ETH zürich

Spline-based regression and decomposition of spatio-temporal spherical fields with seasonal and trend components.

Guillemet Vincent

Supervisors

Prof. Dr. Habib Ammari
ETH Zurich, Dep. of Mathematics

Dr. Simeoni Matthieu
EPF Lausanne, Imaging

Dr. Fageot Julien
EPF Lausanne, LCAV

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

SPLINE-BASED REGRESSION AND DECOMPOSITION OF SPATIO-TEMPORAL SPHERICAL FIELDS WITH SEASONAL AND TREND COMPONENTS

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	First name(s):
GUILLEMET	VINCENT PATRICK LAWRENCE

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	Signature(s)
Zürich, 14/02/2022	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Abstract

We study the composite problem of recovering a spatio-temporal scalar field from finitely many, possibly noised samples, while dismantling this same field into a periodic and trending one. In doing so we bring a model that allows reconstruction, estimation, possibly prediction and interpretation. We start by settling the needed framework to then define a new optimization problem on a product space of measures, one accounting for the periodic part of the field, the other for the trending part. This new problem is shaped like a infinite-dimensional TV penalised regression problem, with a product search space and regularized with a couple of differential operators, one for the season and one for the trend. Then we prove a representer theorem for the problem, characterizing its solution set as the non-empty closed convex hull of its extreme points with the latter having the shape of a sum of one seasonal integrated Dirac flow and one trending integrated Dirac flow with the sum of the seasonal and trending innovations being smaller or equal to the number of samples. We then build up a discretized version of the problem and prove that under very weak assumptions we have pointwise convergence and convergence of the norms, from the solutions of the discretized problem to a solution of the original problem, for separately the seasonal and the trending parts. Moreover we show how to strengthen this convergence into a uniform one. We finish by providing a Python code implementing the tools to solve the discretized problem, by doing so we add a numerical analysis of the problem and its robustness to sparsity, noise, over-determination and we finally apply our methods on a real meteorological data set.

Acknowledgments

I would like to warmly thank Fageot Julien and Simeoni Matthieu for their supervision and mentor-ship. The present work is the realisation of 5 months of work, which at each decisional intersection under their advice took a very fruitful turn. Paved is this realisation of many results inspired by their prior and current work, specifically quoted as if or not. Here lies the dynamic mixture of my work, their experience and knowledge, our time. I am also indebted to Ammari Habib, who made the event of this project coming to life, possible. Notwithstanding the above, my work, this work is the direct child of me personal life well being, framing and substancing it are my friends, my loved one Katharina, my mother Marie-Christine, that I cherish unconditionally, and last but not least my dog Winter.

Contents

1 Introduction

One day a practitioner, a science enjoyer, started querying data about an interesting phenomenon at a given place on our planet. After a short period of time he convinced his friends, the science enjoyers, to do the same. Few years have passed by and the group of friends have now data about this phenomenon, data against time, but also against the position on our planet, where it has been queried. Forgetting the delay the first enjoyer took to convince his friends, they now have:

$$Phenomenon(p_i, t_j) \in \mathbb{R}. \quad p_i \in \mathbb{S}^2, i \leq N. \quad t_j \in \mathbb{R}, j \leq M.$$

Nevertheless, they are not happy, their data is noised because the hand of second guy was shaking and the fourth one was drunk. Being a countable group of friends, their data is spatially and temporally sparse, there are holes. They don't like holes. They want more, they want a denoised continuous model. To do so they decide to use the model developed in [Gui20]. In their possession is now a continous spatio-temporal, eventually denoised, interpolation of the *Phenomenon*. They plot it, using their favorite Python package, and realise that the *Phenomenon* is essentially the superposition of a seasonal part and a trending part, which we will call for now on f_{seas} and f_{trend}. In the group, the general wondering is now, whether it is possible, or not, to estimate at the same time f_{seas} and f_{trend}. Doing so would allow them to understand how the *Phenomenon* fluctuate against time, its amplitude, period. If there is a general trending behavior and eventually would allow them to do prediction on the future behavior of the *Phenomenon*. They thus seek interpretability. One of our science enjoyer is highly knowledgeable in the field of Functional Inverse Problem, has read for example [Jea20; Mat20] and propose the following setting, the one that we will study in the semester paper. Given data, that is the empirical evaluation of a spatio-temporal scalar function:

$$f(p_i, t_j) \in \mathbb{R}. \quad p_i \in \mathbb{S}^2, i \leq N. \quad t_j \in \mathbb{R}, j \leq M.$$

We want to find f_{seas} and f_{trend} such $f_{seas} + f_{trend}$ is a good fit to f.

▶ **Remark 1.1.** In the following we will often have to evaluate a function, say g, on a set of points of the form (p_i, t_j). For notational convenience, we give the definitions:

$$\mathrm{PT} := \begin{bmatrix} (p_1, t_1) & \dots & (p_1, t_M) \\ & \dots & \\ (p_N, t_1) & \dots & (p_N, t_M) \end{bmatrix}. \qquad g(\mathrm{PT}) := \begin{bmatrix} g(p_1, t_1) & \dots & g(p_1, t_M) \\ & \dots & \\ g(p_N, t_1) & \dots & g(p_N, t_M) \end{bmatrix}.$$

◀

A good fit will then be quantified by a cost functional of the form

$$(f_{seas}, f_{trend}) \mapsto \mathrm{C}(f(\mathrm{PT}), f_{seas}(\mathrm{PT}) + f_{trend}(\mathrm{PT})) \in \mathbb{R}^+.$$

It is easy to see that minimizing such a functional is in general ill-posed, in particular, in the sense that infinitely many functions (with arbitrary regularity and compact support) can achieve the 0 value. Then, the extensive literature and in particular the success of the method presented in [Gui20] suggest to penalize the problem in order to discard a couple (f_{seas}, f_{trend}) with a "bad" behavior. To do so we had to the cost functional a penalization functional of the form:

$$\mathrm{W}(\|\mathcal{D}_s f_{seas}\|, \|\mathcal{D}_t f_{trend}\|^{'}),$$

and get:

$$(f_t, f_s) \mapsto \mathrm{C}(f(\mathrm{PT}), f_s(\mathrm{PT}) + f_t(\mathrm{PT})) + \mathrm{W}(\|\mathcal{D}_s f_s\|, \|\mathcal{D}_t f_t\|^{'}) \in \mathbb{R}^+. \tag{1.1}$$

