Master Internship Report

Spline-based Reconstruction of Periodic Signals with Sparse Innovations

Author: Adrian Jarret

Supervisors: Dr. Matthieu Simeoni
Dr. Julien Fageot

Internship taking place in the Laboratoire des Communications AudioVisuelles
at Ecole Polytechnique Fédérale de Lausanne

November 27, 2020
Spline-based Reconstruction of Periodic Signals with Sparse Innovations

by Adrian Jarret

This report summarizes the research work achieved during this master internship. Optimization-based problems have become of great interest for signal approximation purposes, as they achieved good accuracy results while being extremely flexible and versatile. In this work, we put our focus on the context of periodic signals sampled with spatial measurements. The optimization problems are penalized thanks to the total variation norm, using a specific class of (pseudo-)differential operators to use well-chosen reconstruction functions. We introduce three algorithms and their adaptation to this specific use case. The first one is a discrete grid-based $\ell_1$ method, the second is called CPGD, which relies on the estimation of discrete innovations within the FRI framework, and the last one is the Frank-Wolfe algorithm. We put the emphasis on that later algorithm as we underline its greedy behavior. We consider a refined version of this algorithm, that leads to very encouraging outputs.

We pursue the algorithmic treatment with a theoretical analysis of the optimization problem, based on convexity and duality theory. We draw away two conditions that, if verified by the differential operator, would inform on the shape of the solutions and even ensure uniqueness, in some cases. Moreover, we demonstrate the general expression of periodic one-dimensional exponential splines, as a first step to verify the latter conditions for exponential operators.
Introduction

Signal reconstruction relying on optimization problems has been of great interest in the field of signal processing and compressed sensing lately. Beyond obtaining accurate experimental results, the underlying optimization part features both interesting theoretical aspects and a wide range of varying parameters. Inspired from domains as broad as radioastronomy, environmental data monitoring or acoustic sensing, this project focuses on the specific context of periodic signal reconstruction with non-uniform spatial sampling.

Historically, such problems have been treated by way of discretization strategies, leading to powerful frameworks and approximations that are still used nowadays. One of the main singularities of this work is that it tries to tackle the signal reconstruction directly in a continuous setting. It actually corresponds to the most natural way of formulating the signal approximation problems of interest. It induces more flexibility in the reconstruction procedure, however it also requires a more conscious and careful statement of the problem and of the involved solutions.

In such a context, penalization is necessary for "well-posedness" of the problem, but more than that it is also crucial to orient the solution. In such a context, the "well-posedness" of the problem is usually obtained through penalization, as it has proven to perform really well over the numerous use cases. Past the practical advantages, penalization is also crucial here to orient the solutions. The design of the regularization component is very subtle and completely transforms the shape of the reconstructed signals. In this project, we exclusively study total variation penalization, which is the continuous counterpart of the well-known $\ell_1$ norm (also known as LASSO regression or basis pursuit depending on the application domain). It happens that this framework is efficient to promote sparse solutions, and thus is extensively used to this end. Beyond the fact that sparsity is usually a property of great interest in approximation (according to Occam’s razor principle), in our context the use of the total variation norm also allows to decorrelate the measurement procedure and the general form of the solutions. This property is extremely helpful with our irregular spatial measurements, as the recovered solutions become independent (in some sense) from the sampling strategy. Additionally, we make use of (pseudo-)differential operators in the penalization framework. These operators basically model the shape of the output by controlling the smoothness and the spatial extension of any solution.

The periodicity and the spatial sampling are directly inspired from actual data sensing, as any directional signal can be seen as a spatially periodic signal around the measurement device. The theoretic analysis of such periodic functions and operators is thoroughly achieved thanks to the Fourier analysis, and happens to be simpler than the non-periodic case. Indeed, it nearly reduces to the analysis of compactly-supported functions and the behaviors at infinity are no longer a trouble. The periodic context has mostly gained attention recently, and now belongs to the modern fields of study.

In a few words, this work focuses on the reconstruction task of an input signal through the prism of optimization-based methods. The problems are expressed in a continuous periodic setting and the data is obtained through arbitrary spatial sampling. The core of this framework, which enables its adaptability and precision, lies in
the penalization strategy, thanks to the total variation norm coupled to specific well-chosen operators. The idea of this document is to give the broadest understanding of the reconstruction framework.

The contributions of this work are of two kinds:

- We provide the implementation of the three studied strategies: equispaced grid, CPGD, and Frank-Wolfe, within a unified algorithmic framework. They cover the context of periodic one-dimensional reconstruction, with a wide range of penalization operators. The framework allows to easily generate synthetic samples, run and compare the methods with any set of custom reconstruction parameters. As far as the Frank-Wolfe strategy is concerned, this project has been through deep analysis of its greedy behavior and studies a refined version, namely the complete reweighting, about which very few literature is available.

- For the theoretical aspects, this project has taken great care of carefully explaining the dual analysis of the generalized continuous basis pursuit problem. We have obtained a generic approach to prove that, depending on the involved operator, the solutions are splines and might be unique. Our work paves the way for a systematic study of the (non-)uniqueness of TV-based optimization for spatial sampling and pseudo-differential regularization operators. We start to investigate the case of exponential operators, and successfully prove that the corresponding dual certificates always admit a finite number of saturation points. This result relies the general expression of the periodic one-dimensional exponential splines that we demonstrate.

In Chapter 1, we present the mathematical framework to properly establish the optimization problems. In Chapter 2, we go over three different algorithmic strategies to actually solve the aforementioned optimization problems. We precisely describe their mechanics and the adaptations to our context. Chapter 3 illustrates these algorithms with practical experiments and use cases. Finally, Chapter 4 is part of an ongoing theoretical effort to deduce properties on the solution set of the studied problems. We strongly believe that it is possible to characterize the shape of the solutions as well as to prove uniqueness in some cases, as experiments let it appear.

Related Works

**Discrete ℓ₁ methods** Inverse problems have been expressed and treated for a long time in a discrete setting, involving either vector variable as well as infinite dimensional variables. The use of ℓ₁-norm penalization was initially justified for sparsity promoting, in comparison with the frequently used Tikhonov regularization (ℓ₂-norm). The statistics community refers to the obtained problem as LASSO, for least absolute shrinkage and selection operator [1], meanwhile it is commonly referred to under the name of basis pursuit (BP) within the signal processing world [2]. These problems and their resolution have given birth to the field of compressed sensing. More recently, theoretical studies have been proposed [3, 4] to address the question of uniqueness, in the case of regular ℓ₁ penalization as well as when a penalization operator is used (generalized LASSO). Fast and accurate algorithmic strategies, relying on proximal methods, have been proposed in the last decade to numerically approximate solutions (primal dual splitting algorithm, FISTA, ... ) [5, 6]. Although recent, they have turned the ℓ₁ penalized problems into an extensively used tool.
Optimization over measure spaces  When aiming at approximating real life signals, one might express optimization problems in the continuous setting, with data living in a infinite dimensional space (spaces of functions, for instance). It allows more flexibility in the reconstruction, compared to the finite dimensional case, as the reconstructed data are no longer constrained to live within the available dimensions of the problem. Using continuous data might also lead to more natural expression of the optimization problems, with straightforward objective cost functionals. It often amounts to recover data whose shape is a stream of Dirac $w = \sum \alpha_k \delta_{x_k}$. It enters the framework of so-called functions with *finite rate of innovations* (FRI), which is the continuous counterpart of the discrete notion of sparsity. Some reconstruction algorithms propose to leverage this sparsity property to directly estimate the finite number of reconstruction positions with help of Prony’s methods [7, 8]. In functional spaces, the generalization of $\ell_1$-norm is the *total variation* norm, which is commonly used to penalize such optimization problems, as in [9]. It is used for sparse spikes deconvolution or super-resolution problems, as in [10, 11]. There are sometimes referred to as *continuous basis pursuit* [12, 13] or *generalized minimal extrapolation* [14]. These problems can numerically be solved or approximated by means of versions of the *matching pursuit* algorithm (that generalizes in a continuous setting) [15] or with the *conditional gradient method* [9, 16].

Spline-based approximation  Piecewise constant functions are the most well-known splines, but it is actually possible to define splines more generally. A spline is essentially a piecewise-defined function, whose parts are shifted elements from the null space of a (pseudo)-differential operator. The piecewise constant spline is obtained with the derivative operator, and using other operators (with satisfying properties) would lead to other type of spline. The operator acts as a generalization of the derivative and allows to hand-tailor the obtained spline. It has been proven that introducing such a generalized derivative operator into the total variation penalization term of an optimization problem conduces to solutions expressed in terms of splines [17]. More interestingly, the reconstruction splines are the splines issued from the operator used in the regularization term, independently of the sampling procedure. This results certifies that it is very consistent to try to approximate any input signal with a sum of splines. Fundamentally, this new penalization paradigm substitutes splines to the previously used Dirac impulses in the reconstruction strategy. It turns out to be really convenient when the data does not fit well with Dirac reconstruction, in the case of spatial sampling for instance. An interesting case over the real line is studied in [18] and it has been generalized to the periodic setting (over the multidimensional torus) [19] or over the sphere [20].
Chapter 1

Problem statement

An extensive presentation of the mathematical context for periodic functional recovery with total variation penalization has been achieved by Fageot and Simeoni in their paper on periodic splines [19]. In this section, we summarize the key points of the mathematical framework, and we refer to their work for a more exhaustive treatment.

1.1 Search Space and Splines

1.1.1 A few words of topology

We denote by $S(T)$ the space of infinitely differentiable $T$-periodic functions over $\mathbb{R}$. $S'(T)$ is the topological dual of $S(T)$, which is the space of periodic generalized functions. The duality product between $S(T)$ and $S'(T)$ is usually noted as:

$$\langle \cdot, \cdot \rangle : S'(T) \times S(T) \to \mathbb{R}, \quad (w, f) \mapsto \int_T f \, dw.$$  \hspace{1cm} (1.1)

The space of continuous periodic functions over $T$ is denoted as $C(T)$. We define the supremum of infinite norm over $C(T)$ as $\|\varphi\|_{\infty} = \sup_{t \in T} |\varphi(t)|$. Endowed with this norm, $(C(T), \|\cdot\|_{\infty})$ forms a Banach space whose dual space is isomorphic to the space of periodic finite Radon measures $M(T)$. Thus $M(T)$ is a Banach space when it is endowed with the dual norm (also coined as total variation norm):

$$\forall w \in M(T), \quad \|w\|_{TV} = \sup_{\varphi \in C(T), \|\varphi\|_{\infty} \leq 1} \langle w, \varphi \rangle$$ \hspace{1cm} (1.2)

with $\langle w, \varphi \rangle = \int_T \varphi \, dw$. As $S(T)$ is dense in $C(T)$ with respect to the supremum norm $\|\cdot\|_{\infty}$, it is possible to extend the total variation norm to $S'(T)$ and thus we obtain

$$M(T) = \left\{ w \in S'(T), \quad \sup_{\varphi \in S(T), \|\varphi\|_{\infty} \leq 1} \langle w, \varphi \rangle < \infty \right\}.$$ \hspace{1cm} (1.3)

The dual space of $M(T)$ with respect to the strong topology is complex and strictly larger than $C(T)$. However, when $M(T)$ is endowed with its weak-* topology, its dual space is actually $C(T)$. Then, when $C(T)$ is respectively endowed with its weak topology, they play symmetrical role and they are both the dual space of each other. In the following, we will keep with this "reflexive" topology, that is more convenient than the one induced by the TV-norm.

It is interesting to point out that the TV-norm is actually a generalization of the $L^1$-norm for integrable functions. Indeed, for any $f \in L^1(T) \cap S(T)$, the two norms coincide : $\|f\|_1 = \int_T |f(t)| \, dt = \|f\|_{TV}$. Moreover, let us introduce the $T$-periodic
Chapter 1. Problem statement

Dirac impulse \( \text{III} = \sum_{i \in \mathbb{Z}} \delta(-iT) \), coined as Dirac comb. The two following relations hold

\[ \| \text{III} \|_{TV} = 1 \]
\[ \left\| \sum_{k=1}^{K} \alpha_k \text{III}(\cdot - t_k) \right\|_{TV} = \| \alpha \|_1 \]

for \( K \in \mathbb{N}^* \) and any \( \alpha \in \mathbb{R}^K \), \((t_1, \ldots, t_K) \in \mathbb{T}^K \).

1.1.2 Pseudo-Differential Operators

We consider the linear, shift-invariant operators that are continuous from \( S'(\mathbb{T}) \) to itself denoted as \( \mathcal{L}_{SI}(S'(\mathbb{T})) \). It is possible to leverage the periodic property of the objects to perform an extensive study of these operators, as it is done in [21].

The same way as any generalized periodic function can be uniquely decomposed over the Fourier basis \( e_k : x \mapsto e^{i k x} \in S(\mathbb{T}) \) for \( k \in \mathbb{Z} \), one may uniquely decompose any operator \( D \in \mathcal{L}_{SI}(S'(\mathbb{T})) \) over the same Fourier basis. That way, any operator is equivalently characterized by its Fourier sequence. This allows a theoretical expression of the operators as well as explicit numerical approximations of such operators.

In this project, we only consider invertible operators, such that they have trivial null space and admit an inverse operator. We can cite the exponential operator \( D = (D + \alpha \text{Id})^\gamma \) or the Sobolev operator \( D = (\alpha^2 \text{Id} - D^2)^{\gamma/2} \) both of order \( \gamma > 1 \). Any invertible operator \( D \in \mathcal{L}_{SI}(S'(\mathbb{T})) \) belongs to the so-called spline-admissible operators [19, Definition 2], and thus we define

**Definition 1** (Periodic \( D \)-spline). Any function \( f \in S'(\mathbb{T}) \) is said to be a \( D \)-spline if there exists \( K \geq 0 \) weights \((\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \) and \( K \) positions \((t_1, \ldots, t_K) \in \mathbb{T}^K \) such that

\[ D f = \sum_{k=1}^{K} \alpha_k \text{III}(\cdot - t_k). \] (1.4)

The pairs \((\alpha_k, t_k)\) are called the innovations of the spline while the \((t_k)\) are the knots.

As we work with invertible operators, any \( D \)-spline \( f \) can be rewritten as

\[ f = D^{-1} \left\{ \sum_{k=1}^{K} \alpha_k \text{III}(\cdot - t_k) \right\} = \sum_{k=1}^{K} \alpha_k \psi_D(\cdot - t_k) \] (1.5)

where \( \psi_D = D^{-1}\{\text{III}\} \) is called the Green’s function of the operator \( D \), defined as follows:

**Definition 2** (Green’s function). Let \( D \) be a spline admissible operator. Any function \( \psi \in S'(\mathbb{T}) \) such that \( D \psi = \text{III} \) is called a Green’s function of \( D \). In the case of an invertible operator \( D \), one obtains the expression \( \psi_D = D^{-1}\{\text{III}\} \).

Some examples of Green’s functions with exponential and Sobolev operators are given in figures 1.1 and 1.2. The reader should refer to [19, Section 5] for more classes of spline admissible operators.
Chapter 1. Problem statement

1.2 Spatial Sampling for Recovery Problems

This project aims at performing signal approximation from (possibly noisy) measurements of the signal. The measurements we are concerned about are direct measurements, which is also known as spatial sampling.

Let $f_0$ be a $T$-periodic real-valued function $f_0 : T \rightarrow \mathbb{R}$. Let $y_0 = (y_1, \ldots, y_L)$ a $L$-tuple of (potentially noisy) samples of $f_0$ such that

$$y_0 = \Phi(f_0) + w = (f_0(\theta_1) + w_1, \ldots, f_0(\theta_L) + w_L)$$

for $(\theta_1, \ldots, \theta_L)$ in $[0, T]$ the spatial samples and $w$ a measurement noise (a Gaussian white noise for instance, in our context). We aim at finding a function $f \in S'(T)$ that minimizes $E(y_0, \Phi(f))$ with $E$ a convex, proper, lower semi-continuous cost functional (quadratic cost $E(x, y) = \|x - y\|^2$ for instance). As $S'(T)$ is infinite dimensional, this problem needs a regularization. Indeed, without penalization, one could always consider adding an element of the null space of the measurement operator to any solution to create a new solution with the exact same functional cost. A solution commonly proposed is to use the TV norm in combination with a (pseudo-)differential operator [18, 20]. The TV norm is known to promote sparsity in the solution. With the differential operator, the sparsity is to be searched in the generalized derivative of the solution instead of in the solution itself.

