|X_{\boldsymbol{0}}|^q] = \infty$. The Borel-Cantelli lemma then implies that $|X_{\boldsymbol{k}}|^p \geq \langle \boldsymbol{k} \rangle^{dp/q}$ for infinitely many $\boldsymbol{k}$ a.s. Back to (5.135), this implies that $\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} = \infty$ a.s.

□

5.3.5 Moment Estimates for the Lévy White Noise

The proof of Theorem 5.1 will be based on new estimates for the moments $\mathbb{E}[|\langle w, \varphi \rangle|^p]$ of Lévy white noises. In Section 5.3.5, we consider the case $p = 2m$ where $m \geq 1$ is an integer. This will be critical when dealing with Lévy white noises with finite moments. In Section 5.3.5, we determine lower bounds for the moments, which is the main technicality for the negative Besov regularity results of Lévy white noises.

Moment Estimates for $p = 2m \geq 2$

We estimate the evolution of the even moments of the wavelet coefficients of a Lévy white noise w with the scale j. Most of the moment estimates in the literature deal with pth moments with the restriction $p \leq 2$ [408, 431, 409, 412], and we shall see that the extension to higher-order moments calls for some technicalities.

Proposition 5.12. *Let w be a Lévy white noise with finite moments and $m \geq 1$ be an integer. We assume that the moment index of w satisfies $p_{\max} > 2m$. Then, there exists a constant $C > 0$ such that, for every $j \in \mathbb{N}$, $\boldsymbol{G} \in \mathcal{G}^j$, and $\boldsymbol{k} \in \mathbb{Z}^d$,*

$$\mathbb{E}[\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle^{2m}] \leq C 2^{jd(m-1)}. \tag{5.138}$$

Proof. Consider a test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and set $X = \langle w, \varphi \rangle$. The characteristic function of X is [381, Proposition 2.12]

$$\widehat{\mathscr{P}}_X(\xi) = \exp\left(\int_{\mathbb{R}^d} \Psi(\xi\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right) := \exp(\Psi_\varphi(\xi)). \tag{5.139}$$

The functions $\widehat{\mathscr{P}}_X$ and Ψ_φ are $(2m)$-times differentiable because $\mathbb{E}[X^{2m}] < \infty$ [451, Theorem 1.5.1]. Their Taylor expansions give the moments and the cumulants of X, respectively. In particular, we have that $\mathbb{E}[X^{2m}] = (-1)^m \widehat{\mathscr{P}}_X^{(2m)}(0)$. Using the Faà di Bruno formula with the composite function $\xi \mapsto \widehat{\mathscr{P}}_X(\xi) = \exp(\Psi_\varphi(\xi))$, we express the $(2m)$th derivative of $\widehat{\mathscr{P}}_X$ as

$$\widehat{\mathscr{P}}_X^{(2m)}(\xi) = \left(\sum_{n_1,\dots,n_{2m} : \sum_u u n_u = 2m} \frac{(2m)!}{n_1! \dots n_{2m}!} \prod_{v=1}^{2m} \left(\frac{\Psi_\varphi^{(v)}(\xi)}{v!} \right)^{n_v} \right) \widehat{\mathscr{P}}_X(\xi). \tag{5.140}$$

Exploiting that $\Psi_{\varphi}^{(v)}(0) = \left(\int_{\mathbb{R}^d} (\varphi(\boldsymbol{x}))^v \mathrm{d}\boldsymbol{x}\right) \Psi^{(v)}(0)$ for $\xi = 0$ [48, Proposition 9.11], we obtain the bound,

$$\left|\widehat{\mathscr{P}}_X^{(2m)}(0)\right| \leq C' \sum_{n_1,\dots,n_{2m}:\sum_u u n_u = 2m} \prod_{v=1}^{2m} \left(\int_{\mathbb{R}^d} |\varphi(\boldsymbol{x})|^v \, \mathrm{d}\boldsymbol{x}\right)^{n_v} \tag{5.141}$$

with $C' > 0$ a constant. We now apply (5.141) to $\varphi = \psi_{j,\boldsymbol{G},\boldsymbol{k}}$. Since we have

$$\int_{\mathbb{R}^d} |\psi_{j,\boldsymbol{G},\boldsymbol{k}}(\boldsymbol{x})|^v \, \mathrm{d}\boldsymbol{x} = 2^{jdv/2} \int_{\mathbb{R}^d} \left|\psi_{\boldsymbol{G}}(2^j \boldsymbol{x} - \boldsymbol{k})\right|^v \mathrm{d}\boldsymbol{x} = 2^{jd(v/2-1)} \int_{\mathbb{R}^d} |\psi_{\boldsymbol{G}}(\boldsymbol{x})|^v \, \mathrm{d}\boldsymbol{x}, \tag{5.142}$$

we deduce from (5.141) the new bound

$$\begin{aligned} \mathbb{E}[\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle^{2m}] = \left|\widehat{\mathscr{P}}^{(2m)}_{\langle w,\psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle}(0)\right| &\leq C'' \sum_{n_1,\dots,n_{2m}:\sum_u u n_u = 2m} \prod_{v=1}^{2m} 2^{jdn_v(v/2-1)} \\ &= C'' \sum_{n_1,\dots,n_{2m}:\sum_u u n_u = 2m} 2^{jd\sum_{1\leq v\leq 2m}(n_v(v/2-1))}, \end{aligned} \tag{5.143}$$

where C'' is a new constant independent from $j, G, \boldsymbol{k}$. Finally, since $\sum_v n_v v = 2m$ and $\sum_v n_v \geq 1$, we have $\sum_v (n_v(v/2-1)) \leq (m-1)$. Therefore, we obtain (5.138) for an adequate $C > 0$. □

Lower Bound for Moment Estimates

In our previous moment estimates, we gave upper bounds for the quantity $\mathbb{E}[|\langle w, \varphi\rangle|^p]$ (see not only Propositions 5.10 and 5.12, but also Theorem 2 in [431]). This allows one to identify in which Besov space is w. We now address the following problem: Can we bound $\mathbb{E}[|\langle w, \varphi\rangle|^p]$ from below with the moments of φ? Theorem 5.2 answers positively to this question and is crucial for the proof of Theorem 5.1.

Theorem 5.2. *Let w be a Lévy white noise whose indices satisfy $0 < \underline{\beta}_\infty \leq \beta_\infty < p_{\max}$, $\psi \neq 0$ a bounded, and compactly supported test function, and p an integrability* *parameter such that* $0 < p < \beta_\infty$*. Then, for* $\epsilon > 0$ *small enough, there exists constants* $A, B > 0$ *independent from* $j \in \mathbb{N}$ *and* $\boldsymbol{k} \in \mathbb{Z}^d$ *such that*

$$A2^{-j\epsilon}2^{jdp(1/2-1/\underline{\beta}_\infty)} \leq \mathbb{E}[|\langle w, \psi_{j,\boldsymbol{k}}\rangle|^p] \leq B2^{j\epsilon}2^{jdp(1/2-1/\beta_\infty)} \tag{5.144}$$

for any $j \in \mathbb{N}$ *and* $\boldsymbol{k} \in \mathbb{Z}^d$*, where we recall that* $\psi_{j,\boldsymbol{k}} = 2^{jd/2}\psi(2^j \cdot -\boldsymbol{k})$*.*

Proof. First of all, the shift parameter $\boldsymbol{k}$ in (5.144) can be omitted since w is stationary. We also remark that the upper bound of (5.144) has already been proven [431, Corollary 1], where the conditions $p < \beta_\infty < p_{\max}$ are required. Actually, [431] does not consider the index $p_{\max}$ and distinguishes between the conditions $\beta_\infty < \beta_0 < 2$ and $\beta_\infty \leq \beta_0 = 2$ with finite variance. These two scenarios cover the condition $\beta_\infty < p_{\max}$ of Theorem 5.2. Hence, we focus on the lower bound.

Because $p < \beta_\infty \leq 2$, one can use the representation of the pth moment of $\langle w, \varphi\rangle$, that can be found in [451, Theorem 1.5.9] (with $n = 0$ and $\delta = p$):

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] = c_p \int_{\mathbb{R}} \frac{1 - \Re\{\widehat{\mathscr{P}}_{\langle w,\varphi\rangle}(\xi)\}}{|\xi|^{p+1}}\mathrm{d}\xi \tag{5.145}$$

for some explicit constant $c_p > 0$.The relation (5.145) is often used for moment estimates, for instance in [408, 412]. We then remark that

$$\begin{aligned}\Re\{\widehat{\mathscr{P}}_{\langle w,\varphi\rangle}(\xi)\} \leq \left|\widehat{\mathscr{P}}_{\langle w,\varphi\rangle}(\xi)\right| = \left|\widehat{\mathscr{P}}_w(\xi\varphi)\right| = \left|\exp\left(\int_{\mathbb{R}^d} \Psi(\xi\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right)\right| \\ = \exp\left(\int_{\mathbb{R}^d} \Re\{\Psi(\xi\varphi(\boldsymbol{x}))\}\mathrm{d}\boldsymbol{x}\right).\end{aligned} \tag{5.146}$$

The test function φ is chosed to be non-identically zero, hence there exists some constant $m > 0$ such that $\mathrm{Leb}(|\varphi| \geq m\|\varphi\|_\infty) > 0$. We fix such a constant m and observe that

$$\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty} \neq 0. \tag{5.147}$$

The sector condition (5.76) implies that one can found $c_\Psi > 0$ such that $-\Re\{\Psi(\xi)\} = |\Re\{\Psi\}|(\xi) \geq c_\Psi |\Psi(\xi)|$. Thus, one has that $\left|\widehat{\mathscr{P}}_{\langle w,\varphi\rangle}(\xi)\right| \leq \exp\left(-c_\Psi \int_{\mathbb{R}^d} |\Psi(\xi\varphi(\boldsymbol{x}))|\,\mathrm{d}\boldsymbol{x}\right)$, and then

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] \geq c_p \int_{\mathbb{R}} \frac{1 - \mathrm{e}^{-c_\Psi \int_{\mathbb{R}^d}|\Psi(\xi\varphi(\boldsymbol{x}))|\mathrm{d}\boldsymbol{x}}}{|\xi|^{p+1}}\mathrm{d}\xi. \tag{5.148}$$

Now, by definition of the index $\underline{\beta}_{\infty}$, the function $|\Psi|$ is dominating $|\xi|^{\underline{\beta}_{\infty}-\delta}$ at infinity for an arbitrarily small δ such that $0 < \delta < \underline{\beta}_{\infty}$. We fix such δ and set $\beta = (\underline{\beta}_{\infty} - \delta)$. This domination, together with the continuity of the functions $|\Psi|$ and $|\cdot|^{\beta} \mathbb{1}_{|\cdot|>1}$ over $[1, \infty)$, imply the existence of $C > 0$ such that, for any $\xi \in \mathbb{R}$, $|\Psi(\xi)| \geq C \, |\xi|^{\beta} \, \mathbb{1}_{|\xi|>1}$. In particular, we have that

$$\begin{aligned}\int |\Psi(\xi\varphi(\boldsymbol{x}))| \, \mathrm{d}\boldsymbol{x} \geq C \, |\xi|^{\beta} \int_{\mathbb{R}^d} |\varphi(\boldsymbol{x})|^{\beta} \, \mathbb{1}_{|\xi\varphi(\boldsymbol{x})|>1} \mathrm{d}\boldsymbol{x} & \\ \overset{(i)}{\geq} C\mathbb{1}_{|\xi|>1/m\|\varphi\|_\infty} \, |\xi|^{\beta} \int_{\mathbb{R}^d} |\varphi(\boldsymbol{x})|^{\beta} \, \mathbb{1}_{|\varphi(\boldsymbol{x})|>m\|\varphi\|_\infty} \mathrm{d}\boldsymbol{x} & \\ = C\mathbb{1}_{|\xi|>1/m\|\varphi\|_\infty} \, |\xi|^{\beta} \, \|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta}^{\beta}, & \end{aligned} \tag{5.149}$$

where the (i) uses that $\mathbb{1}_{|\xi\varphi(\boldsymbol{x})|>1} \geq \mathbb{1}_{|\varphi(\boldsymbol{x})|>m\|\varphi\|_\infty} \mathbb{1}_{|\xi|>1/m\|\varphi\|_\infty}$, where m is such that (5.147) holds.

Combining (5.148) and (5.149), we therefore have

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] \geq c_p \int_{\mathbb{R}} \frac{1 - \mathrm{e}^{-c_\Psi C \mathbb{1}_{|\xi|>1/m\|\varphi\|_\infty} |\xi|^{\beta} \|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta}^{\beta}}}{|\xi|^{p+1}} \mathrm{d}\xi. \tag{5.150}$$

We use the change of variable $u = (c_\Psi C)^{1/\beta} \|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta} \xi$ to obtain

$$\mathbb{E}[|\langle w, \varphi\rangle|^p] \geq C' \|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta}^{p} \int_{\mathbb{R}} \frac{1 - \exp\left(- |u|^{\beta} \, \mathbb{1}_{|u|>C'' \frac{\|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_\beta}{\|\varphi\|_\infty}} \right)}{|u|^{p+1}} \mathrm{d}u \tag{5.151}$$

for some constants $C', C'' > 0$.We now set $\varphi = \psi_j = 2^{jd/2} \psi(2^j \cdot)$, and observe that, with simple changes of variable,

$$\|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta}^{p} = 2^{jdp\left(\frac{1}{2}-\frac{1}{\beta}\right)} \|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_{\beta}^{p}, \tag{5.152}$$

$$\frac{\|\varphi\mathbb{1}_{|\varphi|>m\|\varphi\|_\infty}\|_{\beta}}{\|\varphi\|_\infty} = \frac{2^{jd\left(\frac{1}{2}-\frac{1}{\beta}\right)} \|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_{\beta}}{2^{jd/2}\|\psi\|_\infty} = 2^{-jd/\beta} \frac{\|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_{\beta}}{\|\psi\|_\infty}. \tag{5.153}$$

In particular,

$$\mathbb{1}_{|u|>C''2^{-jd/\beta}\frac{\|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_\beta}{\|\psi\|_\infty}} \geq \mathbb{1}_{|u|>C''\frac{\|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_\beta}{\|\psi\|_\infty}} \tag{5.154}$$

for any $j \in \mathbb{N}$. Hence, we deduce using (5.151) with $\varphi = \psi_j$ that

$$\mathbb{E}[|\langle w, \psi_j\rangle|^p] \geq B2^{jdp\left(\frac{1}{2}-\frac{1}{\beta}\right)} \tag{5.155}$$

with $B > 0$ a constant given by

$$B = C'\|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_\beta^p \int_{\mathbb{R}} \frac{1-\exp\left(-|u|^\beta \mathbb{1}_{|u|>C''\frac{\|\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty}\|_\beta}{\|\psi\|_\infty}}\right)}{|u|^{p+1}}\mathrm{d}u. \tag{5.156}$$

Remark that $B \neq 0$ because $\psi\mathbb{1}_{|\psi|>m\|\psi\|_\infty} \neq 0$ due to (5.152) and (5.147) . To conclude, we remark that, for $\epsilon > 0$ fixed, one can find $\delta > 0$ small enough such that $2^{-j\epsilon}2^{jd\left(\frac{1}{2}-\frac{1}{\underline{\beta}_\infty}\right)} \leq 2^{jd\left(\frac{1}{2}-\frac{1}{\underline{\beta}_\infty-\delta}\right)} = 2^{jd\left(\frac{1}{2}-\frac{1}{\beta}\right)}$ for any $j \in \mathbb{N}$, which gives the lower bound in (5.144). □

5.3.6 Lévy White Noise with Finite Moments

We consider Lévy white noises whose all the moments are finite, which means that $p_{\max} = \infty$. Their specificity is that one can use the finiteness of the pth moments of the wavelet coefficients of the Lévy white noise w for any $p > 0$. Thanks to the moment estimates in Section 5.3.5, we have all the tools to deduce the Besov regularity of white noises with finite moments.

Proposition 5.13. *Fix $0 < p \leq \infty$ and $\tau, \rho \in \mathbb{R}$. Let w be a Lévy white noise with finite moments and Blumenthal-Getoor indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$. Then, w is*

- *almost surely in $B_p^\tau(\mathbb{R}^d;\rho)$ if $\tau < d/\max(p,\beta_\infty) - d$ and $\rho < -d/p$, for $0 < p \leq 2$, p an even integer, or $p = \infty$; and*
- *almost surely not in $B_p^\tau(\mathbb{R}^d;\rho)$ if $\tau > d/\max(p,\underline{\beta}_\infty) - d$ or $\rho \geq -d/p$ for every $0 < p \leq \infty$.*

As was the case for the Gaussian and compound Poisson cases, Proposition 5.13 allows to deduce the asymptotic growth rate of Lévy white noises with finite moments. Moreover, we obtain lower and upper bounds for the local smoothness in terms of the Blumenthal-Getoor indices of the Lévy white noise.

Corollary 5.5. *Let* $0 < p \leq \infty$ *and* w *be a Lévy white noise with finite moments and Blumenthal-Getoor indices* $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$*. Then, we have that*

$$\frac{d}{\max(p, \beta_\infty)} - d \leq \tau_p(w) \leq \frac{d}{\max(p, \underline{\beta}_\infty)} - d \quad \text{and} \quad \rho_p(w) \geq -\frac{d}{p}. \tag{5.157}$$

For $p \in (0,2)$*,* p *an even integer, or* $p = \infty$*, we moreover have that*

$$\rho_p(w) = -\frac{d}{p}. \tag{5.158}$$

Proof of Proposition 5.13. We only treat the case $p < \infty$. For $p = \infty$, the result is obtained using embeddings (with $p = 2m$, $m \to \infty$ for positive results and $p \to \infty$ for negative results) following the same arguments than for the Gaussian case in Proposition 5.8.

If $\tau < d/\max(p, \beta_\infty) - d$ **and** $\rho < -d/p$**.** We first remark that for $p \leq 2$, we have that $\rho < -\frac{d}{\min(p,2,p_{\max})} = -\frac{d}{\min(p,p_{\max})}$, and the result is a consequence of our previous work [431, Theorem 3]. We can therefore assume that $p = 2m$ with $m \geq 1$ an integer. Then, $p = 2m \geq 2 \geq \beta_\infty$, and the conditions on τ and ρ become $\tau < \frac{d}{2m} - d$ and $\rho < -\frac{d}{2m}$. Due to Proposition 5.12, we have

$$\begin{aligned} \mathbb{E}\left[\|w\|^{2m}_{B^\tau_{2m}(\mathbb{R}^d;\rho)}\right] &= \sum_{j\in\mathbb{N}} 2^{j(2m\tau - d + dm)} \sum_{\boldsymbol{G}\in\mathcal{G}^j} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{2m\rho} \mathbb{E}[\langle w, \psi_{j,\boldsymbol{G},\boldsymbol{k}}\rangle^{2m}] \\ &\leq C 2^d \sum_{j\in\mathbb{N}} 2^{j(2m\tau - d + 2dm)} \frac{1}{2^{jd}} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{2m\rho}. \end{aligned} \tag{5.159}$$

Then, $\frac{1}{2^{jd}} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle 2^{-j}\boldsymbol{k}\rangle^{2m\rho} \underset{j\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} \langle \boldsymbol{x}\rangle^{2m\rho} \mathrm{d}\boldsymbol{x} < \infty$ since $2m\rho < -d$, and the series in (5.159) is finite if and only if $2m\tau - d + 2dm < 0$, what we assumed to be true.

Finally, we have shown that $\mathbb{E}\left[\|w\|^{2m}_{B^{\tau}_{2m}(\mathbb{R}^d;\rho)}\right] < \infty$, so that $w \in B^{\tau}_{2m}(\mathbb{R}^d;\rho)$ almost surely.

Case $\tau > d/p - d$. This part of the proof is actually valid for any Lévy white noise. It uses the decomposition $w = w_1 + w_2$ with w_1 a nontrivial compound Poisson white noise and w_2 a Lévy white noise with finite moments (see Proposition 5.3, here w_2 combines the Gaussian part and the finite-moment part of the Lévy-Itô decomposition). The main idea is that the jumps of the compound Poisson part are by themselves enough to make the Besov norm infinite.

One writes that $w_1 \overset{(\mathcal{L})}{=} \sum_{k\in\mathbb{Z}} a_k \delta(\cdot - \boldsymbol{x}_k)$, as in (5.126). Then, almost surely, all the a_k are nonzero. Let $m > 0$ be such that $\mathscr{P}(|a_0| \geq m) > 0$. One can assume without loss of generality that $m = 1$ (otherwise, consider the white noise w/m). Then, there exists almost surely $k \in \mathbb{Z}$ such that $|a_k| \geq 1$ (it suffices to apply the Borel-Cantelli argument using the independence of the a_k). For simplicity, one reorders the jumps such that this is achieved for $k = 0$, so that $|a_0| \geq 1$.

We first introduce preliminary notations. Let $[a, b]$ be a finite interval on which the mother Daubechies wavelet is strictly positive. In particular, $\min_{x\in[a,b)} |\psi_M(x)| := c > 0$. Then, setting $K = [a, b)^d$ and $C = c^d$, we have that $|\psi_{\boldsymbol{M}}(\boldsymbol{x})| \geq C$ for every $\boldsymbol{x} \in K$, with $\boldsymbol{M} = (M, \ldots, M)$. For each scale $j \geq 0$, we define $\boldsymbol{k}_j \in \mathbb{Z}^d$ as the unique multi-integer such that $2^j\boldsymbol{x}_0 - \boldsymbol{k}_j \in [a, a + 1)^d$. If $b \geq a + 1$, then $2^j\boldsymbol{x}_0 - \boldsymbol{k}_j \in K$. Otherwise, due to the law of the location $\boldsymbol{x}_0$, there is almost surely an infinity of scale $j \geq 0$ such that $2^j\boldsymbol{x}_0 - \boldsymbol{k}_j \in [a, b)^d = K$. We denote by J the (random) ensemble of such j. Then, for each $j \in J$, using that $|a_0| \geq 1$ and $2^j\boldsymbol{x}_0 - \boldsymbol{k}_j \in [a, b)^d = K$, we deduce that

$$\left|a_0\psi_{\boldsymbol{M}}(2^j\boldsymbol{x}_0 - \boldsymbol{k}_j)\right| \geq C. \tag{5.160}$$

Moreover, on each finite interval, there are almost surely finitely many jumps $\boldsymbol{x}_k$. In particular, the random variable $\inf_{k\in\mathbb{Z}\setminus\{0\}} \|\boldsymbol{x}_k - \boldsymbol{x}_0\|$ is a.s. strictly positive. This implies that there exists a (random) integer $j_0 \in \mathbb{N}$ such that $2^{j_0}\|\boldsymbol{x}_k - \boldsymbol{x}_0\| > \operatorname{diam}(\operatorname{Supp}(\psi_{\boldsymbol{M}}))$ for any $k \neq 0$, where $\operatorname{diam}(B)$ is the diameter of a Borelian set $B \subset \mathbb{R}^d$, understood as the Lebesgue measure of its closed convex hull. We therefore have that $\psi_{\boldsymbol{M}}(2^j \boldsymbol{x}_k - \boldsymbol{k}_j) = 0$ for any $j \geq j_0$ and any $k \neq 0$. From these preparatory considerations, one has a.s. that, for $j \geq j_0$, $j \in J$,

$$\left|\langle w_1, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| = \left|\sum_{k \in \mathbb{Z}} a_k \psi_{\boldsymbol{M}}(2^j \boldsymbol{x}_k - \boldsymbol{k}_j)\right| = \left|a_0 \psi_{\boldsymbol{M}}(2^j \boldsymbol{x}_0 - \boldsymbol{k}_j)\right| \geq C. \tag{5.161}$$

Let now focus on the Lévy white noise w_2. Since w_2 has a finite variance, we have that, using the Markov inequality,

$$\mathscr{P}\left(\left|\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \geq \frac{C}{2}\right) \leq \frac{4\mathbb{E}\left[\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle^2\right]}{C^2} = \frac{4\sigma_0^2 \|\psi_{\boldsymbol{M}}\|_2^2}{C^2} 2^{-jd}, \tag{5.162}$$

where σ_0^2 is the variance of w_2 such that $\mathbb{E}[\langle w_2, \varphi\rangle^2] = \sigma_0^2 \|\varphi\|_2^2$ for any test function φ. In particular,

$$\sum_{j \in \mathbb{N}} \mathscr{P}\left(\left|\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \geq \frac{C}{2}\right) \leq \frac{4\sigma_0^2 \|\psi_{\boldsymbol{M}}\|_2^2}{C^2} \sum_{j \in \mathbb{N}} 2^{-jd} < \infty.$$

From a new Borel-Cantelli argument, we know that, almost surely, only finitely many j satisfy $\left|\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \geq C/2$. In particular, there exists a (random) integer j_1 such that, for any $j \geq j_1$, $\left|\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \leq C/2$. Combining this to (5.161), we then have that, for $j \in J$ such that $j \geq \max(j_0, j_1)$,

$$\left|\langle w, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \geq \left|\langle w_1, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| - \left|\langle w_2, \psi_{\boldsymbol{M}}(2^j \cdot - \boldsymbol{k}_j)\rangle\right| \geq C - C/2 = C/2. \tag{5.163}$$

Note moreover that, by definition of $\boldsymbol{k}_j$, $2^j \boldsymbol{x}_0 - \boldsymbol{k}_j \in [a, a+1)^d$, hence we have that $\|\boldsymbol{x}_0 - 2^{-j}\boldsymbol{k}_j\|_\infty \leq M/2^j$ where $M = \max(|a|, |a+1|)$. Then, there exists a (random) integer $j_2 \geq 1$ such that, for every $j \geq j_2$, $M/2^j \leq \|\boldsymbol{x}_0\|_\infty$. We have moreover that $\|2^{-j}\boldsymbol{k}_j\|_2 \leq d^{1/2}\|2^{-j}\boldsymbol{k}_j\|_\infty \leq d^{1/2}(M/2^j + \|\boldsymbol{x}_0\|_\infty)$. Therefore, recalling that $\rho < 0$, for every $j \geq j_2$, we have

$$\langle 2^{-j}\boldsymbol{k}_j\rangle^{\rho p} \geq (1 + d(M/2^j + \|\boldsymbol{x}_0\|_\infty)^2)^{\rho p/2} \geq (1 + 4d\|\boldsymbol{x}_0\|_\infty^2)^{\rho p/2}. \tag{5.164}$$

Putting the pieces together, we can now lower bound the Besov norm of w by keeping only the mother wavelet $\psi_{\boldsymbol{M}}$, a scale $j \in J$ such that $j \geq \max(j_0, j_1, j_2)$, and the corresponding shift parameter $\boldsymbol{k}_j$. Then, combining (5.163) and (5.164), we obtain the almost-sure lower bound

$$\begin{aligned}\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)} &\geq 2^{j(\tau p-d+dp)}\langle 2^{-j}\boldsymbol{k}_j\rangle^{p\rho}\left|\langle w,\psi_{\boldsymbol{M}}(2^j\cdot-\boldsymbol{k}_j)\rangle\right|^p\\ &\geq 2^{j(\tau p-d+dp)}\,(C/2)^p\,(1+4d\|\boldsymbol{x}_0\|^2_\infty)^{\rho p/2}. \qquad (5.165)\end{aligned}$$

This is valid for any $j\in J$ such that $j\geq\max(j_0,j_1,j_2)$ and because J is infinite and $(\tau p-d+dp)>0$, one concludes that $\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)}=\infty$ almost surely.