To get to the previous equation we used few ingredients. First we replaces *seas* and *trend* by s and t, what we will do systematically in the following, in order to get a "light" notation. Secondly the penalization functional W takes two parameters, which contrast with the usual Functional Inverse Problem setting. We do so as we will need to balance the penalization of the *seasonal* and *trending* term separately, not just their sum. Thirdly we used the notion of differential operators, the $\mathcal{D}$'s, and norms, the $\|\cdot\|$'s. Even though these notion will be discussed in further details, lets have here a simple but interested look to them, in order to complete the picture. The use of norms is very intuitive in this setting, as it allows us to quantify the heaviness of a function. Nevertheless one needs to choose it wisely so that the induced geometry makes first our problem as easy to solve as possible and secondly reflects well what the practitioner means by "heaviness", "roughness" or just bad behavior. Extensive work have been done in this direction [Wil10], and the sparsity-promoting TV, Total Variation norm is a natural choice to start with. Finally the use of differential operators can be motivated in several ways, among them we give two. First by thinking about the curvature of a $\mathbb{R} \to \mathbb{R}$ mapping, given by the second derivative, one can penalize using this notion. More generally using the Laplacian. Secondly one can think of a function, lets say [0,1], as a superposition of basis functions. Taking a basis function to be a *Green*'s function and then differentiating it, yields a *Dirac* mass with a certain amplitude, the innovation. A field can thus be informally seen as the superposition of possibly infinitely many integrated innovations and penalizing using the number of innovations and their amplitude is meaningful. This paper is devoted to the theoretical and numerical study of Equation (1.1).

2 Contributions

We here list the contributions in the natural time event ordering. Our whole contribution consist generally in creating, analysing, implementing informatically, a framework that allows to do the decomposition of a scalar spatio-temporal sampled field into a continuous one as the superposition of a seasonal and trending components. To do so we first define a new search space $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ defined as the cartesian product between a search space for the seasonal component $\mathcal{M}_{\mathcal{D}_s}$ and a search space for the trending component $\mathcal{M}_{\mathcal{D}_t}$. We equip this product search space with a linear combination of TV norms and then analyse its properties, find it to be a Banach space, characterize the extremal points of the balls inside and its pre-dual. Secondly we propose an optimization problem, which we think is meaning full to serve our purpose, in the form of

$$\underset{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{\operatorname{argmin}} \{C(\mathbf{Y}, \Phi(F,G)) + \alpha\|\mathcal{D}_t F\|_{TV} + \beta\|\mathcal{D}_s G\|_{TV}\}.$$

Thirdly, we then provide a representer theorem, that characterizes the solution set of the previous problem as the non empty *weak** closed convex hull of its extreme points and provide the general shape of these extreme points. Fourthly, in order to have a practical theory we first set the study of a discretized version of the previous problem. We thus show that under weak assumptions, a sequence of solutions (F^n, G^n) to the discretized problem with increasingly refined discretrized search space has a sub-sequence, here also named (F^n, G^n), that converges in the following ways to a solution $(F^\star, G^\star)$ of the original problem.

$$\lim_{n\to\infty} (F^n, G^n)(p,t) = (F^\star, G^\star)(p,t) \quad \forall (p,t) \in \mathbb{S}^2 \times \mathbb{R}.$$

$$\lim_{n\to\infty} \|\mathcal{D}_t F^n\|_{TV} = \left\|\mathcal{D}_t F^\star\right\|_{TV}, \quad \lim_{n\to\infty} \|\mathcal{D}_s G^n\|_{TV} = \left\|\mathcal{D}_s G^\star\right\|_{TV}.$$

Where the second convergence is, to the best of our knowledge, the first of its kind to be established in this context and can be thought in the following. The weight put respectively on the seasonal and the trending part of a solution to the discretized problem will convergence to an equilibrium, that is given by the solution to the original problem. Over and above we provide sufficient conditions to strengthen the previous convergence into a local uniform one. Fifthly we use an informatical implementation of our own of the problem in order to provide a terse numerical analysis which features the following. A basic reconstruction task test on synthetic data which will show the efficiency of our method to dismantle the different parts of the synthetic field. A couple of robustness task tests on a synthetic field, accounting for sparsed data, noised data, over-determined model and we find that our model is robust against all the previously named tests. Finally we provide an example based on a real meteorological data set, and thus bring evidence that our method can be applied on real-world problem.

Part I

Theory of spatio-temporal field dismantling

3 A review of classical spaces

In this chapter we will review famous functional spaces, that will be useful for our later study of Equation (1.1). The review is neither exhaustive nor complete and is here for the sake of clarity, completeness. Being classical we will allow ourselves to go fast. Moreover in the following we will use the notation $X \in \{\mathbb{S}^1, \mathbb{S}^2, \mathbb{R}\}$ to represent the fact that X can be defined as being $\mathbb{S}^1$ or $\mathbb{S}^2$ or $\mathbb{R}$.

▶ **Definition 3.1 (*Schwartz* space).** For $X \in \{\mathbb{S}^1, \mathbb{S}^2, \mathbb{R}, \mathbb{S}^2 \times \mathbb{R}, \mathbb{S}^2 \times \mathbb{S}^1\}$

$$\mathcal{S}(X) := \{\phi \in C^\infty(X) | \quad \forall \alpha \in \mathbb{N}^{\dim(X)} \forall \beta \in \mathbb{N}^{\dim(X)} \quad \left\| x^\alpha \mathrm{D}^\beta \psi \right\|_\infty < \infty\}. \tag{3.1}$$

Where α and β are multi-index integers, D^β corresponds to the operator that differentiates β_i times in the i_{th} coordinate, x^α corresponds to $\Pi_{i=1}^{\dim(X)} x_i^{\alpha_i}$. The space $\mathcal{S}(X)$ is equipped with the coarsest topology that makes the semi-norms $\left\| x^\alpha \mathrm{D}^\beta \cdot \right\|_\infty$ continuous. ◀

▶ **Definition 3.2 (Tempered distribution).** For $X \in \{\mathbb{S}^1, \mathbb{S}^2, \mathbb{R}, \mathbb{S}^2 \times \mathbb{R}, \mathbb{S}^2 \times \mathbb{S}^1\}$ $\mathcal{S}'(X)$ is defined to be the topological dual of $\mathcal{S}(X)$, that is, the set of all linear continuous (with respect to the topology on $\mathcal{S}(X)$) functionals. The space $\mathcal{S}'(X)$ is equipped with the *weak** topology. That is, the coarsest that makes the following mappings continuous:

$$\forall \phi \in \mathcal{S}(X), \|\cdot\|_\phi : f \mapsto f(\phi). \tag{3.2}$$

We then say for $\{f^n\}_{n=1}^\infty \in \mathcal{S}'(X)$ and $f \in \mathcal{S}'(X)$ that f^n is *weak** convergent or convergent in the topology to f if:

$$\forall \phi \in \mathcal{S}(X) : \quad \lim_{n \to \infty} (f^n - f)(\phi) = 0, \tag{3.3}$$

then we write $\lim\limits_{n \to \infty} f^n = f$. ◀

▶ **Definition 3.3.** For $X \in \{\mathbb{S}^1, \mathbb{S}^2, \mathbb{R}, \mathbb{S}^2 \times \mathbb{R}, \mathbb{S}^2 \times \mathbb{S}^1\}$

$$C_0(X) := \{\phi \in C(X) | \quad \lim_{|x| \to \infty} f(x) = 0\}. \tag{3.4}$$