For any $\lambda > 0$ and $D$ a periodic linear sample-admissible operator with trivial null space, we will focus on the following problem:

$$\text{Find : } f^* \in \arg \min_{f \in \mathcal{M}(T)} \|y_0 - \Phi(f)\|_2^2 + \lambda \|Df\|_{TV} \quad (1.6)$$

The parameter $\lambda$ moderates the relative significance of both these terms.
Chapter 1. Problem statement

However, written as (1.6), two details need to be carefully addressed. First, \( \|Df\|_{TV} \) might not be finite for all \( f \) in \( \mathcal{M}(T) \). It is actually possible to show that the maximal search space is

\[
\mathcal{M}_D(T) = \{ f \in \mathcal{S}'(T), \|Df\|_{TV} < \infty \}
\] (1.7)

which reduces the search space for solutions. Secondly, the operator \( D \) needs to be sampling-admissible, in the sense defined by [19, Section 6.2]. It amounts for the Green’s function \( \psi_D = D^{-1} \Pi \in \mathcal{S}'(T) \) of the operator to be continuous (which is not true for any operator). The authors of [19] present some sufficient conditions for sample-admissibility, in particular focusing on the spectral growth of the operator (when it exists). Sampling-admissibility is necessary in our context as it allows working with spatial measurements. Note that it is possible to use less restrictive operators with other forms of measurements (Fourier sampling for instance).

The well-posed problem with a sampling-admissible operator \( D \) becomes

\[
\min_{f \in \mathcal{M}_D(T)} \|y_0 - \Phi(f)\|_2^2 + \lambda \|Df\|_{TV}.
\] (1.8)

This problem is referred to under many different names and none of them tend to make a consensus for the moment. We will retain generalized Beurling LASSO (B-LASSO), but one might encounter functional penalized basis pursuit, generalised continuous basis pursuit [12] or even sparse spikes deconvolution [10].

As the linear measurements \( f \mapsto \Phi_{t}(f) = f(\theta_t) = \int_T f(t) \, d\Pi(t - \theta_t) \) are linearly independent, we know from representer theorem [19, Theorem 4] that the solution set of (1.8) is a non empty convex set spanned by the extreme points of the form:

\[
f_{ext} = \sum_{k=1}^{K} \beta_k \psi_D(\cdot - t_k) = D^{-1} \left( \sum_{k=1}^{K} \beta_k \Pi(\cdot - t_k) \right)
\] (1.9)

with a positive integer \( K \leq L \) (the number of samples) and distinct \( t_k \in T \) called knots of the spline.

Given this expression, one can make the link with the definition of sampling admissibility. Indeed, to perform spatial sampling, it is necessary for the sampled function to be continuous, for consistency (equality between the right hand side and left hand side limits at each point). Thus, as \( \psi_D \) is continuous for a sampling admissible operator, such extreme points solutions (1.9) can actually be sampled.
Chapter 2

Algorithmic treatment

The generalized Beurling LASSO problem is now well defined and we know that its solution set is non empty and convex. We also know that the extreme points of this set are splines with sparse innovations. This knowledge comes really handful in a computational approximation framework, as the reconstruction task turns into estimating the knots and the intensities of those innovations. Different algorithmic strategies are presented in the following chapter, based on this consideration. They rely on known algorithms but our particular context, specifically the use of differential operators and the periodic setting, has required subtle adaptation work. This is the main part of the contribution to the project during this internship. In everything that follows, the period is set to $T = 2\pi$. All the programming is based on Python programming language and the splines are simulated thanks to the package Periodispline [22] available on GitHub.

2.1 Equispaced Grid Approximation

In all areas of approximation, the always increasing computing power of modern devices keeps the discretization frameworks both computationally possible but also remarkably accurate. The first step in our research is logically to bring these methods and our problem face to face. The periodicity is rather convenient here as the space to discretize is known and bounded to the period.

2.1.1 Reconstruction with fixed knots

Let us first present a closed problem that will help understand the discretization scheme and be used again later on. Let suppose we know a set of $I > 0$ reconstruction knots $(k_1, \ldots, k_I) \in \mathbb{T}^I$. We recall that the extreme points of the solution set are expressed as (1.9), such that any solution $f$ here would be

$$\forall t \in \mathbb{T}, \quad f(t) = \sum_{i=1}^{I} \beta_i \psi_D(t - k_i)$$

with unknown weights $\beta \in \mathbb{R}^I$.

We want to plug such a solution into the problem (1.8). Remembering that the Green's function satisfies $D\psi_D = \Pi$, the regularization term $\|Df\|_{TV}$ becomes

$$\|Df\|_{TV} = \left\| \sum_{i=1}^{I} \beta_i D\psi_D(\cdot - k_i) \right\|_{TV} = \left\| \sum_{i=1}^{I} \beta_i \Pi(\cdot - k_i) \right\|_{TV} = \|\beta\|_1.$$
The sampling operator evaluation turns into

\[ \Phi(f) = \begin{pmatrix} f(\theta_1) \\ \vdots \\ f(\theta_L) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{K} \beta_k \psi_D(\theta_1 - t_k) \\ \vdots \\ \sum_{k=1}^{K} \beta_k \psi_D(\theta_L - t_k) \end{pmatrix} = H\beta \]

with \( H \in \mathbb{R}^{L \times I} \) the matrix whose entries are

\[ H[\ell, i] = \psi_D(\theta_\ell - k_i). \]  

(2.1)

Then, given a set of reconstruction knots \((k_i)_{i=1,\ldots,I}\), the B-LASSO problem is transformed into a regular LASSO problem whose unknown is the vector of weights \( \beta \in \mathbb{R}^I \). We obtain the following convex problem:

\[ \min_{\beta \in \mathbb{R}^I} \|y - H\beta\|_2^2 + \lambda \|\beta\|_1 \]  

(2.2)

This problem can be solved quite efficiently nowadays thanks to recent proximal algorithms, even for large values of \( I \). We use the accelerated proximal gradient descent algorithm, also known as FISTA [6], for computational treatment.

### 2.1.2 Discretization of the period

We have seen that when the reconstruction knots are known, it is straightforward to determine the weights and thus solve the problem. However, there is no prior for the knots to be known beforehand. Let us remind that we are looking for solution in the shape of (1.9). One shall notice that it is equivalent to consider a knot associated to a null weight or no knot at all. Based on this remark, an idea is to discretize the whole interval \( T = [0, 2\pi] \) with a large number \( N \) of equispaced knots \( \hat{t}_n = (n - 1)T/N \) for \( n = 1, \ldots, N \) and to search for the \( N \) associated weights \( \beta_n \). The final number of knots will be the number of non zero coefficients of the weight vector \( \beta \). This method has proven to be efficient in the non periodic setting, even with non invertible differential operator [23, Section V.B].

Let suppose the source signal is a function \( f_0 = D^{-1}m_0 \) with \( m_0 \) being a sparse signal with \( K \) knots. Sparse signal refers to a stream of Dirac defined as linear combination of periodic Dirac impulse, such that \( m_0 = \sum_{k=1}^{K} \alpha_k \Pi(-t_k) \). We want \( N \) to be significantly larger than \( K \) to make sure that the grid is fine enough and consequently is able to precisely approximate the solution. We obtain the same problem as (2.2) with a matrix \( H \in \mathbb{R}^{L \times N} \) where \( N \gg K \). The explicit expression of this matrix is:

\[ H = \begin{pmatrix} \psi_D(\theta_1 - \hat{t}_1) & \cdots & \psi_D(\theta_1 - \hat{t}_N) \\ \vdots & \ddots & \vdots \\ \psi_D(\theta_L - \hat{t}_1) & \cdots & \psi_D(\theta_L - \hat{t}_N) \end{pmatrix}. \]  

(2.3)

Let \( y_0 = \Phi(f_0) \), the problem of interest now becomes:

\[ \min_{\beta \in \mathbb{R}^N} \|y_0 - H\beta\|_2^2 + \lambda \|\beta\|_1 \]  

(2.4)

with a vector \( \beta \) of size \( N \). It completely fits with the application setting of the Accelerated Proximal Gradient Descent algorithm (APGD), and more precisely to the special case of FISTA (see [20, Algorithm 7.6]).
**APGD**

Consider the problem

\[
\text{Find: } \mathbf{u}^* \in \arg \min_{\mathbf{u} \in \mathbb{R}^N} F(\mathbf{u}) + G(\mathbf{u}) \tag{2.5}
\]

with \( F : \mathbb{R}^N \to \mathbb{R} \) a convex and differentiable function with \( \gamma \)-Lipschitz continuous gradient and \( G : \mathbb{R}^N \to \mathbb{R} \) a proper, convex and lower semi-continuous function with an easy to compute proximal operator. The difficulty come from the fact that \( G \) is not differentiable. The APGD algorithm works as follows:

**Algorithm 1** Accelerated Proximal Gradient Descent \((\beta_0, \tau, \delta, \varepsilon)\)

1: Initialization : \((\mathbf{z}_0, \mathbf{x}_0) \leftarrow (\beta_0, \beta_0)\)

2: for \( n \geq 1 \) do

3: \( \mathbf{z}_n = \text{prox}_{\tau G}(\mathbf{x}_{n-1} - \tau \nabla F(\mathbf{x}_n)) \)

4: \( \mathbf{x}_n = \mathbf{z}_n + \frac{n-1}{n+\delta} (\mathbf{z}_n - \mathbf{z}_{n-1}) \)

5: if \( \|\mathbf{x}_n - \mathbf{x}_{n-1}\|_2 \leq \varepsilon \|\mathbf{x}_{n-1}\|_2 \) then

6: \( \mathbf{x}_n \) solution of the problem. Stop.

7: else

8: \( n \leftarrow n + 1 \)

9: end if

10: end for

11: return \((\mathbf{x}_n)_{n \in \mathbb{N}}\)

In our case, \( F \) is the data fidelity term and \( G \) is the regularization term. Let us get into details about the steps and parameters involved.

**Initialization**

The initialization step is straightforward. If the value of the first iterate is not specified, it is set to \( \beta_0 = 0_N \). The two iterable variables \( \mathbf{x}_n \) and \( \mathbf{z}_n \) are then initialized with value \( \beta_0 \). During our experiments, we have used \( 0_N \) as initial iterate.

**Proximal step**

The APGD algorithm makes use of the proximal operator associated to the 1-norm. It actually admits a closed-form expression and the following equality holds [20]:

\[
\forall \alpha \in \mathbb{R}, \forall \mathbf{x} \in \mathbb{R}^N, \quad \text{prox}_{\alpha\|\cdot\|_1}(\mathbf{x}) = \arg \min_{\mathbf{z} \in \mathbb{R}^N} \frac{1}{2\alpha} \|\mathbf{x} - \mathbf{z}\|_2^2 + \|\mathbf{z}\|_1 = (\text{soft}_\alpha(x_i))_{i=1,...,N} \tag{2.6}
\]

with \( \text{soft}_\alpha \) being the soft thresholding function

\[
\forall \alpha \in \mathbb{R}, \forall x \in \mathbb{R}, \quad \text{soft}_\alpha(x) = \max(|x| - \alpha, 0) \text{ sgn}(x). \tag{2.7}
\]

The function \( \text{sgn} \) is the *sign function* such that \( x \cdot \text{sgn}(x) > 0 \) for \( x \in \mathbb{R}^* \), \( \text{sgn}(0) = 0 \). These simple expressions lead to an easy computation of the proximal step.
Step size

The data fidelity term $\mathcal{F} : \beta \rightarrow \| y_0 - H\beta \|_2^2$ is a smooth function with $\gamma$-Lipschitz continuous derivative. We can easily show that

$$\gamma = 2 \| H^T H \|_2 = 2 \| H \|_2^2 = 2 \lambda_{\text{max}}(H^T H)$$

where $\lambda_{\text{max}}(M)$ is the maximum eigenvalue of the matrix $M$. Then, the step size $\tau$ is chosen to maximize the convergence speed according to $\tau \leq \frac{1}{\gamma}$. In our case, we chose the greatest value possible, so that $\tau = \frac{1}{7}$.

Iteration

Then, we obtain the following expression for the iteration step:

$$\begin{cases}
x_{n+1} \leftarrow \text{soft}_{\tau \lambda}(x_n - 2\tau H^T (Hx_n - y)) \\
z_{n+1} \leftarrow z_{n+1} + \frac{n-1}{n+\delta} (z_{n+1} - z_n)
\end{cases} \quad (2.8)$$

The parameter $\delta$ is an empirical acceleration parameter, whose value is usually taken around $\delta = 75$. This choice is motivated by the good results obtained in [24], which highlights significant acceleration for $\delta \in [50, 100]$.

Stopping criterion

When the norm of the difference between two iterates of $x_n$ becomes small enough compared to the norm of the previous iterate, one considers that the process has converged and can stop. This is called the relative improvement, ruled by a parameter $\varepsilon$ that is usually set quite low (for instance we use $\varepsilon \approx 10^{-4}$).

Remark 1 (Value of $\lambda$). In such a problem, the regularization parameter $\lambda$ is at the heart of sparse recovery. Theoretically, as long as $\lambda > 0$, the problem is well-posed, and its value is not considered any further. However, when running actual code, setting the value of $\lambda$ is crucial to exploit sparsity promoting algorithms to their best potential. Practically, $\lambda$ (or more precisely a $\lambda$ related quantity) answers to question: “To what extent am I ready to pay (in sparsity) to allow more accurate approximation?” Low values mean sparsity is cheap, whereas high values lead to extremely sparse solution, up to a critical point where even a unique parameter solution costs too much and the only affordable trade-off is a null approximation.

In the current case of equations (2.8)) starting with an initial value $x_0 = 0$, one obtains:

$$z_1 = \text{soft}_{\tau \lambda}(2\tau H^T y)$$

which leads to 0 when

$$\lambda \geq 2 \max_{i=1,\ldots,N} |H^T y|_i = 2 \| H^T y \|_\infty.$$ 

By linear combinations, the next iterates would then be stuck at 0. Even with another non null initial value $x_0$, we notice the same collapsing behavior of the iterates towards 0 after a certain number of iterations.

Consequently, it seems interesting to scale the value of $\lambda$ to $\| H^T y \|_\infty$, which is a practical measurement that depends both on the sampling operator (that includes sampling positions and reconstruction spline) and the sampled values. More precisely,
\( \lambda \) should verify:

\[
\lambda \in [0, \lambda_{\text{max}}], \quad \text{with} \quad \lambda_{\text{max}} = 2 \| H^T y \|_\infty
\] (2.9)

We usually set \( \lambda = \sigma \| H^T y \|_\infty \) with \( \sigma \in \mathbb{R} \) close to 0.1.

Remark 2 (Sparsity index). The solution obtained following this procedure does not necessarily meet the sparsity index guaranteed by the representer theorem, i.e. sparsity index of the solution lower than \( L \). [23, Theorem 7] ensures that one can fulfill this sparsity constraint using the simplex algorithm over the linear problem

\[
\min_{\beta \in \mathbb{R}^N} \| \beta \|_1 \quad \text{subject to} \quad H \beta = H \beta^*
\] (2.10)

with \( \beta^* \) solution of the FISTA algorithm. It is interesting to point out, however we will not focus on this point any further, as we are more interested in the quality on the approximation than the sparsity of the solutions.

### 2.2 Finite Rate of Innovation

Another relevant type of algorithms for the approximation of problem (1.8) are referred to as finite rate of innovation algorithms (FRI) [7]. In the definition, the FRI is a way of measuring the quantity of information contained per time interval for a signal. For instance, to approximate a signal such as (1.9), it is necessary to estimate the \( t_k \) and \( \beta_k \) values, which leads to a FRI of \( \rho = \frac{2}{T} \) with \( T = 2\pi \) the period of the signal. Practically, for a given periodic signal, it is necessary to perform at least the FRI number \( \gamma \) of observations per unit of time during a period to collect enough information and reconstruct the signal.

Algorithms coined as FRI decouple the problem of reconstructing a spline into two steps. The first step consists in estimating the knots of the spline, the second one in estimating the weights given the knots position. As we have seen, this second step can efficiently be carried out thanks to FISTA as it amounts to the reconstruction problem with fixed knots (2.2). Let us then focus to the estimation of the positions of the knots. We apply a refined version of FRI presented in [8] called CPGD for Cadzow Plug and Play Gradient Descent, that have been specifically designed to reduce the noise dependency of the original FRI methods.

The FRI framework is based on the Fourier analysis of the input signal. In order to retrieve the positions of the knots, the algorithms first estimate the Fourier series coefficients of the input signal (up to some order). We first present how spatial samples of the signal can lead to these coefficients, then we analyze how the knowledge of the Fourier series allow the computation of the knots.