Case $0<p<\underline{\beta}_\infty$ **and** $(d/\underline{\beta}_\infty-d)<\tau<(d/p-d)$**.** Assume that, under those assumptions, we prove that $w\notin B^\tau_p(\mathbb{R}^d;\rho)$ a.s. Then, together the case $\tau>(d/p-d)$ considered below and using embeddings, we deduce the expected result for $\tau>p(d/\max(\underline{\beta}_\infty,p)-d)$.

As soon as $f\notin B^\tau_p(\mathbb{R}^d;\rho)$ for some $p>0$, we also have that $f\notin B^{\tau+\epsilon}_q(\mathbb{R}^d;\rho)$ for any $q>p$ and $\epsilon>0$ (see Figure 5.9). A crucial consequence for us is that it suffices to work with arbitrarily small p in order to obtain the negative result we expect. We assume here that

$$p<\underline{\beta}_\infty/2,\quad p<\frac{\underline{\beta}_\infty\beta_\infty}{2(\beta_\infty-\underline{\beta}_\infty)}. \qquad (5.166)$$

Note that the right inequality in (5.166) simply means that $p<\infty$ (*i,e.*, no restriction) when $\underline{\beta}_\infty=\beta_\infty$. We fix $k_0\in\mathbb{N}\backslash\{0\}$ such that, for any gender $\boldsymbol{G}$, the functions $\Psi_{0,G,k_0\boldsymbol{k}}$ have disjoint support for every $\boldsymbol{k}\in\mathbb{Z}^d$. Then, at fixed $\boldsymbol{G}$ and j, the random variables $(\langle w,\Psi_{j,G,\boldsymbol{k}}\rangle)_{\boldsymbol{k}\in k_0\mathbb{Z}^d}$ are independent. By restricting the range of $\boldsymbol{k}$ and the gender to $\boldsymbol{G}=\boldsymbol{M}$, we have that

$$\|w\|^p_{B^\tau_p(\mathbb{R}^d;\rho)}\geq C\sum_{j\in\mathbb{N}}2^{j(\tau p-d+dp/2)}\sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d,0\leq k_i<k_02^j}\left|\langle w,\psi_{j,\boldsymbol{M},\boldsymbol{k}}\rangle\right|^p, \qquad (5.167)$$

with $C=\inf_{\|\boldsymbol{x}\|_\infty\leq k_0}\langle\boldsymbol{x}\rangle^\rho>0$ is such that $\langle 2^{-j}\boldsymbol{k}\rangle\geq C$ for any $\boldsymbol{k}\in k_0\{0,\dots 2^j-1\}^d$ and any $j\geq0$. We set $X_{j,\boldsymbol{k}}=2^{jd\left(\frac{1}{\underline{\beta}_\infty}-\frac{1}{2}\right)}\langle w,\psi_{j,\boldsymbol{M},\boldsymbol{k}}\rangle$ and

$$M_{j,p}:=2^{-jd}\sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d,0\leq k_i<k_02^j}\left|X_{j,\boldsymbol{k}}\right|^p,$$

which is an average among 2^{jd} independent random variables.

Recall that $p < \underline{\beta}_\infty/2$. Moreover, since all the moments are finite, $p_{\max} = \infty > \beta_\infty$. Hence, one can apply Theorem 5.2 with integrability parameters $q = p$ and $q = 2p$, respectively. There exists $\epsilon > 0$ that can be choosen arbitrarily small and constants m_q, M_q such that

$$m_q 2^{-j\epsilon} 2^{jqd\left(\frac{1}{2} - \frac{1}{\underline{\beta}_\infty}\right)} \leq \mathbb{E}\left[|\langle w, \psi_{j,\boldsymbol{M},\boldsymbol{k}}\rangle|^q\right] \leq M_q 2^{j\epsilon} 2^{jqd\left(\frac{1}{2} - \frac{1}{\beta_\infty}\right)} \tag{5.168}$$

for any $j \in \mathbb{N}, \boldsymbol{k} \in \mathbb{Z}^d$. In particular, with our notations, we have that

$$m_q 2^{-j\epsilon} \leq \mathbb{E}[M_{j,q}] \leq M_q 2^{j\epsilon} 2^{jqd\left(\frac{1}{\underline{\beta}_\infty} - \frac{1}{\beta_\infty}\right)} \tag{5.169}$$

for any $j, \boldsymbol{k}$ and for $q = p$ or $q = 2p$. Then, we control the variance of $M_{j,q}$ as follows:

$$\begin{aligned} \mathrm{Var}(M_{j,p}) = \mathbb{E}\left[(M_{j,p} - \mathbb{E}[M_{j,p}])^2\right] &\overset{(i)}{=} 2^{-jd}\mathbb{E}\left[2^{-jd}\left(\sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d, 0\leq k_i < k_0 2^j} (|X_{j,\boldsymbol{k}}|^p - \mathbb{E}[|X_{j,\boldsymbol{k}}|^p])\right)\right. \\ &\overset{(ii)}{=} 2^{-jd}\mathbb{E}\left[2^{-jd} \sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d, 0\leq k_i < k_0 2^j} ((|X_{j,\boldsymbol{k}}|^p - \mathbb{E}[|X_{j,\boldsymbol{k}}|^p]))^2\right] \\ &\overset{(iii)}{\leq} 2^{-jd}\mathbb{E}\left[2^{-jd} \sum_{\boldsymbol{k}\in k_0\mathbb{Z}^d, 0\leq k_i < k_0 2^j} |X_{j,\boldsymbol{k}}|^{2p}\right] = 2^{-jd}\mathbb{E}[M_{j,2p}] \\ &\overset{(iv)}{\leq} 2^{-jd} 2^{j\epsilon} 2^{2jpd\left(\frac{1}{\underline{\beta}_\infty} - \frac{1}{\beta_\infty}\right)}, \end{aligned} \tag{5.170}$$

where we used that $\mathbb{E}[M_{j,p}] = \mathbb{E}[|X_{j,\boldsymbol{k}}|^p]$ for every $\boldsymbol{k}$ in (i), the independence of the $X_{j,\boldsymbol{k}}$ in (ii), the relation $\mathrm{Var}(X) \leq \mathbb{E}[X^2]$ in (iii), and the right side of (5.169) with $q = 2p$ in (iv).

We then apply the Chebyshev's inequality $\mathscr{P}(|X - \mathbb{E}[X]| \geq x) \leq \mathrm{Var} X/x^2$ to

$X = M_{j,p}$ and $x = 2^{-j\epsilon} m_p/2$ to get

$$\begin{aligned}\mathscr{P}(|M_{j,p} - \mathbb{E}[M_{j,p}]| \geq 2^{-j\epsilon} m_p/2) &\leq \frac{4\mathrm{Var}(M_{j,p})}{m_p^2} 2^{2j\epsilon} \\ &\leq \frac{4M_{2p}}{m_p^2} 2^{j\left(d(2p(1/\underline{\beta}_\infty - 1/\beta_\infty)-1)+3\epsilon\right)} \end{aligned} \tag{5.171}$$

where we used (5.170) in the last inequality. Due to the second inequality in (5.166), we have that

$$2p\left(\frac{1}{\underline{\beta}_\infty} - \frac{1}{\beta_\infty}\right) - 1 < 0. \tag{5.172}$$

Hence, the exponent $d(2p(1/\underline{\beta}_\infty - 1/\beta_\infty) - 1) + 3\epsilon$ in (5.171) is strictly negative for ϵ small enough, what we assume from now. Therefore, $\sum_j \mathscr{P}(|M_{j,p} - \mathbb{E}[M_{j,p}]| \geq 2^{-j\epsilon} m_p/2) < \infty$. From the Borel-Cantelli lemma, only a finite number of such j can therefore satisfy the relation $|M_{j,p} - \mathbb{E}[M_{j,p}]| \geq 2^{-j\epsilon} m_p/2$. A consequence is then that there exists almost surely a random $J \in \mathbb{N}$ such that for every $j \geq J$,

$$M_{j,p} \geq \mathbb{E}[M_{j,p}] - |M_{j,p} - \mathbb{E}[M_{j,p}]| \geq m_p 2^{-j\epsilon} - \frac{m_p}{2} 2^{-j\epsilon} = \frac{m_p}{2} 2^{-j\epsilon}, \tag{5.173}$$

where we used the lower bound in (5.169) with $q = p$. We deduce that

$$\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} \geq C \sum_{j \geq J} 2^{jp(\tau - d + d/\underline{\beta}_\infty)} M_{j,p} \geq \frac{Cm_p}{2} \sum_{j \geq J} 2^{jp(\tau - d + d/\underline{\beta}_\infty - \epsilon)}.$$

For ϵ small enough, we have that $(\tau + d - d/\underline{\beta}_\infty - \epsilon) > 0$ and, therefore, that $\|w\|^p_{B_p^\tau(\mathbb{R}^d;\rho)} = \infty$ almost surely.

If $\rho \geq -d/p$. This case has been treated in full generality in Proposition 5.9. □

5.3.7 Lévy White Noise: the General Case

This section gives us the opportunity to consolidate the results and to deduce the general case from the previous ones. We say that a Lévy white noise w is *non-Gaussian* if its Lévy measure is not identically zero. In particular, w can have a Gaussian part in the Lévy-Itô decomposition (see Proposition 5.3). Proposition 5.14 characterizes the Besov regularity of non-Gaussian Lévy white noises, the Gaussian white noise having already been treated in Section 5.3.3. We conclude this section with the proof of Theorem 5.1.

Proposition 5.14. *Fix $0 < p \leq \infty$ and $\tau, \rho \in \mathbb{R}$. Consider a non-Gaussian Lévy white noise w with Blumenthal-Getoor and moment indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$ and $0 < p_{\max} \leq \infty$. Then, w is*

- *almost surely in $B_p^\tau(\mathbb{R}^d;\rho)$ if $\tau < d/\max(p,\beta_\infty) - d$ and $\rho < -d/\min(p,p_{\max})$, for $0 < p \leq 2$, p an even integer, or $p = \infty$; and*
- *almost surely not in $B_p^\tau(\mathbb{R}^d;\rho)$ if $\tau > d/\max(p,\underline{\beta}_\infty) - d$ or $\rho > -d/\min(p,p_{\max})$ for every $0 < p \leq \infty$.*

Proposition 5.14 reveals new information on the local smoothness and the asymptotic growth rate of non-Gaussian Lévy white noises.

Corollary 5.6. *Let $0 < p \leq \infty$ and w be a Lévy white noise with finite moments and Blumenthal-Getoor indices $0 \leq \underline{\beta}_\infty \leq \beta_\infty \leq 2$. Then, we have that*

$$\frac{d}{\max(p,\beta_\infty)} - d \leq \tau_p(w) \leq \frac{d}{\max(p,\underline{\beta}_\infty)} - d \quad \text{and} \quad \rho_p(w) = -\frac{d}{\min(p,p_{\max})}. \tag{5.174}$$

Proof of Proposition 5.14. According to Proposition 5.3, $w = w_1 + w_2$ with w_1 a compound Poisson white noise, w_2 a Lévy white noise with finite moments (which can include a Gaussian part). Moreover, due to Proposition 5.5, we have that

$$\beta_\infty = \beta_\infty(w_2) \geq \beta_\infty(w_1) = 0, \text{ and} \tag{5.175}$$

$$p_{\max} = p_{\max}(w_1) \leq p_{\max}(w_2) = \infty. \tag{5.176}$$

Case $\tau < \left(\frac{d}{\max(p,\beta_\infty)} - d\right)$ and $\rho < -\frac{d}{\min(p,p_{\max})}$. Then, $\tau < \left(\frac{d}{p} - d\right)$ and $\rho < -\frac{d}{\min(p,p_{\max}(w_1))}$ so that $w_1 \in B_p^\tau(\mathbb{R}^d;\rho)$ due to Proposition 5.11. Similarly,

$w_2 \in B_p^\tau(\mathbb{R}^d;\rho)$ due to Proposition 5.13. Finally, $w = w_1 + w_2 \in B_p^\tau(\mathbb{R}^d;\rho)$.

Case $\tau > \left(\frac{d}{\max(p,\beta_\infty)} - d\right)$. The arguments of Proposition 5.13 for this case are still valid for w.

Case $\rho > -\frac{d}{\min(p,p_{\max})}$. The case $\rho \geq -d/p$ has been treated in Proposition 5.9. We therefore already know that $w \notin B_p^\tau(\mathbb{R}^d;\rho)$ if $\tau > d/\max(p,\beta_\infty) - d$ or if $\rho \geq -d/p$. The only remaining case is when $p > p_{\max}$, $\tau < d/\max(p,\beta_\infty) - d$, and $-d/p_{\max} < \rho < -d/p$. In this case, $w_2 \in B_p^\tau(\mathbb{R}^d;\rho)$ from Proposition 5.13, while $w_1 \notin B_p^\tau(\mathbb{R}^d;\rho)$ with Proposition 5.11 due to the condition $\rho > -d/p_{\max}$. Finally, $w \notin B_p^\tau(\mathbb{R}^d;\rho)$ a.s. as the sum of an element a.s. in $B_p^\tau(\mathbb{R}^d;\rho)$ and an element a.s. not in $B_p^\tau(\mathbb{R}^d;\rho)$. □

Finally, we can translate our results in terms of the local smoothness $\tau_p(w)$ and the asymptotic growth rate $\rho_p(w)$ of Lévy white noises.

Proof of Theorem 5.1. The values of $\tau_p(w)$ and $\rho_p(w)$ are directly deduced from Propositions 5.8, 5.11, and 5.14. Positive results ($w \in B_p^\tau(\mathbb{R}^d;\rho)$) directly give lower bounds for $\tau_p(w)$ and $\rho_p(w)$, while negative results ($w \notin B_p^\tau(\mathbb{R}^d;\rho)$) provide upper bounds. For the Gaussian and compound Poisson cases, studied separatly, there are no restrictions on $0 < p \leq \infty$. The case of a general non-Gaussian Lévy white noise is deduced from Proposition 5.14. □

5.3.8 Discussion and Examples

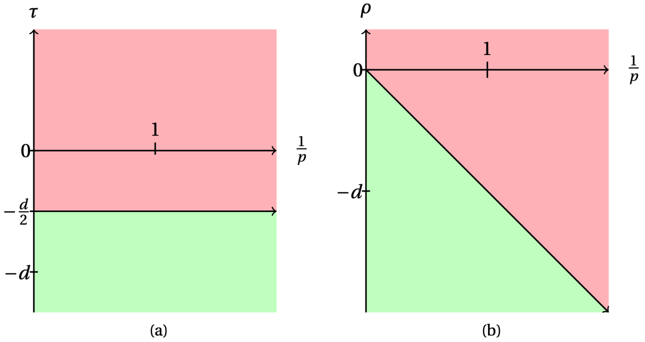
Application to Subfamilies of Lévy White Noises

We apply Theorem 5.1 to deduce the local smoothness and asymptotic growth rate of specific Lévy white noises. We consider Gaussian, symmetric-α-stable [452], symmetric Gamma (including Laplace) [453], compound Poisson, inverse Gaussian [455],

Table 5.3: Lévy White Noises and their Indices

White noise	Parameters	$\widehat{\mathscr{P}}_X(\xi)$	$\beta_\infty = \underline{\beta}_\infty$	$p_{\max}$
Gaussian	$\sigma^2 > 0$	$\mathrm{e}^{-\sigma^2\omega^2/2}$	2	∞
Cauchy [452]	$\gamma > 0$	$\mathrm{e}^{-\gamma\lvert\xi\rvert}$	1	1
SαS [452]	$0 < \alpha < 2$	$\mathrm{e}^{-\lvert\xi\rvert^\alpha}$	α	α
sum of SαS	$0 < \alpha_1, \alpha_2 \leq 2$	$\mathrm{e}^{-\lvert\xi\rvert^{\alpha_1} - \lvert\xi\rvert^{\alpha_2}}$	$\max(\alpha_1, \alpha_2)$	$\min(\alpha_1, \alpha_2)$
Laplace [453]	$\sigma^2 > 0$	$(1 + \sigma^2\xi/2)^{-1}$	0	∞
symmetric Gamma [453]	$\sigma^2, \lambda > 0$	$(1 + \sigma^2\xi/2)^{-\lambda}$	0	∞
compound Poisson with finite moments	$\lambda > 0$ P	$\mathrm{e}^{\lambda(\widehat{P}(\xi) - 1)}$	0	∞
layered stable [454]	$0 < \alpha_1, \alpha_2 < 2$	see (5.177)	α_1	α_2
inverse Gaussian [455]	-	$\mathrm{e}^{1-(1-2\mathrm{j}\xi)^{1/2}}$	$1/2$	∞
with finite moment and Gaussian part	(μ, σ^2, ν)	$\mathrm{e}^{\psi(\xi)}$ with ψ given by (5.5)	2	∞

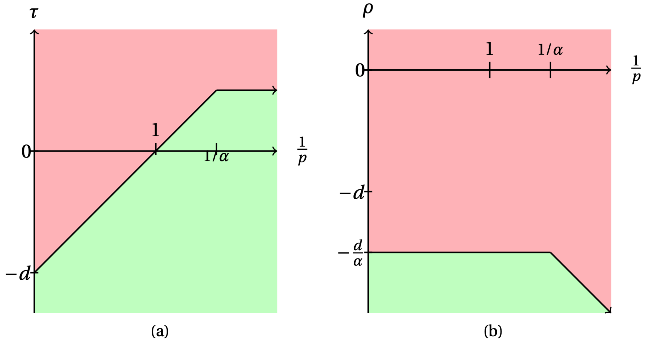
and layered stable white noises [454]. All the underlying laws are known to be infinitely divisible [454, 170]. In Table 5.3, we define the different families in terms of the characteristic function of $X = \langle w, \mathbb{1}_{[0,1]^d} \rangle$ and give adequate references. Most



(a)

(b)

Figure 5.10: Gaussian white noise



(a)

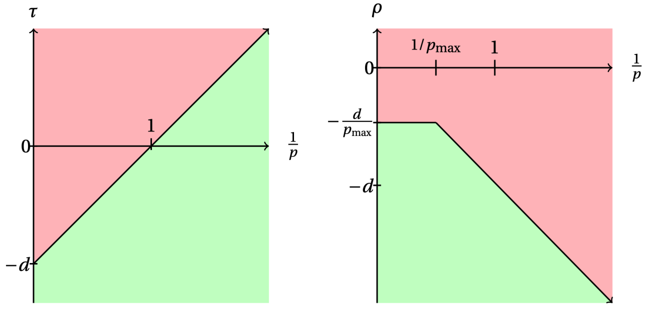
(b)

Figure 5.11: SαS white noise with $\alpha = 2/3$

of these families together with the convention we are following in this work are introduced and detailed in [442, Section 5.1] and [381, Section 2.1.3].

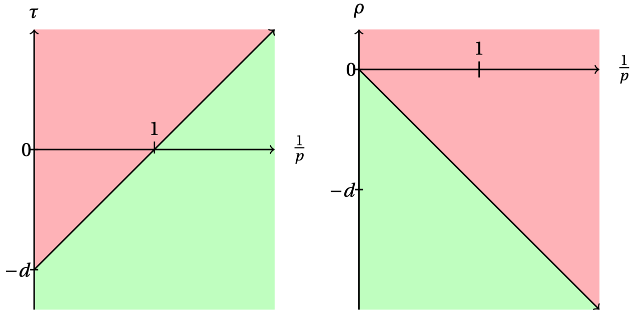
The layered stable white noises have the particularity of describing the complete spectrum of possible couples $(\alpha_1, \alpha_2) = (\beta_\infty, p_{\max}) \in (0, 2)^2$. The characteristic exponent of a layered stable white noise is

$$\Psi_{\alpha_1,\alpha_2}(\xi) = \int_{\mathbb{R}} (\cos(t\xi) - 1) \left(\mathbb{1}_{|t|\leq 1} |t|^{-(\alpha_1+1)} + \mathbb{1}_{|t|>1} |t|^{-(\alpha_2+1)} \right) \mathrm{d}t. \qquad (5.177)$$



(a) (b)

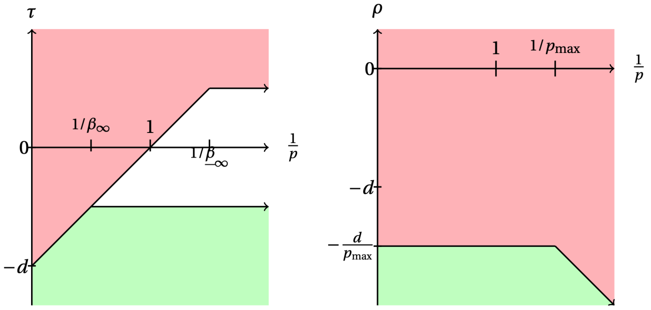
Figure 5.12: compound-Poisson white noise with $p_{\max} = 2$



(a) (b)

Figure 5.13: Symmetric-Gamma white noise

We also provide a visualization of our results in terms of Triebel diagrams. In Figures 5.10 to 5.14, we plot the local smoothness $\frac{1}{p} \mapsto \tau_p(w)$ and asymptotic growth rate $\frac{1}{p} \mapsto \rho_p(w)$ for different Lévy white noises (with the exception of $\tau_p(w)$ which is not fully determined for the general case in Figure 5.14; here, we represent the lower and upper bounds of (5.53)). A given noise is almost surely in a Besov space $B_p^\tau(\mathbb{R}^d;\rho)$ if the points $(1/p,\tau)$ and $(1/p,\rho)$ are in the lower shaded green regions. *A contrario*, the Lévy white noise is almost surely not in $B_p^\tau(\mathbb{R}^d;\rho)$ if



(a) (b)

Figure 5.14: Lévy white noise with $2/3 = \underline{\beta}_\infty < \beta_\infty = 2$ and $p_{\max} = 2/3$

$(1/p, \tau)$ or $(1/p, \rho)$ are in the upper shaded red region. In Figure 5.14, the white region corresponds to the case where we do not know if the Lévy white noise is or is not in the corresponding Besov spaces, a situation that is examplified in Section 5.3.8 and discussed in Section 5.3.9. In this diagrams, we moreover assume that our lower bound (5.54) is valid for any $p > 0$, including no even integers when $p \geq 2$. This conjectural point is discussed in Section 5.3.9.

Lévy White Noises with Distinct Blumenthal-Getoor Indices

In Theorem 5.1, we obtained lower and upper bounds for the local regularity of a Lévy white noise (see (5.53)). This bounds are equal if and only if $\underline{\beta}_\infty = \beta_\infty$. This equality is valid for all the examples presented in Section 5.3.8. It is however possible to construct characteristic exponents with indices that take any values $(\underline{\beta}_\infty, \beta_\infty) \in [0,2]^2$ with the obvious constraint that $\underline{\beta}_\infty \leq \beta_\infty$. This is done in [456, Examples 1.1.14, 1.1.15], where the authors introduce

$$\Psi_{0,\beta_2,M}(\xi) = \sum_{k\geq 1} 2^{\beta_2 M^k - k}(\cos(2^{-M^k}\xi) - 1), \tag{5.178}$$

$$\Psi_{\beta_1,\beta_2,M}(\xi) = \int_{|t|\leq 1} \frac{\cos(t\xi) - 1}{|t|^{\beta_1+1}} \mathrm{d}t + \Psi_{\beta_2,M}(\xi), \tag{5.179}$$

with $0 < \beta_1 \leq \beta_2 < 2$ and $M > 2/(2 - \beta_2)$, and show that $\Psi_{0,\beta_2,M}$ ($\Psi_{\beta_1,\beta_2,M}$, respectively) is a characteristic exponent with Blumenthal-Getoor indices $\underline{\beta}_\infty = 0$ and $\beta_\infty = \beta_2$ ($\underline{\beta}_\infty = \beta_1$ and $\beta_\infty = \beta_2$, respectively).

5.3.9 Summary

We have obtained new results on the localization of Lévy white noises in weighted Besov spaces, summarized in Theorem 5.1. This includes the identification of the local smoothness in many cases (including the examples presented in Section 5.3.8), and lower and upper bounds for the general case. We also identify the asymptotic growth rate of the Lévy white noise in many situations, significantly improving known results.

5.4 Special Case: Compound-Poisson Processes

By invoking the results of Section 5.3, we can characterize the compressibility of sparse stochastic processes [443]. More precisely, it can be shown that the decay rate of the best M-term approximation error of a process s is intimately linked to its local smoothness exponent $\tau_p(s)$. This, in particular, implies that the best M-term approximation error of compound-Poisson processes decays faster than any polynomial [457, Corollary 6.2].

In this work, we refine previous results by presenting a direct method for quantifying the wavelet compressibility of compound Poisson processes. To that end, we expand the given random process over the Haar wavelet basis and we analyse its asymptotic approximation properties. By only considering the nonzero wavelet coefficients up to a given scale, what we call the greedy approximation, we exploit the extreme sparsity of the wavelet expansion that derives from the piecewise-constant nature of compound Poisson processes. More precisely, we provide lower and upper bounds for the mean squared error of greedy approximation of compound Poisson processes. We are then able to deduce that the greedy approximation error has a sub-exponential and super-polynomial asymptotic behavior. Finally, we provide numerical experiments to highlight the remarkable ability of wavelet-based dictionaries in achieving highly compressible approximations of compound Poisson processes.

5.4.1 Context

The class of Lévy processes allows for various compressibility behaviors: the Brownian motion is the less compressible, while the compound Poisson ones are at the other extreme. This compressibility hierarchy has been recently revealed in two different theoretical frameworks.

In the first one, the compressibility is measured via the speed of convergence of the best M-term approximation in wavelet bases. The decay rate of the best M-term error is known to be directly linked to the Besov regularity [458, 459], which has been quantified for a broad class of Lévy processes [424, 425, 429, 430, 431, 325].

Hence, the compressibility of Lévy processes has already been characterized using this approach [460, 443] and synthesized in [381, Chapter 6]. In a nutshell, state-of-the-art results show that the best M-term quadratic approximation error of the Brownian motion behaves asymptotically[12] like $1/M$, while the same quantity decays faster than any polynomial for compound Poisson processes [443, Theorems 4 and 5].

In the second framework, the compressibility of a Lévy process is quantified in the information theoretic sense through the entropy of the underlying Lévy white noise, as in [461, 462]. These two frameworks are complementary and based on totally different tools, but they are consistent and lead to the same compressibility hierarchy.

This work contributes to the analysis of the compressibility of Lévy processes, focusing on the compound Poisson and Gaussian cases. We consider the Haar wavelet approximations of these random processes and quantify the decay rate of their approximation error in the mean squared sense.