In the case where X is bounded, this is just the set of continuous functions. Moreover, together with the infinity norm $\|\cdot\|_\infty$ and its topology, one can prove that $C_0(X)$ is a *Banach* space. ◀

▶ **Definition 3.4.** For $X \in \{\mathbb{S}^1, \mathbb{S}^2, \mathbb{R}, \mathbb{S}^2 \times \mathbb{R}, \mathbb{S}^2 \times \mathbb{S}^1\}$ $\mathcal{M}(X)$ is defined to be the dual space of $C_0(X)$. Being the dual of a *Banach* space, it is itself a *Banach* space. ◀

This previous definition is motivated by the $Riesz - Markov - Kakutani$ theorem, that will be stated and used later on.

▶ **Remark 3.5.** One can see from the previous definitions that

$$\mathcal{M}(X) \subset \mathcal{S}'(X). \tag{3.5}$$

◀

We now have defined the principal spaces and can start putting things together. The goal for now is to define a search space for f_s and a search space for $2f_t$. Remark that for the seasonality of f_{seas} we want:

$$\exists T : \quad \forall (p,t) \in \mathbb{S}^2 \times \mathbb{R} \quad f_s(p,t) = f_s(p, t+T). \tag{3.6}$$

Thus $f_{seas}(\cdot,\cdot) : \mathbb{S}^2 \times \mathbb{R} \to \mathbb{R}$ can be reduced to $f_{seas}(\cdot, \cdot * T) : \mathbb{S}^2 \times \mathbb{S}^1 \to \mathbb{R}$. By doing this manipulation the other way around, provided that we know T, from a function $g(\cdot,\cdot) : \mathbb{S}^2 \times \mathbb{S}^1 \to \mathbb{R}$ we get a candidate for f_{seas} by $g(\cdot, \cdot/T) : \mathbb{S}^2 \times \mathbb{R} \to \mathbb{R}$. A consequence of this remark is that in the following candidates for f_{trend} will have as domain $\mathbb{S}^2 \times \mathbb{R}$ and candidates for f_{seas} will have as domain $\mathbb{S}^2 \times \mathbb{S}^1$. In order to make up our mind, lets suppose for now that $f_{seas} \in \mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^1)$ and $f_{trend} \in \mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^1)$. One of the principal feature of the paper is to consider, as we will see later, only separable differential operators. Meaning that their action in time and space can be separated, formally by a tensor product. A consequence of this choice is that it is enough to define them on the following spaces.

▶ **Definition 3.6.** For $X \in \{\mathbb{S}^1, \mathbb{R}\}$

$$\mathcal{S}(\mathbb{S}^2) \otimes \mathcal{S}(X) := \left\{ \sum_{i=1}^{I} \gamma_i a_i \otimes b_i \in \mathcal{S}(\mathbb{S}^2 \times X) \middle| \quad I \in \mathbb{N}, \{a_i\}_{i=1}^I \subset \mathcal{S}(\mathbb{S}^2), \{b_i\}_{i=1}^I \subset \mathcal{S}(X), \{\gamma_i\}_{i=1}^I \subset \mathbb{R} \right\}. \tag{3.7}$$

$$\mathcal{S}'(\mathbb{S}^2) \otimes \mathcal{S}'(X) :=$$
$$\left\{ \sum_{j=1}^{J} \gamma_j g_j \otimes h_j \in \mathcal{S}'(\mathbb{S}^2 \times X) \middle| \quad I \in \mathbb{N}, \{g_j\}_{j=1}^J \subset \mathcal{S}'(\mathbb{S}^2), \{h_j\}_{j=1}^J \subset \mathcal{S}'(X), \{\gamma_j\}_{j=1}^J \subset \mathbb{R} \right\}. \tag{3.8}$$

◀

We can link the previously defined tensor product spaces to our search spaces by the following conjecture.

▶ **Lemma 3.7.** For $X \in \{\mathbb{S}^1, \mathbb{R}\}$

1. $\mathcal{S}(\mathbb{S}^2 \times X) \cong \mathcal{S}(\mathbb{S}^2) \hat{\otimes} \mathcal{S}(X)$.
2. $\mathcal{S}'(\mathbb{S}^2 \times X) \cong \mathcal{S}'(\mathbb{S}^2) \hat{\otimes} \mathcal{S}'(X)$.

◀

Few remarks are to be made here. First of all the hat on the tensor product stands for the completion of the underlying space. Then, the spaces that are seen on the RHS of the previous equations are nuclear, thus there is only one completion possible despite the fact that one can define different topologies on tensor product spaces [Fra06]. Moreover, for more information about this kind of results or to see what motivates them, please refer to [Fra06].

4 Separable differential operators

We saw with Equation (1.1) that we need differential operators to regularize our problem. The discussion a the end of the previous chapter indicates that we will only make use of separable differential operators. To do so we will briefly recall what is a differential operator on $\{\mathbb{R}, \mathbb{S}^2, \mathbb{S}^1\}$ and then discuss our meaning of separability and how it is apply to structure Equation (1.1).

▶ **Definition 4.1 (Intuitively).** For $X \in \{\mathbb{S}^1, \mathbb{S}^2\}$ a differential operator $\mathcal{D}$ on $\phi \in \mathcal{S}(X)$ is defined as a linear operator that acts on the harmonics of ϕ. It can be proved that as long as the weight that $\mathcal{D}$ puts on harmonics grow not faster than polynomialy, the operator is well defined on $\mathcal{S}(X)$. A differential operator in the usual sense, like the first derivation, will put weight that is increasing (like n for the first derivation, n^2 for the second derivation ...) while increasing the order of the harmonic. In this way, if the weights that $\mathcal{D}$ is putting are all different from 0 and do not decrease faster than any polynomialy, we can define a new operator $\mathcal{D}^{-1}$ with weights defined to be the multiplicative inverse of those of $\mathcal{D}$. From the previous discussion $\mathcal{D}^{-1}$ is a valid differential operator, an "integral" operator, and is found without surprise to be the inverse operator of $\mathcal{D}$. A complete discussion and a true definition can be found in [Jea20]. A nice plot is available in Figure 4.2. ◀

▶ **Definition 4.2 (Intuitively).** For $X = \mathbb{R}$ the definition of a differential operator is essentially the same. Up to the difference that we do not have *Fourier* series, previously called harmonics, but the so called *Fourier* transform. For $\phi \in \mathbb{R}$, one can then see, informally again, $\mathcal{D}$ as a linear operator which is weighting the frequencies $\int_{-\infty}^{\infty} \phi(x) e^{2\pi i x\xi} dx$ of ϕ. Again, the weight function, often called *Frequency response*, should not grow faster than polynomialy in ∞. Under the same condition as in definition 4.1, we can have invertibility of $\mathcal{D}$.