#### 2.2.1 Linking Fourier coefficients to the samples

In order to comply with the formulation of the problem solved by CPGD algorithm, we need to link the spatial samples \( y \) to the truncated Fourier series of a Dirac stream linked to the input signal. Let us reformulate our problem so that this requirement is fulfilled.

Let suppose the source signal \( f_0 \) is a \( K_0 \) sparse signal such that

\[
y_0 = \Phi(f_0) = \Phi(D^{-1} m_0) \quad \text{and} \quad m_0 = \sum_{k=1}^{K_0} \beta_{0,k} \Theta(\cdot - t_{0,k})
\] (2.11)
that we want to approximate with a $K$ sparse signal
\[
f = \sum_{k=1}^{K} \beta_k \psi_D(\cdot - t_k) = D^{-1} \left( \sum_{k=1}^{K} \beta_k \psi_D(\cdot - t_k) \right) = D^{-1} m
\] (2.12)

with the corresponding stream of periodic Dirac $m$. It is possible to compute the infinite Fourier transform series of $m$, noted $\hat{m} = \mathcal{F}(m)$, as:
\[
\hat{m}_n = (\mathcal{F}(m))_n = \sum_{k=1}^{K} \beta_k \exp(-j2\pi nt_k), \quad n \in \mathbb{Z}.
\] (2.13)

Computationally, we will consider the truncated Fourier series $\hat{m}$ up to coefficient $M \in \mathbb{N}$ as an approximation of $\mathcal{F}(m)$:
\[
\hat{m} = \mathcal{P}_M \mathcal{F}(m) \in \mathbb{C}^{2M+1}
\] (2.14)

with:
\[
\mathcal{P}_M : \mathbb{C}^Z \rightarrow \mathbb{C}^{2M+1}, \quad (a_n)_{n \in \mathbb{Z}} \mapsto (a_{-M}, \ldots, a_M)
\] (2.15)

With the Fourier basis noted as $e_n(t) = \exp(j2\pi nt/T)$, the approximated reverse operation of building back the signal $m$ using its truncated Fourier transform series becomes
\[
m \approx \mathcal{F}_M^\dagger(\hat{m}) := \sum_{n=-M}^{M} \hat{m}_n e_n.
\] (2.16)

Naturally, the greater $M$ the better the approximation holds. Then, one can approximate the $y$ noiseless samples of the signal as:
\[
y = \Phi(f) = \Phi D^{-1}(m) \approx \Phi D^{-1} \mathcal{F}_M^\dagger(\hat{m}).
\] (2.17)

For a given $M \in \mathbb{N}$, calculations lead to
\[
\Phi D^{-1} \mathcal{F}_M^\dagger(\hat{m}) = \Phi D^{-1} \left( \sum_{n=-M}^{M} \hat{m}_n e_n \right)
= \Phi \left( \sum_{n=-M}^{M} \hat{m}_n \tilde{D}^{-1} [n] e_n \right)
= \Phi \left( \sum_{n=-M}^{M} \hat{m}_n \hat{\psi}_D [n] e_n \right)
= \left( \sum_{n=-M}^{M} \hat{m}_n \hat{\psi}_D [n] e_n(\theta_l) \right)_{l=1, \ldots, L}
\]
using the fact that the Fourier sequence of the Green’s function $\hat{\psi}_D$ of a spline admissible invertible operator $\mathcal{D}$ coincides with the Fourier sequence of $\mathcal{D}^{-1}$. Eventually, the following approximation holds:
\[
y \approx \mathbf{G} \hat{m}
\] (2.18)
with $G \in \mathbb{C}^{L \times 2M+1}$ the following matrix

$$G[l,n] = \hat{\psi}_D[n] e_n(\theta_l).$$

(2.19)

Note that $G$ can be rewritten as a product of diagonal and Vandermonde matrices such that:

$$G = \text{Diag}[e^{-M(\theta_1)}]_t \text{Vand}^{2M+1}[e_1(\theta_t)]_t \text{Diag} \left[ \hat{\psi}_D[n] \right]_n$$

(2.20)

or, with explicit matrix notations:

$$G = \begin{bmatrix}
  e_{-M}(\theta_1) & 0 \\
  & \ddots \\
  0 & e_{-M}(\theta_L)
\end{bmatrix}
\begin{bmatrix}
  1 & e_1(\theta_1) & e_1(\theta_1)^2 & \ldots & e_1(\theta_1)^{2M} \\
  1 & e_1(\theta_2) & e_1(\theta_2)^2 & \ldots & e_1(\theta_2)^{2M} \\
  & \vdots & \ddots & \ddots & \vdots \\
  1 & e_1(\theta_L) & e_1(\theta_L)^2 & \ldots & e_1(\theta_L)^{2M}
\end{bmatrix}
\begin{bmatrix}
  \hat{\psi}_D[-M] & 0 \\
  & \ddots \\
  0 & \hat{\psi}_D[M]
\end{bmatrix}$$

Finally, the noisy version of equation (2.18) linearly links the measurements $y$ to the Fourier series coefficients of the stream of Dirac (2.14) and transforms the problem to the CPGD context presented in [8]:

$$y \approx G\hat{x} + w$$

(2.21)

with $w$ a random centered noise.

### 2.2.2 Estimating knots positions

**Annihilating filter**

In order to estimate the knots position, the FRI framework leverages a tool called annihilating filter. As presented in [8], this filter is represented by a vector $h \in \mathbb{C}^{K+1}$ such that:

- if we consider the 0-padded sequence $h \in \mathbb{C}^Z$ of $h$, the convolution product $\hat{x} \ast h$ with the (infinite) sequence of Fourier coefficients of $m$ is null: $(\hat{m} \ast h)_n = 0, \quad n \in \mathbb{Z},$

- the roots of the $z$-transform of $h$ are $u_k = \exp(-j2\pi t_k/T)$, in one-to-one relation with the knots positions $t_k$.

From the convolution product $(\hat{m} \ast h) \in \mathbb{C}^Z$, one can extract the $2M - K + 1$ coefficients that would result from the finite-sequences convolution product between $\hat{m}$ and $h$, which can be computed as:

$$(\hat{m} \ast h) = T_K(\hat{m})h$$

(2.22)

with the *Toeplitzification operator* $T_P$ defined for any integer $0 \leq P \leq M$ as:

$$T_P: \begin{cases}
\mathbb{C}^{2M+1} & \rightarrow \mathbb{C}^{(2M-P+1) \times (P+1)} \\
\mathbf{v} & \mapsto [T_P(\mathbf{v})]_{i,j} := x_{-M+P+i-j}
\end{cases}$$

(2.23)

It has been shown (in [7]) that, as long as $M \geq K$, the matrix $T_K(\hat{m})$ has rank $K$, such that the equation

$$T_K(\hat{m})\hat{h} = 0 \in \mathbb{C}^{2M-K+1}$$

(2.24)

always admits non trivial solutions for $\hat{h} \in \mathbb{C}^{K+1}$. Actually, when $P$ is such that $K \leq P \leq M$, the matrix $T_P(\hat{m}) \in \mathbb{C}^{(2M-P+1) \times (P+1)}$ is still of rank $K$, which leads
to the conclusion that
\[ T_P(\hat{m}) \tilde{h} = 0 \in \mathbb{C}^{2M-P+1} \] (2.25)
still admits non trivial solutions for \( \tilde{h} \in \mathbb{C}^{K+1} \) (note the change of \( K \) to \( P \)).

**Fourier series estimation**

Then, as the \( G \) matrix presented earlier (2.20) is injective, the authors propose to approximate the \( \hat{m} \) Fourier coefficients of \( m \) as the solution of the minimization problem:
\[
\min_{z \in \mathbb{C}^{2M+1}} \| Gz - y \|_2^2 \quad \text{subject to} \quad \{ \text{rank} \ T_K(z) \leq K \} \quad (2.26)
\]

However, as a sort of oversampling to reduce noise influence, this problem is solved using a slightly more restrictive constraint. Using the fact that \( T_P(\hat{m}) \) has also rank \( K \) for \( K \leq P \leq M \), the constraint is replaced by \( \{ \text{rank} \ T_M(z) \leq K \} \) (using \( P = M \), bigger matrix), which yields the best empirical performances, leading to:
\[
\min_{z \in \mathbb{C}^{2M+1}} \| Gz - y \|_2^2 \quad \text{subject to} \quad \{ \text{rank} \ T_M(z) \leq K \}. \quad (2.27)
\]

**Knots recovery**

So far, the rank constraint has been used to ensure the existence of a non trivial vector \( h \) but this vector is not calculated when solving (2.27). Indeed, it is not necessary to solve the problem and retrieve the truncated Fourier series coefficients \( \hat{m} \). Nonetheless, it takes part in the calculation of the knots locations (in relation with the roots of the \( z \)-transform). Consequently, once \( \hat{m} \) has been estimated as the solution of problem (2.27), \( h \) is found as the solution of equation (2.24) (using Total Least Squares for example). Note that here, \( P \) is set to the value \( K \) which leads to \( K + 1 \) coefficients of \( h \) and so \( K \) locations of the knots as expected. Solving the equation with a different value of \( P \) would lead to an inaccurate number of solutions.

**2.2.3 Solving the Fourier Series inverse problem**

Due to its rank constraint, the problem (2.27) is quite specific and needs to be treated carefully. It is reformulated in an unconstrained way using the indicator function \( \iota \) whose values live in \( \{0, +\infty\} \). Let \( \mathcal{H}_K \) be the set of complex matrices of rank at most \( K \):
\[
\mathcal{H}_K := \left\{ M \in \mathbb{C}^{(2M+1-P)\times(P+1)} \mid \text{rank} \ M \leq K \right\} \quad (2.28)
\]

The unconstrained version of (2.27) becomes
\[
\min_{z \in \mathbb{C}^{2M+1}} \| Gz - y \|_2^2 + \iota_{\mathcal{H}_K}(T_M(z)) \quad (2.29)
\]
which is the sum of a convex and differentiable term (data fidelity term) and a non-convex (as \( \mathcal{H}_K \) is not convex) and non differentiable term (constraint term).

As the data fidelity term is \( \gamma \)-Lipschitz continuously differentiable, it is possible to optimize (2.29) using proximal gradient descent (PGD). The Lipschitz constant is given by:
\[
\gamma = 2 \| G^*G \| = 2\lambda_{\max}(G^*G) \quad (2.30)
\]
that can be computed as the greatest eigenvalue of the matrix. The non convexity of the constraint term does not ensure that the proximal operator would return unique solutions, but it is not a problem, as any proximal point can be used interchangeably at this step.

However, the proximal step implies at each iteration to solve the following non convex optimization problem:

$$
\tilde{z} \in \text{prox}_{\tau B}(u) = \arg\min_{z \in \mathbb{C}^{2M+1}} \left\{ \frac{1}{2\tau} \|u - z\|^2 + \tau \mathcal{H}_K(T_P(z)) \right\}, \quad u \in \mathbb{C}^{2M+1}
$$

with $B(\cdot) = \tau \mathcal{H}_K(T_P(\cdot))$. The authors of [8] introduce the diagonal matrix $\Gamma \in \mathbb{C}^{(2M+1) \times (2M+1)}$ as:

$$
\Gamma_{i,i} = \min (i, P + 1, 2M + 2 - i), \quad i = 1, \ldots, 2M + 1.
$$

and the weighting matrix $W \in \mathbb{C}^{(2M+1-P) \times (P+1)}$ as:

$$
W = T_P \text{Diag} \left( \Gamma^{-1/2} \right)
$$

such that the proximal operator becomes:

$$
\text{prox}_{\tau H}(x) = T_P^\dagger \Pi_{T_P \cap \mathcal{H}_K}^W T_P(x)
$$

where $\Pi_{T_P \cap \mathcal{H}_K}^W$ is the projection onto $T_P \cap \mathcal{H}_K$ with respect to the $W$-weighted Frobenius norm, $T_P$ being the space of Toeplitz matrices of size $(2M+1-P) \times (P+1)$. This operator does not admit any simple closed-form expression, however it can efficiently be approximated by the method of alternating projections (MAP):

$$
\Pi_{T_P \cap \mathcal{H}_K}^W \approx \left[ \Pi_{T_P} \Pi_{\mathcal{H}_K}^W \right]^n
$$

for some $n \in \mathbb{N}$ big enough. We finally detail the expressions of the involved projection operators:

- the projection onto $T_P \subset \mathbb{C}^{(2M+1-P) \times (P+1)}$ of rectangular Toeplitz matrices admits a simple closed-form expression:

$$
\Pi_{T_P} = T_P T_P^\dagger = T_P \Gamma^{-1} T_P^\ast
$$

- the projection onto $\mathcal{H}_K$ of matrices of rank at most $K$ with respect to the $W$-weighted norm is actually more complex, but can be approximated to the first order with the regular projection onto $\mathcal{H}_K$, such that :

$$
\Pi_{\mathcal{H}_K}^W(X) \approx \Pi_{\mathcal{H}_K}(X) = U \Lambda_K V^H, \quad X \in \mathbb{C}^{(2M+1-P) \times (P+1)}
$$

where $X = U \Lambda V^H$ is the singular value decomposition of $X$ and $\Lambda_K$ is the diagonal matrix of sorted singular values truncated to the $K$ strongest ones.

All the operators involved at any step of the algorithm are either expressed or approximated with simple closed-form expressions. The final version is then fully usable and could be summarized as in algorithm 2.

Remark 3 (On the number $M$ of Fourier coefficients). It is important to note that the maximum number of innovations that can be recovered depends on the practical implementation of the Green functions. Indeed, CPGD aims at finding $\hat{m}$ which is composed of $2M + 1$ coefficients of the Fourier series of the input signal. We
Algorithm 2 CPGD \((K, P, M, \varepsilon)\)

1: Solve (2.29) using PGD :
2: Initialize : \((z_0, x_0) \leftarrow (0_{2M+1}, 0_{2M+1})\)
3: for \(n \geq 1\) do
4: \(z_n = \text{prox}_{\tau G} (x_{n-1} - 2\tau G^* (Gx_{n-1} - y_0))\)
5: \(x_n = T^\dagger \Pi_{P \cap H_K} T_M (z_n)\)
6: if \(\|x_n - x_{n-1}\|_2 \leq \varepsilon \|x_{n-1}\|_2\) then
7: \(x_n\) solution of the problem. Stop.
8: else
9: \(n \leftarrow n + 1\)
10: end if
11: end for
12: Find \(h \in \mathbb{C}^{K+1}\) by solving (2.25) with TLS
13: Find the roots \((u_k)_{k=1,...,K}\) of the \(z\)-transform of \(h\)
14: Deduce the \(K\) knot positions \((t_k)_{k=1,...,K}\)

used the python package Periodispline [22] which also leverages the Fourier series decomposition to define and evaluate the Green’s functions. In most of the cases, there is no reason for the Fourier series of the considered Green’s functions to be bandlimited: they theoretically admit an infinite number of coefficients. However, this is not the case for their actual implementation. For finite storage reason, the series of a given Green’s function \(\psi_D\) is truncated to the first main consecutive coefficients, up to a cutoff frequency (that preserves most of the energy of the function, set as a given percentage of the maximal absolute value frequency). Let \(M_c\) be the index of the highest frequency stored for the function’s representation series. The extreme point solution (1.9) is expressed as a linear combination of shifted versions of the reconstruction Green’s function, and thus is implemented with a finite bandwidth in the Fourier frequency space. It leads to the conclusion that one shall not expect the CPGD algorithm to accurately reconstruct more than \(2M_c + 1\) coefficients of the periodic Dirac stream \(x\), practically summarized as \(M \leq M_c\). Stated another way, one of the two following statements shall hold:

- Given a reconstruction Green’s function \(\psi_D\) (or equivalently a regularization operator \(D\)), the maximum frequency index of the Fourier series to reconstruct \(M\) is bounded by the cutoff frequency of \(\psi_D\).

- Given an objective reconstruction frequency index \(M\), one shall make sure that enough coefficients \(M_c \geq M\) are considered to evaluate the reconstruction Green function (it is possible to increase the number of coefficients \(M_c\) used by changing the precision in the calculation of \(\psi_D\)).

2.2.4 Recovering the weights of the knots

After estimating the \(K\) positions of the reconstruction knots only the associated weights remain unknown. It is the exact same situation of section 2.1.1 which is already streamlined. The same FISTA algorithm will lead to the weights \(\beta\) within few iterations. Indeed, when compared to the grid reconstruction, the number of weights to estimate has reduced from \(N >> K_0\) to only \(K\).