More precisely, we focus on quantities such as

$$\mathbb{E}\left[\|s - \mathrm{P}_M\{s\}\|_2^2\right] \tag{5.180}$$

where s is a compound Poisson process or the Brownian motion and $\mathrm{P}_M : L_2([0,1]) \to L_2([0,1])$ is a possibly nonlinear approximation operator based on $M \geq 1$ Haar wavelet coefficients of the input function. We compare various approximation schemes, depending on which wavelet coefficients are chosen. The two best-known schemes are the linear and the best M-term approximation, albeit both suffer from practical limitations. On one hand, the linear scheme does not capture the sparsity that might be inherent in the signal of interest (see Proposition 5.16). On the other hand, in order to *exactly* implement a compression scheme based on the best M-term approximation of the random process, one needs to have access to the full infinite set of wavelet coefficients. Without additional knowledge on the wavelet expansion, the implementation may become cumbersome and not memory efficient, if not impossible. This is why alternative approximation schemes have

[12] More precisely, one can deduced from [443] that the wavelet approximation error of the Brownian motion decays almost surely faster that $1/M^{1-\epsilon}$ and slower than $1/M^{1+\epsilon}$ for any $\epsilon > 0$ when $M \to \infty$.

Table 5.4: The decay rate of the MSE (5.180) of linear, greedy, and best M-term approximation schemes for compound Poisson processes and the Brownian motion.

	Brownian Motion	Compound Poisson
Linear	$\frac{1}{M}$	$\frac{1}{M}$
Greedy	$\frac{1}{M}$	$\boldsymbol{\mathbb{E}_{N\sim\mathrm{Pois}(\lambda)}\left[2^{-\frac{M}{N}}\right]}$
Best	$\frac{1}{M}$	$\ll M^{-k}, \forall k \in \mathbb{N}$

been proposed, most notably the “tree approximation” scheme which has brought significant attention in the literature [463, 464, 465, 466].

In the same spirit, we consider a very simple *greedy approximation scheme*, in which only the first M nonzero wavelet coefficients are preserved. This scheme is well-suited to compound Poisson processes, for which most of the wavelet coefficients are zero due to their piecewise constancy.

Our main result is to provide lower and upper-bounds for the greedy approximation error in the mean-squared sense (Theorem 5.3). It essentially states that the mean-square error of the Haar greedy approximation of the compound Poisson process s behaves roughly as

$$\mathbb{E}\left[\|s - \mathrm{P}_M^{\mathrm{greedy}}\{s\}\|_2^2\right] \approx \mathbb{E}\left[2^{-\frac{M}{N}}\right] \tag{5.181}$$

where N is the (random) number of jumps of s (see (5.231) for the precise meaning of (5.181)). This allows us to deduce that the mean-square error decays faster than any polynomial, and slower than any exponential (Theorem 5.4). We also perform a similar analysis for the linear approximation of the compound Poisson process, as well as for the linear and greedy approximations of the Brownian motion. This highlights the specificity of the compound Poisson processes: the greedy approximation dramatically outperforms the linear scheme for compound Poisson processes, contrary to the Gaussian case. We summarize this situation in Table I, where the main contribution is highlighted in bold.

Finally, we illustrate our theoretical findings with numerical examples in various cases. Specifically, we highlight that the approximation error obtained within our method is close to the best M-term approximation. Moreover, we highlight the role of the wavelet dictionary by comparing the linear and best M-term schemes for compound Poisson processes and the Brownian motion in a Fourier-type dictionary corresponding to the discrete cosine transform (DCT). These empirical observations raise interesting theoretical questions which we briefly expose and can be exploited in future works.

5.4.2 Mathematical Background

In this work, we shall consider compound Poisson processes and white noises that are zero-mean, finite variance (which is equivalent to say that the jumps themselves are zero-mean with finite variance), and whose probability law of jumps $\mathcal{P}$ has a PDF (in particular, it has no atoms, what will be used in our analysis). The prototypical example is a compound Poisson process with Gaussian jump heights. Throughout this work, we shall write the ordered jump positions of compound Poisson processes with the letter x, and the unordered ones with the letter τ.

We consider the family of Haar wavelets whose mother and father wavelets are respectively

$$\psi = \mathbb{1}_{[0,\frac{1}{2}]} - \mathbb{1}_{[\frac{1}{2},1]} \quad \text{and} \quad \phi = \mathbb{1}_{[0,1]}. \tag{5.182}$$

Haar wavelets are known to form an orthonormal basis for $L_2(\mathbb{R})$ [346]. This means that any function $f \in L_2(\mathbb{R})$ admits the unique expansion

$$f(\cdot) = \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(\cdot) + \sum_{k\in\mathbb{Z}} \langle f, \phi_k\rangle \phi_k(\cdot), \tag{5.183}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $L_2(\mathbb{R})$, defined as $\langle \varphi_1, \varphi_2 \rangle = \int_{\mathbb{R}} \varphi_1(x)\varphi_2(x)\mathrm{d}x$.

The simple characteristics and implementation of Haar wavelets make them favorable in practice [467, 468]. They are also compactly supported, which is of great importance in our analysis, due to the *whiteness* property of Lévy white noises (see above). Last but not least, the family consists of piecewise constant functions. Hence, it is natural to represent compound Poisson processes (that are themselves almost surely piecewise constant) in this basis.

In the sequel, we restrict both the random processes and the wavelet transforms to $[0, 1]$ and study the local compressibility of compound Poisson processes over this compact interval.

Due to the support localization of the Haar wavelets, we readily see that the family $\boldsymbol{\Psi} = \{\psi_{j,k}\}_{j\geq 0, 0\leq k\leq 2^j-1} \cup \{\phi\}$ forms an orthonormal basis of $L_2([0, 1])$, hence the Lévy process s can be almost surely written as

$$s = \langle s, \phi\rangle\phi + \sum_{j\geq 0}\sum_{k=0}^{2^j-1} \langle s, \psi_{j,k}\rangle\psi_{j,k}. \tag{5.184}$$

We consider three different approximation schemes for square-integrable functions over $[0, 1]$: the linear, best M-term, and greedy approximations. Our main goal and contribution is to precisely quantify the approximation power of the greedy scheme. Before defining the approximation schemes, let us introduce the natural indexing of wavelets by defining the indexing function $\mathrm{Ind} : \boldsymbol{\Psi} \to \mathbb{N}$ as

$$\mathrm{Ind}(\phi) = 0, \quad \mathrm{Ind}(\psi_{j,k}) = 2^j + k, \tag{5.185}$$

for all $j \geq 0$ and $k = 0, \ldots, 2^j - 1$.

Definition 5.6. *Let $f \in L_2([0, 1])$. We denote by*

- $\mathrm{P}_M^{\mathrm{lin}}(f)$*, the* **linear** *approximation of f, that is obtained by keeping the* **first** *M wavelet coefficients (with respect to the indexing function* Ind*) of f in the expansion* (5.184)*.*
- $\mathrm{P}_M^{\mathrm{best}}(f)$*, the* **best** *$M$-term approximation of f, that is obtained by keeping the M* **largest** *wavelet coefficients of f.*

The first scheme in Definition 5.6 is called linear due to the fact that $\mathrm{P}_M^{\mathrm{lin}}(f)$ depends linearly on f. However, the best M-term approximation is adaptive to the signal and is therefore nonlinear. One can hope that the adaptiveness of the best M-term approximation significantly improve the quality of the approximation when compared with the linear one, what appears to be the case for some classes of functions [458].

As an alternative approach, we consider a compression scheme for compound Poisson processes that can be performed in an online fashion with respect to the stream of the wavelet coefficients. The main idea is to exploit the tremendous sparsity of the expansion of compound Poisson processes over the Haar wavelet basis, that is done by retaining only the nonzero wavelet coefficients and is called the *greedy approximation.*

Definition 5.7. *Let $f \in L_2([0,1])$. We denote by $\mathrm{P}_M^{\mathrm{greedy}}(f)$, the* **greedy** *approximation of f, where only the M* **first nonzero** *wavelet coefficients are preserved (the ordering being understood with respect to the indexing function* Ind *in* (5.185)*).*

As for the best M-term approximation, the greedy approximation of f is nonlinear with respect to f. However, it is greedy in the sense that it can be computed by simply looking at the ordered wavelets coefficients. Hence, it does not necessitate to observe the complete set of wavelet coefficients, contrary to the best M-term approximation. It therefore shares the simplicity of the linear scheme and the adaptiveness of the optimal scheme (the best M-term).

The three approximation schemes introduced in this section clearly satisfy the relations

$$\|f - \mathrm{P}_M^{\mathrm{best}}(f)\|_2 \leq \|f - \mathrm{P}_M^{\mathrm{greedy}}(f)\|_2 \leq \|f - \mathrm{P}_M^{\mathrm{lin}}(f)\|_2 \tag{5.186}$$

for any function $f \in L_2([0,1])$ and any $M \geq 0$.

Let s be a Lévy process. To quantify the performance of an approximation scheme, we consider the mean-squared error (MSE), which we denote by $\mathrm{MSE}_M^{\mathrm{method}}$ for the approximation scheme method $\in \{\mathrm{lin}, \mathrm{greedy}, \mathrm{best}\}$ and is defined as

$$\mathrm{MSE}_M^{\mathrm{method}} = \mathbb{E}\left[\|s - \mathrm{P}_M^{\mathrm{method}}(s)\|_{L_2}^2\right]. \tag{5.187}$$

It is clear from (5.186) that

$$\mathrm{MSE}_M^{\mathrm{best}} \leq \mathrm{MSE}_M^{\mathrm{greedy}} \leq \mathrm{MSE}_M^{\mathrm{lin}}. \tag{5.188}$$

5.4.3 Preliminary Results

In Lemma 5.4, we characterize the law of the minimal distance Δ between two consecutive jumps of a compound Poisson process.

Lemma 5.4. *Consider a compound Poisson process s with parameters $(\mathcal{P},\lambda)$ and a fixed interval $[a,b]$. Denote by N, the number of points in $[a,b]$ and $\mathbf{x}=(x_1,\dots,x_N)$, the ordered set of jumps of s that are in $[a,b]$. With the convention $x_0=a$, we define the random variable Δ as*

$$\Delta=\begin{cases}(b-a), & N=0\\ \min\limits_{1\leq i\leq N}(x_i-x_{i-1}), & N\geq 1.\end{cases} \tag{5.189}$$

Then, almost surely, $\Delta\leq (b-a)/N$, and for any $n\geq 1$ and $\delta\in[0,(b-a)/n]$, we have

$$\mathbb{P}(\Delta\geq\delta|N=n)=\left(1-n\frac{\delta}{b-a}\right)^n. \tag{5.190}$$

Proof. We first remark that the inequality $\Delta\leq(b-a)/N$ is obviously true when $N=0$, since $\Delta=(b-a)$ in this case. As for $N\geq 1$, we have by definition of Δ that $\Delta\leq x_i-x_{i-1}$, for all $i=1,\dots,N$, with the convention that $x_0=a$. By summing up these equality for for all values of i, we obtain that

$$N\Delta\leq x_N-x_0\leq b-a. \tag{5.191}$$

This yields that $\Delta\leq(b-a)/N$.

For the second part, we define the random vector $\mathbf{d}=(d_1,\dots,d_n)\in[0,1]^n$ as

$$d_i=\frac{x_i-x_{i-1}}{b-a},\quad i=1,2,\dots,n, \tag{5.192}$$

By rewriting (5.192) in the vectorial form, we obtain that

$$\mathbf{d}=\mathbf{H}\mathbf{x}-\frac{a}{b-a}\mathbf{e}_1, \tag{5.193}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0) \in \mathbb{R}^n$, $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{H} \in \mathbb{R}^{n \times n}$ is the lower-bidiagonal matrix

$$\mathbf{H} = \frac{1}{b-a} \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}. \tag{5.194}$$

Now, due to (5.19) and the change of variables (5.193), the PDF of $\mathbf{d}$ is

$$p_{\mathbf{d}}(\boldsymbol{v}|N = n) = n! \mathbb{1}_{\boldsymbol{v} \in [0,1]^n, \|\boldsymbol{v}\|_1 \leq 1}, \tag{5.195}$$

where $\|\mathbf{v}\|_1 = |v_1| + \cdots + |v_n| = v_1 + \cdots + v_n$ for $\mathbf{v} \in [0, 1]^n$. In addition, from the definition of Δ, the probability of $\{\Delta \geq x\}$ for any $x \in [0, (b-a)/n]$ can be computed as

$$\begin{aligned} \mathbb{P}(\Delta \geq x | N = n) &= \mathbb{P}\left(\cap_{i=1}^n \{d_i \geq x/(b-a)\} | N = n\right) \\ &= \int_{[\frac{x}{b-a}, 1]^n} n! \mathbb{1}_{\|\boldsymbol{v}\|_1 \leq 1} \mathrm{d}\boldsymbol{v} \\ &= n! \int_{[0, 1 - \frac{x}{b-a}]^n} \mathbb{1}_{\|\boldsymbol{u}\|_1 \leq 1 - n\frac{x}{b-a}} \mathrm{d}\boldsymbol{u}, \end{aligned} \tag{5.196}$$

where the latter is obtained via the change of variable $u_i = v_i - \frac{x}{b-a}$ for $i = 1, \ldots, n$. We remark that if $u_i \geq 0$ for $i = 1, \ldots, n$ and $\|\boldsymbol{u}\|_1 \leq 1 - nx/(b-a)$, then we would have $u_i \leq 1 - nx/(b-a) \leq 1 - x/(b-a)$ for any $i = 1, \ldots, n$. In other words, the upper-limit of the integral in (5.196) is redundant and can be replaced with $+\infty$.

Doing so, we obtain that

$$
\begin{aligned}
\mathbb{P}(\Delta \geq x | N = n) &= n! \int_{[0,+\infty)^n} \mathbb{1}_{\|\boldsymbol{u}\|_1 \leq 1 - n\frac{x}{b-a}} \mathrm{d}\boldsymbol{u} \\
&\overset{(i)}{=} \frac{n!}{2^n} \int_{\mathbb{R}^n} \mathbb{1}_{\|\boldsymbol{u}\|_1 \leq 1 - n\frac{x}{b-a}} \mathrm{d}\boldsymbol{u} \\
&= \frac{n!}{2^n} \mathrm{Leb}\left(\left\{\|\boldsymbol{u}\|_1 \leq \left(1 - n\frac{x}{b-a}\right)\right\}\right) \\
&= \frac{n!}{2^n}\left(1 - n\frac{x}{b-a}\right)^n \mathrm{Leb}\left(\{\|\boldsymbol{u}\|_1 \leq 1\}\right), \qquad (5.197)
\end{aligned}
$$

where (i) is due to the symmetry of the integrand with respect to the sign of **u** and where Leb denotes the Lebesgue measure. Finally, we use a known result stating that the volume of the ℓ_1 unit ball in $\mathbb{R}^n$ is $2^n/n!$ [469]. This yields to

$$
\mathbb{P}(\Delta \geq x | N = n) = \frac{n!}{2^n}\left(1 - n\frac{x}{b-a}\right)^n \frac{2^n}{n!} = \left(1 - n\frac{x}{b-a}\right)^n.
$$

□

The probability law of the Haar wavelet coefficients of s has been characterized in [470], where their characteristic functions have been explicitly computed. Here, we study the law of wavelet coefficients using the properties of the underlying Lévy white noise. In order to achieve this goal, we introduce the auxiliary functions defined for $t \in [0,1]$ as

$$
\tilde{\phi}(t) = (1-t)\mathbb{1}_{[0,1]}(t), \quad \text{and} \qquad (5.198)
$$

$$
\tilde{\psi}_{j,k}(t) = \begin{cases} 2^{j/2}(k/2^j - t), & t \in [\frac{k}{2^j}, \frac{k+1/2}{2^j}) \\ 2^{j/2}(t - (k+1)/2^j), & t \in [\frac{k+1/2}{2^j}, \frac{k+1}{2^j}) \\ 0, & \text{otherwise}, \end{cases} \qquad (5.199)
$$

for any $j \geq 0$ and $k = 0, \dots, 2^j - 1$. We conclude this part with Proposition 5.15, that expresses the Haar wavelet coefficients of s using the underlying Lévy white noise and the auxiliary functions (5.198) and (5.199).

Proposition 5.15. *Let s be a Lévy process. Then, for any $j \geq 0$ and $0 \leq k \leq 2^j - 1$, we have*

$$\langle s, \psi_{j,k} \rangle = \langle w, \tilde{\psi}_{j,k} \rangle, \qquad \langle s, \phi \rangle = \langle w, \tilde{\phi} \rangle, \tag{5.200}$$

where w is the Lévy white noise such that $\mathrm{D}s = w$.

Proof. A simple computation reveals that $-\mathrm{D}\tilde{\psi}_{j,k} = \psi_{j,k}$. Hence, using the known identity $\mathrm{D}^* = -\mathrm{D}$, we have that

$$\langle s, \psi_{j,k} \rangle = \langle s, -\mathrm{D}\tilde{\psi}_{j,k} \rangle = \langle \mathrm{D}s, \tilde{\psi}_{j,k} \rangle = \langle w, \tilde{\psi}_{j,k} \rangle. \tag{5.201}$$

With a similar idea, we remark that $\mathrm{D}\tilde{\phi} = \delta - \phi$. Combining with $\langle s, \delta \rangle = s(0) = 0$, we have that

$$\langle s, \phi \rangle = \langle s, \delta - \mathrm{D}\tilde{\phi} \rangle = s(0) + \langle \mathrm{D}s, \tilde{\phi} \rangle = \langle w, \tilde{\phi} \rangle. \tag{5.202}$$

□

In Proposition 5.16, we determine the $\mathrm{MSE}_M^{\mathrm{lin}}$ of any Lévy process that has finite variance.

Proposition 5.16. *Let s be a Lévy process with finite variance σ_0^2. Then, for every $M \geq 1$, we have*

$$\mathrm{MSE}_M^{\mathrm{lin}} = \frac{\sigma_0^2}{12} \frac{1}{2^J} \left(2 - \frac{m}{2^J} \right), \tag{5.203}$$

where $J = \lfloor \log_2 M \rfloor$ and $m = M - 2^J \in \{0, \dots, 2^J - 1\}$. In particular, for every $M \in 2^{\mathbb{N}}$, we have that

$$\mathrm{MSE}_M^{\mathrm{lin}} = \frac{\sigma_0^2}{6M}. \tag{5.204}$$

Proof. One observes from Definition 5.6 that

$$\mathrm{P}_M^{\mathrm{lin}}(s) = \langle s, \phi \rangle \phi + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \langle s, \psi_{j,k} \rangle \psi_{j,k} + \sum_{k=0}^{m-1} \langle s, \psi_{J,k} \rangle \psi_{J,k}. \tag{5.205}$$

This together with (5.184) yields that

$$s - \mathrm{P}_M^{\mathrm{lin}}(s) = \sum_{j \geq J+1} \sum_{k=0}^{2^j-1} \langle s, \psi_{j,k} \rangle \psi_{j,k} + \sum_{k=m}^{2^J-1} \langle s, \psi_{J,k} \rangle \psi_{J,k}. \tag{5.206}$$

Haar wavelets that are supported in $[0, 1]$, form an orthonormal basis for $L_2([0, 1])$. Using this, we express the approximation error based on the wavelet coefficients, as

$$\|s - \mathrm{P}_M^{\mathrm{lin}}(s)\|_{L_2}^2 = \sum_{j \geq J+1} \sum_{k=0}^{2^j-1} |\langle s, \psi_{j,k} \rangle|^2 + \sum_{k=m}^{2^J-1} |\langle s, \psi_{J,k} \rangle|^2. \tag{5.207}$$

By taking expectation over both sides and by using Proposition 5.15, we have that

$$\begin{aligned}
\mathbb{E}[\|s - \mathrm{P}_M^{\mathrm{lin}}(s)\|_{L_2}^2] &= \sum_{j \geq J+1} \sum_{k=0}^{2^j-1} \mathbb{E}[|\langle s, \psi_{j,k} \rangle|^2] + \sum_{k=m}^{2^J-1} \mathbb{E}[|\langle s, \psi_{J,k} \rangle|^2] \\
&= \sum_{j \geq J+1} \sum_{k=0}^{2^j-1} \mathbb{E}[|\langle w, \tilde{\psi}_{j,k} \rangle|^2] + \sum_{k=m}^{2^J-1} \mathbb{E}[|\langle w, \tilde{\psi}_{J,k} \rangle|^2] \qquad (5.208) \\
&= \sum_{j \geq J+1} \sum_{k=0}^{2^j-1} \sigma_0^2 \|\tilde{\psi}_{j,k}\|_{L_2}^2 + \sum_{k=m}^{2^J-1} \sigma_0^2 \|\tilde{\psi}_{J,k}\|_{L_2}^2. \qquad (5.209)
\end{aligned}$$

Finally, we replace $\|\tilde{\psi}_{j,k}\|_{L_2}^2 = \frac{2^{-2j}}{12}$ (obtained via a direct computation; see Figure 5.15 for visualisation) for all $j \geq 0$ and $k = 0, \dots, 2^j - 1$ in the summation above to deduce that

$$\begin{aligned}
\mathbb{E}\left[\|s - \mathrm{P}_M^{\mathrm{lin}}(s)\|_{L_2}^2\right] &= \frac{\sigma_0^2}{12} \left(\sum_{j \geq J+1} \sum_{k=0}^{2^j-1} 2^{-2j} + \sum_{k=m}^{2^J-1} 2^{-2J} \right) \\
&= \frac{\sigma_0^2}{12M} \left(\frac{1}{2^J} + \frac{2^J - m}{2^{2J}} \right) \\
&= \frac{\sigma_0^2}{12M} \frac{1}{2^J} \left(2 - \frac{m}{2^J} \right). \qquad (5.210)
\end{aligned}$$

□

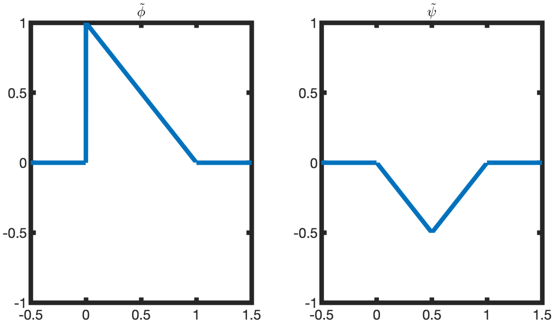


Figure 5.15: Auxiliary functions $\tilde{\phi}$ and $\tilde{\psi}_{j,k}$ for $j = 0, 1, 2$ and $k = 0, \dots, 2^j - 1$.

Proposition 5.16 shows that the linear approximations of Lévy processes with finite variance share the same mean-square error. Let us also remark that if s is a Brownian motion, then the random variables $X_{j,k} = \langle s, \psi_{j,k} \rangle = \langle w, \tilde{\psi}_{j,k} \rangle$ are all Gaussian. Hence,

$$\mathbb{P}(\exists j, k : X_{j,k} = 0) \leq \sum_{j \geq 0} \sum_{k=0}^{2^j - 1} \mathbb{P}(X_{j,k} = 0) = 0, \tag{5.211}$$

and all the countably many wavelet coefficients are almost surely nonzero and hence, the linear and greedy schemes coincide, as stated in Corollary 5.7.

Corollary 5.7. *Let s be a Brownian motion. Then, for any $M \geq 0$, we have the almost sure relation*

$$\mathrm{MSE}_M^{\mathrm{greedy}} = \mathrm{MSE}_M^{\mathrm{lin}}. \tag{5.212}$$

When the wavelet coefficients are sparse (*i.e.* when at each scale, only a few of them are nonzero), the linear and greedy approximation schemes are no longer identical. In Proposition 5.17, we study the sparsity of the wavelet coefficients of compound Poisson processes. Precisely, we first characterize when a specific wavelet coefficient vanishes, depending on the presence of jumps. Using this primary result, we provide upper and lower bounds for the minimal (random) scale at which at least M wavelet coefficients are nonzero.

Proposition 5.17. *Let s be a compound Poisson process whose law of jumps admits a PDF with zero-mean and finite variance.*

1. *For all $j \geq 0$ and $k = 0, \dots, 2^j - 1$, denote $K_{j,k}$ as the random number of jumps of s in the support of $\psi_{j,k}$. Then, we almost surely have*

$$\langle s, \psi_{j,k} \rangle = 0 \quad \Leftrightarrow \quad K_{j,k} = 0. \tag{5.213}$$

In other words, the symmetric difference between the events $\langle s, \psi_{j,k} \rangle = 0$ and $K_{j,k} = 0$ has probability zero.

2. *Consider the wavelet expansion* (5.184) *of s and denote by N_J, the random number of nonzero wavelet coefficients with scale no larger than J. Furthermore, condition to $\{N \geq 1\}$, let J_M be the smallest random value of J such that $N_J \geq M$; that is, J_M is characterized by $N_{J_M - 1} < M \leq N_{J_M}$. Then, we have*

$$\left\lceil \frac{M-2}{N} \right\rceil \leq J_M \leq \left\lfloor \frac{M-1}{N} + \log \Delta^{-1} \right\rfloor, \tag{5.214}$$

where the random variable Δ is defined in (5.189).

Proof. **Item 1)** Assume that $K_{j,k} = 0$. This means that s is constant over the support of $\psi_{j,k}$, taking the fixed (random) value s_0. By recalling that

$$\int_{\mathbb{R}} \psi_{j,k}(x) \mathrm{d}x = 0, \tag{5.215}$$

for all $j \geq 0$ and $k = 0, \dots, 2^j - 1$, we deduce that the corresponding wavelet coefficient is $\langle s, \psi_{j,k} \rangle = s_0 \int_{\mathbb{R}} \psi_{j,k}(x) \mathrm{d}x = 0$.

For the converse, we show that, condition to the event $\{K_{j,k} = K\}$ for an arbitrary (but fixed) integer $K \geq 1$, we have that

$$\mathbb{P}\left(\langle s, \psi_{j,k} \rangle = 0 | K_{j,k} = K \right) = 0. \tag{5.216}$$

Consider the jumps that are inside the support of $\psi_{j,k}$ and denote their (unordered) locations and heights by $\{\tilde{\tau}_1, \dots, \tilde{\tau}_K\}$ and $\{\tilde{a}_1, \dots, \tilde{a}_K\}$, respectively. Due to (5.200), we have that

$$\langle s, \psi_{j,k} \rangle = \langle w, \tilde{\psi}_{j,k} \rangle = \sum_{i=1}^{K} \tilde{a}_i \tilde{\psi}_{j,k}(\tilde{\tau}_i). \tag{5.217}$$

We recall that the jump locations $\tilde{\tau}_i$ are i.i.d. with a uniform law. Moreover, the jump heights $\tilde{a}_i$ are independent of $\tilde{\tau}_i$s and are themselves i.i.d. copies of a random variable that admits PDF. This implies that the random variables $Z_i = \tilde{a}_i \tilde{\psi}_{j,k}(\tilde{\tau}_i)$ for $i = 1, \ldots, K$ are also i.i.d. and their law has a PDF too, which we denote by p_Z. Finally, the random variable $\langle s, \psi_{j,k} \rangle$ also has PDF (that is the K times convolution of p_Z with itself) and thus, is nonzero with probability one (no atoms).