◀

▶ **Example 4.3.** We place ourself in the setting where $X = \mathbb{S}^1$, $\mathcal{D} = \mathrm{Id} + \frac{\partial^2}{\partial^2\theta}$ and

$$f(\theta) = 0.1404935e^{-i2\pi 5\theta} + 0.08388894e^{-i2\pi 4\theta} + 0.11365338e^{-i2\pi 3\theta} + 0.1460625e^{-i2\pi 2\theta} + \\ 0.11968168e^{-i2\pi 1\theta} + 0.07155425 + 0.06261077e^{i2\pi 1\theta} + 0.02268224e^{i2\pi 2\theta} +$$

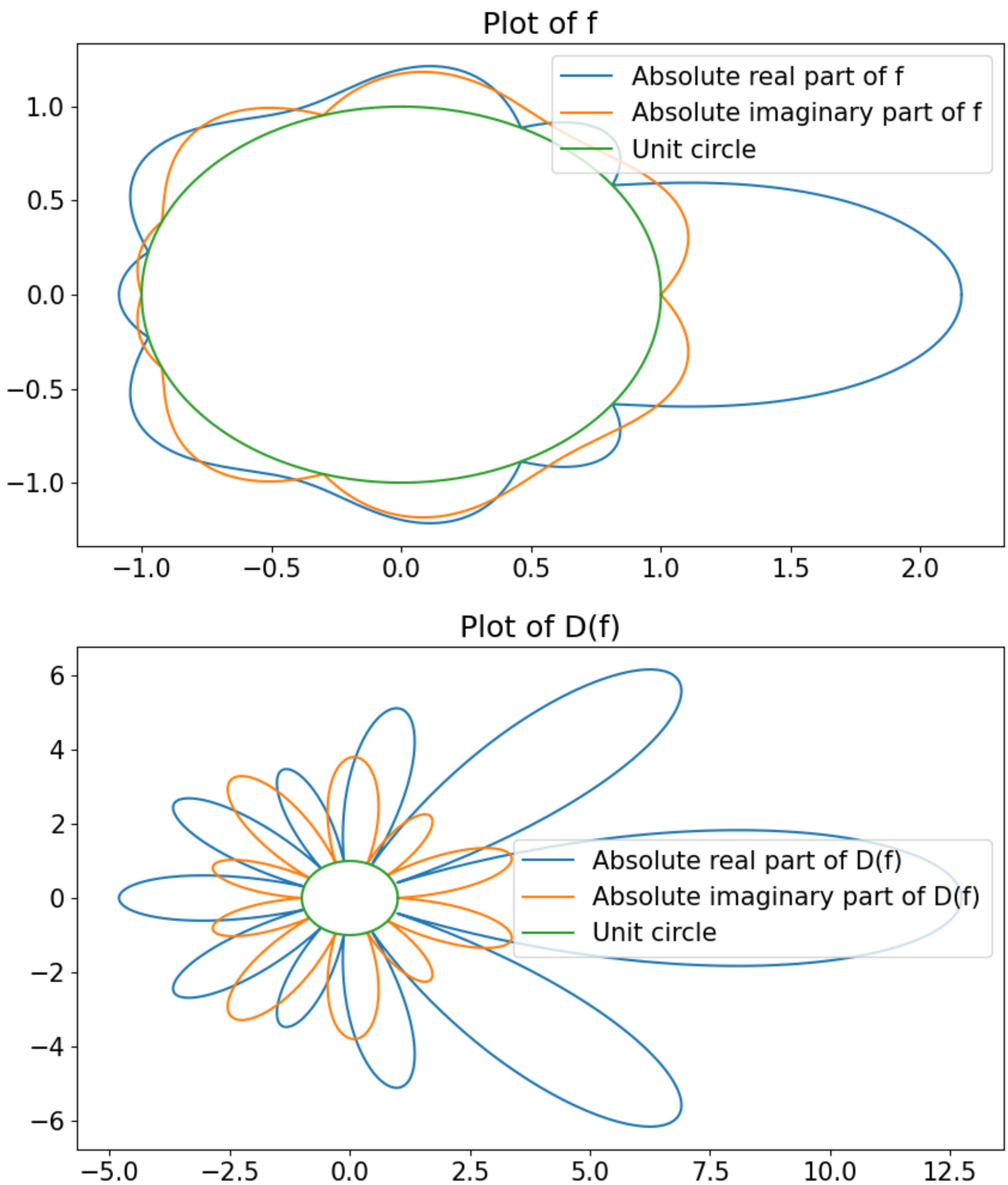


Figure 4.1: In the figure we can see different plots, those of $f, \mathcal{D}(f)$.