Remark 4 (Choice of number \(K\) of knots). In the CPGD framework, the reconstruction method needs to be specified the number of knots \(K\) to consider. The final solution
returned by the algorithm has at most $K$ knots, but in practice it is made of less as the weights estimation step tends to set some knots to 0. Indeed, this is formulated as a LASSO problem, and consequently returns sparse solutions. A solution to give more flexibility to the whole reconstruction procedure is to set this maximal number of reconstruction knots to its maximal value $K = P = M$. That way, the weighting step has available a larger number of knots and can be more selective in the choice of knots to keep for the final solution.

### 2.3 Frank-Wolfe Algorithm

Another way of solving problem (1.8) is to adapt an optimization algorithm called Frank-Wolfe algorithm [25] or Conditional Gradient Method [9]). Frank-Wolfe algorithm (FW) is built to efficiently solve a problem written as:

$$\min_{m \in C} f(m)$$

where $C$ is a weakly compact convex set of a Banach space and $f$ is a differentiable convex function. It is very interesting when the Banach structure of the search space does not allow the existence of any gradient method. In our particular case, $C$ will be a subspace of $\mathcal{M}$ but the objective function we used so far does not comply with the requirements of such a framework. However, with a little effort, it is possible to sort out an equivalent formulation that is FW-friendly. Let us detail the algorithm generic formulation in algorithm 3.

**Algorithm 3 Frank-Wolfe Algorithm**

1: for $k = 0, \cdots, n$ do
2:     Minimize : $c^{[k]} \in \arg\min_{c \in C} f(m^{[k]}) + df(m^{[k]})[c - m^{[k]}]$.
3:     if $df(m^{[k]})[c^{[k]} - m^{[k]}] = 0$ then
4:         $m^{[k]}$ solution of (2.36). Stop.
5:     else
6:         Step size search: $\gamma^{[k]} \leftarrow \frac{2}{\|c^{[k]} - m^{[k]}\|}$ or $\gamma^{[k]} \in \arg\min_{\gamma \in [0,1]} f(m^{[k]} + \gamma (c^{[k]} - m^{[k]}))$
7:         Update: $m^{[k+1]} \leftarrow m^{[k]} + \gamma^{[k]}(c^{[k]} - m^{[k]})$
8:     end if
9: end for

### 2.3.1 Reformulation of the original problem

The same way as before, consider a source sparse signal as defined in (2.11). Inspired from Q. Denoyelle’s work in [16], we want to express the measurement operator $\Phi$ as an integral operator that applies to the stream of periodic Dirac (that we previously denoted as $m_0(t) = \sum_k \beta_{0,k} \Pi(t - t_{0,k})$). More precisely, we want to write:

$$y_0 = \Phi(f_0) = \tilde{\Phi}(m_0)$$

thus:

$$\tilde{\Phi} = \Phi D^{-1}.$$
Again, let us use a generic $K$-sparse solution function as (2.12). The following holds:

$$y = \{ \Phi D^{-1} \} (m)$$

$$= \Phi D^{-1} \left( \sum_{k=1}^{K} \beta_k \Pi (\cdot - t_k) \right)$$

$$= \Phi (f)$$

$$= \Phi \left( \sum_{k=1}^{K} \beta_k \psi_D (\cdot - t_k) \right)$$

$$= \left( \sum_{k=1}^{K} \beta_k \psi_D (\theta \ell - t_k) \right)_{\ell=1,...,L}$$

$$= \left( \int_{T} \psi_D (\theta \ell - t) \, dm(t) \right)_{\ell=1,...,L}$$

$$= \left( \int_{T} \varphi_\ell (t) \, dm(t) \right)_{\ell=1,...,L}$$

with

$$\forall t \in T, \quad \varphi_\ell (t) = \psi_D (\theta \ell - t). \quad (2.39)$$

Then the integral operator appears as:

$$\forall \mu \in \mathcal{M}(T), \quad \tilde{\Phi} (\mu) = \int_{T} \varphi(t) \, d\mu(t) \quad (2.40)$$

and the new measurement operator eventually becomes:

$$\tilde{\Phi} : \left\{ \begin{array}{ll} \mathcal{M}(T) & \rightarrow \mathbb{R}^L \\ \mu(t) & \mapsto \left( \int_{T} \varphi_\ell (t) \, d\mu \right)_{\ell=1,...,L} \end{array} \right. \quad (2.41)$$

Particularly, if $\mu(t) = \sum_{i=1}^{I} \alpha_i \Pi (t - t_i)$, for any $I \in \mathbb{N}$, $(\alpha_i)_{i=1,...,I} \in \mathbb{R}^I$, then:

$$\tilde{\Phi} (\mu) = \left( \sum_{i=1}^{I} \alpha_i \psi_D (\theta \ell - t_i) \right)_{\ell=1,...,L} \quad (2.42)$$

In order to comply with the context presented in [16], it is necessary to make sure that $\varphi \in \mathcal{C}^{(2)}$ (see [16, Definition 4]).

Another few lines of calculations lead to its adjoint operator that will be useful later on:

$$\tilde{\Phi}^* : \left\{ \begin{array}{ll} \mathbb{R}^L & \rightarrow \mathcal{C}_0 (T) \\ h & \mapsto \sum_{\ell=1}^{L} h_\ell \psi_D (\theta \ell - \cdot) \end{array} \right. \quad (2.43)$$
Indeed, we can write:

\[
\langle \tilde{\Phi} \mu, h \rangle_{\mathbb{R}^L} = \langle \mu, \tilde{\Phi}^\ast h \rangle_{\mathcal{M}, \mathcal{M}_0}
\]

\[
\sum_{\ell=1}^L h_\ell \int_T \psi_D(\theta_\ell - t) \, d\mu(t) = \int_T \{ \tilde{\Phi}^\ast h \} (t) \, d\mu(t)
\]

Remembering that \( Df = m \), problem (1.8) can be expressed in the equivalent following way:

\[
\min_{m \in \mathcal{M}(T)} P_\lambda(m) := \left\| y_0 - \tilde{\Phi}(m) \right\|_2^2 + \lambda \|m\|_{TV}. \quad (2.44)
\]

This can be stated another way around. Thanks to the invertibility of \( D \), the set of solutions of (2.44) simply is \( D^{-1} \) applied to the set of solutions of (1.8):

\[
\text{Sol}(2.44) = D^{-1}\{\text{Sol}(1.8)\}. \quad (2.45)
\]

However, \( P_\lambda \) is not differentiable as it and thus this expression is not well suited for directly applying the Frank-Wolfe algorithm. Problem (2.44) can be turned into an equivalent differentiable problem by means of an epigraphical lift, as demonstrated in the following lemma.

**Lemma 1. The B-Lasso problem**

\[
\min_{m \in \mathcal{M}(T)} P_\lambda(m) := \left\| y_0 - \tilde{\Phi}(m) \right\|_2^2 + \lambda \|m\|_{TV}
\]

is equivalent to

\[
\min_{(s,m) \in C} \tilde{P}_\lambda(s, m) := \left\| y_0 - \tilde{\Phi}(m) \right\|_2^2 + \lambda s. \quad (2.46)
\]

with \( C = \{(s,m) \in \mathbb{R}_+ \times \mathcal{M}(T) : \|m\|_{TV} \leq s \leq M = \|y_0\|_2^2 / \lambda \} \).

**Proof.** Consider a solution \( m^\ast \) of \( P_\lambda \) on \( \mathcal{M}(T) \), then

\[ P_\lambda(m^\ast) \leq P_\lambda(0) = \lambda M \]

with \( M = \|y_0\|_2^2 / \lambda \). Moreover,

\[ P_\lambda(m^\ast) \leq P_\lambda(m) \leq \tilde{P}_\lambda(s, m) \]

for any \( m \in \mathcal{M} \) and \( s \geq \|m\|_1 \). The equality is reached when \( m = m^\ast \) and \( s = \|m^\ast\| \) for instance.

In their paper, Denoyelle et al. also prove that this latter problem (2.46) is well-posed, and \( \tilde{P}_\lambda \) reaches a minimum on \( C \), which concludes justifying the use of the algorithm.
2.3.2 Practical implementation of the algorithm

The Frank-Wolfe algorithm leverages the differentiable property of the objective function to iteratively minimize a first order linear approximation. It keeps track of a running iterate $m^{[k]}$ that best approximates $m$ after $k$ steps.

We apply the algorithm as previously described with the objective function $f$ being $\tilde{P}_\lambda$, thus the current iterate is actually an element $(s^{[k]}, m^{[k]}) \in C$. The next iterate $(s^{[k+1]}, m^{[k+1]})$ is consequently calculated as a convex combination of $(s^{[k]}, m^{[k]})$ and the solution $\epsilon^{[k]} = (t^{[k]}, u^{[k]})$ of the minimization of the first order approximation of $\tilde{P}_\lambda$ given by:

$$\left(t^{[k]}, u^{[k]}\right) \in \arg\min_{c = (t,u) \in C} \tilde{P}_\lambda(t,u) + d\tilde{P}_\lambda(s^{[k]}, m^{[k]}) \left(t - s^{[k]}, u - m^{[k]}\right).$$

(2.47)

With $\gamma^{[k]}$ the combination weight, the following holds:

$$m^{[k+1]} = m^{[k]} + \gamma^{[k]} \left(u^{[k]} - m^{[k]}\right) = \left(1 - \gamma^{[k]}\right) m^{[k]} + \gamma^{[k]} u^{[k]}$$

(2.48)

$$s^{[k+1]} = s^{[k]} + \gamma^{[k]} \left(t^{[k]} - s^{[k]}\right) = \left(1 - \gamma^{[k]}\right) s^{[k]} + \gamma^{[k]} t^{[k]}$$

(2.49)

New spike creation

It is important to point out that the approximation step (2.47) aims at finding the minimum of a linear functional, thus this minimal point is necessary either $(0,0)$ or located on the border of the set $C$. Extreme points of $C$ are of the form $(M, \pm M \, \text{III} (-t))$ for $t \in T$. Consequently, what the minimisation step does is basically adding a new periodic Dirac impulse to the approximation iterate, which is then reweighed.

First, let us detail the calculation of the differential of $\tilde{P}_\lambda$. One can see the data fidelity term as the composition between the two functions:

$$R: \begin{cases} \mathbb{R}^L \to \mathbb{R} \\ x \mapsto \|y_0 - x\|_2^2 \end{cases} \quad \text{and} \quad \tilde{\Phi}: \begin{cases} \mathcal{M}(T) \to \mathbb{R}^L \\ \mu(t) \mapsto \tilde{\Phi}_\mu \end{cases}$$

that we can easily determine the differential of. Indeed, $R$ is a quadratic form and $\tilde{\Phi}$ is linear, thus is equal to its differential. We obtain, for any $a \in \mathbb{R}^L$ and $m \in \mathcal{M}(T)$:

$$dR(a): \begin{cases} \mathbb{R}^L \to \mathbb{R} \\ h \mapsto \langle y - a, h \rangle_{\mathbb{R}^L} \end{cases} \quad \text{and} \quad d\tilde{\Phi}(\mu): \begin{cases} \mathcal{M}(T) \to \mathbb{R}^L \\ \mu'(t) \mapsto \tilde{\Phi}_\mu' \end{cases}$$

Given $R \circ \tilde{\Phi}: \mathcal{M}(T) \to \mathbb{R}$, and for any $\mu \in \mathcal{M}(T)$, $d \left( R \circ \tilde{\Phi} \right) (\mu) : \mathcal{M}(T) \to \mathbb{R}$, we obtain:

$$d \left( R \circ \tilde{\Phi} \right) (\mu) \cdot \mu' = \left(dR \left( \tilde{\Phi}_m \right) \circ d\tilde{\Phi}(\mu) \right) \cdot \mu' = \langle y_0 - \tilde{\Phi}_\mu, \tilde{\Phi}_\mu' \rangle_{\mathbb{R}^L}$$

that can be nicely rewritten as:

$$d \left( R \circ \tilde{\Phi} \right) (\mu) \cdot \mu' = \langle \tilde{\Phi}^* (y_0 - \tilde{\Phi}_\mu), \mu' \rangle_{\mathcal{C}_0(T), \mathcal{M}(T)} = \int_T \tilde{\Phi}^* (y_0 - \tilde{\Phi}_\mu) \, d\mu'$$

(2.50)

Finally, the closed form expression is:

$$d\tilde{P}_\lambda(s, \mu): \quad (s', \mu') \mapsto \int_T 2\tilde{\Phi}^* (\tilde{\Phi}_\mu - y_0) \, d\mu' + \lambda s'.$$
Then, finding a minimizer of (2.47) having the shape of a single peak periodic Dirac amounts to finding a point
\[ t^{[k]} \in \arg \min_{t \in \mathbb{T}} \left( \pm \frac{2}{\lambda} \left( \Phi^* (\Phi m^{[k]} - y_0) \right) (t) + 1 \right) \lambda M \]  
(2.52)
or equivalently
\[ t^{[k]} \in \arg \max_{t \in \mathbb{T}} \left( |\eta^{[k]}(t)| - 1 \right) \]  
(2.53)
with
\[ \eta^{[k]}(t) := \frac{2}{\lambda} \left( \Phi^* (y_0 - \Phi m^{[k]}) \right) (t) \]  
(2.54)
This function \( \eta^{[k]} \) is called empirical dual certificate. Any given iterate \( m^{[k]} \) being expressed as a sum of (at most, depends on some \( \epsilon \)-weighted Dirac) \( k \) periodic Dirac, i.e. \( m^{[k]} = \sum_{i=1}^{k} \alpha_i^{[k]} \Pi \cdot (-t_i^{[k]}) \), the function \( \eta^{[k]} \) admits the closed form expression:
\[ \eta^{[k]}(t) = \frac{2}{\lambda} \sum_{\ell=1}^{L} \left( y_\ell - \sum_{j=1}^{L} \sum_{i=1}^{k} \alpha_i^{[k]} \psi_D(\theta_j - t_i^{[k]}) \right) \psi_D(\theta_\ell - t). \]  
(2.55)

In practice, it is not sufficient to know the position of the next periodic Dirac impulse, it is also necessary to know if it is positively or negatively weighted. The first naive idea basically solves two different minimization problems, each one related to one of the two situations. The one leading to the lower value of the objective function determines the sign of the new impulse to add to the current iterate. Positive impulse leads to the problem:
\[ \arg \min_{t \in \mathbb{T}} -\eta^{[k]}(t) \]  
(2.56)
By linearity of \( \Phi^* \), negatively weighted impulses lead to the opposite problem.

However, it is actually possible to complete this step by solving only one minimization problem. Let us consider any intermediate step index \( k \) of the algorithm. We write :
\[ L^{[k]} = d\tilde{P}_\lambda(s^{[k]}, m^{[k]}) : (\mathbb{R}, \mathcal{M}(\mathbb{T})) \rightarrow \mathbb{R} \]  
(2.57)
Let \( \epsilon \in \{-1, +1\} \), then, for any \( t \in \mathbb{T} \), we have:
\[ L^{[k]}(M, \epsilon M \Pi t) = 2\epsilon M \left( \Phi^* (\Phi m^{[k]} - y_0) \right) (t) + \lambda M \]
\[ = -\epsilon M \lambda \eta^{[k]}(t) + \lambda M \]
\[ = \lambda M \left( 1 - \epsilon \eta^{[k]}(t) \right) \]  
(2.58)

However, when \( t = t_{k+1} \) is the solution of (2.53), there exist \( \epsilon^{[k+1]} \in \{-1, +1\} \) such that the pair \( (M, \epsilon^{[k+1]} M \Pi t_{k+1}) \) minimizes \( L^{[k]} \) over \( C \). \( \epsilon^{[k+1]} \) simply is the sign of the next periodic Dirac impulse. In particular, being the minimum:
\[ L^{[k]}(M, \epsilon^{[k+1]} M \Pi t_{k+1}) \leq L^{[k]}(0, 0) = 0 \]
Thus:
\[ \epsilon^{[k+1]} \eta^{[k]}(t_{k+1}) \geq 1 > 0 \]  
(2.59)
Which leads to :
\[ \epsilon^{[k+1]} = \text{sgn}(\eta^{[k]}(t_{k+1})) \]  
(2.60)
Eventually, the algorithm first numerically solves (2.53) to estimate the position of the next impulse, and then evaluates the empirical dual certificate at this latter position to determine the sign of the impulse to append to the current iterate.

Combination weight

The weight \( \gamma^{[k]} \) of the convex combination can be expressed as decreasing with \( k \) in such a manner that the first iterates have more significance than the last ones in the final expression of the solution. For instance, one can choose:

\[
\gamma^{[k]} = \frac{2}{k + 2}
\]

The convergence is actually quite long and with very low accuracy.