Item 2) Recall that N is the total number of jumps of s over $[0, 1]$. Due to (5.213) and the fact that the wavelets $\psi_{j,k}$ for $k = 0, \ldots, 2^j - 1$ have disjoint support, at each scale $j \geq 1$, at most N wavelet coefficients are nonzero. On the other hand, the support of any wavelet function of scale j is of size 2^{-j}. Hence, due to the definition of Δ, the number of jumps in the support of $\psi_{j,k}$ is either one or is upper-bounded by the length of the interval divided by the minimum distance ($= 2^{-j}\Delta^{-1}$). In other words, the support of each wavelet of scale j contains at most $\max(1, 2^{-j}\Delta^{-1})$ jumps.

Denote by n_j, the number of nonzero wavelet coefficients in the jth scale. Using the previous observation, we deduce for all $j \geq 1$ that

$$\frac{N}{\max(1, 2^{-j}\Delta^{-1})} = N \min(1, 2^j \Delta) \leq n_j \leq N. \tag{5.218}$$

As for $j = 0$ (mother and father wavelets), we deduce similar to Item 1) that condition to $N \geq 1$, we have $n_0 = 2$.

By defining $J_{\text{lim}} = \lfloor \log_2(\Delta^{-1}) \rfloor$, one readily verifies that for $j \leq J_{\text{lim}}$, we have $\min(1, 2^j \Delta) = 2^j \Delta$. By contrast, $\min(1, 2^j \Delta) = 1$ for $j \geq J_{\text{lim}}$. Using these simple observations and by summing up lower-bounds of (5.218) for $j = 1, \ldots, J$ (together with $n_0 = 2$), we obtain, since $\sum_{j=0}^{J} n_j = N_J$, that

$$2 + N\Delta(2^{J+1} - 2) \leq N_J, \quad \forall J \leq J_{\text{lim}}, \tag{5.219}$$

$$2 + N\Delta(2^{J_{\text{lim}}+1} - 2) + (J - J_{\text{lim}})N \leq N_J, \quad \forall J \geq J_{\text{lim}}. \tag{5.220}$$

To simplify the first lower-bound, we use the inequality $2^x \geq x$ for $x = J+1+\log_2 \Delta$, which results to

$$2 - 2N\Delta + N(J + 1 + \log_2 \Delta) \leq N_J, \quad \forall J \leq J_{\mathrm{lim}}. \tag{5.221}$$

As for the the second lower-bound, we use

$$J_{\mathrm{lim}} \leq \log_2(\Delta^{-1}) \leq J_{\mathrm{lim}} + 1 \tag{5.222}$$

to obtain that

$$2 + N\Delta(\Delta^{-1} - 2) + (J + \log_2 \Delta)N \leq N_J, \quad \forall J \geq J_{\mathrm{lim}}. \tag{5.223}$$

It is now readily to verify that the two lower-bounds in (5.221) and (5.223) are indeed equal and hence, we have that

$$2 + N(J + 1 - 2\Delta + \log_2 \Delta) \leq N_J, \quad \forall J \geq 0. \tag{5.224}$$

Finally, using $\Delta N \leq 1$, we conclude that

$$N(J + 1 + \log_2 \Delta) \leq N_J, \quad \forall J \geq 0. \tag{5.225}$$

We follow the same principle to obtain an upper-bound for N_J as well. By summing up upper-bounds of (5.218) for $j = 1, \dots, J$, together with $n_0 = 2$, we obtain that

$$N_J \leq 2 + NJ, \quad \forall J \geq 0. \tag{5.226}$$

Now, by the definition of J_M, we know that $N_{J_M} \geq M$. Combining it with (5.226) applied to $J = J_M$ yields that

$$M \leq N_{J_M} \leq NJ_M + 2, \tag{5.227}$$

which implies the lower-bound

$$J_M \geq \frac{M-2}{N}. \tag{5.228}$$

Similarly, from the definition of J_M, we have $N_{J_M - 1} \leq M - 1$. This together with (5.225) applied to $J = J_M - 1$ gives

$$N(J_M + \log_2 \Delta) \leq N_{J_M - 1} \leq M - 1, \tag{5.229}$$

from which we deduce the upper-bound

$$J_M \leq \frac{M-1}{N} + \log_2 \Delta^{-1}. \tag{5.230}$$

We complete the proof of (5.214) by combining (5.228) and (5.230), knowing that $J_M \in \mathbb{N}$. □

5.4.4 Main Result

We now present our main result on characterizing the asymptotic behavior of the greedy approximation of compound Poisson processes.

Theorem 5.3. *Let s be a compound Poisson process with Poisson parameter $\lambda > 0$ whose law of jumps admits a PDF with zero-mean and finite variance. Then for every $M \in \mathbb{N}$, we have that*

$$C_1 M^{-1} \mathbb{E}[2^{-\frac{M}{N}}] \leq \mathrm{MSE}_M^{\mathrm{greedy}} \leq C_2 M \mathbb{E}[2^{-\frac{M}{N}}], \tag{5.231}$$

where N is a Poisson random variable with parameter λ, and $C_1, C_2 > 0$ are some constants.

Proof of Theorem 5.3. Let $\sigma_0^2 < +\infty$ be the variance of the process s. We divide the proof and show each side of the inequality (5.231) separately.

Upper-bound: First, we show that for any $n \geq 1$, we have

$$\mathbb{E}\left[\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2 | N = n\right] \leq \frac{\sigma_0^2 n}{6\lambda} 2^{-\frac{M-2}{n}}. \tag{5.232}$$

Let us then work conditionally to $N = n$. From Proposition 5.17, we have (condition to $N = n$) that $J_M \geq \lceil \frac{M-2}{n} \rceil$. This implies that

$$\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2 \leq \|s - \mathrm{P}^{\mathrm{lin}}_{2^{\lceil \frac{M-2}{n} \rceil}}(s)\|_2. \tag{5.233}$$

Taking expectation from both sides yields

$$\mathbb{E}\left[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2|N=n\right] \leq \mathbb{E}\left[\|s-\mathrm{P}^{\mathrm{lin}}_{2^{\lceil\frac{M-2}{n}\rceil}}(s)\|_2^2|N=n\right]$$
$$= \sum_{j\geq\lceil\frac{M-2}{n}\rceil}\sum_{k=0}^{2^j-1}\mathbb{E}[|\langle s,\psi_{j,k}\rangle|^2|N=n], \qquad (5.234)$$

On the other hand, condition to $N=n$, we have the equality in law

$$w=\mathrm{D}s=\sum_{i=1}^{n}a_i\delta(\cdot-\tau_i), \qquad (5.235)$$

where $\{\tau_i\}_{i=1}^n$ is the sequence of unordered jumps of s in $[0,1]$ and $\{a_i\}_{i=1}^n$ is the sequence of corresponding heights. Therefore, we have that

$$\langle s,\psi_{j,k}\rangle=\langle w,\tilde{\psi}_{j,k}\rangle=\sum_{i=1}^{n}Z_i, \qquad (5.236)$$

where the random variables $Z_i=a_i\tilde{\psi}_{j,k}(\tau_i)$ are i.i.d. copies of a zero-mean random variable. We recall that the law of jumps a_i has zero-mean and variance σ_0^2/λ. Hence, the second-order moment of Z_i can be computed as

$$\mathbb{E}[Z_i^2|N=n]=\mathbb{E}[a_i^2\tilde{\psi}_{j,k}(\tau_i)^2|N=n]\overset{(i)}{=}\mathbb{E}[a_i^2|N=n]\mathbb{E}[\tilde{\psi}_{j,k}(\tau_i)^2|N=n]$$
$$=\frac{\sigma_0^2}{\lambda}\int_{\mathbb{R}}\tilde{\psi}_{j,k}(x)^2p_{\tau_i|N=n}(x)\mathrm{d}x\overset{(ii)}{=}\frac{\sigma_0^2}{\lambda}\int_0^1\tilde{\psi}_{j,k}(x)^2\mathrm{d}x=\frac{\sigma_0^2}{\lambda}\|\tilde{\psi}_{j,k}\|_2^2=\frac{\sigma_0^2\times2^{-2j}}{12\lambda}, \qquad (5.237)$$

where we used the independence of a_i and τ_i in (i) and the uniform law of τ_i in (ii) and finally, we replaced $\|\tilde{\psi}_{j,k}\|_{L_2}^2=\frac{2^{-2j}}{12}$ in the last equality. Now, due to the independence of the Z_i, we deduce that

$$\mathbb{E}[\langle s,\psi_{j,k}\rangle^2|N=n]=\sum_{i=1}^{n}\mathbb{E}[Z_i^2|N=n]=n\frac{\sigma_0^2\times2^{-2j}}{12\lambda}. \qquad (5.238)$$

By substituting (5.238) in (5.234), we obtain that

$$\mathbb{E}\left[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2|N=n\right] \le \sum_{j\ge\lceil\frac{M-2}{n}\rceil}\sum_{k=0}^{2^j-1} n\frac{\sigma_0^2 2^{-2j}}{12\lambda} = \frac{\sigma_0^2 n}{12\lambda}\sum_{j\ge\lceil\frac{M-2}{n}\rceil} 2^{-j} = \frac{\sigma_0^2 n}{6\lambda}2^{-\lceil\frac{M-2}{n}\rceil} \le \tag{5.239}$$

By taking the expectation, we obtain that

$$\begin{aligned}\mathbb{E}\left[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2\right] &\le \sum_{n=1}^{\infty}\frac{\sigma_0^2 n}{6\lambda}2^{-\frac{M-2}{n}}\mathbb{P}(N=n)\\ &= \sum_{n=1}^{M}\frac{\sigma_0^2 n}{6\lambda}2^{-\frac{M-2}{n}}\mathbb{P}(N=n) + \sum_{n=M+1}^{\infty}\frac{\sigma_0^2 n}{6\lambda}2^{-\frac{M-2}{n}}\mathbb{P}(N=n)\\ &\le \frac{2\sigma_0^2}{3\lambda}M\mathbb{E}[2^{-\frac{M}{N}}] + \frac{2\sigma_0^2}{3\lambda}\sum_{n=M+1}^{\infty}n\mathbb{P}(N=n),\end{aligned} \tag{5.240}$$

where in the last inequality, we have used $2^{-\frac{M}{n}} \le 1$ and $2^{\frac{2}{n}} \le 4$ for all values of $M, n \ge 1$. Now, by invoking the relation $n\mathbb{P}(N=n) = n\mathrm{e}^{-\lambda}\lambda^n/n! = \lambda\mathbb{P}(N=n-1)$ for any $n \ge 1$, we deduce that

$$\sum_{n=M+1}^{\infty}n\mathbb{P}(N=n) = \sum_{n=M+1}^{\infty}\mathbb{P}(N=n-1) = \lambda\mathbb{P}(N\ge M). \tag{5.241}$$

On one hand, from the Chernov bound we have

$$\mathbb{P}(N\ge M) \le \mathbb{E}[\mathrm{e}^{tN}]\mathrm{e}^{-tM} = \mathrm{e}^{\lambda(\mathrm{e}^t-1)}\mathrm{e}^{-tM}, \quad \forall t>0, \tag{5.242}$$

where we used

$$\mathbb{E}[\mathrm{e}^{tN}] = \sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\mathrm{e}^{tn}\lambda^n}{n!} = \sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{(\lambda\mathrm{e}^t)^n}{n!} = \mathrm{e}^{\lambda(e^t-1)}. \tag{5.243}$$

Using (5.242) with $t = \mathrm{ln}2$ (such that $\mathrm{e}^t = 2$) yields

$$\sum_{n=M+1}^{\infty}n\mathbb{P}(N=n) \le \mathbb{E}[\mathrm{e}^{tN}]\mathrm{e}^{-tM} \le \lambda\mathrm{e}^{\lambda}2^{-M}. \tag{5.244}$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2\right] &\leq \frac{2\sigma_0^2}{3\lambda}M\mathbb{E}[2^{-\frac{M}{N}}]+\frac{2\sigma_0^2}{3}\mathrm{e}^{\lambda}2^{-M}\\ &= M\mathbb{E}[2^{-\frac{M}{N}}]\frac{2\sigma_0^2}{3\lambda}\left(1+\frac{\lambda\mathrm{e}^{\lambda}2^{-M}}{M\mathbb{E}[2^{-\frac{M}{N}}]}\right)\\ &\leq M\mathbb{E}[2^{-\frac{M}{N}}]\frac{2\sigma_0^2}{3\lambda}\left(1+\frac{\lambda\mathrm{e}^{\lambda}2^{-M}}{2^{-M}\mathbb{P}(N=1)}\right)\\ &= M\mathbb{E}[2^{-\frac{M}{N}}]\frac{2\sigma_0^2}{3\lambda}(1+\mathrm{e}^{2\lambda}), \end{aligned} \tag{5.245}$$

which is the announced upper-bound with the constant $C_2=\frac{2\sigma_0^2}{3\lambda}(1+\mathrm{e}^{2\lambda})$.

Lower-bound: Similar to the upper-bound, we show that for any $n\geq 1$, we have the inequality

$$\mathbb{E}[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2|N=n]\geq\frac{\sigma_0^2}{48\mathrm{e}\lambda}n^{-1}2^{-\frac{M-1}{n}}, \tag{5.246}$$

which immediately implies the announced lower-bound.

We treat the case $N=1$ separately. Condition to $N=1$, both wavelet coefficients of order zero (associated to mother and father wavelets) are nonzero. Moreover, for any $j\geq 1$, there is exactly one wavelet coefficient of scale j that is nonzero. This implies that $J_M=M-2$ and in addition, we have that

$$\mathbb{E}[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2|N=1]=\mathbb{E}[\|s-\mathrm{P}_{2^{M-1}}^{\mathrm{lin}}(s)\|_2^2|N=1]. \tag{5.247}$$

Similar to the proof of Proposition 5.16 and together with (5.238), we deduce that

$$\begin{aligned}\mathbb{E}[\|s-\mathrm{P}_{2^{M-1}}^{\mathrm{lin}}(s)\|_2^2|N=1] &= \sum_{j\geq(M-1)}\sum_{k=0}^{2^J-1}\mathbb{E}[\langle s,\psi_{j,k}\rangle^2|N=1]=\sum_{j\geq(M-1)}2^j\frac{\sigma_0^2\times 2^{-2j}}{12\lambda}\\ &= \frac{\sigma_0^2}{6\lambda}2^{-(M-1)}\geq\frac{\sigma_0^2}{48\lambda\mathrm{e}}2^{-(M-1)}, \end{aligned} \tag{5.248}$$

which together with (5.247) proves (5.246) in this case.

Consider an arbitrary integer $n \geq 2$ and let us work conditionally to $N = n$. From the definition of J_M, we almost surely have that

$$\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2 \geq \|s - \mathrm{P}_{2^{J_M+1}}^{\mathrm{lin}}(s)\|_2. \tag{5.249}$$

This together with the right inequality of (5.214) implies almost surely that

$$\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2 \geq \|s - \mathrm{P}_{2^{\lfloor \frac{M-1}{n} + \log_2 \Delta^{-1} \rfloor + 1}}^{\mathrm{lin}}(s)\|_2. \tag{5.250}$$

By defining $\delta = (2n^2 - 2n + 2)^{-1} > 0$ (the precise value will be used later) and $\overline{J} = \lfloor \frac{M-1}{n} + \log_2(\delta^{-1}) \rfloor + 1$, we observe that

$$\begin{aligned}
\mathbb{E}\left[\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2 | N = n\right] &\geq \mathbb{E}\left[\|s - \mathrm{P}_{2^{\lfloor \frac{M-1}{n} + \log_2 \Delta^{-1} \rfloor + 1}}^{\mathrm{lin}}(s)\|_2^2 | N = n\right] \\
&\geq \mathbb{E}\left[\mathbb{1}_{\Delta \geq \delta}\|s - \mathrm{P}_{2^{\lfloor \frac{M-1}{n} + \log_2 \Delta^{-1} \rfloor + 1}}^{\mathrm{lin}}(s)\|_2^2 | N = n\right] \\
&\geq \mathbb{E}\left[\mathbb{1}_{\Delta \geq \delta}\|s - \mathrm{P}_{2^{\overline{J}}}^{\mathrm{lin}}(s)\|_2^2 | N = n\right] \\
&= \mathbb{E}[\mathbb{1}_{\Delta \geq \delta} \sum_{j \geq \overline{J}} \sum_{k=0}^{2^j - 1} \langle s, \psi_{j,k} \rangle^2 | N = n] \\
&= \sum_{j \geq \overline{J}} \sum_{k=0}^{2^j - 1} \mathbb{E}[\mathbb{1}_{\Delta \geq \delta} \langle s, \psi_{j,k} \rangle^2 | N = n].
\end{aligned} \tag{5.251}$$

Similar to the upper-bound, we consider the jumps of s in $[0, 1]$ and we denote their (unordered) locations and heights by $\tau_1, \dots, \tau_n$ and $a_1, \dots, a_n$, respectively. With regard to the convention $\tau_0 = 0$, we consider the random variable $\tilde{\Delta} = \min_{0 \leq i < j < n-1} |\tau_i - \tau_j|$ and consequently, the event

$$E = \{\tilde{\Delta} \geq \delta\} \cap \{0 \leq \tau_1, \dots, \tau_{n-1} \leq 1/2 - \delta\}. \tag{5.252}$$

We observe that condition to $E \cap \{N = n\}$, we have that

$$\Delta = \min\left(\tilde{\Delta}, \min_{1 \leq i \leq n-1} (\tau_n - \tau_i)\right) \geq \min(\tilde{\Delta}, \delta) \geq \delta. \tag{5.253}$$

This implies that condition to $N = n$, we have

$$\mathbb{1}_E \mathbb{1}_{[1/2,1]}(\tau_n) \leq \mathbb{1}_{\Delta \geq \delta}. \tag{5.254}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{\Delta\geq\delta}\langle s, \psi_{j,k}\rangle^2 | N = n\right] &= \mathbb{E}\left[\mathbb{1}_{\Delta\geq\delta}\left(\sum_{i=1}^n a_i \tilde{\psi}_{j,k}(\tau_i)\right)^2 | N = n\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^n a_i \mathbb{1}_{\Delta\geq\delta} \tilde{\psi}_{j,k}(\tau_i)\right)^2 | N = n\right] \\
&\overset{(i)}{=} \sum_{i=1}^n \mathbb{E}\left[\left(a_i \mathbb{1}_{\Delta\geq\delta}\tilde{\psi}_{j,k}(\tau_i)\right)^2 | N = n\right] \\
&\overset{(ii)}{=} \sum_{i=1}^n \mathbb{E}[a_i^2] \mathbb{E}\left[\mathbb{1}_{\Delta\geq\delta}\tilde{\psi}^2_{j,k}(\tau_i) | N = n\right],
\end{aligned} \tag{5.255}$$

where we used the independence (condition to $N = n$) of jumps τ_i and heights a_i of s in (i) and we used the independence of a_i from N and Δ as well the fact that the law of a_i has zero mean in (ii). By substituting $\mathbb{E}[a_i^2] = \frac{\sigma_0^2}{\lambda}$ and invoking (5.254), we obtain

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{\Delta\geq\delta}|\langle s, \psi_{j,k}\rangle|^2 | N = n\right] &= n\frac{\sigma_0^2}{\lambda}\mathbb{E}[\mathbb{1}_{\Delta\geq\delta}\tilde{\psi}^2_{j,k}(\tau_n)|N = n] \\
&\geq n\frac{\sigma_0^2}{\lambda}\mathbb{E}[\mathbb{1}_E \mathbb{1}_{\tau_n\in[1/2,1]}\tilde{\psi}^2_{j,k}(\tau_n)|N = n] \\
&= n\frac{\sigma_0^2}{\lambda}\mathbb{P}[E|N = n]\mathbb{E}[\mathbb{1}_{x_n\in[1/2,1]}\tilde{\psi}^2_{j,k}(x_n)|N = n],
\end{aligned} \tag{5.256}$$

where the latter is deduced from the independence of E and $\{1/2 \leq \tau_n \leq 1\}$ (condition to $N = n$). By using Lemma 5.4 with $a = 0$ and $b = 1/2 - \delta$ (we remind that $\delta = (2n^2 - 2n + 2)^{-1}$), we can compute the conditional probability of the event

E as

$$\mathbb{P}(E|N=n) = \left(1-(n-1)\frac{\delta}{1/2-\delta}\right)^{(n-1)}$$
$$= \left(1-(n-1)\frac{(2n^2-2n+2)^{-1}}{1/2-(2n^2-2n+2)^{-1}}\right)^{(n-1)} = \left(1-n^{-1}\right)^{(n-1)}. \tag{5.257}$$

Now, using Lemma 5.4 and the above computation, we have that

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{\Delta\geq\delta}|\langle s,\psi_{j,k}\rangle|^2|N=n\right] &\geq n\frac{\sigma_0^2}{\lambda}\mathbb{P}(E|N=n)\int_{\frac{1}{2}}^{1}\tilde{\psi}_{j,k}^2(x)\mathrm{d}x\\ &= \frac{\sigma_0^2 n}{\lambda}(1-n^{-1})^{(n-1)}\|\tilde{\psi}_{j,k}\mathbb{1}_{[1/2,1]}\|_2^2\\ &\overset{(i)}{=} \frac{\sigma_0^2 n}{\lambda}(1-n^{-1})^{(n-1)}\mathbb{1}_{k\geq 2^{j-1}}\|\tilde{\psi}_{j,k}\|_2^2\\ &\overset{(ii)}{\geq} \frac{\sigma_0^2 n}{\lambda}\mathrm{e}^{-1}\mathbb{1}_{k\geq 2^{j-1}}\|\tilde{\psi}_{j,k}\|_2^2,\end{aligned} \tag{5.258}$$

where (i) simply exploits that $\tilde{\psi}_{j,k}\mathbb{1}_{[1/2,1]} = 0$ for $k \leq 2^{j-1}-1$ together with $\tilde{\psi}_{j,k}\mathbb{1}_{[1/2,1]} = \tilde{\psi}_{j,k}$ for $k \geq 2^{j-1}$ and (ii) uses $(1-n^{-1})^{(n-1)} \geq \mathrm{e}^{-1}$. Going back to (5.251), we obtain for any $n \geq 2$ that

$$\begin{aligned}\mathbb{E}[\|s-\mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2|N=n] &\geq \sum_{j\geq\overline{J}}\sum_{k=0}^{2^j-1}\frac{\sigma_0^2 n}{\lambda}\mathrm{e}^{-1}\|\tilde{\psi}_{j,k}\|_2^2\mathbb{1}_{k\geq 2^{j-1}} = \sum_{j\geq\overline{J}}\frac{\sigma_0^2 n}{\lambda}e^{-1}\|\tilde{\psi}_{j,k}\|_2^2 2^{j-1} \overset{(i)}{=}\\ &\overset{(ii)}{\geq} \frac{\sigma_0^2}{24\mathrm{e}\lambda}n2^{-\frac{M-1}{n}}\delta \overset{(iii)}{=} \frac{\sigma_0^2}{48\mathrm{e}\lambda}\frac{n}{n^2-n+1}2^{-\frac{M-1}{n}}\\ &\overset{(iv)}{\geq} \frac{\sigma_0^2}{48\mathrm{e}\lambda}n^{-1}2^{-\frac{M-1}{n}},\end{aligned} \tag{5.259}$$

where (i) uses $\|\tilde{\psi}_{j,k}\|_{L_2}^2 = 2^{-2j}/12$, (ii) simply follows from $\lfloor\frac{M-1}{n}+\log_2\delta^{-1}\rfloor \leq \frac{M-1}{n}+\log_2\delta^{-1}$, (iii) uses the value of $\delta = (2n^2+2n-2)^{-1}$, and (iv) that $\frac{n}{n^2-n+1} \geq \frac{1}{n}$, due to $n^2 - n + 1 \leq n^2$ for any $n \geq 1$. Finally, we take the overall expectation to deduce that

$$\begin{aligned}\mathbb{E}[\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2] &\geq \sum_{n=1}^{\infty} \frac{\sigma_0^2}{48\mathrm{e}\lambda} n^{-1} 2^{-\frac{M-1}{n}} \mathbb{P}(N = n) \\ &\geq \frac{\sigma_0^2}{48\mathrm{e}\lambda} \sum_{n=1}^{M} n^{-1} 2^{-\frac{M}{n}} \mathbb{P}(N = n) \\ &\geq \frac{\sigma_0^2}{48\mathrm{e}\lambda} M^{-1} \sum_{n=1}^{M} 2^{-\frac{M}{n}} \mathbb{P}(N = n). \end{aligned} \tag{5.260}$$

We note that

$$\sum_{n=1}^{M} 2^{-\frac{M}{n}} \mathbb{P}(N = n) \geq 2^{-M} \mathbb{P}(N = 1) = \lambda \mathrm{e}^{-\lambda} 2^{-M}. \tag{5.261}$$

Moreover, we use (5.242) to deduce that

$$\sum_{n=M+1}^{\infty} 2^{-\frac{M}{n}} \mathbb{P}(N = n) \leq \mathbb{P}(N \geq M + 1) \leq \lambda \mathrm{e}^{\lambda} 2^{-M}. \tag{5.262}$$

Combining the two inequalities with (5.260) yields

$$\begin{aligned}\mathbb{E}[\|s - \mathrm{P}_M^{\mathrm{greedy}}(s)\|_2^2] &\geq \frac{\sigma_0^2}{48\mathrm{e}\lambda} M^{-1} \mathbb{E}[2^{-\frac{M}{N}}] \frac{\sum_{n=1}^{M} 2^{-\frac{M}{n}} \mathbb{P}(N = n)}{\mathbb{E}[2^{-\frac{M}{N}}]} \\ &\geq \frac{\sigma_0^2}{48\mathrm{e}\lambda} M^{-1} \mathbb{E}[2^{-\frac{M}{N}}] \frac{\lambda \mathrm{e}^{-\lambda} 2^{-M}}{\lambda \mathrm{e}^{-\lambda} 2^{-M} + \lambda \mathrm{e}^{\lambda} 2^{-M}} = M^{-1} \mathbb{E}[2^{-\frac{M}{N}}] \frac{\sigma_0^2}{48\mathrm{e}\lambda(1 + \mathrm{e}^{2\lambda})}, \end{aligned} \tag{5.263}$$

which yields the desired lower-bound with the constant $C_1 = \frac{\sigma_0^2}{48\mathrm{e}\lambda(1+\mathrm{e}^{2\lambda})}$. □

Theorem 5.3 provides lower and upper bounds for the greedy approximation error of any finite-variance compound Poisson process. In Theorem 5.4, we use these bounds to deduce sub-exponential super-polynomial behaviors for the greedy approximation error of compound-Poisson processes.

Theorem 5.4. *Let s be a compound Poisson process whose law of jumps admits a PDF with zero-mean and finite variance. Then the greedy approximation error $\mathrm{MSE}_M^{\mathrm{greedy}}$ of s follows a sub-exponential and super-polynomial asymptotic behavior. Precisely, for any $k \in \mathbb{N}$, we have that*

$$\lim_{M\to+\infty} M^k \mathrm{MSE}_M^{\mathrm{greedy}} = 0, \tag{5.264}$$

and for any $\alpha > 0$,

$$\lim_{M\to+\infty} e^{\alpha M} \mathrm{MSE}_M^{\mathrm{greedy}} = +\infty. \tag{5.265}$$

Proof. It is sufficient to prove that the quantity $\mathbb{E}[2^{-\frac{M}{N}}]$ has sub-exponential and super-polynomial asymptotic behavior.