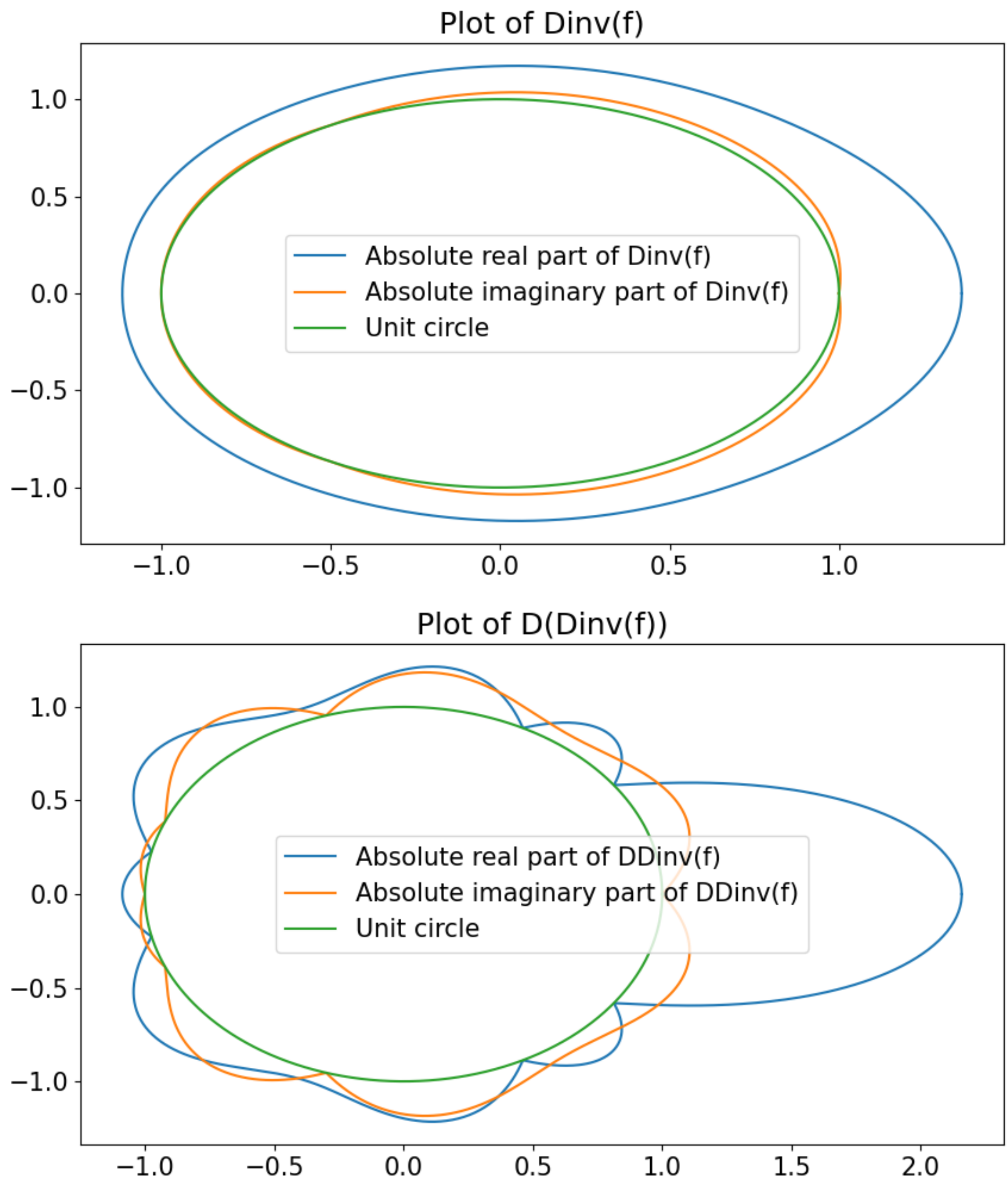


Figure 4.2: In the figure we can see different plots, those of $\mathcal{D}^{-1}(f)$, $\mathcal{D}^{-1}(\mathcal{D}(f))$. Providing evidence that $\mathcal{D}$ is indeed invertible.

.

$$0.13878584\mathrm{e}^{\mathrm{i}2\pi 3\theta} + 0.0883289\mathrm{e}^{-\mathrm{i}2\pi 4\theta} + 0.04736365\mathrm{e}^{\mathrm{i}5\pi 1\theta}.$$

◀

We know have all the tools needed to state the awaited notion of separability. The choice of why separability can be a good idea has already been made in [Gui20] and we will see a bit later in this paper why in some situations it is not ideal. We should nevertheless pay attention to the next definition, and how easy it is to define spatio-temporal differentiable operators, provided that one has separability.

▶ **Definition 4.4.** For $X \in \{\mathbb{S}^1, \mathbb{R}\}$, a differentiable operator D on $\mathcal{S}(\mathbb{S}^2)$ and a differentiable operator L on $\mathcal{S}(X)$. We define the separable differential operator $\mathrm{D} \otimes \mathrm{L}$ by its action on $\mathcal{S}(\mathbb{S}^2) \otimes \mathcal{S}(X)$. For $\phi = \sum_{i=1}^{I} \gamma_i a_i \otimes b_i \in \mathcal{S}(\mathbb{S}^2) \otimes \mathcal{S}(X)$:

$$\mathrm{D} \otimes \mathrm{L}\left(\sum_{i=1}^{I} \gamma_i a_i \otimes b_i\right) := \sum_{i=1}^{I} \gamma_i \mathrm{D} \otimes \mathrm{L}(a_i \otimes b_i) := \sum_{i=1}^{I} \gamma_i \mathrm{D}(a_i)\mathrm{L}(b_i). \tag{4.1}$$

◀

In addition, using point of lemma 3.7 and definition 4.4 one can extends the action of $\mathrm{D} \otimes \mathrm{L}$ from $\mathcal{S}(\mathbb{S}^2) \otimes \mathcal{S}(X)$ to $\mathcal{S}(\mathbb{S}^2 \times X)$. Nevertheless, as said before, the general search space we are considering now is a space of distribution. We thus have to extend the definition of a differential operator on a *Schwartz* space into a distributional space. It is very classical to do it in the following way, using the adjoint operator, here denoted by $\star$.

▶ **Definition 4.5.** For $X \in \{\mathbb{S}^1, \mathbb{R}\}$, a differentiable operator D on $\mathcal{S}(\mathbb{S}^2)$ and a differentiable operator L on $\mathcal{S}(X)$. We define:

$$\mathrm{D}g(a) := g(\mathrm{D}^\star a)) \qquad \forall g \in \mathcal{S}'(\mathbb{S}^2) \quad \forall a \in \mathcal{S}(\mathbb{S}^2). \tag{4.2}$$

$$\mathrm{L}h(b) := h(\mathrm{L}^\star b)) \qquad \forall h \in \mathcal{S}'(X) \quad \forall b \in \mathcal{S}(X). \tag{4.3}$$

$$\begin{aligned}\mathrm{D} \otimes \mathrm{L}\left(\sum_{j=1}^{J} \gamma_j g_j \otimes h_j\right)(\phi) := \sum_{j=1}^{J} \gamma_j \mathrm{D} \otimes \mathrm{L}(g_j \otimes h_j)(\phi) := \sum_{i=1}^{J} \gamma_i g_j \otimes h_j(\mathrm{D}^\star \otimes \mathrm{L}^\star(\phi)),\\ \sum_{j=1}^{J} \gamma_j g_j \otimes h_j \in \mathcal{S}'(\mathbb{S}^2) \otimes \mathcal{S}'(X), \quad \phi \in \mathcal{S}(\mathbb{S}^2 \times X).\end{aligned} \tag{4.4}$$

◀

In addition, using point 2 of Lemma 3.7 and equation 4.4 of Definition 4.5, one can extends the action of $\mathrm{D}\otimes\mathrm{L}$ form $\mathcal{S}'(\mathbb{S}^2)\otimes\mathcal{S}'(X)$ to $\mathcal{S}'(\mathbb{S}^2\times X)$. Our most profound desire is now realised, we have a definition of (separable) differential operator on $\mathcal{S}'(\mathbb{S}^2\times X)$. Such an operator can live in three different regimes depending on its nullspace.

- The nullspace has dimension 0.
- The nullspace has finite dimension, different from 0.
- The nullspace has infinite dimension.