Another more efficient way is to find \( \gamma^{[k]} \) that minimizes the expression

\[
\gamma^{[k]} \in \arg \min_{\gamma \in [0, 1]} \tilde{\Phi}_{\lambda} \left( \| m^{[k]} \|_{TV} + \gamma (M - \| m^{[k]} \|_{TV}), m^{[k]} + \gamma (u^{[k]} - m^{[k]}) \right)
\]

This method conduces to better results in less iterations as the iterates are tailored accordingly to the problem, not with general coarse formula.

Additionally, we can show that it is possible to compute the exact value of \( \gamma^{[k]} \) at each step. First let us emphasize a simple property that was not explicitly stated in [16]. Let us call the first variable \( s \) of (2.46) the regularization variable, and denote by \( s^{[k]} \) the iterates when applying the FW algorithm. Then, the following holds:

**Lemma 2.** For any iteration index \( k \in \mathbb{N} \),

\[
s^{[k]} = \| m^{[k]} \|_{TV}
\]

**Proof.** If we consider two finite weighted sums of Dirac combs \( g_1 \) and \( g_2 \), with distinct knots in \( T \) and no common knot between the two, then \( \| g_1 + g_2 \|_{TV} = \| g_1 \|_{TV} + \| g_2 \|_{TV} \). Then, for any \( \gamma \in [0, 1] \), at any iteration step \( k \), we can write

\[
\| m^{[k]} + \gamma (u^{[k]} - m^{[k]}) \|_{TV} = (1 - \gamma) \| m^{[k]} \|_{TV} + \gamma \| u^{[k]} \|_{TV}
\]

(as long as the new Dirac impulse position does not belong to the previous set of positions, which is supposed to be guaranteed by the FW theory). Starting the iterates of FW with \((0, 0)\) at iteration \( k = 0 \) (also true at iteration \( k = 1 \) with \((\gamma^{[1]} M, \pm \gamma^{[1]} M \|_{TV})\)), equation (2.62) is verified. Equation (2.62) being true at any given iteration \( k \), then:

\[
s^{[k+1]} = \left( 1 - \gamma^{[k]} \right) s^{[k]} + \gamma^{[k]} M = \left( 1 - \gamma^{[k]} \right) \| m^{[k]} \|_{TV} + \gamma^{[k]} \| u^{[k]} \|_{TV} = \| m^{[k+1]} \|_{TV},
\]

it is still true at iteration \( k + 1 \), which proves the property by induction.

This result allows to explicitly compute the solution.

**Proposition 1.** The unique solution to (2.61) is given by

\[
\gamma^{[k]} = \frac{\left[ \left( \tilde{\Phi} m^{[k]} - y_0 \right)^T \left( \tilde{\Phi} u^{[k]} + y_0 \right) + \| y_0 \|_2^2 - \| \tilde{\Phi} m^{[k]} \|_2^2 \right] - \lambda [M - \| m^{[k]} \|_{TV}]}{2 \| \tilde{\Phi} m^{[k]} - \tilde{\Phi} u^{[k]} \|_2^2}.
\]

(2.63)
Proof. Indeed, one can rewrite the expression of the objective function of (2.61) as follows:

\[
\begin{aligned}
\tilde{P}_\lambda \left( s^{[k]} + \gamma (M - s^{[k]}), m^{[k]} + \gamma (u^{[k]} - m^{[k]}) \right) \\
= \tilde{P}_\lambda \left( \|m^{[k]}\|_{TV} + \gamma (M - \|m^{[k]}\|_{TV}), m^{[k]} + \gamma (u^{[k]} - m^{[k]}) \right) \\
= \|y_0 - \gamma \tilde{\Phi} u^{[k]} - (1 - \gamma) \tilde{\Phi} m^{[k]}\|^2_2 + \lambda (\gamma M + (1 - \gamma) \|m^{[k]}\|_{TV}) \\
= \gamma^2 \left[ \|\tilde{\Phi} u^{[k]}\|^2_2 + \|\tilde{\Phi} m^{[k]}\|^2_2 - 2 \left( \tilde{\Phi} u^{[k]} \right)^T \tilde{\Phi} m^{[k]} \right] \\
+ 2\gamma \left[ - \left( \tilde{\Phi} m^{[k]} \right)^T \left( \tilde{\Phi} u^{[k]} + y_0 \right) + \|y_0\|^2_2 - \|\tilde{\Phi} m^{[k]}\|^2_2 \right] + \gamma \lambda \left[ M - \|m^{[k]}\|_{TV} \right] \\
= \gamma^2 \left[ \|\tilde{\Phi} m^{[k]} - \tilde{\Phi} u^{[k]}\|^2_2 \right]
\end{aligned}
\]

The expression is a convex second degree polynomial, the minimum of which can be computed thanks to the coefficients. It leads to the aforementioned expression.

\[\square\]

Stopping criterion

As stated earlier, the new spike creation implies solving problem (2.53) in order to estimate the position of the new Dirac impulse. Interestingly enough, the actual value of the maximum of the empirical dual certificate \( \max_{t \in T} |\eta^{[k]}(t)| = \|\eta^{[k]}\|_{\infty} \) is very useful and so is also regarded. Denoyelle (still in [16], remark 5) has proved that this value is related to the natural stopping criterion

\[
(\bar{s}^{[k]}, m^{[k]}) \in \arg \min_{c \in C} \left\{ \text{d} \tilde{P}_\lambda \left( s^{[k]}, m^{[k]} \right) | c \right\} \tag{2.64}
\]

in the way that, if \((\bar{s}^{[k]}, m^{[k]})\) satisfies (2.64), then the following equation holds:

\[
\|\eta^{[k]}\|_{\infty} = 1 \tag{2.65}
\]

Thus, after calculating the new Dirac impulse position, verifying the stopping criterion becomes a trivial step. Given an arbitrary satisfying precision tolerance \( \varepsilon \), the algorithm simply verifies if the lastly calculated value of the empirical certificate function is \( \varepsilon \)-close from 1.

In practice, when the minimization step is successfully solved, the successive extremum values of the certificate are monotonously decreasing, although two behaviors have been observed. Either the sequence decays fast and distinctly reaches a value lower than 1, either it has an asymptotic and slow convergence towards 1 by upper value. In the first situation, the couple \((0,0)\) is actually the solution of (2.47). Considering this couple as the new spikes and plugging it in the convex combination of solution has the effect of decreasing the current weights of the iterate and eventually increasing the supremum norm of the empirical certificate. It usually becomes greater.
than 1 and the convergence continues until it reaches the stopping criterion. In the second situation, the stopping criterion is naturally reached after a finite number of iterations.

**Choice of the regularization parameter**

Let us recall the expression of the empirical dual certificate:

$$\eta^{(k)}(t) = \frac{2}{\lambda} \langle \tilde{\Phi}^\ast(y_0 - \tilde{\Phi}m^{[k]}) \rangle(t)$$

Then, at step $k = 0$, $m^{[0]} = 0$ such that:

$$\eta^{[0]}(t) = \frac{2}{\lambda} \tilde{\Phi}^\ast(y_0)(t)$$  \hspace{1cm} (2.66)

Consequently, in the same way as the equispaced knots strategy, the value of the regularization parameter $\lambda$ is also limited by an extreme value of

$$\lambda_{\text{max}} := 2 \max_{t \in \mathbb{T}} |\tilde{\Phi}^\ast(y_0)(t)| = 2 \mathbf{\|\tilde{\Phi}^\ast(y_0)\|}_\infty.$$ \hspace{1cm} (2.67)

Any value of $\lambda$ larger than $\lambda_{\text{max}}$ would force the algorithm to stop at the first iteration. In practice, we also set $\lambda = \sigma \mathbf{\|\tilde{\Phi}^\ast(y_0)\|}_\infty$ with $\sigma \approx 0.1$, which conduces to the best empirical results.

To put everything in a nutshell, the Frank-Wolfe algorithm approximates the sum of periodic Dirac

$$m(t) = \sum_{k=1}^{K} \beta_k \Pi(t - t_k)$$

using an iterative method. Each iteration of the loop adds a new weighted periodic Dirac impulse to the current approximate, the weights being decreased along the iterations to ensure that the first retrieved Dirac impulses are given more significance in the reconstruction.

**2.3.3 All-knots simultaneous reweighting strategy**

An assessment we made during our experimentation is that the regular FW strategy is greedy. Indeed, each iteration step exhibits a new Dirac impulse to add to the current iterate by maximizing a given cost function over the problem (namely the empirical dual certificate). This greedy approach might be successful, and even efficient, when the problem is well presented (noiseless, with regular measurement points). However, it might also lead to very unsatisfactory approximations. Indeed, this greedy approach is a “forward” strategy, that is only able to produce more complexity in the solution by increasing the size of the reconstruction support. It is not able to perform retro actions on the previously evaluated weights or positions.

One way to circumvent this issue comes from the analysis of the procedure. The regular FW in practice retrieves the innovations by first finding the position, and then, separately, estimating the weight. While the position estimation step is really subtle, from deep dual analysis, the weight estimation appears quite simple. The convex combination relies on the fact that all the previous weights have been correctly estimated, and gives really small margin for the new impulse to be significant with respect to the others. Then, the same way it is done with the sliding version [16, Algorithm 2] or in [26], a solution is to re-estimate the weights of all the knots simultaneously. More precisely, at any step $k$, the current iterate is $m^{[k]}$ and the new spike to be added is at position $u^{[k]} \in \mathbb{T}$ from (2.47). We solve the LASSO problem associated to the fixed knots $\text{Supp } (m^{[k]} \cup u^{[k]})$ to simultaneously estimate all the weights $\beta^{[k]}$. 

This strategy is very interesting, as it allows to *a posteriori* cancel any previously placed spike, thanks to the sparsity promoting behavior of the LASSO. It gives the algorithm the right to make mistakes and correct them pretty efficiently. In term of evolution of the value of the objective function \( \hat{P} \) along the iterations, it guarantees a decreasing behavior, as the same problem is solved step after step, with a larger support of recovery allowed (the set of possible reconstruction knots is increasing). To go even faster and make use of the largest information possible, the FISTA algorithm is initialized with the regular convex combination weighted iterate. It leads to significant gains, both in term of accuracy and duration of computation. The both strategies are really easy to compare as they share the same stopping criterion.

A consequent part of this internship has been focused on designing this reweighting strategy and analyzing its results. It leaded to a fine understanding of the regular FW mechanics as well as its weaknesses.
Chapter 3

Practical Experiments

One of the explicit goals of this internship was to obtain a working algorithmic framework that is efficiently able to run the three presented algorithms. Thanks to the rigorous analysis of their respective behavior, this objective has been fulfilled and we now dispose of a fresh implementation of each of these algorithms. In this chapter we thus present the common experimental setting that we have proposed for experimentation and show some examples of reconstruction.

3.1 Experimental Setting

Our experimental framework generates measurement data \( y_0 \) from a synthetic input signal \( f_0 = D^{-1}m_0 \). \( m_0 \) is a random sparse signal with \( K_0 \) knots and random associated weights. Then \( y_0 \) is fed to the algorithms as the measurement data. The response from the algorithms is given as list of \( K \) innovations that characterizes the sparse reconstructed signal that approximates \( f_0 \) by solving problem (1.8).

Practical statement of the problem

Let us dive into details on the input parameters involved.

- **The sampled spline** \( f_0 \) is built from a differential operator \( D_0 \) that needs to be specified as an input. In our experiment, we set its number of knots \( K_0 = 4 \), and let its innovation be taken randomly (uniform distribution for the knots over \( T \) and normal distribution for the weights). The number \( L \) of measurements is set as \( 8K_0 + 1 \) (so that \( M \) from CPGD is \( M = 4K_0 \)).

- **The reconstruction spline** \( f \) also needs to be built from an operator \( D \). Actually, there is no reason for \( D \) to be equal to \( D_0 \), as we suppose there is no prior knowledge on the input data. One might guess the type of operator \( D_0 \) involved from the input signal, but the specific operator’s parameters are impossible to estimate beforehand. In our experiences, we nevertheless let \( D = D_0 \), but it is important to keep in mind that they might differ.

- **The noise level** needs to be specified as a *peak signal-to-noise ratio* (PSNR) in decibel. It is involved in the samples generation as \( y_0 = \Phi(f_0) + w \) with \( w \) drawn according to \( \mathcal{N}(0, \omega^2) \) with

\[
\omega = \max_{\ell=1, \ldots, L} |\Phi(f_0)| \times \exp\left(-\frac{\text{PSNR}}{10}\right).
\]

We set the PSNR to 20dB for most of the experiences, but this value is often changed to study noise influence.
• The regularization parameter is chosen as a fraction of a maximal value $\lambda = \sigma \lambda_{\text{max}}$ previously calculated. This parameter needs to increase with the value of the noise, but experiments have shown that $\sigma \approx 0.1$ is usually a good rule of thumb for moderate noise levels of about 20dB.

Stopping criterion
As we have already explained, the stopping criterion differs with the algorithm. However, the algorithm FISTA is used at many occurrences, either as the main algorithm (equispaced grid strategy) or as a step to estimate the weights (for CPGD and FW with reweighting). For all of these different cases, the same relative improvement tolerance $\varepsilon$ is set to be equal to $10^{-4}$.

For CPGD, the stopping criterion for the estimation of the truncated Fourier series coefficients is also a relative improvement precision (see algorithm 2). It is also set to $\varepsilon = 10^{-4}$.

As the FW algorithm is concerned, the stopping criterion is slightly different and hard to relate to the other ones. Indeed, it is based on the maximum absolute value reached by the empirical dual certificate. If this value is $\nu$-close from 1 the algorithm stops. Actually, when $\nu$ is lower than $10^{-2}$, the regular FW takes significantly longer to converge (when it does). Consequently, we do not try to go lower and set $\nu = 10^{-2}$ as it already provides satisfactory results.

Metrics
Until now, the metrics have not been much into consideration. The question of a relevant metric really is task-dependent. As we are not concerned yet with any application, we made the choice of considering a bunch of different indicators that we think are relevant for the reconstruction. These indicators include:

• the computation time,
• the final objective value of the optimization problem (we evaluate (1.8) with the returned solution),
• the final sparsity index of the solution,
• the quadratic error between the source spline and the reconstructed one (expressed as the root of the relative squared error), which is an a priori quantity as it needs the full input signal to be computed,
• the quadratic error between the measurement sampled and the expected value of the reconstructed spline on the measurement points, which can be computed a posteriori as it only needs the samples,
• the fact that the algorithm has explicitly converged or not.

The final value of the objective function has come of great help when comparing the different method. Indeed, it is a common quantity that is agnostic of the reconstruction algorithm, and thus allows to measure to what extent each method performs with respect to the minimization task.
3.2 Explicit Results

3.2.1 Grid-based solutions

Let us set the size of the reconstruction grid to \( N = 300 \) and the penalization parameter \( \lambda = \sigma \lambda_{\text{max}} \). The noise level is set to 20 dB and the operator are exponential \( D = D_0 = (D + \alpha \text{Id})^\gamma \) with \( \gamma = 2 \) and \( \alpha = 3 \). We present different reconstructions with various values of \( \sigma \) in fig. 3.1 and table 3.1.

![Grid-based reconstruction with different values of \( \sigma \)](image)

**Figure 3.1:** Grid-based reconstruction with different values of \( \sigma \)

| \( \sigma \) | iterations | duration | converged | objective | splines rrse | samples rrse |
|-----------|------------|----------|-----------|-----------|--------------|--------------|
| 0.01      | 2000       | 0.196    | False     | 1.393     | 0.280        | 0.219        |
| 0.1       | 2000       | 0.221    | False     | 5.266     | 0.230        | 0.334        |
| 0.2       | 2000       | 0.197    | False     | 8.616     | 0.318        | 0.394        |
| 0.3       | 2000       | 0.188    | False     | 11.587    | 0.400        | 0.463        |

Table 3.1: Metrics for grid-based reconstruction

As expected, when \( \sigma \) tends to 0, the reconstruction behaves as an interpolation framework, where the solution tries to go through every sampled point. Obviously, the error between the measurements and the predicted values of the spline at the same positions (referred to as “samples rrse” in the table) tends to be smaller in this case. However, it is not correlated with the error between the source spline and the approximation one (confer to “splines rrse” in the table).

Note that none of the cases have converged within 2000 iterations. It is due to the conditioning number of the measurement matrix \( \mathbf{H} \) being quite large, as the matrix is full and of size \( N \times N \). Nevertheless, the non-convergence is not a real concern, as the improvement is already quite low and the solution does not vary in the following iterations.