Super-polynomiality: First note that there exists an integer number $N_0 \in \mathbb{N}$ such that for every $n \geq N_0$, we have $\mathbb{P}(N = n) \leq 2^{-n}$. We then consider the following decomposition for any $M \geq N_0 + 1$

$$\mathbb{E}[2^{-\frac{M}{N}}] = \sum_{n=1}^{N_0-1} \mathbb{P}(N = n) 2^{-\frac{M}{n}} + \sum_{n=N_0}^{M-1} \mathbb{P}(N = n) 2^{-\frac{M}{n}} + \sum_{n=M}^{\infty} \mathbb{P}(N = n) 2^{-\frac{M}{n}}. \tag{5.266}$$

We separately show that each term of the previous decomposition decays faster than the inverse of any polynomial as $M \to \infty$.

For the first term, simply due to $\mathbb{P}(N = n) \leq 1$, we have that

$$M^k \sum_{n=1}^{N_0-1} \mathbb{P}(N = n) 2^{-\frac{M}{n}} \leq (N_0 - 1) M^k 2^{-\frac{M}{N_0}} \longrightarrow 0, \tag{5.267}$$

as $M \to \infty$. Regarding the second term, we use the bound $\mathbb{P}(N = n) \leq 2^{-n}$ for $n \geq N_0$ to deduce that

$$\forall n \geq N_0 : \mathbb{P}(N = n) 2^{-\frac{M}{n}} \leq 2^{-n-\frac{M}{n}} \leq 2^{-2\sqrt{M}}, \tag{5.268}$$

where in the last inequality, we have used $x+y \geq 2\sqrt{xy}$ with $x=n$ and $y=M/n$. Hence,

$$M^k \sum_{n=N_0}^{M-1} \mathbb{P}(N=n) 2^{-\frac{M}{n}} \leq M^{k+1} 2^{-2\sqrt{M}} \to 0, \tag{5.269}$$

as $M \to \infty$. Finally for the last term, we use (5.242) with $t=1$ to obtain that

$$M^k \sum_{n=M}^{\infty} \mathbb{P}(N=n) 2^{-\frac{M}{n}} \leq M^k \sum_{n=M-1}^{\infty} \mathbb{P}(N=n) = M^k \mathbb{P}(N \geq M) \leq M^k \mathrm{e}^{\lambda(\mathrm{e}-1)} \mathrm{e}^{-M} \to 0, \tag{5.270}$$

as $M \to \infty$.

Sub-exponentiality: To show the sub-exponential behavior, we fix $\alpha > 0$ and for all $n_0 \geq 2$, we note that

$$\mathrm{e}^{\alpha M} \mathbb{E}[2^{-\frac{M}{N}}] \geq 2^{\log_2(\mathrm{e})\alpha M} \mathbb{P}(N=n_0) 2^{-\frac{M}{n_0}} = \mathbb{P}(N=n_0) 2^{(\alpha \log_2(\mathrm{e}) - \frac{1}{n_0})M}. \tag{5.271}$$

Now by fixing n_0 to be a sufficiently large integer so that $\log_2 \mathrm{e}\alpha - \frac{1}{n_0} > 0$, we deduce that the right hand side explodes. □

An enlightening consequence of the super-polynomial behavior of the greedy approximation error is that it demonstrates that our provided lower- and upper-bounds are asymptotically comparable. Specifically from the upper-bound provided in Theorem 5.3, we deduce that

$$\frac{\mathrm{MSE}_M^{\mathrm{greedy}}}{\mathbb{E}[2^{-\frac{M}{N}}]^{(1-\epsilon)}} \leq C_2 M \mathbb{E}[2^{-\frac{M}{N}}]^{\epsilon} \tag{5.272}$$

for any $\epsilon > 0$. Moreover, Theorem 5.4 implies that the quantity $M\mathbb{E}[2^{-\frac{M}{N}}]^{\epsilon}$ tends to 0 as $M \to +\infty$ and is therefore bounded from above. Using a similar argumentation for the lower-bound of Theorem 5.3, we obtain the following corollary.

Corollary 5.8. *For any $\epsilon > 0$, there are positive constants $C_{1,\epsilon}, C_{2,\epsilon} > 0$ such that*

$$C_{1,\epsilon} \mathbb{E}[2^{-\frac{M}{N}}]^{(1+\epsilon)} \leq \mathrm{MSE}_M^{\mathrm{greedy}} \leq C_{2,\epsilon} \mathbb{E}[2^{-\frac{M}{N}}]^{(1-\epsilon)}, \tag{5.273}$$

for all values of $M \geq 1$.

Our theoretical analysis validates the two following observations in a rigorous manner:

- A piecewise constant function with a fixed number of jumps $n \geq 1$ is such that its greedy approximation in the Haar basis roughly behaves like $\mathcal{O}(2^{-M/n})$, which is exponential and therefore decays to 0 faster than any polynomial. Note that the exponential decay is faster for smaller values of n.
- The number of jumps N of a compound Poisson is random. It is almost surely finite but can be arbitrarily large. The concrete effect is that the mean-square error of the greedy approximation roughly behaves like $\mathcal{O}(\mathbb{E}[2^{-M/N}])$. The subexponential behavior of the MSE is then a consequence, as we have shown.

It is worth noting that the characterization provided by Theorem 5.4 is not deducible from earlier works that was based on the machinery of Besov regularity, such as [443]. Previous works focus on the almost sure behavior of the approximation error, while we focus on the mean-square approximation in this work. These are two different regimes and to the best of our knowledge, the possible link between the two has not been investigated.

By contrast, we obtain some information regarding the asymptotic behavior of best M-term approximation error of compound Poisson processes from Theorem 5.4. Indeed, by combining (5.188) and (5.264), one observes that

$$\limsup_{M\to+\infty} M^k \mathrm{MSE}_M^{\mathrm{best}} \leq \lim_{M\to+\infty} M^k \mathrm{MSE}_M^{\mathrm{greedy}} = 0, \tag{5.274}$$

for any $k \in \mathbb{N}$. Using the fact that MSE is non-negative simply implies that

$$\lim_{M\to+\infty} M^k \mathrm{MSE}_M^{\mathrm{best}} = 0. \tag{5.275}$$

The super-polynomial decay of the greedy approximation error shows that this method, despite being very simple and easily implemented, reaches excellent approximation performances. In the next section, we will empirically show that the greedy scheme performs similarly to the practically uncomputable best M-term approximation scheme.

5.4.5 Numerical Illustration

In this section, we provide a numerical demonstration of the main results of this work. First, it is illustrative and reflects the potential practical impact of our theoretical claims in a complementary and empirical manner. Second, it shows that the results obtained for the greedy approximation method are similar to what would be obtained for the best M-term approximation. Finally, it emphasizes that wavelets are able to exploit the inherent sparsity of non-Gaussian signals, which is not the case of traditional Fourier-based approximation schemes. These empirical observations then give rise to open theoretical questions that might be of interest to the community.

To simulate each approximation scheme, we first generate a signal that consists of $2^{10} = 1024$ equispaced samples of a given random process over $[0, 1]$. We then compute its (discrete) Haar wavelet coefficients of scale up to $J_{\max} = 10$. Finally, we create the approximated signal according to the given approximation scheme[13]. We repeat each experiment 1000 times and we report the average to reduce the effect of the underlying randomness (Monte Carlo method). The averaged values are then good approximations of the quantities of interest, that is, the MSEs given by (5.187) for different approximation schemes.

In the first experiment, we compute the MSE of greedy approximation for Brownian motions and compound Poisson processes with different values of $\lambda = 10, 50, 100, 500$ and with Gaussian jumps, as a function of the number M of coefficients that are preserved. We recall that, N being the random number of jumps of the compound Poisson process s over $[0, 1]$, $\lambda = \mathbb{E}[N]$ is the averaged number of jumps. To have a fair comparison, we unify the variance of the random processes in all cases to be $\sigma_0^2 = 1$ (which corresponds to a law of jumps with variance $\sigma_0^2/\lambda = 1/\lambda$ for compound Poisson processes).

[13] For the best M-term approximation, we do not have access to the infinitely many wavelet coefficients but only to the ones up to a given scale ($J_{\max} = 10$ in this case). This means that we only have an approximation of the best M-term for our simulations. However, the variance of the wavelet coefficients decay with the scale j like 2^{-2j} and the coefficients at larger scales are therefore very small with high probability. Our approximation of the best M terms is therefore excellent.

The results are depicted in Figure 5.16, where in each case we plot the MSE in log scale, that is $\log_2(M) \mapsto 10\log_{10}(\mathrm{MSE})$. From Proposition 5.16 and Corollary 5.7, we expect that the MSE of Brownian motion follows a global linear decay in the log scale, while decaying sub-linearly locally. Indeed, for $M = 2^J$, $J \in \mathbb{N}$, we deduce from (5.204) that

$$10\log_{10}(\mathrm{MSE}_M^{\mathrm{greedy}}) = \alpha - \beta J, \tag{5.276}$$

where $\alpha = 10\log_{10}(\sigma_0^2/6)$ and $\beta = 10\log_{10}(2)$ which shows a linear decay with respect to $J = \log_2(M)$. However, in the regime when $J = \lfloor \log_2(M) \rfloor$ is fixed, that is when $2^J \leq M < 2^{J+1}$, we obtain from (5.203) that

$$10\log_{10}(\mathrm{MSE}_M^{\mathrm{greedy}}) = \alpha - \beta(J+1) + \beta\log_2\left(3 - \frac{M}{2^J}\right), \tag{5.277}$$

which shows that the error decays sub-linearly in this regime. These theoretical claims can be observed in Figure 5.16, as well.

In addition, from Theorem 5.4, we know that the MSE of compound Poisson processes in the log scale should asymptotically decay faster than any straight line. This is also observable in Figure 5.16, indicating the dramatic difference between the compressiblity of compound Poisson processes and Brownian motions, as expected.

We moreover remark in Figure 5.16 that the small-scale behavior ($\log_2(M) = J \leq 3$) does not distinguish between different values of λ, but also between compound Poisson processes and the Brownian motion. Again, this empirical fact has a theoretical counterpart: it is linked with the fact that the statistics of finite variance compound Poisson processes are barely distinguishable from the ones of the Brownian motion at coarse scales. This has been formalized in [442] which states, when particularized to our case, that compound Poisson processes with finite variance converge to the Brownian motion when zoomed out and correctly renormalized.

Finally, we observe in Figure 5.16 that as $\lambda \to +\infty$, the greedy approximation of compound Poisson processes converges pointwise to the one of Brownian motion. This empirical observation poses an interesting theoretical question which is also consistent with [393, Theorem 5], which states—when specialized to our problem—that the compound Poisson process with constant variance σ_0^2 and Gaussian jumps converges in law to the Brownian motion when $\lambda \to \infty$.

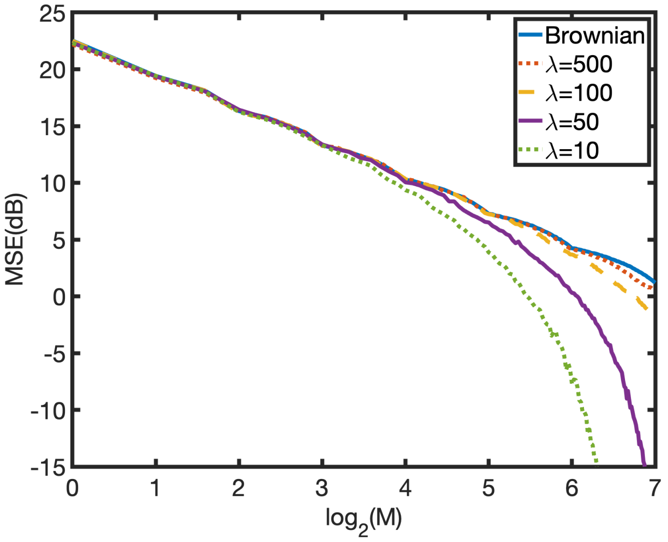


Figure 5.16: Greedy approximations of Brownian motions and compound Poisson processes with different values of λ and Gaussian jumps. We fix the variance to one in all cases.

As we have seen in the introduction, it is particularly satisfactory to characterize the compressibility of Lévy processes via their best M-term approximation error in a given basis. Although our greedy approximation error only provides an upper-bound for the best M-term approximation error, we demonstrate numerically in Figure 5.17 that the two approximation schemes are comparable in the sense of MSE. This is also an important observation, as it reveals that the extremely simple greedy approximation performs almost as good as the best M-term approximation, the latter being a theoretical bound for M-term approximation schemes.

We now investigate the effect of the dictionary in which we perform the approximation scheme. We consider the Haar transform and discrete cosine transform (DCT) for approximating the Brownian motion and compound Poisson processes with Gaussian jumps. The results are depicted in Figure 5.18, where we plot the best M-term approximation error of each setup in the log scale.

We observe that the DCT works slightly better than Haar for the Brownian motion. This is not surprising: The DCT is known to be asymptotically equivalent to the Karhunen-Loève transform (KLT), which is optimal for *Gaussian* stationary processes [351]. It is worth noting that this is also valid for the Brownian motion, which is not stationary but still admits stationary increments.

However, there is a dramatic difference between Haar and DCT for compound Poisson processes. We see in Figure 5.18 that, contrary to the Haar dictionary, the DCT is unable to take advantage of the effective sparsity of compound Poisson processes. This is of course not a surprise and is folklore knowledge, but it has not yet been justified theoretically for the best of our knowledge. This is nevertheless consistent with recent theoretical and empirical results demonstrating that wavelet methods outperform classical Fourier-based methods for the analysis of sparse stochastic processes [48, 443].

5.4.6 Summary

The theoretical and empirical findings of this work are reminiscent to the so-called “Mallat’s heuristic" [471], which states that

> *“Wavelets are the best bases for representing objects composed of singularities, when there may be an arbitrary number of singularities, which may be located in all possible spatial position."*

and which remarkably describes the compound Poisson model.

To do so, we provided a theoretical analysis to characterize the compressibility of compound Poisson processes. To that end, we introduced a simple approximation greedy scheme performed over the Haar wavelet basis. We then provided comparable lower and upper-bounds for the mean-squared approximation error. This enabled us to deduce the sub-exponential super-polynomial asymptotic behavior for the error.

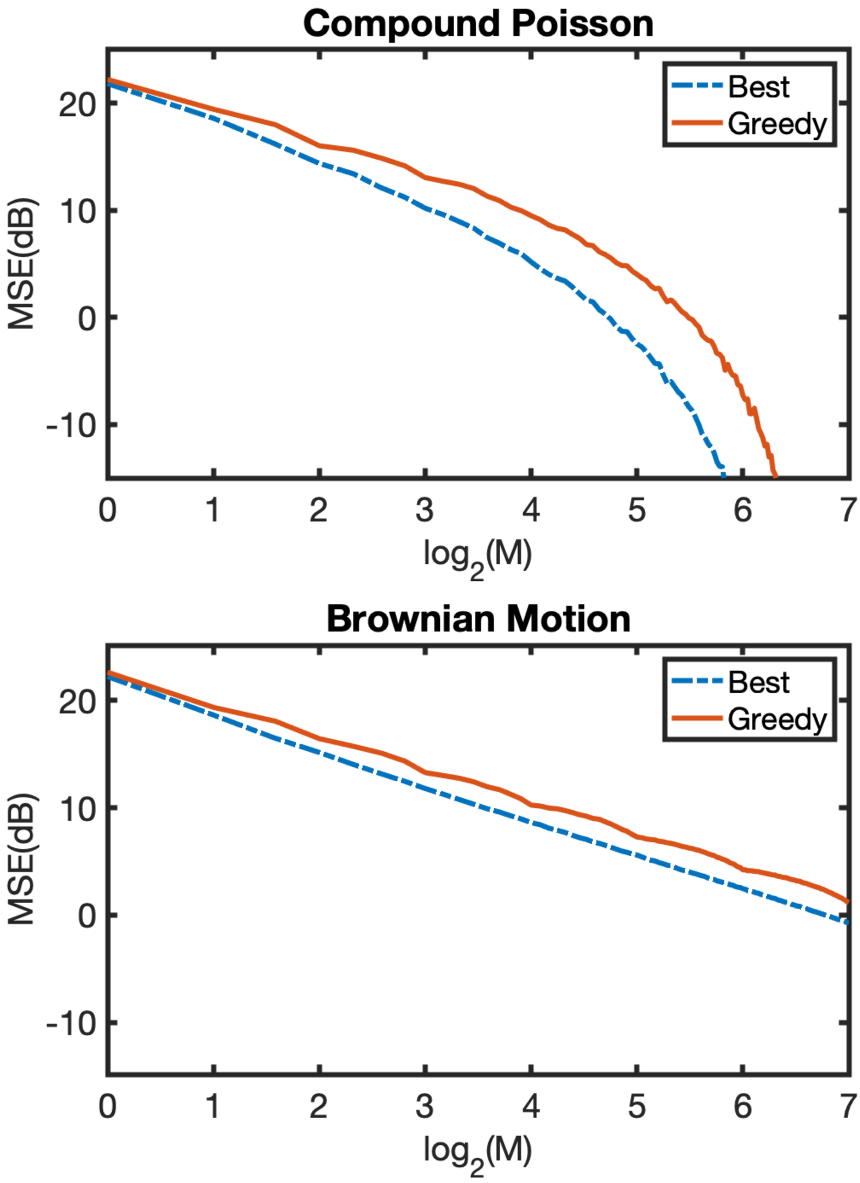
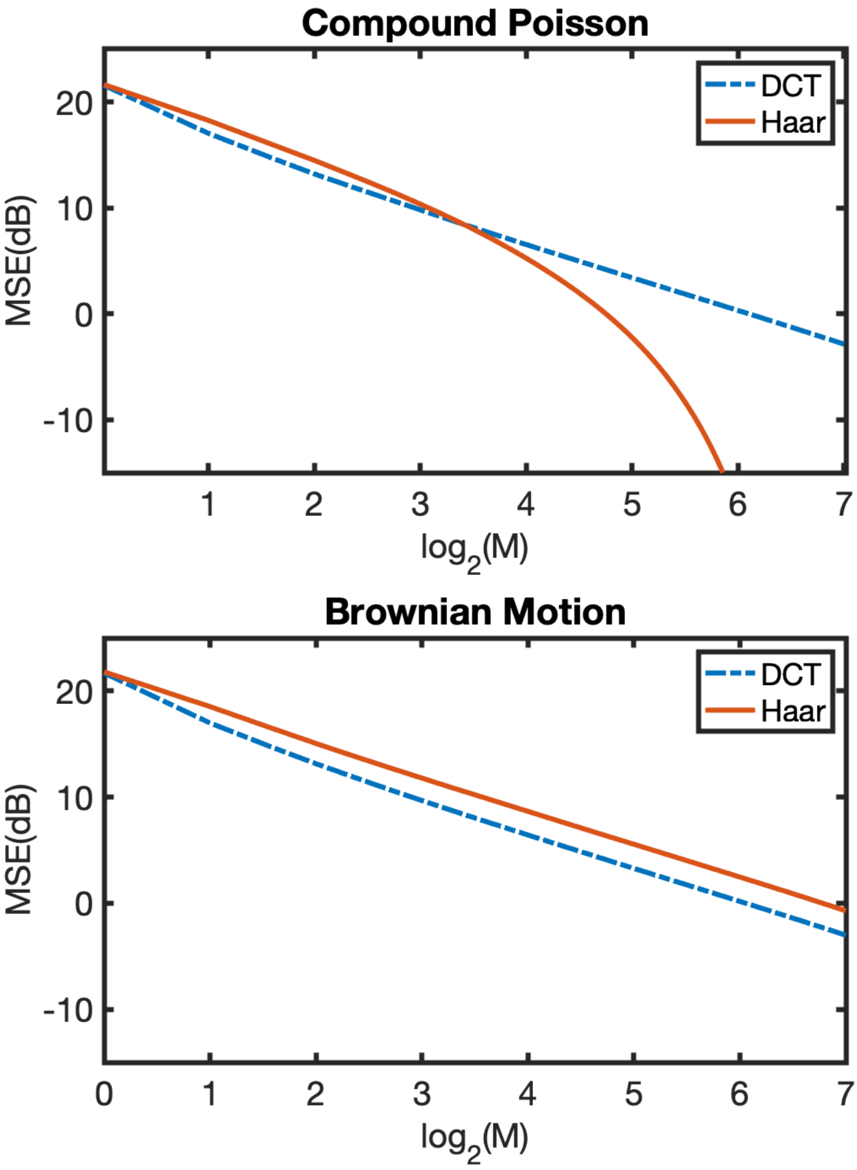


Figure 5.17: Greedy and Best M-term approximation of a compound Poisson process (top) with $\lambda = 10$ and Gaussian jumps and a Brownian motion (bottom). We normalize both processes to have unit variance



Compound Poisson

Brownian Motion

Figure 5.18: Comparison of DCT versus Haar wavelets to optimally represent (best M-term) a compound Poisson process (top) with $\lambda = 10$ and Gaussian jumps with a Brownian motion (bottom). We normalize both random processes to have unit variance.

Chapter 6

Conclusion

In this thesis, we have provided a theoretical foundation for the study of convex optimization problems posed over infinite-dimensional Banach spaces. By utilizing our general framework, we have developed novel schemes for supervised learning and solving linear inverse problems. We have established that splines are optimal in many of the proposed schemes. Consequently, we have studied splines systematically from both deterministic and stochastic points of view.

In what follows, we first summarise the key contributions of this thesis (Section 6.1). We then provide a brief overview of the future research directions (Section 6.2).

6.1 Summary of the Main Contributions

6.1.1 Spline Theory

We have studied multi-spline spaces and, in particular, their shortest-support generators. First, we have proven the optimality of Hermite splines. Specifically, we have shown that they are maximally localized among the pair of functions with similar reproduction ability (Theorem 1.1).

Next, we have considered general multi-spline spaces. We have defined the notion of mB-splines (Definition 1.4) and proven a lower-bound for their support size (Theorem 1.3). Moreover, we have shown that mB-splines generate Riesz bases (Thoerem 1.4). Finally, we have proposed a recursive algorithm to construct mB-splines which demonstrates the sharpness of our announced lower-bound on their support size (Theorem 1.5).

6.1.2 Optimization Theory

We have studied a broad class of convex optimization problems that are posed over infinite-dimensional Banach spaces. Specifically, we have investigated the case where the search space admits a direct-sum or direct-product decomposition. First, we have studied the topological structure of these spaces (Theorem 2.3). Next, we have proven a general representer theorem that characterizes the solution set of the aforementioned optimization problems (Theorem 2.4). Finally, we have adapted the general representer theorem to the case in which a seminorm is used as the regularization functional (Theorem 2.5).

6.1.3 Supervised Learning

Based on the general results of Chapter 2, we have developed novel schemes for supervised learning that inherit a certain notion of “sparsity”.

Multi-Kernel Regression

We have developed a Banach-space formalism for learning with multiple kernels. As the first step, we have characterized the topological structure of the underlying search space (Theorem 3.2). Next, we have identified the class of kernel functions that are admissible in our framework (Theorem 3.3). Finally, we have proven a representer theorem that suggests a sparse and adaptive kernel expansion for the mapping to be learned (Theorem 3.4).

Univariate Learning Models Under Lipschitz Constraint

We have considered the problem of learning one-dimensional mappings under joint sparsity and Lipschitz constraints. We have proposed two formulations that address these constraints together. The first one uses the Lipschitz constant as a regularization term (Theorem 3.6), while the second one involves a sparsity-promoting regularization term paired with a hard constraint on the Lipschitz constant (Theorem 3.7). For both cases, we have presented efficient algorithms for finding the sparsest linear spline solution (Section 3.3.4).

Learning Activation Functions of Deep Neural Networks

We have introduced a functional framework for the learning of activation functions of deep neural networks. We have first identified the connection between our proposed formulation and the global Lipschitz constant of the network (Theorem 3.9). Next, we have proven a representer theorem that guarantees the existence of an optimal neural network with linear spline activation functions (Theorem 3.11). Finally, we have proposed a B-spline based algorithm for training linear spline activation functions that is scalable in time and memory (Section 3.4.4).

Learning Multivariate CPWL Functions

We have introduced the HTV seminorm and proposed its use as a regularization functional for learning multivariate CPWL functions. First, we have studied the duality mappings of Schatten matrix norms (Theorem 3.13). Then, we have rigorously defined the HTV seminorm and illustrated its suitability as a measure of complexity (Section 3.5.2). In particular, we have provided a closed-form expression for the HTV of CPWL functions (Theorem 3.17). To demonstrate the practical relevance of our theoretical findings, we have proposed a computational scheme for learning two-dimensional CPWL mappings with HTV regularization (Section 3.5.3).

6.1.4 Inverse Problems

We have applied our Banach-space optimization theory to the class of linear inverse problems involving multicomponent priors on the signal of interest. We have first studied hybrid models that are the continuous-domain counterparts of redundant dictionaries used in the framework of compressed-sensing (Section 4.2). We have then considered a composite sparse-plus-smooth model for the target signal that deploys adequate regularization functionals for each component (Section 4.3). Finally, we have developed a novel rotation-invariant and sparsity-promoting regularization term for fitting curves to 2D point-clouds (Section 4.4). In all of these cases, we have derived representer theorems that offer parametric forms for the solutions to the respective problems (Theorems 4.1, 4.2 and 4.4, respectively). These theoretical characterizations allowed us to discretize the problems exactly and solve them numerically.

6.1.5 Stochastic Processes

We have studied the family of sparse stochastic processes that is known to be the limit point of compound-Poisson processes. First, we have proposed a novel scheme, based on nonuniform B-splines, for generating gridless trajectories of a broad subfamily of sparse stochastic processes (Section 5.2). Next, we have characterized the Besov regularity of Lévy white noises (Theorem 5.1) which is tightly linked to the compressibility of sparse stochastic processes. Finally, we have refined the existing compressibility rate for compound-Poisson processes by providing a direct method for analyzing their wavelet compressibility (Theorem 5.3).

6.2 Outlook and Future Works

The work presented in this thesis opens several new fields of inquiry. Some of these future lines of research are as follows:

- **Vector-valued L-splines.** In Chapter 1, we have developed a theory for multi-splines that are sums of piecewise-polynomial functions. One can go even further and introduce a vectorial extension of the notion of L-splines. This is motivated from applications where different components of the signal are coupled. Examples include but are not limited to multi-task learning, change-point detection and multimodal signal processing. An open question is to develop a variational framework that demonstrates the optimality of this new family of splines.

- **Sensitivity of infinite-dimensional optimization problems.** While our representer theorems in Chapter 2 characterize the solution set of a broad class of optimization problems, they do not provide information about the sensitivity of the problem with respect to its parameters. In particular, robustness of the optimal solution to small perturbations of the measurement vector is of great importance for practical purposes. A possible approach for such an analysis is to study the mathematical properties of the duality mapping.

- **Computational techniques for finding the optimal solution.** A major challenge in our theoretical framework is to numerically solve the proposed infinite-dimensional problems. While our representer theorems allow us to recast these problems as finite-dimensional ones, the latter are still high-dimensional and nonconvex. To illustrate the relevance of our theory, we have proposed grid-based techniques in this thesis. However, these methods are only feasible for low-dimensional problems and developing gridless optimization techniques in higher dimensions will significantly improve the practical relevance of the research presented in Chapters 3 and 4.