We will not consider the case where D or L have a infinite dimensional nullspace because it is not easily interpretable and pseudo-inverses are hard to define. For the same reason we also will not work with D or L having a finite dimensional nullspace, as it would imply that the operator $\mathrm{D}\otimes\mathrm{L}$ has an infinite dimensional nullspace. In a consequence, in the following, we will always suppose that D and L are invertible. Thus $\mathrm{D}\otimes\mathrm{L}$ will also always be invertible. We will also suppose that D and L are self-adjoint, and so will be $\mathrm{D}\otimes\mathrm{L}$.

► **Definition 4.6.** We say that $\mathcal{D}$ is a spatio-temporal trend-seasonal pseudo-differential operator (STTS differential operator) if it is of the form

$$\mathcal{D} := (\mathcal{D}_t, \mathcal{D}_s) := (\mathrm{D}_t\otimes\mathrm{L}_t, \mathrm{D}_s\otimes\mathrm{L}_s). \tag{4.5}$$

Where s stands for seasonal, t for trend and $\{\mathrm{D}_t, \mathrm{D}_s\}$ are differential operators on $\mathbb{S}^2$, L_s is a differential operator on $\mathbb{S}^1$, L_t is a differential operator on $\mathbb{R}$. The operator $\mathcal{D}$ naturally acts on

$$\mathcal{S}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{S}(\mathbb{S}^2\times\mathbb{S}^1),$$

and using its adjoint the action can be extended, as before, to

$$\mathcal{S}'(\mathbb{S}^2\times\mathbb{R})\times\mathcal{S}'(\mathbb{S}^2\times\mathbb{S}^1).$$

Moreover we will say that $\mathcal{D}$ is spline admissible if

- D_t, L_t, D_t, L_t are invertible.
- D_t, L_t, D_t, L_t are self-adjoint.

Finally, the spline, *Green*'s function associated to the operator $\mathcal{D}$ at the point $(p^1, x, p^2, c)\in\mathbb{S}^2\times\mathbb{R}\times\mathbb{S}^2\times\mathbb{S}^1$. is defined to be:

$$\mathcal{D}^{-1}(\delta_{p^1,x}, \delta_{p^2,c}) \tag{4.6}$$

◀

Using the separable structure of $\mathcal{D}$ we get the refinement:

$$\mathcal{D}^{-1}(\delta_{p^1,x}, \delta_{p^2,c}) = (\mathcal{D}_t(\delta_{p^1,x}), \quad \mathcal{D}_s(\delta_{p^2,c})) = (\mathrm{D}_t \otimes \mathrm{L}_t(\delta_{p^1,x}), \quad \mathrm{D}_s \otimes \mathrm{L}_s(\delta_{p^2,c})) =$$

$$(\mathrm{D}_t \otimes \mathrm{L}_t(\delta_{p^1} \otimes \delta_x), \quad \mathrm{D}_s \otimes \mathrm{L}_s(\delta_{p^2} \otimes \delta_c)) = (\mathrm{D}_t(\delta_{p^1})\mathrm{L}_t(\delta_x), \quad \mathrm{D}_s(\delta_{p^2})\mathrm{L}_s(\delta_c))$$

We will use the following notation

$$\psi^1_{p^1} := \mathrm{D}_t(\delta_{p^1}), \quad \psi^2_{p^2} := \mathrm{D}_s(\delta_{p^2}).$$

$$\zeta_x := \mathrm{L}_t(\delta_x), \quad \phi_c := \mathrm{L}_s(\delta_c).$$

We will use later a superposition of these splines, $(\psi^1_{p^1}\zeta_x, \quad \psi^2_{p^2}\phi_c)$, to approximate a candidate (f_{trend}, f_{seas}). We are now in good position to end this chapter. We have an initial search the space for the trend and seasonal part of our *phenomenon*.We have a notion of splines, later used to shape a candidate for solving Equation (1.1). We have a notion of differential operator acting on this search space, later used in the regularization of a candidate for solving Equation (1.1). To do all the previous things we used the heavy assumption of separability of our operator. It is to be remembered that this assumption forces us, as we do not want a infinite dimensional nullspace, to have an invertible operator. In addition, it can be proved that a spline of the form $\mathrm{D}_t(\delta_{p^1})\mathrm{L}_t(\delta_x)$ decays towards 0 in ∞, provided that the *Frequency Response* of D_t grows strictly faster than a polynomial of order 1. In consequence, if D_t verifies the previous condition, as our nullspace is empty and our spline decays to 0 in ∞, we can not a have candidate to Equation (1.1) that does not decay to 0 in ∞. Therefore by the assumption of separability we potentially get ride of a whole class of interesting splines. It remains, as we will continue to see, that this assumptions is nearly as pleasant theoretically and numerically as the first lightbeams of spring.

5 The search space and its properties

The search space we previously gave is not satisfactory for our research. Indeed, we will need to take the norm of a candidate in the setting of Equation (1.1), and it is yet not possible. We will thus need to restrict our space a little bit and guided from previous work, e.g. [Uns18], we will go in the direction of a space of measure, being a subspace of the space of tempered distribution, equipped with a total variation norm. We will here define what will be our definitive search space and show some useful properties about it. In this chapter we will suppose that $\mathcal{D} = (\mathcal{D}_t, \mathcal{D}_s)$ is a spline admissible trend-seasonal differential operator.

Remember that we need a search space for (f_{trend}, f_{seas}). We will naturally go in the direction of one that is the cartesian product between a search space for f_{trend} and a search space for f_{seas}. First we will define these individual search spaces and then we will link them and discuss which kind of topology, norm, we have to put on the product.

▶ **Definition 5.1.** The search space for f_{trend}:

$$\mathcal{M}_{\mathcal{D}_t} := \mathcal{M}_{\mathcal{D}_t}(\mathbb{S}^2 \times \mathbb{R}) := \{f \in \mathcal{S}'(\mathbb{S}^2 \times \mathbb{R}) | \quad \mathcal{D}_t(f) \in \mathcal{M}(\mathbb{S}^2 \times \mathbb{R})\}$$
$$\text{equipped with the norm } \|f\|_{\mathcal{D}_t} := \|\mathcal{D}_t(f)\|_{TV}. \tag{5.1}$$

The search space for f_{seas}:

$$\mathcal{M}_{\mathcal{D}_s} := \mathcal{M}_{\mathcal{D}_s}(\mathbb{S}^2 \times \mathbb{S}^1) := \{f \in \mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^1) | \quad \mathcal{D}_s(f) \in \mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)\}$$
$$\text{equipped with the norm } \|f\|_{\mathcal{D}_s} := \|\mathcal{D}_s(f)\|_{TV}. \tag{5.2}$$

◀

In the remaining of this paper, the facts that $\mathcal{M}_{\mathcal{D}_t}$ is equipped with $\|\cdot\|_{\mathcal{D}_t}$ and $\mathcal{M}_{\mathcal{D}_s}$ is equipped with $\|\cdot\|_{\mathcal{D}_s}$ will implicitly hold true.