3.2.2 CPGD

Let us run the CPGD algorithm with the same setting (same source spline and same input data). As previously mentioned, the parameters of CPGD are set to \( M = 4K_0 \)
and $K = P = M$. We obtain the following results in fig. 3.2 and table 3.2.

![Figure 3.2: CPGD reconstruction with different values of $\sigma$](image)

| factor | iterations | duration | converged | objective | splines rrse | samples rrse |
|--------|------------|----------|-----------|-----------|--------------|--------------|
| 0.01   | 500        | 70.610   | False     | 1.591     | 0.286        | 0.262        |
| 0.1    | 500        | 10.472   | False     | 5.415     | 0.277        | 0.342        |
| 0.2    | 500        | 4.436    | False     | 8.868     | 0.340        | 0.399        |
| 0.3    | 500        | 4.632    | False     | 11.899    | 0.439        | 0.480        |

Table 3.2: Metrics for CPGD reconstruction

For this exact same example, the CPGD algorithm overall performs worse than the equispaced knots reconstruction. In terms of accuracy and objective function value, it does nearly as good, however the computation times are really long. It is an encouraging result though, as this algorithm is quite new and has not been as well optimized as FISTA has been over the years.

### 3.2.3 Frank-Wolfe algorithm

The same way as before, we apply our implementation of FW to the same input data. In this section, we plot the results for both the original FW and our reweighting version in fig. 3.3, table 3.3 and table 3.4.
In comparison with the other algorithm, FW performs well and presents same quality results. It provides really good accuracy, but at a price of longer computation times (compared to the equispaced grid algorithm). What is really striking however is the improvement obtained thanks to our reweighting strategy. When the regular convex combination weighting strategy makes the convergence really slow ($\sigma = 0.01$ or $\sigma = 0.3$), the reweighting allows significant speed gains. Additionally, the final value of the objective function is either the lowest achieved among the three methods or as good as the one obtained with the first strategy. One might think it is more interesting to chose the grid approximation then, as it is still 10 times faster, this idea needs to be mitigated. Indeed, we are working with a really simple one dimensional context. With more complex situations, like multidimensional frameworks, the grid solution might have more trouble to scale out than FW. More research is necessary in this way to tackle this question.
We would like to emphasize the benefits of this simultaneous complete reweighting strategy within FW with a more striking example (see fig. 3.4, table 3.5 and table 3.6).

**Figure 3.4:** FW reconstruction with different values of $\sigma$ and the two weighting strategies

| factor | iterations | duration | converged | objective | splines | rrse | samples | rrse |
|--------|------------|----------|-----------|-----------|---------|------|---------|------|
| 0.01   | 1957       | 552.351  | True      | 0.156     | 0.310   | 0.195|
| 0.1    | 187        | 33.294   | True      | 0.514     | 0.250   | 0.312|
| 0.2    | 792        | 154.415  | True      | 0.770     | 0.214   | 0.370|
| 0.3    | 47         | 7.535    | True      | 1.016     | 0.321   | 0.440|

**Table 3.5:** Metrics for regular FW reconstruction

| factor | iterations | duration | converged | objective | splines | rrse | samples | rrse |
|--------|------------|----------|-----------|-----------|---------|------|---------|------|
| 0.01   | 66         | 15.981   | True      | 0.128     | 0.306   | 0.172|
| 0.1    | 29         | 5.830    | True      | 0.488     | 0.197   | 0.300|
| 0.2    | 24         | 4.518    | True      | 0.760     | 0.203   | 0.369|
| 0.3    | 19         | 3.542    | True      | 0.986     | 0.257   | 0.426|

**Table 3.6:** Metrics for FW reconstruction with reweighting

In this example, the improvement is considerable, regarding both the computation time and the error between the source and the approximated spline. It essentially comes from the bad estimation of the last negative bump from the regular FW method. When looking at the consecutive appearances of the Diracs along the iterations, one clearly sees that the algorithm is not able to correct the bad placements. On the other hand, the complete reweighting strategy allows to possibly remove or reduce the intensity of those problematic Diracs. Another counter effect of the convex combination weighting is that, iteration after iteration, all the previously added Diracs have their intensity being reduced. Thus, at some point, the algorithm has to reintroduce weight to those Diracs by placing spikes at nearly the exact same position of previous ones. It leads to clusters of medium intensity Diracs at the recovering position, instead of a
single high intensity one, that both slows down the procedure and damages the sparsity index. This phenomenon is also present with reweighting but much less frequent.

To summarize this chapter, we would like to underline the fact that we are provided with three different algorithms that all seem to successfully tackle the same B-LASSO problem. While CPGD does not seem to be mature enough to demonstrate good performances in the above presented context, it is too early to conclude that it should not be considered for the problem. Indeed, one should have noticed that the evaluation of any method strongly is task-dependent. A lot of parameters are susceptible to lead towards one or another algorithm, for instance the noise level, the type of spline and operator involved, the number of available samples... A noteworthy point that arises from this remark is that this approximation framework is very versatile. It is possible to handle of large range of application situations, and the understanding of the algorithms would really benefits from real data applications. Many questions have not been studied, like the choice a reconstruction operator for an unknown input signal or the choice of a good penalization parameter $\lambda$. 
Chapter 4

Theoretical analysis of the solution set

In optimization theory, the convexity of the problems is a first-order property. From a numerical point of view, it makes solving the problems generally significantly easier, with the absence of local optimums. It also is of great help when theoretical analysis is involved. Indeed, convex problems are particularly well fitted for dual analysis. Under certain conditions, it is possible to obtain very insightful results as existence of solutions, optimal value of the objective function or even uniqueness of the solution. Great work has been done in this way, using duality analysis to derive properties in the general case [10] or in more particular contexts [27]. In this chapter, we first conduct the dual analysis of our problem of interest before presenting an interesting context that would lead to uniqueness in the solution set. We link the conditions of such a context to the special case of our invertible pseudo-differential operators.

4.1 Duality Theory

As explained earlier, because of the invertibility of the operator, solving the penalized problem

$$\min_{f \in \mathcal{M}(\Omega)} \|y - \Phi(f)\|_2^2 + \lambda \|Df\|_{TV}$$

(4.1)

is equivalent to solving the following problem expressed over the measure space

$$\min_{m \in \mathcal{M}(\Omega)} P_\lambda(m) := \|y - \tilde{\Phi}(m)\|_2^2 + \lambda \|m\|_{TV}.$$ 

(4.2)

The differential operator and the shape of the solution now lie in the measurement operator $\tilde{\Phi}$. In this section, we will focus on the constrained problem, which is expressed using the same measurement operator. We will call it the primal problem.

$$\min_{m \in \mathcal{M}(\Omega)} \|m\|_{TV}, \quad y_0 = \tilde{\Phi}(m)$$

(4.3)

It is actually a reformulation of the following problem, using the invertibility of the operator.

$$\min_{f \in \mathcal{M}(\Omega)} \|Df\|_{TV}.$$ 

(4.4)

The equivalence between problems (4.2) and (4.3) is actually proven [23, Theorem 5] by the fact that, for any solution $m^*$ of (4.2), there exists a vector $y_\lambda$ such that $m^*$ is
solution of \(4.3\) with measurement vector equal to \(y_\lambda\). Additionally, \(y_\lambda\) is unique and independent from the solution \(m^*\).

We suppose that the measurement vector \(y_0\) is obtained by applying the measurement operator to a given measure \(m_0 \in \mathcal{M}(\mathbb{T})\) such that \(y_0 = \tilde{\Phi}(m_0)\).

### 4.1.1 Derivation of the dual problem

The first step of the analysis is to find the dual problem associated to \((4.3)\). By identifying the good functions and variables, it is possible to directly apply the theory from [28, Chapter 3].

**Proposition 2.** The dual problem of \((4.3)\) is

\[
\sup_{\|\tilde{\Phi}^*p\|_\infty \leq 1} \langle y_0, p \rangle. \tag{4.5}
\]

**Proof.** According to the [28, Chapter 3.4], an optimization problem can be written as

\[
\min_{x \in X} F(x) + G(\Lambda x), \tag{P}
\]

with \(F : X \to \mathbb{R}, G : Y \to \mathbb{R}\) and \(\Lambda : X \to Y\) linear, then one can introduce an antedual variable \(y \in Y\) such that the \(\Phi\) function (from Ekeland and Téman) becomes

\[
\Phi(x, y) = F(x) + G(\Lambda x - y).
\]

The dual problem of \((P)\) is

\[
\sup_{y^* \in Y^*} - (F^*(\Lambda^* y^*) + G^*(-y^*)), \tag{D}
\]

where \(F^* : X^* \to \mathbb{R}\) and \(G^* : Y^* \to \mathbb{R}\) denote the Fenchel conjugates of the functions \(F\) and \(G\), \(\Lambda^* : Y^* \to X^*\) is the adjoint operator of \(\Lambda\).

In our case, we shall use \(F = \|\cdot\|_{TV}, G = \iota_{y_0}\) and \(\Lambda = \tilde{\Phi}\) such that \(X = \mathcal{M}(\mathbb{T})\), \(X^* = C\) and \(Y = Y^* = \mathbb{R}^L\). \(\iota\) is the indicator function over \(\mathbb{R}^L\), which value is 0 if the argument is equal to \(y_0\), \(+\infty\) otherwise. Thus, the dual problem of \((4.3)\) becomes

\[
\sup_{p \in \mathbb{R}^L} - \|\tilde{\Phi}^* p\|_{TV}^* - \iota_{y_0}^*(-p). \tag{4.6}
\]

Let us express the conjugate functions that play a role in the above-mentioned expression.

**Lemma 3.** The Fenchel conjugate function of the TV-norm \(\|\cdot\|_{TV} : C \to \mathbb{R}\) is given by the indicator function \(\iota_{B_\infty}\) of \(B_\infty = \{\eta \in C, \|\eta\|_\infty \leq 1\}\) the unit ball for the infinite norm in \(C\).

Indeed, if \(F = \|\cdot\|_{TV}\), then for \(\eta \in C\):

\[
F^*(\eta) = \sup_{m \in \mathcal{M}} \langle \eta, m \rangle - \|m\|_{TV}
\]

We have the fundamental inequality \(\langle \eta, m \rangle \leq \|\eta\|_\infty \|m\|_{TV}\). Thus, if \(\|\eta\|_\infty \leq 1\) then \(\langle \eta, m \rangle - \|m\|_{TV} \leq (\|\eta\|_\infty - 1)\|m\|_{VT} \leq 0\), and one reach 0 for \(m = 0\). Otherwise, if \(\|\eta\|_\infty > 1\), there exists \(m_0\) such that \(\langle \eta, m_0 \rangle \geq a\|m_0\|_{TV}\) with \(a > 1\) (for instance,
\( m_0 = \mathbb{I}(t - \arg \max \eta) \). Consequently, for \( \lambda > 0 \), \( F^*(\eta) \geq \langle \eta, \lambda m_0 \rangle - \| \lambda m_0 \|_{TV} \geq \lambda a \| m_0 \|_{TV} \xrightarrow{\lambda \to +\infty} +\infty \).

**Lemma 4.** The Fenchel conjugate function of \( G = \iota_{y_0} \) is given by \( G^*(h) = \langle y_0, h \rangle \) for \( h \in \mathbb{R}^L \).

Similarly, we have
\[
G^*(h) = \sup_{z \in \mathbb{R}^L} \langle z, h \rangle - \iota_{y_0}(z)
\]

Directly, the following holds:
\[
\langle z, h \rangle - \iota_{y_0}(z) = \begin{cases} 
\langle y_0, h \rangle & \text{if } z = y_0 \\
-\infty & \text{otherwise}
\end{cases}
\]

which proves the result.

Given these two lemmas, one can rewrite (4.6) as
\[
\sup_{p \in \mathbb{R}^L} -\iota_{\mathbb{B}^\infty}(\tilde{\Phi}^*p) + \langle y_0, p \rangle.
\]

which leads to the given expression. Note that it is also possible to directly derive the dual problem without applying the previous theory (problems (P) and (D)) by introducing the equivalent of the Lagrange dual function (using duality product instead of inner product). It uses the same calculations of the Fenchel conjugate functions.

The dual problem is actually much simpler than the primal one, and we will see in the next paragraphs what information it provides about the solution of the initial problem.

### 4.1.2 Strong duality

The key property that makes dual analysis so insightful for our problem is the strong duality. Relying on [28], Duval and Peyrè proved that strong duality holds between the primal and dual problems presented earlier ([10, Proposition 13]). Let us summarize the implications here.

**Proposition 3.** Strong duality holds between (4.3) and (4.5), such that if there exists a solution \( p^* \) to the dual problem (4.5), then the primal problem (4.3) also admits solutions.

Additionally, the optimal values of both problems coincide, i.e.
\[
\min_{m \in \mathcal{M}(T)} \| m \|_{TV} = \sup_{\| \Phi^*p \|_\infty \leq 1} \langle y_0, p \rangle.
\]

Finally, one can derive extremality conditions. If a solution \( p^* \) exists, then
\[
\tilde{\Phi}^*p^* \in \partial \| m^* \|_{TV}
\]

where \( m^* \) is any solution of the primal (4.3). Conversely, if (4.7) holds, then \((p^*, m^*)\) is a joint couple of solutions to respectively the primal and dual problems.

The strong duality comes really useful for our concern. Indeed, the dual problem is actually simpler than the primal one, and we can adapt the proof from [10, Proposition 2] to ensure existence of solutions.
Proposition 4. Let $m_0 \in \mathcal{M}(\mathbb{T})$ and $y_0 = \tilde{\Phi}m_0 \in \mathbb{R}^L$. There exists a solution to the dual problem

$$\sup_{\|p\|_{\infty} \leq 1} \langle y_0, p \rangle.$$

Proof. As $y_0 = \tilde{\Phi}m_0$ the objective function can be rewritten as

$$\sup_{\|\eta\|_{\infty} \leq 1} \langle m_0, \eta \rangle \leq \|m_0\|_{TV}$$

where $\mathcal{R}(\cdot)$ denotes the range of an operator, and thus is bounded. We can consider $(\eta_n)_{n \in \mathbb{N}}$ a maximizing sequence of the problem. Then $(\eta_n)_{n \in \mathbb{N}}$ lives in the finite dimensional linear space $\mathcal{R}(\tilde{\Phi}^*) = \text{Span}(\tilde{\Phi}_i^*)_{i=1,\ldots,L}$. As $(\eta_n)_{n \in \mathbb{N}}$ is bounded we may extract a subsequence converging to $\eta^* \in C(\mathbb{T})$. $\eta^*$ satisfies both $\|\eta^*\|_{\infty} \leq 1$ and $\eta^* \in \mathcal{R}(\tilde{\Phi}^*)$ so that $\eta^* = \tilde{\Phi}^*p^*$ and $p^* \in \mathbb{R}^L$ is solution of the dual problem.

Thus, as long as the data $y_0$ is sampled from an actual measure $m_0 \in \mathcal{M}(\mathbb{T})$, we have an assurance that the primal problem (4.3) admits solutions. What is even more insightful is that, in this case, it is possible to link the solutions of the primal and the dual problem through the extremality conditions. Duval and Peyré have derived a closed form expression for the subdifferential of the TV-norm [10, Proposition 12], given by:

$$\forall m \in \mathcal{M}(\mathbb{T}), \quad \partial \|m\|_{TV} = \left\{ \eta \in C(\mathbb{T}), \|\eta\|_{TV} \leq 1 \right\} \quad \text{and} \quad \int \eta \, dm = \|m\|_{TV}$$

Then, for problem (4.3), there exists a function $\eta \in C(\mathbb{T})$ coined as dual certificate that verify the three following properties:

(i) $\eta \in \mathcal{R}(\tilde{\Phi}^*) = \text{Span}(\tilde{\Phi}_i^*)_{i=1,\ldots,L}$

(ii) $\|\eta\|_{\infty} \leq 1$

(iii) for any solution $m^*$ of (4.3), $\langle \eta, m^* \rangle = \|m^*\|_{TV}$

Note that such a function $\eta$ might not be unique, and any function satisfying the three above mentioned requirements is called a dual certificate. We could aim for uniqueness by choosing for example the certificate made from the minimal norm $p$ (as made in [10]). The name comes from the fact that the existence of such a function certifies that a solution to the primal problem exists.