- **Convergence of the sparse process generator.** In Chapter 5 (and in particular, Section 5.2), we have proposed a method for generating approximated trajectories of sparse stochastic processes. An open question is to study the convergence rate of our proposed method. We suspect that the compressibility rate of the target process will have an effect on the convergence speed. This conjecture stems from the compressibliity hierarchy that has been developed in the remaining parts of Chapter 5, where we have demonstrated that the compound-Poisson processes are among the most compressible members of the family.

Bibliography

[1] Julien Fageot, Shayan Aziznejad, Michael Unser, and Virginie Uhlmann, “Support and approximation properties of Hermite splines,” *Journal of Computational and Applied Mathematics*, vol. 368, pp. 1–15, 2020.

[2] Alexis Goujon, Shayan Aziznejad, Alireza Naderi, and Michael Unser, “Shortest-support multi-spline bases for generalized sampling,” *Journal of Computational and Applied Mathematics*, vol. 395, pp. 1–18, 2021.

[3] Thomas Debarre, Shayan Aziznejad, and Michael Unser, “Hybrid-spline dictionaries for continuous-domain inverse problems,” *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5824–5836, 2019.

[4] Isaac J Schoenberg, “Contribution to the problem of approximation of equidistant data by analytic functions,” *Quarterly of Applied Mathematics*, vol. 4, no. 2, pp. 112–141, 1946.

[5] I.J. Schoenberg, *Cardinal Spline Interpolation*, SIAM, 1973.

[6] M. Unser, “Splines: A perfect fit for signal and image processing,” *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

[7] Carl de Boor, “On calculating with B-splines,” *Journal of Approximation Theory*, vol. 6, no. 1, pp. 50–62, 1972.

[8] Grace Wahba, *Spline Models For Observational Data*, SIAM, 1990.

[9] Isaac J Schoenberg, "On Spline Interpolation at all Integer Points of the Real Axis," *Séminaire Delange-Pisot-Poitou. Théorie des nombres*, vol. 9, no. 1, pp. 1–18, 1967.

[10] C. De Boor, *A Practical Guide to Splines*, Springer-Verlag, 1978.

[11] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part I—Theory," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–833, 1993.

[12] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part II—Efficient design and applications," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 834–848, 1993.

[13] C. de Boor, R.A. DeVore, and A. Ron, "The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$," *Journal of Functional Analysis*, vol. 119, no. 1, pp. 37–78, 1994.

[14] C. de Boor, R.A. DeVore, and A. Ron, "Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$," *Transactions of the American Mathematical Society*, vol. 341, no. 2, pp. 787–806, 1994.

[15] C. de Boor, R.A. DeVore, and A. Ron, "Approximation orders of FSI spaces in $L_2(\mathbb{R}^d)$," *Constructive Approximation*, vol. 14, no. 4, pp. 631–652, 1998.

[16] K. Jetter and G. Plonka, "A survey on L_2-approximation order from shift-invariant spaces," in *Multivariate Approximation and Applications*, pp. 73–111. Cambridge University Press, 2001.

[17] Michael Unser and Ingrid Daubechies, "On the approximation power of convolution-based least squares versus interpolation," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1697–1711, 1997.

[18] A. Ron, "Factorization theorems for univariate splines on regular grids," *Israel Journal of Mathematics*, vol. 70, no. 1, pp. 48–68, 1990.

[19] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1069–1080, 2001.

[20] Xavier Glorot, Antoine Bordes, and Yoshua Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

[21] P. Lipow and I.J. Schoenberg, "Cardinal interpolation and spline functions. III. Cardinal Hermite interpolation," *Linear Algebra and Its Applications*, vol. 6, pp. 273–304, 1973.

[22] I.J. Schoenberg and A. Sharma, "Cardinal interpolation and spline functions. V. The B-splines for cardinal Hermite interpolation," *Linear Algebra and Its Applications*, vol. 7, no. 1, pp. 1–42, 1973.

[23] R.T. Farouki, "The Bernstein polynomial basis: A centennial retrospective," *Computer Aided Geometric Design*, vol. 29, no. 6, pp. 379–419, 2012.

[24] H. Prautzsch, W. Boehm, and M. Paluszny, *Bézier and B-Spline Techniques*, Springer-Verlag, 2013.

[25] G.E. Farin, *Curves and Surfaces for CAGD: A Practical Guide*, Morgan Kaufmann Publishers, 2002.

[26] W. Böhm, G. Farin, and J. Kahmann, "A survey of curve and surface methods in CAGD," *Computer Aided Geometric Design*, vol. 1, no. 1, pp. 1–60, 1984.

[27] W. Dahmen, B. Han, R.-Q. Jia, and A. Kunoth, "Biorthogonal multiwavelets on the interval: Cubic Hermite splines," *Constructive Approximation*, vol. 16, no. 2, pp. 221–259, 2000.

[28] R.F. Warming and R.M. Beam, "Discrete multiresolution analysis using Hermite interpolation: Biorthogonal multiwavelets," *SIAM Journal on Scientific Computing*, vol. 22, no. 4, pp. 1269–1317, 2000.

[29] V. Uhlmann, J. Fageot, and M. Unser, "Hermite snakes with control of tangents," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2803–2816, 2016.

[30] V. Uhlmann, J. Fageot, H. Gupta, and M. Unser, "Statistical optimality of Hermite splines," in *Proceedings of the Eleventh International Workshop on Sampling Theory and Applications (SampTA'15)*, 2015, pp. 226–230.

[31] Claude E. Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[32] Abdul J. Jerri, “The Shannon sampling theorem–its various extensions and applications: A tutorial review,” *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, 1977.

[33] Michael Unser, “Sampling–50 years after Shannon,” *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[34] Athanasios Papoulis, “Generalized sampling expansion,” *IEEE Transactions on Circuits and Systems*, vol. 24, no. 11, pp. 652–654, 1977.

[35] Akram Aldroubi and Michael Unser, “Sampling procedures in function spaces and asymptotic equivalence with Shannon’s sampling theory,” *Numerical Functional Analysis and Optimization*, vol. 15, no. 1-2, pp. 1–21, 1994.

[36] Michael Unser and Akram Aldroubi, “A general sampling theory for nonideal acquisition devices,” *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2915–2925, 1994.

[37] Robert Hummel, “Sampling for Spline Reconstruction,” *SIAM Journal on Applied Mathematics*, vol. 43, no. 2, pp. 278–288, 1983.

[38] Michael Unser, Akram Aldroubi, and Murray Eden, “Polynomial spline signal approximations: Filter design and asymptotic equivalence with Shannon’s sampling theorem,” *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 95–103, 1992.

[39] Akram Aldroubi, Michael Unser, and Murray Eden, “Cardinal spline filters: Stability and convergence to the ideal sinc interpolator,” *Signal Processing*, vol. 28, no. 2, pp. 127–138, 1992.

[40] Michael Unser and Josiane Zerubia, “A Generalized Sampling Theory without Band-Limiting Constraints,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 959–969, 1998.

[41] Michael Unser and Josiane Zerubia, “Generalized sampling: Stability and performance analysis,” *IEEE Transactions on Signal Processing*, vol. 45, no. 12, pp. 2941–2950, 1997.

[42] AG García, MA Hernández-Medina, and G Pérez-Villalón, "Generalized sampling in shift-invariant spaces with multiple stable generators," *Journal of Mathematical Analysis and Applications*, vol. 337, no. 1, pp. 69–84, 2008.

[43] Volker Pohl and Holger Boche, "U-invariant sampling and reconstruction in atomic spaces with multiple generators," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3506–3519, 2012.

[44] R. Radha, K. Sarvesh, and S. Sivananthan, "Sampling and reconstruction in a shift invariant space with multiple generators," *Numerical Functional Analysis and Optimization*, vol. 40, no. 4, pp. 365–385, 2019.

[45] Akram Aldroubi, "Oblique projections in atomic spaces," *Proceedings of the American Mathematical Society*, vol. 124, no. 7, pp. 2051–2060, 1996.

[46] Karlheinz Gröchenig, José Luis Romero, and Joachim Stöckler, "Sampling theorems for shift-invariant spaces, Gabor frames, and totally positive functions," *Inventiones Mathematicae*, vol. 211, no. 3, pp. 1119–1148, 2018.

[47] Ole Christensen, *An Introduction to Frames and Riesz Bases*, Springer, 2016.

[48] Michael Unser and Pouya D. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, 2014.

[49] Costanza Conti, Lucia Romani, and Michael Unser, "Ellipse-preserving Hermite interpolation and subdivision," *Journal of Mathematical Analysis and Applications*, vol. 426, no. 1, pp. 221–227, 2015.

[50] Costanza Conti, Mariantonia Cotronei, and Tomas Sauer, "Factorization of Hermite subdivision operators preserving exponentials and polynomials," *Advances in Computational Mathematics*, vol. 42, no. 5, pp. 1055–1079, 2016.

[51] Lucia Romani and Alberto Viscardi, "On the refinement matrix mask of interpolating hermite splines," *Applied Mathematics Letters*, vol. 109, 2020.

[52] S. Aziznejad, A. Naderi, and M. Unser, "Optimal spline generators for derivative sampling," in *Proceedings of the Thirteenth International Workshop on Sampling Theory and Applications (SampTA'19)*, Bordeaux, French Republic, July 8-12, 2019, pp. 1–4.

[53] Michael Unser and Shayan Aziznejad, “Convex optimization in sums of Banach spaces,” *Applied and Computational Harmonic Analysis*, vol. 56, pp. 1–25, 2022.

[54] Bernhard Schölkopf and Alexander J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.

[55] Alain Berlinet and Christine Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, vol. 3, Kluwer Academic Boston, 2004.

[56] G. Kimeldorf and G. Wahba, “A correspondence between Bayesian estimation on stochastic processes and smoothing by splines,” *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 495–502, 1970.

[57] A. Badoual, J. Fageot, and M. Unser, “Periodic splines and Gaussian processes for the resolution of linear inverse problems,” *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 6047–6061, 2018.

[58] Ernesto De Vito, Lorenzo Rosasco, Andrea Caponnetto, Michele Piana, and Alessandro Verri, “Some properties of regularized kernel methods,” vol. 5, pp. 1363–1390, 2004.

[59] Holger Wendland, *Scattered Data Approximations*, Cambridge University Press, 2005.

[60] Thomas Hofmann, Bernhart Schölkopf, and Alexander J. Smola, “Kernel methods in machine learning,” *Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.

[61] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola, “A generalized representer theorem,” in *Computational Learning Theory*. 2001, pp. 416–426, Springer Berlin Heidelberg.

[62] David L Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[63] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[64] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[65] Michael Elad, *Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[66] Yonina C Eldar and Gitta Kutyniok, *Compressed sensing: theory and applications*, Cambridge University Press, 2012.

[67] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, 2013.

[68] Michael Unser, Julien Fageot, and Harshit Gupta, “Representer theorems for sparsity-promoting ℓ_1 regularization,” *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5167–5180, 2016.

[69] Emmanuel J Candès and Carlos Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[70] Vincent Duval and Gabriel Peyré, “Exact support recovery for sparse spikes deconvolution,” *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[71] Axel Flinth and Pierre Weiss, “Exact solutions of infinite dimensional total-variation regularized problems,” *Information and Inference: A Journal of the IMA*, vol. 8, no. 3, pp. 407–443, 2019.

[72] Kristian Bredies and Marcello Carioni, “Sparsity of solutions for variational inverse problems with finite-dimensional data,” *Calculus of Variations and Partial Differential Equations*, vol. 59, no. 1, pp. 1–26, 2020.

[73] Michael Unser, Julien Fageot, and John Paul Ward, “Splines are universal solutions of linear inverse problems with generalized tv regularization,” *SIAM Review*, vol. 59, no. 4, pp. 769–793, 2017.

[74] Harshit Gupta, Julien Fageot, and Michael Unser, “Continuous-domain solutions of linear inverse problems with Tikhonov *versus* generalized TV

regularization," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4670–4684, 2018.

[75] Michael Unser, "A unifying representer theorem for inverse problems and machine learning," *Foundations of Computational Mathematics*, vol. 21, no. 4, pp. 941–960, 2021.

[76] Claire Boyer, Antonin Chambolle, Yohann De Castro, Vincent Duval, Frédéric De Gournay, and Pierre Weiss, "On representer theorems and convex regularization," *SIAM Journal of Optimization*, vol. 29, no. 2, pp. 1260–1281, 2019.

[77] Nachman Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[78] Charles A Micchelli, Massimiliano Pontil, and Peter Bartlett, "Learning the kernel function via regularization.," *Journal of Machine Learning Research*, vol. 6, no. 7, 2005.

[79] Mauricio A. Aĺvarez, Lorenzo Rosasco, and Neil D. Lawrence, *Kernels for Vector-Valued Functions: A Review*, 2012.

[80] Mehmet Gönen and Ethem Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[81] Michaël A van Wyk and Tariq S Durrani, "A framework for multiscale and hybrid RKHS-based approximators," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3559–3568, 2000.

[82] Carl de Boor and Robert E. Lynch, "On splines and their minimum properties," *Journal of Mathematics and Mechanics*, vol. 15, no. 6, pp. 953–969, 1966.

[83] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds., pp. 85–100. Springer-Verlag, Berlin, 1977.

[84] Anatoly Yu. Bezhaev and Vladimir A. Vasilenko, *Variational Theory of Splines*, Kluwer Academic/Plenum Publishers, New York, 2001.

[85] A.M. Mosamam and J.T. Kent, “Semi-reproducing kernel Hilbert spaces, splines and increment kriging,” *Journal of Nonparametric Statistics*, vol. 22, no. 6, pp. 711–722, 2010.

[86] M. Elad and A.M. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[87] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[88] J.-L. Starck, M. Elad, and D.L. Donoho, “Redundant multiscale transforms and their application for morphological component separation,” in *Advances in Imaging and Electron Physics*, pp. 287–348. Elsevier, 2004.

[89] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[90] H. Rauhut, K. Schnass, and P. Vandergheynst, “Compressed sensing and redundant dictionaries,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.

[91] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[92] J. Lin, S. Li, and Y. Shen, “Compressed data separation with redundant dictionaries,” *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4309–4315, jul 2013.

[93] Robert E. Megginson, *An Introduction to Banach Space Theory*, Springer, New York, 1998.

[94] Walter Rudin, *Functional Analysis*, McGraw-Hill, New York, 2nd edition, 1991.

[95] Arne Beurling and A. E. Livingston, “A theorem on duality mappings in Banach spaces,” *Arkiv för Matematik*, vol. 4, no. 5, pp. 405–411, 1962.

[96] Ioana Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, vol. 62, Springer Science & Business Media, 2012.

[97] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S Kazimierski, *Regularization Methods in Banach Spaces*, vol. 10, Walter de Gruyter, 2012.

[98] M. Riesz, "Sur les fonctions conjuguées," *Mathematische Zeitschrift*, vol. 27, pp. 218–244, 1927.

[99] Victor Klee, "On a theorem of Dubins," *Journal of Mathematical Analysis and Applications*, vol. 7, no. 3, pp. 425–427, 1963.

[100] Lester E Dubins, "On extreme points of convex sets," *Journal of Mathematical Analysis and Applications*, vol. 5, no. 2, pp. 237–244, 1962.

[101] F. L. Bauer, J. Stoer, and C. Witzgall, "Absolute and monotonic norms," *Numerische Mathematik*, vol. 3, no. 1, pp. 257–264, 1961.

[102] Patrick N. Dowling and Satit Saejung, "Extremal structure of the unit ball of direct sums of Banach spaces," *Nonlinear Analysis: Theory, Methods and Applications*, vol. 68, no. 4, pp. 951–955, 2008.

[103] M. Unser, "A representer theorem for deep neural networks," *Journal of Machine Learning Research*, vol. 20, no. 110, pp. 1–30, 2019.

[104] Michael Unser and Julien Fageot, "Native banach spaces for splines and variational inverse problems," *arXiv preprint arXiv:1904.10818*, 2019.

[105] Shayan Aziznejad and Michael Unser, "Multikernel regression with sparsity constraint," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 201–224, 2021.

[106] Shayan Aziznejad, Harshit Gupta, Joaquim Campos, and Michael Unser, "Deep neural networks with trainable activations and controlled Lipschitz constant," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4688–4699, 2020.

[107] Pakshal Bohra, Joaquim Campos, Harshit Gupta, S. Aziznejad, and Michael Unser, "Learning activation functions in deep (spline) neural networks," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 295–309, 2020.

[108] Shayan Aziznejad and Michael Unser, "Duality mapping for Schatten matrix norms," *Numerical Functional Analysis and Optimization*, vol. 42, no. 6, pp. 679–695, 2021.

[109] Joaquim Campos, Shayan Aziznejad, and Michael Unser, "Learning of continuous and piecewise-linear functions with Hessian total-variation regularization," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 36–48, 2021.

[110] S. Aziznejad, T. Debarre, and M. Unser, "Sparsest univariate learning models under Lipschitz constraint," *IEEE Open Journal of Signal Processing*, pp. 140–154, 2022.

[111] Shayan Aziznejad, Joaquim Campos, and Michael Unser, "Measuring complexity of learning schemes using Hessian-Schatten total variation," *arXiv preprint arXiv:2112.06209*, 2021.

[112] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[113] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, "Overview of supervised learning," in *The elements of statistical learning*, pp. 9–41. Springer, 2009.

[114] Felipe Cucker and Ding Xuan Zhou, *Learning Theory: An Approximation Theory Viewpoint*, vol. 24, Cambridge University Press, 2007.

[115] Luc Devroye, László Györfi, and Gábor Lugosi, *A Probabilistic Theory of Pattern Recognition*, vol. 31, Springer Science & Business Media, 2013.

[116] László Györfi, Michael Kohler, Adam Krzyzak, and Harro Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer Science & Business Media, 2006.

[117] Ingo Steinwart and Andreas Christmann, *Support Vector Machines*, Springer Science & Business Media, 2008.

[118] George Kimeldorf and Grace Wahba, “Some results on Tchebycheffian spline functions,” *Journal of Mathematical Analysis and Applications*, vol. 33, no. 1, pp. 82–95, 1971.

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

[120] V. Vapnik, *Statistical Learning Theory*, vol. 3, Wiley, New York, 1998.

[121] T. Evgeniou, M. Pontil, and T. Poggio, “Regularization networks and support vector machines,” *Advances in Computational Mathematics*, vol. 13, no. 1, pp. 1, 2000.

[122] Andrea Caponnetto and Ernesto De Vito, “Optimal rates for the regularized least-squares algorithm,” *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 331–368, 2007.

[123] Shahar Mendelson and Joseph Neeman, “Regularization in kernel learning,” *The Annals of Statistics*, vol. 38, no. 1, pp. 526–565, 2010.

[124] Ingo Steinwart, Don R Hush, and Clint Scovel, “Optimal rates for regularized least squares regression.,” in *Conference on Learning Theory*, 2009, pp. 79–93.

[125] Mona Eberts and Ingo Steinwart, “Optimal regression rates for SVMs using Gaussian kernels,” *Electronic Journal of Statistics*, vol. 7, pp. 1–42, 2013.

[126] Peter L Bartlett, Andrea Montanari, and Alexander Rakhlin, “Deep learning: A statistical viewpoint,” *Acta Numerica*, vol. 30, pp. 87–201, 2021.

[127] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[128] K. Jin, M. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[129] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2015, pp. 234–241.

[130] Guido F Montufar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio, "On the number of linear regions of deep neural networks," in *Advances in Neural Information Processing Systems*, 2014, pp. 2924–2932.

[131] Ronen Eldan and Ohad Shamir, "The power of depth for feedforward neural networks," in *Proceedings of the 29th Conference on Learning Theory*, New York, USA, 2016, vol. 49, pp. 907–940.

[132] H. N. Mhaskar and T. Poggio, "Deep vs. shallow networks: An approximation theory perspective," *Analysis and Applications*, vol. 14, no. 06, pp. 829–848, 2016.

[133] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao, "Why and when can deep—but not shallow—networks avoid the curse of dimensionality: A review," *International Journal of Automation and Computing*, vol. 14, no. 5, pp. 503–519, 2017.

[134] George Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[135] A. Maas, A. Hannun, and A. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013, vol. 30, p. 3.

[136] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[137] Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee, "Understanding deep neural networks with rectified linear units," in *International Conference on Learning Representations*, 2018.

[138] Tomaso Poggio, Lorenzo Rosasco, Amnon Shashua, Nadav Cohen, and Fabio Anselmi, "Notes on hierarchical splines, DCLNs and i-theory," Tech. Rep., Center for Brains, Minds and Machines (CBMM), 2015.

[139] Randall Balestriero and Richard G Baraniuk, "Mad max: Affine spline insights into deep learning," *Proceedings of the IEEE*, 2020.

[140] Rahul Parhi and Robert D Nowak, "The role of neural network activation functions," *IEEE Signal Processing Letters*, vol. 27, pp. 1779–1783, 2020.

[141] Rahul Parhi and Robert D Nowak, “Banach space representer theorems for neural networks and ridge splines,” *Journal of Machine Learning Research*, vol. 22, no. 43, pp. 1–40, 2021.

[142] Rahul Parhi and Robert D Nowak, “What kinds of functions do deep neural networks learn? insights from variational spline theory,” *arXiv preprint arXiv:2105.03361*, 2021.

[143] Ingo Steinwart, “Sparseness of support vector machines,” *Journal of Machine Learning Research*, vol. 4, no. November, pp. 1071–1105, 2003.

[144] Ingo Steinwart, “Sparseness of support vector machines—Some asymptotically sharp bounds,” in *Advances in Neural Information Processing Systems*, 2004, pp. 1069–1076.

[145] Ingo Steinwart and Andreas Christmann, “Sparsity of SVMs that use the epsilon-insensitive loss,” in *Advances in Neural Information Processing Systems*, 2009, pp. 1569–1576.

[146] V. Roth, “The generalized LASSO,” *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 16–28, 2004.

[147] Lei Shi, Yun-Long Feng, and Ding-Xuan Zhou, “Concentration estimates for learning with ℓ_1-regularizer and data dependent hypothesis spaces,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 2, pp. 286–302, 2011.

[148] Hong-Yan Wang, Quan-Wu Xiao, and Ding-Xuan Zhou, “An approximation theory approach to learning with ℓ_1 regularization,” *Journal of Approximation Theory*, vol. 167, pp. 240–258, 2013.

[149] W. Rudin, *Real and Complex Analysis*, Tata McGraw-hill education, 2006.

[150] Gregory E Fasshauer, Fred J Hickernell, and Qi Ye, “Solving support vector machines in reproducing kernel Banach spaces with positive definite functions,” *Applied and Computational Harmonic Analysis*, vol. 38, no. 1, pp. 115–139, 2015.

[151] H. Zhang, Y. Xu, and J. Zhang, “Reproducing kernel Banach spaces for machine learning,” *Journal of Machine Learning Research*, vol. 10, no. Dec, pp. 2741–2775, 2009.

[152] Haizhang Zhang and Jun Zhang, “Regularized learning in Banach spaces as an optimization problem: Representer theorems,” *Journal of Global Optimization*, vol. 54, no. 2, pp. 235–250, 2012.

[153] Francis Bach, “Breaking the curse of dimensionality with convex neural networks,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 629–681, 2017.

[154] F. R Bach, G. Lanckriet, and M. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” in *Proceedings of the Twenty-First International Conference on Machine Learning*, 2004, pp. 41–48.

[155] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, “Learning the kernel matrix with semidefinite programming,” *Journal of Machine Learning Research*, vol. 5, no. Jan, pp. 27–72, 2004.

[156] Alain Rakotomamonjy, Francis R Bach, Stéphane Canu, and Yves Grandvalet, “SimpleMKL,” *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2491–2521, 2008.

[157] Francis R Bach, “Consistency of the group LASSO and multiple kernel learning,” *Journal of Machine Learning Research*, vol. 9, no. Jun, pp. 1179–1225, 2008.

[158] J. Bazerque and G. Giannakis, “Nonparametric basis pursuit via sparse kernel-based learning,” *arXiv preprint arXiv:1302.5449*, 2013.

[159] J. Gao, P. Kwan, and D. Shi, “Sparse kernel learning with LASSO and Bayesian inference algorithm,” *Neural Networks*, vol. 23, no. 2, pp. 257–264, 2010.

[160] M. Kloft, U. Brefeld, P. Laskov, K. Müller, A. Zien, and S. Sonnenburg, “Efficient and accurate ℓ_p-norm multiple kernel learning,” in *Advances in Neural Information Processing Systems*, 2009, pp. 997–1005.

[161] Marius Kloft, Ulrich Rückert, and Peter L Bartlett, “A unifying view of multiple kernel learning,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 66–81.

[162] Gelfand I and GE Shilov, “Generalized functions, vol. i: Properties and operations,” 1964.

[163] Emmanuel J Candès, Justin Romberg, and Terence Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[164] Kristian Bredies and Hanna Katriina Pikkarainen, “Inverse problems in spaces of measures,” *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[165] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies, “The sliding frank–wolfe algorithm and its application to super-resolution microscopy,” *Inverse Problems*, vol. 36, no. 1, pp. 014001, 2019.

[166] T. Debarre, J. Fageot, H. Gupta, and M. Unser, “B-Spline-based exact discretization of continuous-domain inverse problems with generalized TV regularization,” *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4457–4470, 2019.

[167] Matthieu Simeoni, “Functional penalised basis pursuit on spheres,” *Applied and Computational Harmonic Analysis*, 2021.

[168] B. Simon, “Distributions and their Hermite expansions,” *Journal of Mathematical Physics*, vol. 12, no. 1, pp. 140–148, 1971.

[169] L. Schwartz, *Théorie des distributions*, vol. 2, Hermann Paris, 1957.

[170] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, 1999.

[171] F. Girosi, M. Jones, and T. Poggio, “Priors, Stabilizers and Basis Functions: From Regularization to Radial, Tensor and Additive Splines,” 1993.

[172] A. Smola, B. Schölkopf, and K. Müller, “The connection between regularization operators and support vector kernels,” *Neural Networks*, vol. 11, no. 4, pp. 637–649, 1998.

[173] A. Yuille, "The motion coherence theory," in *Proceedings of the International Conference on Computer Vision*. IEEF. Computer Society Press, 1998, pp. 344–354.

[174] Ingo Steinwart, Don Hush, and Clint Scovel, "An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4635–4643, 2006.

[175] N. Aronszajn and K. T. Smith, "Theory of Bessel potentials I," *Annales de l'Institut Fourier*, vol. 11, pp. 385–475, 1961.

[176] 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, 2009.

[177] E. Soubies, F. Soulez, M.T. McCann, T.-a. Pham, L. Donati, T. Debarre, D. Sage, and M. Unser, "Pocket guide to solve inverse problems with GlobalBioIm," *Inverse Problems*, vol. 35, no. 10, pp. 1–20, 2019.

[178] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.

[179] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis, "Compressed sensing using generative models," in *Proceedings of the 34th International Conference on Machine Learning*, 2017, vol. 70, pp. 537–546.