▶ **Corollary 5.2.** Because $\mathcal{D}_t$ and $\mathcal{D}_s$ are both invertible and $\mathcal{M}(\mathbb{S}^2 \times \mathbb{R})$ and $\mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$ are both *Banach* spaces, it is easy to see that $\mathcal{M}_{\mathcal{D}_t}$ and $\mathcal{M}_{\mathcal{D}_s}$ are again both *Banach* spaces. ◀

To continue we also need the next definition which is motivated by the fact that we initially, in chapter 2, defined the spaces $\mathcal{M}(X)$ to be the duals of the spaces $C_0(X)$. Therefore, as the spaces $\mathcal{M}_{\mathcal{D}_t}$ and $\mathcal{M}(\mathbb{S}^2\times\mathbb{R})$ are tightly linked (congruent), we can expect the space $\mathcal{M}_{\mathcal{D}_t}$ to be the dual of a space defined from $C_0(\mathbb{S}^2\times\mathbb{R})$. The following definition makes explicit this expectation.

▶ **Definition 5.3.**

$$\begin{aligned} C_{\mathcal{D}_t^{-1},0} := C_{\mathcal{D}_t^{-1},0}(\mathbb{S}^2\times\mathbb{R}) := \{f \in \mathcal{S}'(\mathbb{S}^2\times\mathbb{R})| \quad f = \mathcal{D}_t(g) \text{ with } g \in C_0(\mathbb{S}^2\times\mathbb{R})\} \\ \text{equipped with the norm } \|f\|_{\mathcal{D}_t^{-1}} := \left\|\mathcal{D}_t^{-1}(f)\right\|_\infty = \|g\|_\infty. \end{aligned} \tag{5.3}$$

$$\begin{aligned} C_{\mathcal{D}_s^{-1},0} := C_{\mathcal{D}_s^{-1},0}(\mathbb{S}^2\times\mathbb{S}^1) := \{f \in \mathcal{S}'(\mathbb{S}^2\times\mathbb{S}^1)| \quad f = \mathcal{D}_s(g) \text{ with } g \in C_0(\mathbb{S}^2\times\mathbb{S}^1)\} \\ \text{equipped with the norm } \|f\|_{\mathcal{D}_s^{-1}} := \left\|\mathcal{D}_s^{-1}(f)\right\|_\infty = \|g\|_\infty. \end{aligned} \tag{5.4}$$

◀

In the remaining of this paper, the facts that $C_{\mathcal{D}_t^{-1},0}$ is equipped with $\|\cdot\|_{\mathcal{D}_t^{-1}}$ and $C_{\mathcal{D}_s^{-1},0}$ is equipped with $\|\cdot\|_{\mathcal{D}_s^{-1}}$ will implicitly hold true.

▶ **Corollary 5.4.** Because $\mathcal{D}_t^{-1}$ and $\mathcal{D}_s^{-1}$ are both invertible and $C_0(\mathbb{S}^2\times\mathbb{R})$ and $C_0(\mathbb{S}^2\times\mathbb{S}^1)$ are both *Banach* spaces, it is easy to see that $C_{\mathcal{D}_t^{-1},0}$ and $C_{\mathcal{D}_s^{-1},0}$ are again both *Banach* spaces. ◀

▶ **Lemma 5.5.** The following identifications of duals hold:

$$(C_{\mathcal{D}_t^{-1},0}; \quad \|\cdot\|_{\mathcal{D}_t^{-1}})' = (\mathcal{M}_{\mathcal{D}_t}; \quad \|\cdot\|_{\mathcal{D}_t}) \tag{5.5}$$

$$(C_{\mathcal{D}_s^{-1},0}; \quad \|\cdot\|_{\mathcal{D}_s^{-1}})' = (\mathcal{M}_{\mathcal{D}_s}; \quad \|\cdot\|_{\mathcal{D}_s}) \tag{5.6}$$

◀

Proof of Lemma 5.5: We will here only prove equation 5.5 as the proof for 5.6 is very similar. Concerning equation 5.5, the following chain holds true:

$$\mathcal{M}_{\mathcal{D}_t} \underset{\mathcal{D}_t}{\cong} \mathcal{M}(\mathbb{S}^2\times\mathbb{R}) = C_0'(\mathbb{S}^2\times\mathbb{R}) \underset{\mathcal{D}_t^{-1}}{\cong} C'_{\mathcal{D}_t^{-1},0}$$

The first link of the chain is due to $\mathcal{D}_t$ being a congruence (isometric isomorphism). The second link of the chain is a definition (or can be seen more formally as a consequence of the Riesz-Markov-Kakutani theorem). The last link of the chain is due to $\mathcal{D}_t^{-1}$ being a congruence (again isometric isomorphism) and this congruence can be extended using the adjoint operator ($\mathcal{D}_t^{-1}$ here, since $\mathcal{D}_t$ is supposed to be self-adjoint) to a congruence on the dual space. As $\mathcal{D}_{trend}$ and $\mathcal{D}_{trend}^{-1}$ will cancel out, this proves equation 4.5. ■

The reason why it is important to characterize the predual of, for now, our individual search spaces, is that our futur formalization of Equation (1.1) will use the notion of dual and predual as it will be stated as a functional optimization problem. What leverages the importance of lemma 5.5. To sum things up we now have

- The search space for f_{trend} given by $\mathcal{M}_{\mathcal{D}_t}$ with predual $C_{\mathcal{D}_t^{-1},0}$.
- The search space for f_{seas} given by $\mathcal{M}_{\mathcal{D}_s}$ with predual $C_{\mathcal{D}_s^{-1},0}$.

Naturally, we will define the coupled search space for (f_{trend}, f_{seas}) by $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$. Two questions, at least, quickly arise. First, which norm do we put on $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$? Secondly, is it possible, by a clever choice of norm, to find $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ to be the predual of $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$? Lets start by answering the first one. The norm we will put on $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ will be used to penalize our problem, to penalize the couple (f_{trend}, f_{seas}). But we need a certain plasticity in this penalization, in the way that we should be able to penalize more f_{trend} than f_{seas}, if wanted, and vice-versa. Therefore, we propose the following norm

$$\text{For fixed } \alpha \in \mathbb{R}_\star^+, \quad \beta \in \mathbb{R}_\star^+.$$

$$\forall (f, g) \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}, \quad \|(f, g)\|_{\mathcal{D},(\alpha,\beta)} := \alpha \|f\|_{\mathcal{D}_t} + \beta \|g\|_{\mathcal{D}_s}.$$

In order to make the notation lighter, if the context is clear, we will write $\|\cdot\|_{\mathcal{D}}$ and not $\|\cdot\|_{\mathcal{D},(\alpha,\beta)}$. Then, the following lemma answers our second question.