4.1.3 Dual certificate

Let us consider a function $\eta$ that is a dual certificate for problem (4.3). $\eta$ is a continuous periodic function with infinite norm being 1, so it necessarily reaches at least once $\pm 1$ over its period. We can thus consider what we call its saturation points defined as:

$$\text{Sat}(\eta) = \{t \in \mathbb{T}, \eta(t) = \pm 1\}$$

and its signed saturation points:

$$\text{Sat}^\pm(\eta) = \{(t, v) \in \mathbb{T} \times \{-1, +1\}, \eta(t) = v\}.$$
Similarly, we can define the signed support of a measure $m \in \mathcal{M}(\mathbb{T})$ as
\[
\text{Supp}^\pm(m) = (m^{-1}(\mathbb{R}_+^*) \times \{+1\}) \cup (m^{-1}(\mathbb{R}_-^*) \times \{-1\}) \subset \mathbb{T} \times \{\pm 1\}.
\]
(4.11)

Then, one can state [10, Proposition 3]:

**Proposition 5.** Let $m_0 \in \mathcal{M}(\mathbb{T})$ and $y_0 = \tilde{\Phi} m_0$. The following equivalence holds:

$m_0$ is a solution to
\[
\min_{m \in \mathcal{M}(\mathbb{T})} \|m\|_{TV}
\]
with $y_0 = \tilde{\Phi}(m)$

if and only if

\[
\text{Supp}^\pm(m_0) \subset \text{Sat}^\pm(\eta)
\]
with $\eta$ being a dual certificate of the problem.

**Proof.** For the direct implication, let $\eta \in \mathcal{C}(\mathbb{T})$ be a dual certificate and suppose there exists $t_1 \in \text{Supp}^\pm(m_0)$ but $t_1 \notin \text{Sat}^\pm(\eta)$. Then $|\eta(t)| < 1$ and, as $\eta$ is a dual certificate, the following holds:

\[
\langle \eta, m_0 \rangle = \|m_0\|_{TV} = \sup_{\psi \in \mathcal{C}(\mathbb{T}), \|\psi\|_{TV} \leq 1} \int \psi dm_0
\]

However, we could consider a function $\psi_1 \in \mathcal{C}(\mathbb{T})$ such that $\psi_1(t_1) = 1$ and $|\psi_1| \geq |\eta|$ everywhere on $\mathbb{T}$. Then $\|m_0\|_{TV} \geq \langle m_0, \psi_1 \rangle > \langle m_0, \eta \rangle = \|m_0\|_{TV}$ which is impossible.

On the other way, suppose $\text{Supp}^\pm(m_0) \subset \text{Sat}^\pm(\eta)$, then $\langle m_0, \eta \rangle = \|m_0\|_{TV}$ which is the optimal value of the dual problem. By strong duality, is is also the optimal value of the primal problem, and it is reached with $m = m_0$, so $m_0$ is a solution of the primal.

This result is crucial in our study, as any dual certificate constrains the location of the support of any solution to the primal problem. In another manner, if we define the set of certificates

\[
\text{Cert}(y_0) = \left\{ \eta \in \mathcal{R}\left(\tilde{\Phi}^*\right), \ \eta \text{ is a dual certificate of (4.3)} \right\},
\]

the following inclusion holds

\[
\left\{ \bigcup_{m \text{ solution of (4.3)}} \text{Supp}^\pm(m) \right\} \subset \left\{ \bigcap_{\eta \in \text{Cert}(y_0)} \text{Sat}^\pm(\eta) \right\}.
\]
(4.12)

We can go even further and obtain a sufficient condition that ensures uniqueness. For any dual certificate $\eta$, we can consider the set

\[
\mathcal{M}_\eta(\mathbb{T}) = \{ m \in \mathcal{M}(\mathbb{T}), \text{Supp}^\pm(m) \subset \text{Sat}^\pm(\eta) \} \subset \mathcal{M}(\mathbb{T})
\]

of the periodic Radon measures with support included in the saturation points of $\eta$ and thus the restriction $\tilde{\Phi}_{\text{Sat}^\pm(\eta)}$ of $\tilde{\Phi}$ to $\mathcal{M}_\eta(\mathbb{T})$. The existence of a solution to the primal problem ensures that there exists at least a measure $m_1$ that solves (4.3), and necessarily $m_1 \in \mathcal{M}_\eta(\mathbb{T})$. If $\tilde{\Phi}_{\text{Sat}^\pm(\eta)}$ would happen to be injective, then there could not be more than one solution to the problem, $m_1$ would be unique. If Sat($\eta$) is infinite, it actually does not bring much information, as $\tilde{\Phi}_{\text{Sat}^\pm(\eta)}$ remains as complex
as $\tilde{\Phi}$. However, it becomes particularly helpful when $\text{Sat}(\eta)$ is finite, as proving that the measurement function is injective reduces to proving that a certain matrix is full rank. It turns the injectivity question into a finite dimensional one.

### 4.2 Towards Uniqueness of Solutions

So far, with very few hypotheses, it has been possible to state strong properties that directly derive from the dual analysis of the general problem (4.3). Indeed we did not need any assumption on the class or the type of the measurement operator $\tilde{\Phi}$. A really interesting work has already been achieved when a Fourier sampling operator is concerned [27]. The authors have proven that the constrained problem (the same as the one we study) either admits only positive solutions or the solution is unique. They also have obtained uniqueness for a certain type of operator applied to the regularized problem. In this section we want to examine to what extent we can leverage the previous properties in our context of spatial sampling with invertible operators.

#### 4.2.1 Spatial sampling

When performing spatial sampling, we obtain a vector $y \in \mathbb{R}^L$ of the $L$ measurements and thus lives in finite dimensional space. It leads to a quite simple expression of the adjoint measurement operator $\tilde{\Phi}^*$ as:

$$
\tilde{\Phi}^* : \begin{cases}
\mathbb{R}^L & \rightarrow \mathcal{C}(T) \\
 h & \mapsto \sum_{\ell=1}^{L} h_\ell \psi_D(\theta_\ell - \cdot) 
\end{cases}. \tag{4.13}
$$

In practice, we are able to plot (and even sometimes to give a closed-form expression) of the Green’s functions involved in (4.13) for some well defined pseudo differential operators $D$. On the examples we have investigated on - namely exponential, Sobolev, Wendland and Matérn operators - they behave quite regularly and admit a finite number of saturation points (see the plots from Fageot and Simeoni in [19]).

Then, as any dual certificate lives in $\mathcal{R}(\tilde{\Phi}^*)$, it is expressed as a linear combination of (reversed) shifted Green’s function, and thus we can expect it to have a finite number of saturation points. This would be quite interesting in the perspective of finding a solution to the primal problem, as one can state the following property:

**Proposition 6.** Let $\eta \in \mathcal{C}$ be a dual certificate function associated with the primal and dual problems (4.3) and (4.5). If $\eta$ admits a finite number of saturation points, then any solution of the primal problem is a discrete measure with a number of impulses no greater than the number of saturation points.

Even stronger, the solution set is a convex and compact subset of the finite-dimensional vector space

$$
\text{Span} (\{ \Pi(\cdot - \tau) : \tau \in \text{Sat}(\eta) \}) ,
$$

and, for any solution $m^* = \sum_{\tau \in \text{Sat}(\eta)} \beta_\tau \Pi(\cdot - \tau)$, the sign of the weights $\beta_\tau$ are given by the sign of the associated saturation points $\tau \in \text{Sat}^\pm(\eta)$.

Stated another way around, any solution of the primal problem expressed in the continuous setting is a spline with a finite rate of innovations. This property naturally
derives from proposition 5 on the inclusion of the support of the solutions.

The uniqueness property discussed earlier could also be made explicit in the case of spatial sampling and discrete measures. Suppose we are in the conditions of proposition 6, let us call $P = |\text{Sat}(\eta)|$ for $\eta$ a dual certificate. The space of measures $\mathcal{M}_\eta(T)$ with support included in the saturation points of $\eta$ is then directly characterized by the $P$ weights of the innovations, and thus is isomorphic to $\mathbb{R}^P$. More precisely, if $\text{Sat}(\eta) = \{\tau_1, \ldots, \tau_P\}$, one can write:

$$\tilde{\Phi}_{\text{Sat}^\pm(\eta)}: \begin{cases} \mathbb{R}^P & \rightarrow \mathbb{R}^L \\ \beta & \mapsto \tilde{\Phi} \left( \sum_{p=1}^{P} \beta_p \text{III}(\cdot - \tau_p) \right) = \left( \sum_{p=1}^{P} \beta_p \psi_D(\theta_\ell - \tau_p) \right)_{\ell=1,\ldots,L} \end{cases}.$$  

(4.14)

We may reformulate:

**Proposition 7.** Let $\eta$ be a dual certificate with a finite set of saturation points $\text{Sat}(\eta) = \{\tau_1, \ldots, \tau_P\}$. If $\tilde{\Phi}_{\text{Sat}^\pm(\eta)}$ (as defined in (4.14)) is injective then the solution of the primal problem is unique. In terms of matrix, it is equivalent for the following matrix $\Psi_\tau \in \mathbb{R}^{L \times P}$ to have full rank:

$$\Psi_\tau = (\psi_D(\theta_\ell - \tau_p))_{\ell,p} = \begin{pmatrix} \psi_D(\theta_1 - \tau_1) & \ldots & \psi_D(\theta_1 - \tau_P) \\ \vdots & \ddots & \vdots \\ \psi_D(\theta_L - \tau_1) & \ldots & \psi_D(\theta_L - \tau_P) \end{pmatrix}. \quad (4.15)$$

Necessarily, if the number $P$ of saturation points of a certificate is greater than $L$, the matrix (4.15) could not satisfy the full rank condition. Thus this property is only interesting when there exists a dual certificate with $L$ or less saturation points.

The conditions stated for both the previous theorems essentially depend on the operator $D$. The first one amounts to estimating the set of saturation points that a linear combination of shifted Green's function with infinite norm of 1 can reach over a period. As the combination is finite, the actual question would be: can such a function saturate to 1 in more than a finite number of points? The second condition is linked to the rank of a finite number of vector evaluations of the Green's function. Thus, the central point of these properties clearly is the pseudo differential operator.

### 4.2.2 Application to the exponential operator

During our practical experiences, we extensively have made use of exponential splines. As Fageot and Simeoni define them [19], they are linked to the operator $D = (D + \alpha \text{Id})^\gamma$, with $\gamma > 1$ so that the Green's function is continuous and thus the operator is sample-admissible [19, Section 6.2]. Let us consider integer $\gamma = N \in \mathbb{N}$, such that $N \geq 2$. We shall ask if such an operator, which is invertible, satisfies the requirements of proposition 6 and proposition 7.

For the sake of readability and without loss of generality, we will now consider that $T = [0,1]$, i.e. the period of the considered objects is 1. Any function $g \in S(\mathbb{R})$ over the real line can be periodized, and thus be seen as a periodic generalized function $\tilde{g} \in S(T) \subset S'(\mathbb{R})$, with the transformation

$$\forall r \in \mathbb{R}, \quad \tilde{g}(r) = g(r - \lfloor r \rfloor).$$
This operation truncates the initial function $g$ and simply loops back to the value at the beginning of the period. It does not ensure any property as continuity or regularity for the function $\tilde{g}$. In the following, we identify both $g$ and $\tilde{g}$ as the periodized element and only note $\{t \mapsto g(t)\} \in S'(\mathbb{T})$ when the variable is in $\mathbb{T}$. If the variable $r$ is taken in $\mathbb{R}$, we specifically write $g(r - \lfloor r \rfloor)$.

With $N = 2$, it is actually possible to derive a closed-form expression for $\psi_D$ knowing that $D\psi_D = \mathbb{I}$ and using the derivative with respect to the periodic distributions. We obtain:

$$D = (D + \alpha \text{Id})^2 \Rightarrow \forall r \in \mathbb{R}, \psi_D(r) = \left(\frac{1}{1-e^{-\alpha}}(r - \lfloor r \rfloor) + \frac{e^{-\alpha}}{(1-e^{-\alpha})^2}\right)e^{-\alpha(r - \lfloor r \rfloor)}$$

(4.16)

In fact, the following theorem holds:

**Theorem 1.** The Green’s function of the periodic exponential operator $D = (D + \alpha \text{Id})^N$ of order $N \in \mathbb{N} \setminus \{0\}$ is given by

$$\forall r \in \mathbb{R}, \psi_D(r) = P_{N-1}(r - \lfloor r \rfloor)e^{-\alpha(r - \lfloor r \rfloor)}$$

(4.17)

where $P_{N-1}$ is a polynomial of degree $(N - 1)$.

The coefficients of $P_{N-1}$ are known and can be computed recursively, starting from the leading coefficient. Let $a_k$ be the coefficients of $P_{N-1}$ according to

$$P_{N-1}(X) = a_{N-1}X^{N-1} + \cdots + a_1X + a_0,$$

then the following holds:

$$\forall k \in \{0, 1, \ldots, N - 1\}, \quad a_k = \frac{1}{k!}b_k$$

(4.18)

and

$$b_{N-1} = \frac{1}{1-e^{-\alpha}}$$

(4.19)

$$\forall k \in \{2, 3, \ldots, N\}, \quad b_{N-k} = A\left[\sum_{i=1}^{k-1} \frac{1}{(k-i)!}b_{N-i}\right] = A\left[\sum_{i=1}^{k-1} \frac{1}{i!}b_{N-k+i}\right].$$

(4.20)

with

$$A = \frac{e^{-\alpha}}{1-e^{-\alpha}}.$$

The proof requires the statement of a short lemma.

**Lemma 5.** Let $\alpha > 0$, for any polynomial $P$, the function $t \mapsto P(t)e^{-\alpha t} \in S'(\mathbb{T})$ satisfies the following equality

$$(D + \alpha \text{Id}) \left(P(t)e^{-\alpha t}\right) = \left[P(0) - P(1)e^{-\alpha}\right] \mathbb{I} + P'(t)e^{-\alpha t}.$$  

(4.21)

where the derivative $D$ is taken in sense of the periodic distributions and $P'$ is the usual derivative of the polynomial $P$. 
Chapter 4. Theoretical analysis of the solution set

Proof. For any periodic test function \( u \in \mathcal{C}_\infty (\mathbb{T}) \), we can write, using integration by parts:

\[
\left< (D + \alpha \text{Id}) (P(t)e^{-\alpha t}), u \right> = \int_0^1 -P(t)e^{-\alpha t} u'(t) + \alpha P(t)e^{-\alpha t} u(t) dt
\]

\[
= - \left\{ [P(t)e^{-\alpha t} u(t)]_0^1 - \int_0^1 [P'(t) - \alpha P(t)] e^{-\alpha t} u(t) dt \right\} + \int_0^1 \alpha P(t)e^{-\alpha t} u(t) dt
\]

\[
= - [P(t)e^{-\alpha t} u(t)]_0^1 + \int_0^1 P'(t)e^{-\alpha t} u(t) dt
\]

\[
= P(0)u(0) - P(1)e^{-\alpha}u(1) + \int_0^1 P'(t)e^{-\alpha t} dt
\]

\[
= \left< (P(0) - P(1)e^{-\alpha}) \text{III} + P'(t)e^{-\alpha t}, u \right>
\]

which allows to conclude by periodicity of \( u \), as \( u(0) = u(1) \). \( \square \)

Note that this formula is still valid when \( P \) is of degree 0 (a constant polynomial) with \( P' = 0 \). We shall now demonstrate the previous theorem.