[180] Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nati Srebro, "Exploring generalization in deep learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 5947–5956.

[181] Harshit Gupta, Kyong Hwan Jin, Ha Q Nguyen, Michael T McCann, and Michael Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1440–1453, 2018.

[182] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[183] Yu Sun, Jiaming Liu, and Ulugbek S Kamilov, "Block coordinate regularization by denoising," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 908–921, 2020.

[184] Jiaming Liu, Yu Sun, Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S Kamilov, "RARE: Image reconstruction using deep priors learned without ground truth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, 2020.

[185] Zihui Wu, Yu Sun, Alex Matlock, Jiaming Liu, Lei Tian, and Ulugbek S Kamilov, "SIMBA: Scalable inversion in optical tomography using deep denoising priors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1163 – 1175, 2020.

[186] Pakshal Bohra, Alexis Goujon, Dimitris Perdios, Sébastien Emery, and Michael Unser, "Learning lipschitz-controlled activation functions in neural networks for plug-and-play image reconstruction methods," in *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021.

[187] Donald E McClure, "Perfect spline solutions of l_∞ extremal problems by control methods," *Journal of Approximation Theory*, vol. 15, no. 3, pp. 226–242, 1975.

[188] Samuel Karlin, "Interpolation properties of generalized perfect splines and the solutions of certain extremal problems. i," *Transactions of the American Mathematical Society*, vol. 206, pp. 25–66, 1975.

[189] Carl de Boor, "How small can one make the derivatives of an interpolating function?," *Journal of Approximation Theory*, vol. 13, no. 2, pp. 105–116, 1975.

[190] Charles A Micchelli, Th J Rivlin, and Shmuel Winograd, "The optimal recovery of smooth functions," *Numerische Mathematik*, vol. 26, no. 2, pp. 191–200, 1976.

[191] Carl de Boor, "On "best" interpolation," *Journal of Approximation Theory*, vol. 16, no. 1, pp. 28–42, 1976.

[192] A Pinkus, "On smoothest interpolants," *SIAM Journal on Mathematical Analysis*, vol. 19, no. 6, pp. 1431–1441, 1988.

[193] Ulrike von Luxburg and Olivier Bousquet, "Distance-based classification with lipschitz functions.," *Journal of Machine Learning Research*, vol. 5, no. Jun, pp. 669–695, 2004.

[194] Thomas Debarre, Quentin Denoyelle, Michael Unser, and Julien Fageot, "Sparsest piecewise-linear regression of one-dimensional data," *Journal of Computational and Applied Mathematics*, p. 114044, 2021.

[195] I. Savarese, P.and Evron, D. Soudry, and N. Srebro, "How do infinite width bounded norm networks look in function space?," in *Proceedings of the Thirty-Second Conference on Learning Theory*, Alina Beygelzimer and Daniel Hsu, Eds., Phoenix, USA, 2019, vol. 99 of *Proceedings of Machine Learning Research*, pp. 2667–2690, PMLR.

[196] Nik Weaver, *Lipschitz Functions*, chapter Chapter 1, pp. 1–34, World Scientific, 2018.

[197] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[198] A. Kurdila and M. Zabarankin, *Convex Functional Analysis*, Springer Science & Business Media, 2006.

[199] Ryan J Tibshirani, "The lasso problem and uniqueness," *Electronic Journal of statistics*, vol. 7, pp. 1456–1490, 2013.

[200] Stephen Boyd, Neal Parikh, and Eric Chu, *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*, Now Publishers Inc, 2011.

[201] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[202] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on Imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[203] S. Lane, M. Flax, D. Handelman, and J. Gelfand, "Multi-layer perceptrons with B-spline receptive field functions," in *Advances in Neural Information Processing Systems*, 1991, pp. 684–692.

[204] L. Vecci, F. Piazza, and A. Uncini, "Learning and approximation capabilities of adaptive spline activation function neural networks," *Neural Networks*, vol. 11, no. 2, pp. 259–270, 1998.

[205] Stefano Guarnieri, Francesco Piazza, and Aurelio Uncini, "Multilayer feedforward networks with adaptive spline activation function," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 672–683, 1999.

[206] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, "Learning activation functions to improve deep neural networks," in *Proceedings of the International Conference on Learning Representations*, 2015.

[207] S. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: A simple and accurate method to fool deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2574–2582.

[208] Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard, "The robustness of deep networks: A geometrical perspective," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 50–62, 2017.

[209] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30088–30095, 2020.

[210] J. Heinonen, *Lectures on Lipschitz Analysis*, Number 100. University of Jyväskylä, 2005.

[211] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Cognitive Modeling*, vol. 5, no. 3, pp. 1, 1988.

[212] A. Krogh and J. Hertz, "A simple weight decay can improve generalization," in *Advances in Neural Information Processing Systems*, 1992, pp. 950–957.

[213] D. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 6, pp. 797–829, 2006.

[214] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 265–288, 1996.

[215] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[216] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[217] Rajendra Bhatia, *Matrix analysis*, vol. 169, Springer-Verlag New York, 1997.

[218] S. Lefkimmiatis and M. Unser, "Poisson image reconstruction with Hessian Schatten-norm regularization," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4314–4327, 2013.

[219] S. Lefkimmiatis, J.P. Ward, and M. Unser, "Hessian Schatten-norm regularization for linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1873–1888, 2013.

[220] Yuan Xie, Shuhang Gu, Yan Liu, Wangmeng Zuo, Wensheng Zhang, and Lei Zhang, "Weighted Schatten p-norm minimization for image denoising and background subtraction," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4842–4857, 2016.

[221] Shangqi Gao and Qibin Fan, "Robust Schatten-p norm based approach for tensor completion," *Journal of Scientific Computing*, vol. 82, no. 1, pp. 1–23, 2020.

[222] Roger A Horn and Charles R Johnson, *Matrix Analysis*, Cambridge University Press, 2012.

[223] Mark A Davenport and Justin Romberg, “An overview of low-rank matrix recovery from incomplete observations,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 608–622, 2016.

[224] Emmanuel J Candès and Benjamin Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717, 2009.

[225] Emmanuel J Candès, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski, “Phase retrieval via matrix completion,” *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.

[226] Mike E Davies and Yonina C Eldar, “Rank awareness in joint sparse recovery,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[227] Maryam Fazel, Ting Kei Pong, Defeng Sun, and Paul Tseng, “Hankel matrix rank minimization with applications to system identification and realization,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.

[228] Ehsan Asadi, Shayan Aziznejad, Mohammad H Amerimehr, and Arash Amini, “A fast matrix completion method for index coding,” in *Proceedings of the Twenty-Fifth European Signal Processing Conference (EUSIPCO’17)*, Kos Island, Greece, August 28 - September 2 2017, IEEE, pp. 2606–2610.

[229] Homa Esfahanizadeh, Farshad Lahouti, and Babak Hassibi, “A matrix completion approach to linear index coding problem,” in *Proceedings of the Information Theory Workshop (ITW 2014)*, Hobart, TAS, Australia, 2014, IEEE, pp. 531–535.

[230] Fuad Kittaneh, “Inequalities for the Schatten p-norm,” *Glasgow Mathematical Journal*, vol. 26, no. 2, pp. 141–143, 1985.

[231] Fuad Kittaneh, “Inequalities for the Schatten p-norm II,” *Glasgow Mathematical Journal*, vol. 29, no. 1, pp. 99–104, 1987.

[232] Fuad Kittaneh, “Inequalities for the Schatten p-norm III,” *Communications in Mathematical Physics*, vol. 104, no. 2, pp. 307–310, 1986.

[233] Fuad Kittaneh, “Inequalities for the Schatten p-norm IV,” *Communications in Mathematical Physics*, vol. 106, no. 4, pp. 581–585, 1986.

[234] Fuad Kittaneh and Hideki Kosaki, “Inequalities for the Schatten p-norm V,” *Publications of the Research Institute for Mathematical Sciences*, vol. 23, no. 2, pp. 433–443, 1987.

[235] Jean-Christophe Bourin, “Matrix versions of some classical inequalities,” *Linear Algebra and Its Applications*, vol. 416, no. 2-3, pp. 890–907, 2006.

[236] O. Hirzallah, F. Kittaneh, and M.S. Moslehian, “Schatten p-norm inequalities related to a characterization of inner product spaces,” *Mathematical Inequalities and Applications*, vol. 13, no. 2, pp. 235–241, 2010.

[237] Mohammad Sal Moslehian, Masaru Tominaga, and Kichi-Suke Saito, “Schatten p-norm inequalities related to an extended operator parallelogram law,” *Linear Algebra and Its Applications*, vol. 435, no. 4, pp. 823–829, 2011.

[238] Cristian Conde and Mohammad Sal Moslehian, “Norm inequalities related to p-Schatten class,” *Linear Algebra and Its Applications*, vol. 498, pp. 441–449, 2016.

[239] David Wenzel and Koenraad MR Audenaert, “Impressions of convexity: An illustration for commutator bounds,” *Linear Algebra and Its Applications*, vol. 433, no. 11-12, pp. 1726–1759, 2010.

[240] Che-Man Cheng and Chunyu Lei, “On Schatten p-norms of commutators,” *Linear Algebra and Its Applications*, vol. 484, pp. 409–434, 2015.

[241] W So, “Facial structures of Schatten p-norms,” *Linear and Multilinear Algebra*, vol. 27, no. 3, pp. 207–212, 1990.

[242] Denis Potapov and Fedor Sukochev, “Fréchet differentiability of S_p norms,” *Advances in Mathematics*, vol. 262, pp. 436–475, 2014.

[243] Fuad Kittaneh, “On the continuity of the absolute value map in the Schatten classes,” *Linear Algebra and Its Applications*, vol. 118, pp. 61–68, 1989.

[244] Rajendra Bhatia and Fuad Kittaneh, "Cartesian decompositions and Schatten norms," *Linear Algebra and Its Applications*, vol. 318, no. 1-3, pp. 109–116, 2000.

[245] Ping Liu and Yu-wen Wang, "The best generalized inverse of the linear operator in normed linear space," *Linear Algebra and Its Applications*, vol. 420, no. 1, pp. 9–19, 2007.

[246] Charles R Johnson and Roger A Horn, *Matrix Analysis*, Cambridge University Press, 1985.

[247] Feiping Nie, Heng Huang, and Chris Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2012, vol. 26.

[248] Fanhua Shang, Yuanyuan Liu, and James Cheng, "Scalable algorithms for tractable schatten quasi-norm minimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, vol. 30.

[249] Fanhua Shang, Yuanyuan Liu, Fanjie Shang, Hongying Liu, Lin Kong, and Licheng Jiao, "A unified scalable equivalent formulation for schatten quasi-norms," *Mathematics*, vol. 8, no. 8, pp. 1325, 2020.

[250] Paris Giampouras, René Vidal, Athanasios Rontogiannis, and Benjamin Haeffele, "A novel variational form of the schatten-p quasi-norm," *arXiv preprint arXiv:2010.13927*, 2020.

[251] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, "Structure tensor total variation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 1090–1122, 2015.

[252] Zdravko Cvetkovski, *Inequalities: Theorems, Techniques and Selected Problems*, Springer-Verlag Berlin Heidelberg, 2012.

[253] WV Petryshyn, "A characterization of strict convexity of Banach spaces and other uses of duality mappings," *Journal of Functional Analysis*, vol. 6, no. 2, pp. 282–291, 1970.

[254] John Giles, David Gregory, and Brailey Sims, "Geometrical implications of upper semi-continuity of the duality mapping on a Banach space," *Pacific Journal of Mathematics*, vol. 79, no. 1, pp. 99–109, 1978.

[255] Manuel D Contreras and Rafael Payá, "On upper semicontinuity of duality mappings," *Proceedings of the American Mathematical Society*, pp. 451–459, 1994.

[256] Charles J Himmelberg and T Parthasarathy, "Measurable relations," *Fund. Math*, vol. 87, no. 1, pp. 53–72, 1975.

[257] Luigi Ambrosio, Nicola Fusco, and Diego Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, vol. 254, Clarendon Press Oxford, 2000.

[258] Levent Onural, "Impulse functions over curves and surfaces and their applications to diffraction," *Journal of Mathematical Analysis and Applications*, vol. 322, no. 1, pp. 18–27, 2006.

[259] Gerald B Folland, *Real analysis: modern techniques and their applications*, vol. 40, John Wiley & Sons, 1999.

[260] Françoise Demengel, "Fonctions à hessien borné," in *Annales de l'institut Fourier*, 1984, vol. 34, pp. 155–190.

[261] Walter Hinterberger and Otmar Scherzer, "Variational methods on the space of functions of bounded hessian for convexification and denoising," *Computing*, vol. 76, no. 1-2, pp. 109–133, 2006.

[262] Kristian Bredies, Karl Kunisch, and Thomas Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[263] Maïtine Bergounioux and Loic Piffet, "A second-order model for image denoising," *Set-Valued and Variational Analysis*, vol. 18, no. 3-4, pp. 277–306, 2010.

[264] L. Condat and D. Van De Ville, "Three-directional box-splines: Characterization and efficient evaluation," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 417–420, 2006.

[265] A. Entezari, M. Nilchian, and M. Unser, "A Box Spline Calculus for the Discretization of Computed Tomography Reconstruction Problems," *IEEE Transactions on Medical Imaging*, vol. 31, no. 8, pp. 1532–1541, 2012.

[266] Alireza Entezari, Dimitri Van De Ville, and Torsten Möller, "Practical Box Splines for Reconstruction on the Body Centered Cubic Lattice," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 313–328, 2008.

[267] H.R. Kunsch, E. Agrell, and F.A. Hamprecht, "Optimal Lattices for Sampling," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 634–647, 2005.

[268] Wolfgang Dahmen and Charles A. Micchelli, "On the optimal approximation rates for criss-cross finite element spaces," *Journal of Computational and Applied Mathematics*, vol. 10, no. 3, pp. 255–273, 1984.

[269] Hartmut Prautzsch and Wolfgang Boehm, "Box Splines," in *Handbook of Computer Aided Geometric Design*, pp. 255–282. North-Holland, Amsterdam, 2002.

[270] Weihong Guo and Ming-Jun Lai, "Box Spline Wavelet Frames for Image Edge Analysis," vol. 6, no. 3, pp. 1553–1578, 2013.

[271] D. Van De Ville, T. Blu, M. Unser, W. Philips, I. Lemahieu, and R. Van de Walle, "Hex-Splines: A Novel Spline Family for Hexagonal Lattices," *IEEE Transactions on Image Processing*, vol. 13, no. 6, pp. 758–772, 2004.

[272] Carl De Boor, K Höllig, and S. D Riemenschneider, *Box Splines*, Springer, 2011.

[273] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspect of Functional Analysis*, pp. 796–830. Cremonese, Rome, Italy, 1971.

[274] M. Unser, "Approximation Power of Biorthogonal Wavelet Expansions," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 519–527, 1996.

[275] Ryan J. Tibshirani and Jonathan Taylor, "The Solution Path of the Generalized LASSO," *The Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.

[276] Zhouchen Lin, Risheng Liu, and Zhixun Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in *Proceedings of the 24th Conference on Advances in Neural Information Processing Systems*, 2011, vol. 24.

[277] George Bernard Dantzig, Alex Orden, and Philip S. Wolfe, *Notes on Linear Programming: Part I: The Generalized Simplex Method for Minimizing a Linear Form Under Linear Inequality Restraints*, RAND Corporation, 1954.

[278] David G. Luenberger and Yinyu Ye, *Linear and Nonlinear Programming*, vol. 116 of *Operations Research & Management Science*, Springer New York, 2008.

[279] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the Third International Conference on Learning Representations*, San Diego, CA, USA, May 7-9 2015.

[280] Thomas Debarre, Shayan Aziznejad, and Michael Unser, “Continuous-domain formulation of inverse problems for composite sparse-plus-smooth signals,” *IEEE Open Journal of Signal Processing*, vol. 2, pp. 545–558, 2021.

[281] Icíar Lloréns Jover, Thomas Debarre, Shayan Aziznejad, and Michael Unser, “Coupled splines for sparse curve fitting,” *IEEE Transactions on Image Processing*, vol. 31, pp. 4707–4718, 2022.

[282] A. Tikhonov, “Solution of incorrectly formulated problems and the regularization method,” *Soviet Mathematics Doklady*, vol. 4, pp. 1035–1038, 1963.

[283] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.

[284] SD Fisher and JW Jerome, “Spline solutions to L1 extremal problems in one and several variables,” *Journal of Approximation Theory*, vol. 13, no. 1, pp. 73–83, 1975.

[285] D L. Donoho and Philip B Stark, “Uncertainty principles and signal recovery,” *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[286] D L. Donoho and Xiaoming Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[287] D L. Donoho and Michael Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[288] Michael Elad, Mario AT Figueiredo, and Yi Ma, “On the role of sparse and redundant representations in image processing,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.

[289] D L. Donoho and Gitta Kutyniok, “Microlocal analysis of the geometric separation problem,” *Communications on Pure and Applied Mathematics*, vol. 66, no. 1, pp. 1–47, 2013.

[290] J-L Starck, Michael Elad, and D L. Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[291] J-L Starck, Y Moudden, J Bobin, M Elad, and D L. Donoho, “Morphological component analysis,” in *Wavelets XI*. International Society for Optics and Photonics, 2005, vol. 5914, p. 59140Q.

[292] Michael Elad, J-L Starck, Philippe Querre, and D L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[293] Jérôme Bobin, Jean-Luc Starck, Jalal M Fadili, Yassir Moudden, and D L. Donoho, “Morphological component analysis: An adaptive thresholding strategy,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.

[294] Jian-Feng Cai, Stanley Osher, and Zuowei Shen, “Split Bregman methods and frame based image restoration,” *Multiscale Modeling & Simulation*, vol. 8, no. 2, pp. 337–369, 2009.

[295] Ricardo Otazo, E J. Candès, and Daniel K Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[296] Michael Unser and Thierry Blu, "Cardinal exponential splines: Part i-theory and filtering algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1425–1438, 2005.

[297] E J. Candès and Terence Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[298] C. De Mol and M. Defrise, "Inverse imaging with mixed penalties," in *Proceedings URSI EMTS*, Pisa, Italy, 2004, pp. 798–800.

[299] A. Gholami and S.M. Hosseini, "A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals," *Signal Processing*, vol. 93, no. 7, pp. 1945–1960, 2013.

[300] Valeriya Naumova and Steffen Peter, "Minimization of multi-penalty functionals by alternating iterative thresholding and optimal parameter choices," *Inverse Problems*, vol. 30, no. 12, pp. 125003, 2014.

[301] M. Grasmair, T. Klock, and V. Naumova, "Adaptive multi-penalty regularization based on a generalized Lasso path," *Applied and Computational Harmonic Analysis*, vol. 49, no. 1, pp. 30–55, 2018.

[302] V. Debarnot, P. Escande, T. Mangeat, and P. Weiss, "Learning low-dimensional models of microscopes," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 178–190, 2021.

[303] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, vol. 22, American Mathematical Society, Providence RI, USA, sep 2001.

[304] L.A. Vese and S.J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, vol. 19, no. 1/3, pp. 553–572, 2003.

[305] L.A. Vese and S.J. Osher, “Image denoising and decomposition with total variation minimization and oscillatory functions,” *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1/2, pp. 7–18, 2004.

[306] D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[307] B. Adcock and A.C. Hansen, “Generalized sampling and infinite-dimensional compressed sensing,” *Foundations of Computational Mathematics*, vol. 16, no. 5, pp. 1263–1323, 2016.

[308] I. Daubechies, M. Defrise, and C. De Mol, “Sparsity-enforcing regularisation and ISTA revisited,” *Inverse Problems*, vol. 32, no. 10, pp. 104001, 2016.

[309] P. Bohra and M. Unser, “Computation of “best” interpolants in the l_p sense,” in *Proceedings of the Forty-Fifth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'20)*, Barcelona, Kingdom of Spain, May 4-8, 2020, pp. 5505–5509.

[310] Leello Dadi, Shayan Aziznejad, and Michael Unser, “Generating sparse stochastic processes using matched splines,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 4397–4406, 2020.

[311] Michael Unser and Thierry Blu, “Generalized smoothing splines and the optimal discretization of the wiener filter,” *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, 2005.

[312] O. Hori and S. Tanigawa, “Raster-to-vector conversion by line fitting based on contours and skeletons,” in *Proceedings of the Second International Conference on Document Analysis and Recognition (ICDAR '93)*, Tsukuba, Japan, October 20-22, 1993, pp. 353–358.

[313] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[314] R. Delgado-Gonzalo, V. Uhlmann, D. Schmitter, and M. Unser, “Snakes on a plane: A perfect snap for bioimage analysis,” *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 41–48, 2015.

[315] Anil K Jain, Yu Zhong, and Marie-Pierre Dubuisson-Jolly, “Deformable template models: A review,” *Signal Processing*, vol. 71, no. 2, pp. 109–129, 1998.

[316] M. Jacob, T. Blu, and M. Unser, “Efficient energies and algorithms for parametric snakes,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1231–1244, 2004.

[317] V. Uhlmann, R. Delgado-Gonzalo, C. Conti, L. Romani, and M. Unser, “Exponential Hermite splines for the analysis of biomedical images,” in *Proceedings of the Thirty-Ninth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’14)*, Firenze, Italian Republic, May 4-9, 2014, pp. 1650–1653.

[318] R. Delgado-Gonzalo and M. Unser, “Spline-based framework for interactive segmentation in biomedical imaging,” *IRBM—Ingénierie et Recherche Biomédicale / BioMedical Engineering and Research*, vol. 34, no. 3, pp. 235–243, 2013.

[319] P. Thévenaz and M. Unser, “Snakuscules,” *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 585–593, 2008.

[320] M Grossman, “Parametric curve fitting,” *The Computer Journal*, vol. 14, no. 2, pp. 169–172, 1971.

[321] Michael Plass and Maureen Stone, “Curve-fitting with piecewise parametric cubics,” in *Proceedings of the Tenth Annual Conference on Computer Graphics and Interactive Techniques*, Detroit, Michigan, USA, July 25-29, 1983, pp. 229–239.

[322] Y. Hu and M. Jacob, “Higher degree total variation (HDTV) regularization for image recovery,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2559–2571, 2012.

[323] J. Fageot and M. Simeoni, “TV-based reconstruction of periodic functions,” *Inverse Problems*, vol. 36, no. 11, pp. 115015, 2020.

[324] C. Fernandez-Granda, “Super-resolution of point sources via convex programming,” *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, pp. 251–303, 2016.

[325] Shayan Aziznejad and Julien Fageot, "Wavelet analysis of the Besov regularity of Lévy white noise," *Electronic Journal of Probability*, vol. 25, pp. 1–38, 2020.

[326] Shayan Aziznejad and Julien Fageot, "Wavelet compressibility of compound Poisson processes," *IEEE Transactions on Information Theory*, vol. 68, no. 4, pp. 2752–2766, 2022.

[327] R. Gray and L. Davisson, *An Introduction to Statistical Signal Processing*, Cambridge University Press, 2004.

[328] A. Webb, *Statistical Pattern Recognition*, John Wiley & Sons, 2003.

[329] N. Ahmed and K.R. Rao, *Orthogonal Transforms for Digital Signal Processing*, Springer Berlin Heidelberg, 1975.

[330] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[331] Anuj Srivastava, Ann B Lee, Eero P Simoncelli, and S-C Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.

[332] David Mumford and Agnès Desolneux, *Pattern Theory: The Stochastic Analysis of Real-World Signals*, A. K. Peters/CRC Press, Aug. 2010.

[333] B. Pesquet-Popescu and J. Lévy Véhel, "Stochastic fractal models for image processing," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 48–62, 2002.

[334] J. Fageot, E. Bostan, and M. Unser, "Wavelet statistics of sparse and self-similar images," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2951–2975, 2015.

[335] F. Sciacchitano, *Image reconstruction under non-Gaussian noise*, Ph.D. thesis, Technical University of Denmark (DTU), 2017.

[336] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Los Alamitos, USA, 1999, vol. 1, pp. 541–547.

[337] P. Gupta, A.K. Moorthy, R. Soundararajan, and A.C. Bovik, “Generalized Gaussian scale mixtures: A model for wavelet coefficients of natural images,” *Signal Processing: Image Communication*, vol. 66, pp. 87–94, 2018.

[338] Stephane Mallat, *A Wavelet Tour of Signal Processing*, Elsevier, Sept. 1999.

[339] Jean-Luc Starck, Fionn Murtagh, and Jalal M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, May 2010.

[340] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The LASSO and Generalizations*, Chapman and Hall/CRC, 2015.

[341] J. Starck, D. Donoho, J. Fadili, and A. Rassat, *Sparsity and the Bayesian Perspective*, Astronomy & Astrophysics, 2013.

[342] I. Rish and G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*, CRC press, 2014.

[343] A. Amini, M. Unser, and F. Marvasti, “Compressibility of Deterministic and Random Infinite Sequences,” *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5193–5201, 2011.

[344] A. Amini and M. Unser, “Sparsity and Infinite Divisibility,” *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2346–2358, 2014.

[345] B. Adcock, A.C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: A new theory for compressed sensing,” *Forum of Mathematics, Sigma*, vol. 5, 2017.

[346] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

[347] S. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[348] Y. Meyer, *Wavelets and Operators*, vol. 37 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, 1992.

[349] Charilaos Christopoulos, Athanassios Skodras, and Touradj Ebrahimi, “The JPEG2000 still image coding system: An overview,” *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2000.

[350] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transfom,” *IEEE Transactions on Computers*, vol. 23, no. 1, pp. 90–93, 1974.

[351] M. Unser, “On the approximation of the discrete Karhunen-Loève transform for stationary processes,” *Signal Processing*, vol. 7, no. 3, pp. 231–249, 1984.

[352] Thomas Kailath, “The innovations approach to detection and estimation theory,” *Proceedings of the IEEE*, vol. 58, no. 5, pp. 680–695, 1970.

[353] M. Unser, P. D. Tafti, and Q. Sun, “A Unified Formulation of Gaussian Versus Sparse Stochastic Processes—Part I: Continuous-Domain Theory,” *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1945–1962, Mar. 2014.

[354] David Mumford and Basilis Gidas, “Stochastic models for generic images,” *Quarterly of Applied Mathematics*, vol. 59, no. 1, pp. 85–111, 2001.

[355] E. Bostan, U. S. Kamilov, M. Nilchian, and M. Unser, “Sparse Stochastic Processes and Discretization of Linear Inverse Problems,” *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2699–2710, 2013.

[356] M Alper Kutay, Athina P Petropulu, and Catherine W Piccoli, “On modeling biomedical ultrasound RF echoes using a power-law shot-noise model,” *IEEE Transactions on Iltrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 4, pp. 953–968, 2001.

[357] Stephen M Kogon and Dimitris G Manolakis, “Signal modeling with self-similar α-stable processes: The fractional Lévy stable motion model,” *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 1006–1010, 1996.

[358] José R Gallardo, Dimitrios Makrakis, and Luis Orozco-Barbosa, “Use of α-stable self-similar stochastic processes for modeling traffic in broadband networks,” *Performance Evaluation*, vol. 40, no. 1-3, pp. 71–98, 2000.