Proof of Theorem 1. Let \( D = (D + \alpha \text{Id})^N \) for a given \( N \in \mathbb{N}, N \geq 2 \). Let us verify that \( t \mapsto P_{N-1}(t)e^{-\alpha t} \) is solution of the Green’s function equation \( D\psi_D = \text{III} \). By iteratively applying \( N \) times the formula (4.21), one obtains

\[
(D + \alpha \text{Id})^N \left[ P_{N-1}(t)e^{-\alpha t} \right] = \sum_{k=0}^{N-1} \left( P_{N-1}^{(k)}(0) - P_{N-1}^{(k)}(1)e^{-\alpha} \right) (D + \alpha \text{Id})^{N-1-k} \{\text{III}\}.
\]

(4.22)

Note that, as a polynomial, \( P_{N-1} \) satisfies:

\[
P_{N-1}^{(k)}(0) = (k!) a_k
\]

\[
P_{N-1}^{(k)}(1) = \sum_{i=0}^{N-1-k} \frac{(k+i)!}{i!} a_{k+i}.
\]

Then, for any \( k \in \{0, 1, \ldots, N-2\} \),

\[
P_{N-1}^{(k)}(0) - P_{N-1}^{(k)}(1)e^{-\alpha} = (k!) a_k - e^{-\alpha} \sum_{i=0}^{N-1-k} \frac{(k+i)!}{i!} a_{k+i}
\]

\[
= b_k - e^{-\alpha} \sum_{i=0}^{N-1-k} \frac{1}{i!} b_{k+i}
\]

\[
= (1 - e^{-\alpha}) b_k - e^{-\alpha} \sum_{i=1}^{N-1-k} \frac{1}{i!} b_{k+i}
\]

\[
= (1 - e^{-\alpha}) A \sum_{i=1}^{N-1-k} \frac{1}{i!} b_{k+i} - e^{-\alpha} \sum_{i=1}^{N-1-k} \frac{1}{i!} b_{k+i}
\]

\[
= 0
\]

and for \( k = N-1 \),

\[
P_{N-1}^{(N-1)}(0) - P_{N-1}^{(N-1)}(1)e^{-\alpha} = (N-1)! a_{N-1} (1 - e^{-\alpha}) = 1
\]
Equation (4.22) turns into

\[ D \left( P_{N-1}(t)e^{-\alpha t} \right) = \mathcal{I}. \]

As \( D \) is invertible, the solution is unique and \( \psi_D(t) = P_{N-1}(t)e^{-\alpha t} \) for any \( t \in \mathbb{T} \).

**Remark 5.** In a matrix setting, the coefficients \( b_k \) are obtained as the vector

\[
\mathbf{b} = (b_{N-1} \ b_{N-2} \ \ldots \ b_1 \ b_0)^T
\]

solution of the eigenvector problem

\[ \mathbf{M}\mathbf{b} = \frac{1}{A}\mathbf{b} \quad (4.23) \]

with \( b_{N-1} = 1/(1 - e^{-\alpha}) \) and

\[
\mathbf{M} = \begin{bmatrix}
A^{-1} & 0 & \ldots & 0 \\
1 & 0 & \ldots & 0 \\
1/2! & 1 & 0 & \ldots \\
& 1/2! & \ddots & \ddots \\
& & \ddots & \ddots \\
& & & \ddots & \ddots \\
\frac{1}{(N-2)!} & \ldots & \ldots & 1 & 0 & 0 \\
\frac{1}{(N-1)!} & \frac{1}{(N-2)!} & \ldots & 1/2! & 1 & 0
\end{bmatrix}
\quad (4.24)
\]

\( \mathbf{M} \) is lower triangular so that the computation of \( \mathbf{b} \) is actually straightforward, the coefficients being determined as soon as \( b_{N-1} \) is given.

**Remark 6.** One might notice that the polynomials \( P_N \) are the consecutive derivative one of the other, according to the equation \((P_N)' = P_{N-1}\). Indeed, from equation (4.21), one obtains

\[
(D + \alpha \text{Id})^{N+1} \left( P_N(t)e^{-\alpha t} \right) = (D + \alpha \text{Id})^N \left\{ \left[ P_N(0) - P_N(1)e^{-\alpha} \right] \mathcal{I} \right\}
\]

\[
+ (D + \alpha \text{Id})^N \left\{ (P_N)'(t)e^{-\alpha t} \right\}
\]

\[
= (D + \alpha \text{Id})^N \left\{ (P_N)'(t)e^{-\alpha t} \right\}
\]

\[
= \mathcal{I}
\]

which allows to conclude.

Such a Green’s function behaves pretty well over the period. It has continuous derivative up to the order \((N - 2)\) (see [19, Section 5.1]), so it is always continuous for \( N \geq 2 \), and it admits a unique maximum. Now if we pay attention to the expression of the adjoint operator (4.13), we notice that any dual certificate \( \eta \) would be expressed as a linear combination of such shifted Green’s functions \( \psi_D(\theta - \cdot) \). It is actually not a spline as the Green’s functions are evaluated in the “reverse” sense (it would need \((\cdot - \theta)\) to be a spline, and the Green’s functions are not symmetric with respect to their knot with an exponential operator). The question that arises from now on is: can such a finite linear combination of (reversed) shifted Green’s function admit an infinite number of saturation points (given that its infinite norm is equal to 1)? For the sake of generality, we will consider the case of regular splines instead of certificates, but the conclusions would directly transfer as the reversing operation does not affect this property.
Let us analyze how such a spline behaves. For \( D = (D + \alpha \operatorname{Id})^N \) with \( N \geq 2 \), consider the function \( f \in \mathcal{M}_D(\mathbb{T}) \) defined as

\[
\forall t \in \mathbb{T}, \quad f(t) = \sum_{\ell=1}^L \beta_\ell \psi_D(t - \theta_\ell)
\]

such that \( \|f\|_\infty = 1 \). As \( \psi_D \) is continuous, \( f \) is also continuous.

First let us rewrite how the shifting operation modifies the expression of the Green’s function. Consider \( \theta \in \mathbb{T} \), then

\[
\forall t \in \mathbb{T}, \quad t - \theta - |t - \theta| = \begin{cases} 
  t - \theta & \text{if } \theta \leq t < 1 \\
  t + 1 - \theta & \text{if } 0 \leq t < \theta
\end{cases}
\]

which leads to

\[
\forall t \in \mathbb{T}, \quad \psi_D(t - \theta) = P(t - \theta - |t - \theta|) e^{-\alpha(t - \theta - |t - \theta|)}
\]

\[
= \begin{cases} 
  P(t - \theta) e^{\alpha \theta} e^{-\alpha t} & \text{if } \theta \leq t < 1 \\
  P(t + 1 - \theta) e^{\alpha (\theta - 1)} e^{-\alpha t} & \text{if } 0 \leq t < \theta
\end{cases}
\]

\[
= \begin{cases} 
  P_1(t) e^{-\alpha t} & \text{if } \theta \leq t < 1 \\
  P_2(t) e^{-\alpha t} & \text{if } 0 \leq t < \theta
\end{cases}
\]

with \( P_1(\theta) = P_2(\theta) = P(0) \) by continuity of \( \psi_D \). Thus, any shifted Green’s function can be expressed as a piecewise polynomial over the period multiplied by a decreasing exponential. From this statement, we can deduce the following result:

**Proposition 8.** Let \( N \in \mathbb{N} \setminus \{0, 1\} \) and \( D = (D + \alpha \operatorname{Id})^N \). Any continuous \( D \)-spline \( f = \sum_{\ell=1}^L \beta_\ell \psi_D(\cdot - \theta_\ell) \) with infinite norm \( \|f\|_\infty = 1 \) admits a finite number \( P \) of saturation points \( t \in \mathbb{T} \) such that \( |f(t)| = 1 \), and \( P \leq LN \).

**Proof.** Consider the \( L \)-partitioning of \( \mathbb{T} \) in the sequence of the \( L \) intervals \((\theta_\ell, \theta_{\ell+1}]\) with the convention \( \theta_{L+1} = \theta_1 \). On each subinterval, \( f \) can actually be written as \( f_\ell(t) = P_\ell(t) e^{-\alpha t} \) with \( P_\ell \) being a polynomial of degree \((N - 1)\). That way, \( f \) admits at most \((N - 1)\) saturation points on each interval. Indeed, the norm constraint enforces that any saturation point is a local extremum of \( f_\ell \) so that \( f_\ell' \) is 0 at these points. \( f_\ell' \) is written as a product between a polynomial of degree \((N - 1)\) and \( e^{-\alpha t} \), and thus cannot have more than \((N - 1)\) zeros. By summing over all the subintervals and adding the \( L \) innovation positions, there are at most \( L + LN(N - 1) = LN \) saturation points over \( \mathbb{T} \).

Note that the proof of the previous proposition exhibits an upper bound for the number of saturation points, but it might not be sharp at all, and it is probably possible to determine a much lower bound.

Numerous questions remain still not answered. So far, we have not succeeded in proving Proposition 7 applied to the case of exponential Green’s function. However, the research is ongoing and we believe that it is possible to determine a set conditions according to which an operator \( D \) would ensure uniqueness in the solution set. Moreover, the same analysis of the expression of periodic splines can probably be pursued with the other common differential operators (like the Sobolev operator).
The other way around, one may also ask if it is possible, starting from a function that would satisfy proposition 6 and proposition 7, to build an associated spline-admissible operator. In that case, such an operator would ensure that the solution is unique and is always a spline.
Conclusion and acknowledgments

The explicit objectives of this internship were to understand, develop and adapt the three algorithms, namely the equispaced knots, Cadzow Plug-and-Play Gradient Descent and Frank-Wolfe, to the context of reconstruction of periodic signal with sparse innovations and non-uniform spatial sampling. These goals have successfully been achieved. We even came out with an improved version of the Frank-Wolfe algorithm that outperforms the regular one in all the tested scenarios. On the way, we have developed our understanding of the mechanics at play. It has led us to really interesting theoretical questions, involving duality and convex optimization theory.

On that topic, we have designed a demonstration framework that would allow to both characterize the shape of the elements of the solution set and to ensure uniqueness of the solution. We have succeeded to start applying this framework to the case of exponential operators. In particular, we came up with the general expression of the Green’s function associated to the one-dimensional periodic exponential operator. However, this work remains uncompleted, and various ideas are on the verge to be considered in that direction. In further works, it would be really meaningful to focus on other operators such as Sobolev or Wendland. We strongly believe that many results would logically come out of the current achieved work on the topic.

As far as I am concerned, this internship has been extremely enriching and educative. Although essentially theoretic, it has brought me to discover compress sensing, which is a field I had very limited experience with. Beyond mathematical material and programming skills, these six months of research work have also taught me how to pursue an investigation work in an academic environment, by planning my working objectives and making sure I keep the motivation and interests high enough to make progress. This is a significant personal improvement compared to my previous research experience two years ago. The maturity I have acquired from that former internship has definitely helped me to make this current experience more productive and rewarding.

Doing a PhD is the logical follow-up for this work, and this is what I am willing to enroll myself in. The academic research environment fits well with my personality and my way of learning and discovering. I am also interested in continuing this project. A lot of progress can be made in very different directions, and I would really enjoy turning it into some concrete contributions. I have already applied to a doctoral school at EPFL. The choices of an advisor and a topic are really crucial and I would like to be conscientious in making them. That is why I will also make use of the following weeks to browse thesis proposals and to carefully explore other opportunities. Nonetheless, I have really appreciated working at EPFL, and specifically in LCAV, with my supervisors and their colleagues from other laboratories and I am sure it would provide a great environment to produce a high quality successful PhD thesis.

During this internship, I have been closely supervised by both Matthieu and Julien. I would like to deeply thank them for the help, patience and attention they provided.
about my work, and I am sincerely looking forward to keeping this fruitful working environment with them.
Bibliography

[1] R. Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: Journal of the Royal Statistical Society (Series B) 58 (1996), pp. 267–288.

[2] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. “Atomic Decomposition by Basis Pursuit”. In: SIAM Journal on Scientific Computing 20.1 (1998), pp. 33–61. DOI: 10.1137/S1064827596304010. eprint: https://doi.org/10.1137/S1064827596304010. URL: https://doi.org/10.1137/S1064827596304010.

[3] Ryan J. Tibshirani. “The lasso problem and uniqueness”. In: Electron. J. Statist. 7 (2013), pp. 1456–1490. DOI: 10.1214/13-EJS815. URL: https://doi.org/10.1214/13-EJS815.

[4] Alnur Ali and Ryan J. Tibshirani. The Generalized Lasso Problem and Uniqueness. 2019. arXiv: 1805.07682 [math.ST].

[5] Laurent Condat. “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms”. In: Journal of Optimization Theory and Applications 158.2 (2013), pp. 460–479. DOI: 10.1007/s10957-012-0245-9. URL: https://hal.archives-ouvertes.fr/hal-00609728.

[6] Amir Beck and Marc Teboulle. “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems”. In: SIAM J. Imaging Sciences 2 (Jan. 2009), pp. 183–202. DOI: 10.1137/080716542.

[7] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. “Sparse Sampling of Signal Innovations”. In: IEEE Signal Processing Magazine 25.2 (2008), pp. 31–40.

[8] Matthieu Simeoni, Adrien Besson, Paul Hurley, and Martin Vetterli. CPGD: Cadzow Plug-and-Play Gradient Descent for Generalised FRI. 2020. arXiv: 2006.06374 [math.OC].

[9] Kristian Bredies and Hanna Katriina Pikkarainen. “Inverse problems in spaces of measures”. en. In: ESAIM: Control, Optimisation and Calculus of Variations 19.1 (2013), pp. 190–218. DOI: 10.1051/cocv/2011205. URL: http://www.numdam.org/item/COCV_2013__19_1_190_0.

[10] Vincent Duval and Gabriel Peyré. Exact Support Recovery for Sparse Spikes Deconvolution. 2014. arXiv: 1306.6909 [math.OC].

[11] Emmanuel Candes and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-Resolution. 2012. arXiv: 1203.5871 [cs.IT].

[12] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. “Recovery of Sparse Translation-Invariant Signals With Continuous Basis Pursuit”. In: IEEE Transactions on Signal Processing 59.10 (2011), pp. 4735–4744. DOI: 10.1109/TSP.2011.2160058.
Bibliography

[13] Vincent Duval and Gabriel Peyré. “Sparse Spikes Super-resolution on Thin Grids II: the Continuous Basis Pursuit”. In: Inverse Problems 33.9 (Aug. 2017). DOI: 10.1088/1361-6420/aa7fce. URL: https://hal.archives-ouvertes.fr/hal-01389956.

[14] Yohann de Castro and Fabrice Gamboa. Exact Reconstruction using Beurling Minimal Extrapolation. 2012. arXiv: 1103.4951 [math.ST].

[15] Karin C Knudson, Jacob Yates, Alexander Huk, and Jonathan W Pillow. Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger. 2014. URL: https://proceedings.neurips.cc/paper/2014/file/c06d06da9666a219db15cf575aff2824-Paper.pdf.

[16] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy. 2018. arXiv: 1811.06416 [math.NA].

[17] Michael Unser, Julien Fageot, and John Paul Ward. “Splines Are Universal Solutions of Linear Inverse Problems with Generalized TV Regularization”. In: SIAM Review 59.4 (Jan. 2017), pp. 769–793. ISSN: 1095-7200. DOI: 10.1137/16m1061199. URL: http://dx.doi.org/10.1137/16M1061199.

[18] Thomas Debarre, Quentin Denoyelle, Michael Unser, and Julien Fageot. Sparsest Continuous Piecewise-Linear Representation of Data. 2020. arXiv: 2003.10112 [math.OC].

[19] Julien Fageot and Matthieu Simeoni. TV-based Reconstruction of Periodic Functions. 2020. arXiv: 2006.14097 [math.OC].

[20] Matthieu Martin Jean-Andre Simeoni. “Functional Inverse Problems on Spheres: Theory, Algorithms and Applications”. In: (2020), p. 326. DOI: 10.5075/epfl-thesis-7174. URL: http://infoscience.epfl.ch/record/275337.

[21] Anais Badoual, Julien Fageot, and Michael Unser. “Periodic Splines and Gaussian Processes for the Resolution of Linear Inverse Problems”. In: IEEE Transactions on Signal Processing 66.22 (Nov. 2018), pp. 6047–6061. ISSN: 1941-0476. DOI: 10.1109/TSP.2018.2873514. URL: http://dx.doi.org/10.1109/TSP.2018.2873514.

[22] Matthieu Martin Jean-Andre Simeoni. Periodispline. https://github.com/mattieumee/periodispline. Accessed: 2020-07-21.

[23] Harshit Gupta, Julien Fageot, and Michael Unser. “Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization”. In: IEEE Transactions on Signal Processing 66.17 (2018), pp. 4670–4684.

[24] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Activity Identification and Local Linear Convergence of Forward–Backward-type methods. 2018. arXiv: 1503.03703 [math.OC].

[25] Marguerite Frank and Philip Wolfe. “An algorithm for quadratic programming”. In: Naval Research Logistics Quarterly 3.1-2 (1956), pp. 95–110. DOI: 10.1002/nav.3800030109. Eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800030109. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800030109.

[26] Jean-Baptiste Courbot and Bruno Colicchio. “A Fast Homotopy Algorithm for Gridless Sparse Recovery”, working paper or preprint. Sept. 2020. URL: https://hal.archives-ouvertes.fr/hal-02940848.
[27] Julien Fageot, Thomas Debarre, and Quentin Denoyelle. *On the Uniqueness of Inverse Problems with Fourier-domain Measurements and Generalized TV Regularization*. 2020. arXiv: 2009.11855 [math.OC].

[28] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics, 1999. DOI: 10.1137/1.9781611971088. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611971088. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611971088.