[359] Nick Laskin, Ioannis Lambadaris, Fotios C Harmantzis, and Michael Devetsikiotis, “Fractional lévy motion and its application to network traffic modeling,” *Computer Networks*, vol. 40, no. 3, pp. 363–375, 2002.

[360] K. Itô, “Stationary random distributions,” *Kyoto Journal of Mathematics*, vol. 28, no. 3, pp. 209–223, 1954.

[361] I.M. Gel’fand, “Generalized random processes,” *Doklady Akademii Nauk SSSR*, vol. 100, pp. 853–856, 1955.

[362] I.M. Gel’fand and N.Y. Vilenkin, *Generalized Functions. Vol. 4: Applications of Harmonic Analysis*, Academic Press, New York-London, 1964.

[363] P. Cartier, “Processus aléatoires généralisés,” *Séminaire Bourbaki*, vol. 8, pp. 425–434, 1963.

[364] X. Fernique, “Processus linéaires, processus généralisés,” *Annales de l’Institut Fourier*, vol. 17, pp. 1–92, 1967.

[365] P.J. Brockwell and J. Hannig, “CARMA (p, q) generalized random processes,” *Journal of Statistical Planning and Inference*, vol. 140, no. 12, pp. 3613–3618, 2010.

[366] D. Berger, “Lévy driven linear and semilinear stochastic partial differential equations,” *arXiv preprint arXiv:1907.01926*, 2019.

[367] D. Berger, “Lévy driven CARMA generalized processes and stochastic partial differential equations,” *Stochastic Processes and their Applications*, vol. 130, no. 10, pp. 5865–5887, 2020.

[368] A. Abdesselam, “A second-quantized Kolmogorov–Chentsov theorem via the operator product expansion,” *Communications in Mathematical Physics*, pp. 1–54, 2020.

[369] R.C. Dalang and T. Humeau, “Random field solutions to linear SPDEs driven by symmetric pure jump Lévy space-time white noises,” *Electronic Journal of Probability*, vol. 24, 2019.

[370] K. Itô, *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, vol. 47, SIAM, 1984.

[371] J.B. Walsh, “An introduction to stochastic partial differential equations,” in *École d’Été de Probabilités de Saint Flour XIV-1984*, pp. 265–439. Springer, 1986.

[372] E. Bostan, J. Fageot, U.S. Kamilov, and M. Unser, "MAP estimators for self-similar sparse stochastic models," in *Proceedings of the Tenth International Workshop on Sampling Theory and Applications (SampTA13), Bremen, Germany*, 2013, pp. 197–199.

[373] E. Clarkson and H.H. Barrett, "Characteristic functionals in imaging and image-quality assessment: tutorial," *JOSA A*, vol. 33, no. 8, pp. 1464–1475, 2016.

[374] F. Trèves, *Topological Vector Spaces, Distributions and Kernels*, Academic Press, New York-London, 1967.

[375] H. Biermé, O. Durieu, and Y. Wang, "Generalized random fields and Lévy's continuity theorem on the space of tempered distributions," *Commun. Stoch. Anal.*, vol. 12, no. 4, pp. Article 4, 427–445, 2018.

[376] R.A. Minlos, "Generalized random processes and their extension in measure," *Trudy Moskovskogo Matematicheskogo Obshchestva*, vol. 8, pp. 497–518, 1959.

[377] B. Simon, *Functional integration and quantum physics*, vol. 86, Academic press, 1979.

[378] Julien Fageot, Arash Amini, and Michael Unser, "On the Continuity of Characteristic Functionals and Sparse Stochastic Modeling," *Journal of Fourier Analysis and Applications*, vol. 20, no. 6, pp. 1179–1211, 2014.

[379] Robert C Dalang, Thomas Humeau, et al., "Lévy processes and lévy white noise as tempered distributions," *The Annals of Probability*, vol. 45, no. 6B, pp. 4389–4418, 2017.

[380] Julien Fageot, "On tempered discrete and l\'evy white noises," *arXiv preprint arXiv:2201.00797*, 2022.

[381] J. Fageot, *Gaussian versus Sparse Stochastic Processes: Construction, Regularity, Compressibility*, EPFL thesis no. 7657 (2017), 231 p., Swiss Federal Institute of Technology Lausanne (EPFL), 2017.

[382] Julien Fageot and Thomas Humeau, "The domain of definition of the lévy white noise," *Stochastic Processes and their Applications*, vol. 135, pp. 75–102, 2021.

[383] B.S. Rajput and J. Rosinski, “Spectral representations of infinitely divisible processes,” *Probability Theory and Related Fields*, vol. 82, no. 3, pp. 451–487, 1989.

[384] M. Griffiths and M. Riedle, “Modelling Lévy space-time white noises,” *arXiv preprint arXiv:1907.04193*, 2019.

[385] R.C. Dalang and J.B. Walsh, “The sharp Markov property of Lévy sheets,” *The Annals of Probability*, pp. 591–626, 1992.

[386] D. Daley and D. Vere-Jones, *An Introduction to The Theory of Point Processes: Volume II: General Theory and Structure*, Springer Science & Business Media, 2007.

[387] A. Amini, P. Thévenaz, J.P. Ward, and M. Unser, “On the linearity of Bayesian interpolators for non-Gaussian continuous-time AR(1) processes,” *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5063–5074, 2013.

[388] A. Amini, U. S. Kamilov, E. Bostan, and M. Unser, “Bayesian Estimation for Continuous-Time Sparse Stochastic Processes,” *IEEE Transactions on Signal Processing*, vol. 61, no. 4, 2013.

[389] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, “MMSE Estimation of Sparse Lévy Processes,” *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 137–147, 2013.

[390] Simon J Godsill and G Yang, “Bayesian inference for continuous-time ARMA models driven by non-Gaussian Lévy processes,” in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings.* IEEE, 2006, vol. 5, pp. V–V.

[391] Bernt Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 2nd ed edition, 1989.

[392] Sylvain Rubenthaler, “Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process,” *Stochastic Processes and their Applications*, vol. 103, no. 2, pp. 311–349, Feb. 2003.

[393] Julien Fageot, Virginie Uhlmann, and Michael Unser, "Gaussian and sparse processes are limits of generalized Poisson processes," *Applied and Computational Harmonic Analysis*, 2018.

[394] P. J. Brockwell, "Lévy-Driven CARMA Processes," *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 1, pp. 113–124, 2001.

[395] Philip E. Protter, *Stochastic Integration and Differential Equations*, Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin Heidelberg, 2 edition, 2005.

[396] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[397] Michael Unser and Pouya Dehghani Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 989–1006, 2010.

[398] Michael Unser, "Cardinal exponential splines: part ii-think analog, act digital," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1439–1449, 2005.

[399] Lennart Bondesson, "On Simulation from Infinitely Divisible Distributions," *Advances in Applied Probability*, vol. 14, no. 4, pp. 855–869, 1982.

[400] Luc Devroye, "Nonuniform random variate generation," in *Handbooks in Operations Research and Management Science*, vol. 13, pp. 83–121. Elsevier, 2006.

[401] Isaac Jacob Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions," in *IJ Schoenberg Selected Papers*, pp. 3–57. Springer, 1988.

[402] F. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[403] H. Triebel, *Function Spaces and Wavelets on Domains*, vol. 7 of *EMS Tracts in Mathematics*, European Mathematical Society (EMS), Zürich, 2008.

[404] Y. Meyer, F. Sellan, and M.S. Taqqu, “Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion,” *Journal of Fourier Analysis and Applications*, vol. 5, no. 5, pp. 465–494, 1999.

[405] P. Pad and M. Unser, “Optimality of operator-like wavelets for representing sparse AR(1) processes,” *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4827–4837, 2015.

[406] M. Hairer, “A theory of regularity structures,” *Inventiones mathematicae*, vol. 198, no. 2, pp. 269–504, 2014.

[407] M. Hairer and C. Labbé, “The reconstruction theorem in Besov spaces,” *Journal of Functional Analysis*, vol. 273, no. 8, pp. 2578–2618, 2017.

[408] C.S. Deng and R.L. Schilling, “On shift Harnack inequalities for subordinate semigroups and moment estimates for Lévy processes,” *Stochastic Processes and Their Applications*, vol. 125, pp. 3851–3878, 2015.

[409] F. Kühn, “Existence and estimates of moments for Lévy-type processes,” *Stochastic Processes and Their Applications*, vol. 127, no. 3, pp. 1018–1041, 2017.

[410] Franziska Kühn, “Lévy matters. VI,” *Lecture Notes in Mathematics*, vol. 2187, 2017.

[411] G. Laue, “Remarks on the relation between fractional moments and fractional derivatives of characteristic functions,” *Journal of Applied Probability*, pp. 456–466, 1980.

[412] H. Luschgy and G. Pagès, “Moment estimates for Lévy processes,” *Electronic Communications in Probability*, vol. 13, pp. 422–434, 2008.

[413] L.N. Slobodeckiı, “Generalized Sobolev spaces and their application to boundary problems for partial differential equations,” *Leningrad. Gos. Ped. Inst. Ucen. Zap*, vol. 197, pp. 54–112, 1958.

[414] E. Di Nezza, G. Palatucci, and E. Valdinoci, “Hitchhiker’s guide to the fractional Sobolev spaces,” *arXiv preprint arXiv:1104.4345*, 2011.

[415] Hans Triebel, *Theory of Function Spaces*, Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel, 2010.

[416] M. Kabanava, “Tempered Radon measures,” *Revista Matemática Complutense*, vol. 21, no. 2, pp. 553–564, 2008.

[417] J. Bertoin, *Lévy Processes*, vol. 121, Cambridge University Press, 1998.

[418] Z. Ciesielski, “Orlicz spaces, spline systems, and brownian motion,” *Constructive Approximation*, vol. 9, no. 2-3, pp. 191–208, 1993.

[419] Z. Ciesielski, G. Kerkyacharian, and B. Roynette, “Quelques espaces fonctionnels associés à des processus gaussiens,” *Studia Mathematica*, vol. 107, no. 2, pp. 171–204, 1993.

[420] T. Hytönen and M.C. Veraar, “On Besov regularity of Brownian motions in infinite dimensions,” *Probability and Mathematical Statistics*, vol. 28, no. 1, pp. 143–162, 2008.

[421] B. Roynette, “Mouvement brownien et espaces de Besov,” *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 43, no. 3-4, pp. 221–260, 1993.

[422] P. Sjögren, “Riemann sums for stochastic integrals and L_p moduli of continuity,” *Probability Theory and Related Fields*, vol. 59, no. 3, pp. 411–424, 1982.

[423] M.C. Veraar, “Correlation inequalities and applications to vector-valued gaussian random variables and fractional brownian motion,” *Potential Analysis*, vol. 30, no. 4, pp. 341–370, 2009.

[424] R.L. Schilling, “On Feller processes with sample paths in Besov spaces,” *Mathematische Annalen*, vol. 309, no. 4, pp. 663–675, 1997.

[425] R.L. Schilling, “Growth and Hölder conditions for the sample paths of Feller processes,” *Probability Theory and Related Fields*, vol. 112, no. 4, pp. 565–611, 1998.

[426] R.L. Schilling, “Function spaces as path spaces of Feller processes,” *Mathematische Nachrichten*, vol. 217, no. 1, pp. 147–174, 2000.

[427] B. Böttcher, R.L. Schilling, and J. Wang, *Lévy Matters III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, vol. 2099, Springer, 2014.

[428] V. Herren, "Lévy-type processes and Besov spaces," *Potential Analysis*, vol. 7, no. 3, pp. 689–704, 1997.

[429] M.C. Veraar, "Regularity of Gaussian white noise on the d-dimensional torus," *Marcinkiewicz centenary volume*, vol. 95, pp. 385–398, 2011.

[430] J. Fageot, M. Unser, and J.P. Ward, "On the Besov regularity of periodic Lévy noises," *Applied and Computational Harmonic Analysis*, vol. 42, no. 1, pp. 21 – 36, 2017.

[431] J. Fageot, A. Fallah, and M. Unser, "Multidimensional Lévy white noise in weighted Besov spaces," *Stochastic Processes and Their Applications*, vol. 127, no. 5, pp. 1599–1621, 2017.

[432] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia PA, USA, 1992.

[433] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge University Press, 2009.

[434] A. Durand and S. Jaffard, "Multifractal analysis of Lévy fields," *Probability Theory and Related Fields*, vol. 153, no. 1-2, pp. 45–96, 2012.

[435] T. Mori, "Representation of linearly additive random fields," *Probability theory and related fields*, vol. 92, no. 1, pp. 91–115, 1992.

[436] R.J. Adler, D. Monrad, R.H. Scissors, and R. Wilson, "Representations, decompositions and sample function continuity of random fields with independent increments," *Stochastic Processes and thei Applications*, vol. 15, no. 1, pp. 3–30, 1983.

[437] R.M. Blumenthal and R.K. Getoor, "Sample functions of stochastic processes with stationary independent increments," *Journal of Mathematics and Mechanics*, vol. 10, pp. 493–516, 1961.

[438] S. Jaffard, "The multifractal nature of Lévy processes," *Probability Theory and Related Fields*, vol. 114, no. 2, pp. 207–227, 1999.

[439] C. Chong, R.C. Dalang, and T. Humeau, "Path properties of the solution to the stochastic heat equation with Lévy noise," *Stochastics and Partial Differential Equations: Analysis and Computations*, vol. 7, no. 1, pp. 123–168, 2019.

[440] F. Kühn and R.L. Schilling, "On the domain of fractional Laplacians and related generators of Feller processes," *Journal of Functional Analysis*, vol. 276, no. 8, pp. 2397–2439, 2019.

[441] M. Rosenbaum, "First order p-variations and Besov spaces," *Statistics & Probability Letters*, vol. 79, no. 1, pp. 55–62, 2009.

[442] J. Fageot and M. Unser, "Scaling limits of solutions of linear stochastic differential equations driven by Lévy white noises," *Journal of Theoretical Probability*, vol. 32, no. 3, pp. 1166–1189, 2019.

[443] Julien Fageot, Michael Unser, and John Paul Ward, "The n-term approximation of periodic generalized Lévy processes," *Journal of Theoretical Probability*, vol. 33, pp. 180–200, 2020.

[444] W.E. Pruitt, "The growth of random walks and Lévy processes," *The Annals of Probability*, vol. 9, no. 6, pp. 948–956, 1981.

[445] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.

[446] P.A. Cioica and S. Dahlke, "Spatial Besov regularity for semilinear stochastic partial differential equations on bounded Lipschitz domains," *International Journal of Computer Mathematics*, vol. 89, no. 18, pp. 2443–2459, 2012.

[447] P.A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling, "Adaptive wavelet methods for elliptic stochastic partial differential equations," *BIT Numerical Mathematics*, vol. 52, no. 3, pp. 589–614, 2012.

[448] D.E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers, Differential Operators*, vol. 120 of *Cambridge Tracts in Mathematics*, Cambridge University Press, Cambridge, 2008.

[449] H.-J. Schmeisser and H. Triebel, *Topics in Fourier Analysis and Function Spaces*, Wiley Chichester, 1987.

[450] M. Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 989–1006, 2011.

[451] N.G. Ushakov, *Selected topics in characteristic functions*, Walter de Gruyter, 2011.

[452] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Stochastic Modeling. Chapman & Hall, New York, 1994.

[453] S. Koltz, T.J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations*, Boston, MA: Birkhauser, 2001.

[454] C. Houdré and R. Kawai, "On layered stable processes," *Bernoulli*, vol. 13, no. 1, pp. 252–278, 2007.

[455] O.E. Barndorff-Nielsen, "Processes of normal inverse Gaussian type," *Finance and Stochastics*, vol. 2, no. 1, pp. 41–68, 1997.

[456] W. Farkas, N. Jacob, and R.L. Schilling, "Function spaces related to continuous negative definite functions: ψ-Bessel potential spaces," *Dissertationes Math. (Rozprawy Mat.)*, vol. 393, pp. 1–62, 2001.

[457] J. Fageot, *Gaussian versus Sparse Stochastic Processes: Construction, Regularity, Compressibility*, EPFL thesis no. 7657 (2017), 231 p., Swiss Federal Institute of Technology Lausanne (EPFL), 2017.

[458] R. Devore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[459] G. Garrigós and E. Hernández, "Sharp Jackson and Bernstein inequalities for N-term approximation in sequence spaces with applications," *Indiana University mathematics journal*, vol. 53, no. 6, pp. 1741–1764, 2004.

[460] J.P. Ward, J. Fageot, and M. Unser, "Compressibility of symmetric-α-stable processes," in *Proceedings of the Eleventh International Workshop on Sampling Theory and Applications (SampTA'15)*, Washington DC, USA, 2015, pp. 236–240.

[461] H. Ghourchian, A. Amini, and A. Gohari, "How compressible are innovation processes?," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4843–4871, 2018.

[462] J. Fageot, A. Fallah, and T. Horel, "Entropic compressibility of Lévy processes," *arXiv preprint arXiv:2009.10753*, 2020.

[463] Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald DeVore, "Tree approximation and optimal encoding," *Applied and Computational Harmonic Analysis*, vol. 11, no. 2, pp. 192–226, 2001.

[464] Richard G Baraniuk, "Optimal tree approximation with wavelets," in *Wavelet Applications in Signal and Image Processing VII*. International Society for Optics and Photonics, 1999, vol. 3813, pp. 196–207.

[465] Peter Binev and Ronald DeVore, "Fast computation in adaptive tree approximation," *Numerische Mathematik*, vol. 97, no. 2, pp. 193–217, 2004.

[466] Richard G Baraniuk, Ronald A DeVore, George Kyriazis, and Xiang Ming Yu, "Near best tree approximation," *Advances in Computational Mathematics*, vol. 16, no. 4, pp. 357–373, 2002.

[467] Piotr Porwik and Agnieszka Lisowska, "The Haar-wavelet transform in digital image processing: Its status and achievements," *Machine Graphics and Vision*, vol. 13, no. 1/2, pp. 79–98, 2004.

[468] Radomir S Stanković and Bogdan J Falkowski, "The Haar wavelet transform: Its status and achievements," *Computers & Electrical Engineering*, vol. 29, no. 1, pp. 25–44, 2003.

[469] Xianfu Wang, “Volumes of generalized unit balls,” *Mathematics Magazine*, vol. 78, no. 5, pp. 390–395, 2005.

[470] J. Fageot, E. Bostan, and M. Unser, “Statistics of wavelet coefficients for sparse self-similar images,” in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP’14)*, Paris, French Republic, 2014, pp. 6096–6100.

[471] D.L. Donoho, “Unconditional bases are optimal bases for data compression and for statistical estimation,” *Applied and Computational Harmonic Analysis*, vol. 1, no. 1, pp. 100–115, 1993.

Curriculum Vitæ

Shayan Aziznejad

Personal Information

Address: BM 4138, EPFL, CH-1015 Lausanne

Tel: +41 (0)78 693 14 67

Language: Persian (Native), Azerbaijani (Native), English (Fluent), French (Intermediate)

E-mail: sh.aziznejad@gmail.com

Homepage: https://shayanaziznejad.github.io

Education

EPFL, Lausanne, Switzerland — August 2017 - June 2022

- Ph.D., Electrical Engineering, Biomedical Imaging Group
- Advisor: Prof. Michael Unser

Scoula Normale Superiore, Pisa, Italy — October 2021 - December 2021

- Visiting Ph.D. in Prof. Luigi Ambrosio's research group

Sharif University of Technology, Tehran, Iran — September 2012 - June 2017

- B.Sc., Electrical Engineering, Communication Systems
- B.Sc., Pure Mathematics **(Double Major Program)**

Allameh Helli High School (branch of NODET*) — September 2008 - June 2012

- Diploma in Mathematics and Physics

***National Organization for Developement of Exeptional Talents**

Honors and Awards

- SNSF* Postdoc Mobility Fellowship — December 2021

  ***Swiss National Science Foundation**

- Best Student Paper Award, ICASSP'19* — May 2019

  ***44th IEEE International Conference on Acoustics, Speech, and Signal Processing**

- Gold Medal, National Mathematical Olympiad, Iran — September 2011

Research Interests

- Nonparametric Regression
- Optimization Theory
- Functional Analysis
- Probability Theory

Publications

Preprints

[P1] **S. Aziznejad**, J. Campos, M. Unser, "Measuring Complexity of Learning Schemes Using Hessian Total-Variation," *arXiv preprint arXiv:2112.06209*, 2021.

Journal Papers

[J15] I. Lloŕens Jover, T. Debarre, **S. Aziznejad**, M. Unser, "Coupled Splines for Sparse Curve Fitting," IEEE Transactions on Image Processing , 2022.

[J14] **S. Aziznejad***, T. Debarre*, M. Unser, "Sparsest Univariate Learning Models Under Lipschitz Constraint," IEEE Open Journal of Signal Processing, 2022.

[J13] **S. Aziznejad**, J. Fageot, "Wavelet Compressiblity of Compound Poisson Processes," IEEE Transactions on Information Theory, 2022.

[J12] M. Unser, **S. Aziznejad**, ”Convex Optimization in Sums of Banach Spaces,” Applied and Computational Harmonic Analysis, 2022.

[J11] J. Campos, **S. Aziznejad**, M. Unser, ”Learning Continuous and Piecewise-Linear Functions with Hessian Total-Variation Regularization,” IEEE Open Journal of Signal Processing, 2021.

[J10] T. Debarre, **S. Aziznejad**, M. Unser, ”Continuous-Domain Formulation of Inverse Problems for Composite Sparse-Plus-Smooth Signals,” IEEE Open Journal of Signal Processing, 2021.

[J9] **S. Aziznejad**, M. Unser, ”Duality Mapping for Schatten Matrix Norms,” Numerical Functional Analysis and Optimization, 2021.

[J8] A. Goujon, **S. Aziznejad**, A. Naderi, M. Unser, ”Shortest Multi-Spline Bases for Generalized Sampling,” Journal of Computational and Applied Mathematics, 2021.

[J7] **S. Aziznejad**, M. Unser, ”Multikernel Regression with Sparsity Constraint,” SIAM Journal on Mathematics of Data Science, 2021.

[J6] **S. Aziznejad**, J. Fageot, ”Wavelet Analysis of the Besov Regularity of Lévy White Noises,” Electronic Journal of Probability, 2020.

[J5] P. Bohra, J. Campos, H. Gupta, **S. Aziznejad**, M. Unser, ”Learning Activations in Deep (Spline) Neural Networks,” IEEE Open Journal of Signal Processing, 2020.

[J4] **S. Aziznejad**, H. Gupta, J. Campos, M. Unser, ”Deep Neural Networks with Trainable Activations and Controlled Lipschitz Constant,” IEEE Transactions on Signal Processing, 2020.

[J3] L. Dadi*, **S. Aziznejad***, M. Unser, ”Generating Sparse Stochastic Processes Using Matched Splines,” IEEE Transactions on Signal Processing, 2020. *Equal contribution

[J2] J. Fageot, **S. Aziznejad**, M. Unser, V. Uhlmann, ”Support and Approximation Properties of Hermite Splines,” Journal of Computational and Applied Mathematics, 2020.

[J1] T. Debarre, **S. Aziznejad**, M. Unser, ”Hybrid-Spline Dictionaries for Continuous-Domain Inverse Problems,” IEEE Transactions on Signal Processing, 2019.

Conference Proceedings

[C6] M. Pourya, **S. Aziznejad**, M. Unser, D. Sage, ”GraPhiC: Graph-Based Hierarchical Clustering for Single-Molecule Localization Microscopy,” International Symposium on Biomedical Imaging (ISBI'21), 2021.

[C5] **S. Aziznejad**, E. Soubies, M. Unser, ”Dictionary Learning with Statistical Sparsity in the Presence of Noise,” European Signal Processing Conference (EUSIPCO'20), 2020.

[C4] **S. Aziznejad**, A. Naderi, M. Unser, ”Optimal Spline Generators for Derivative Sampling,” International Conference on Sampling Theory and Applications (SampTA'19), 2019.

[C3] **S. Aziznejad**, M. Unser, ”Deep Spline Networks with Control of Lipschitz Regularity,” Best student paper award, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'19), 2019.

[C2] G. Delfi, S. **Aziznejad**, S. Amani, M. Babaie-Zadeh, C. Jutten, ”A generalization of weighted sparse decomposition to negative weights,” IEEE European Signal Processing Conference (EUSIPCO'17), 2017.

[C1] E. Asadi, **S. Aziznejad**, M.H. Amerimehr, A. Amini, ”A fast matrix completion method for index coding,” IEEE European Signal Processing Conference (EUSIPCO'17), 2017.

Conference Abstracts

[A2] T. Debarre, **S. Aziznejad**, M. Unser, ”Sparse Dictionaries for Continuous-Domain Inverse Problems,” Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'19), 2019.

[A1] **S. Aziznejad**, M. Unser, ”Multiple-Kernel Regression with Sparsity Constraints,” Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'19), 2019.

REVIEWERS FOR JOURNALS IEEE Transactions on Signal Processing, SIAM Journal on Mathematics of Data Science, Journal of Machine Learning Research, Journal of Computational and Applied Mathematics, IEEE Signal Processing Letters, IEEE Transactions on Circuits and Systems I, Digital Signal Processing.

SUPERVISED STUDENTS

0.1 Master's Thesis (Full time, 4 months)

- Eliana Renzo — Spring 2021
  Title: Scouting for Clusters in SMLM Images
- Joaquim Campos — Fall 2019
  Title: Higher-Order Regularization Methods for Supervised Learning
- Leello Dadi — Spring 2019
  Title: Generating Sparse Stochastic Processes Using Matched Splines

0.2 Summer Internship (Full time, 8-12 weeks)

- Mehrsa Pouria (with D. Sage) — 2020
  Title: Graph-Based Hierarchical Clustering of SMLM Data
- Tina Behnia (with P. Bohra) — 2020
  Title: Dictionary Learning with $S\alpha S$ Signal and Noise Prior
- Alireza Naderi — 2018
  Title: Shortest Support Generators of Sum of Spline Spaces

0.3 Master Semester Project (16-24 hours per week, 4 months)

- Haojun Zhu (with A. Goujon) — 2021
  Title: Effect of Simple Operations on the Linear Regions of CPWL Functions
- Mickael Gindroz (with T. Pham) — 2021
  Title: Differentiable Approximation of Hessian-Schatten Regularization
- Benoit Knuchel (with T. Debarre) — 2021
  Title: Continuous-Domain Multicomponent Image Reconstruction
- Moulik Choraria (with J. Yoo) — 2020
  Title: Learning Robust Neural Networks via Controlling their Lipschitz Regularity
- Arnaud Panatier (with A. Badoual) — 2018
  Title: Deep Neural Networks: Learning with Splines

TEACHING

TA at EPFL

- Image Processing II — Spring 2018,2019
- Image Processing I — Fall 2017, 2018, 2020
- Signals and Systems II — Spring 2020 , 2021
- Signals and Systems I — Fall 2019

TA at Sharif University

- Game Theory — Spring 2016
- Numerical Optimization — Fall 2015
- Probability and Its Applications — Fall 2015
- Engineering Probability and Statistics — Fall 2014, 2015
- Foundations of Mathematics — Spring 2015

Other

- Math Olympiad Instructor — 2012-2015