▶ **Lemma 5.6.** Let $\mathcal{D}$ be a trend-seasonal spline admissible pseudo differential operator and let α, β be positive real numbers . We have that:

$$C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0} \text{ equipped with the norm } \|(f, g)\|_{\mathcal{D}^{-1}} = max(\frac{1}{\alpha}\|f\|_{\mathcal{D}_{trend}^{-1}}, \frac{1}{\beta}\|g\|_{\mathcal{D}_{seas}^{-1}}),$$
$$(f, g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}, \quad (5.7)$$

$\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ equipped with the norm $\|(f,g)\|_{\mathcal{D}} = \alpha\|f\|_{\mathcal{D}_t} + \beta\|g\|_{\mathcal{D}_s}$,

$$(f,g) \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}, \quad (5.8)$$

are *Banach* spaces. Moreover we have the predual characterization:

$$(C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}; \quad \|\cdot)\|_{\mathcal{D}^{-1}})' = (\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}; \quad \|\cdot\|_{\mathcal{D}}). \quad (5.9)$$

◀

Proof of Lemma 5.6.

Step I: $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is a *Banach* space.

It is easy to prove that $\|\cdot\|_{\mathcal{D}^{-1}}$ defines a norm on $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. By definition, to prove that it is a *Banach* space we only have to show that any Cauchy sequence is convergent in norm to a limit in $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. Take a Cauchy sequence $\{(f_n,g_n)\}_{n=1}^{\infty} \subset C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. Then $\lim\limits_{n\to\infty}\|(f_{n+1}-f_n, g_{n+1}-g_n)\|_{\mathcal{D}^{-1}} = \lim\limits_{n\to\infty} max(\frac{1}{\alpha}\|f_{n+1}-f_n\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|g_{n+1}-g_n\|_{\mathcal{D}_s^{-1}}) = 0$, what implies (because $\frac{1}{\alpha}$ and $\frac{1}{\beta}$ are different from 0) that $\lim\limits_{n\to\infty}\|f_{n+1}-f_n\|_{\mathcal{D}_t^{-1}} = 0$ and $\lim\limits_{n\to\infty}\|g_{n+1}-g_n\|_{\mathcal{D}_s^{-1}} = 0$. Since $C_{\mathcal{D}_t^{-1},0}$ and $C_{\mathcal{D}_s^{-1},0}$ are Banach spaces we have that there exits $(f,g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ such that $\lim\limits_{n\to\infty}\|f-f_n\|_{\mathcal{D}_t} = 0$ and $\lim\limits_{n\to\infty}\|g-g_n\|_{\mathcal{D}_s} = 0$. This implies that $\lim\limits_{n\to\infty}\|(f-f_n, g-g_n)\|_{\mathcal{D}^{-1}} = 0$ and that $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is a Banach space. One can show in exactly the same fashion that $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is a Banach space. What proves the first part of the claim.

Step II: predual characterization.

Recall that the dual space is defined as the set of bounded linear functionals. Here $H \in (C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0})'$ is bounded if the dual norm

$$\sup_{(f,g)\in C_{\mathcal{D}_t^{-1},0}\times C_{\mathcal{D}_s^{-1},0}, \quad \|(f,g)\|_{\mathcal{D}^{-1}}=1} |H(f,g)|$$

is finite. We will prove the characterization 4.9 by double inclusion and we will show that the norm of $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is the same as the dual norm of $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. We first prove that $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s} \subset (C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0})'$. For $H = (F,G) \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ and $(f,g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$, the action of (F,G) on (f,g) is naturally given by $H(f,g) = F(f) + G(g)$ and from the linearity of F and G we get that H is a linear functional. It remains to prove that H is bounded. In the following chain it is implicit that $(f, g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$:

$$\sup_{\|(f,g)\|_{\mathcal{D}^{-1}}=1} |H(f,g)| \underset{5.6.1}{=} \sup_{\|(f,g)\|_{\mathcal{D}^{-1}}=1} |F(f)+G(g)| \underset{5.6.2}{=} \sup_{\|(f,g)\|_{\mathcal{D}^{-1}}=1} F(f)+G(g) \underset{5.6.3}{=}$$

$$\sup_{max(\frac{1}{\alpha}\|f\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|g\|_{\mathcal{D}_s^{-1}})=1} F(f)+G(g) \underset{5.6.4}{=} \sup_{\|f\|_{\mathcal{D}_t^{-1}}=\alpha, \|g\|_{\mathcal{D}_s^{-1}}=\beta} F(f)+G(g) \underset{5.6.5}{=}$$

$$\alpha \sup_{\|f\|_{\mathcal{D}_t^{-1}}=1} F(f) + \beta \sup_{\|g\|_{\mathcal{D}_s^{-1}}=1} G(g) \underset{5.6.6}{=} \alpha \sup_{\|f\|_{\mathcal{D}_t^{-1}}=1} |F(f)| + \beta \sup_{\|g\|_{\mathcal{D}_s^{-1}}=1} |G(g)| \underset{5.6.7}{=}$$

$$\alpha\|F\|_{\mathcal{D}_t} + \beta\|G\|_{\mathcal{D}_s}.$$

Previously, we had that (5.6.1) holds by definition of H; (5.6.2) holds because one can always replace f and/or g by -f and/or -g and we are taking the sup so negative values do not matter; (5.6.3) holds by definition of the norm $\|\cdot\|_{\mathcal{D}^{-1}}$; (5.6.4) holds because the sup is only achieved for the highest values of $\|\cdot\|_{\mathcal{D}_{trend}^{-1}}$ and $\|\cdot\|_{\mathcal{D}_{seas}^{-1}}$; (5.6.5) is a re-normalisation; (5.6.6) holds for the same reason as (5.6.2); (5.6.7) holds by definition of $\|\cdot\|_{\mathcal{D}_t}$ and $\|\cdot\|_{\mathcal{D}_s}$. Therefore, H is a linear bounder functional and we have the inclusion $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s} \subset (C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0})'$. Note that we also proved with this chain that the dual norm of $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is the same as the norm on $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$. Thus, it remains to show the inclusion $(C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0})' \subset \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$. To do so it will be enough to show that $H \in (\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})'$ can be decomposed in $H = (F, G)$ with $F \in \mathcal{M}_{\mathcal{D}_t}$ and $G \in \mathcal{M}_{\mathcal{D}_s}$. Let us define the projection operators:

$$I_{C_{\mathcal{D}_t^{-1},0}} : C_{\mathcal{D}_t^{-1},0} \to C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}. \quad I_{C_{\mathcal{D}_t^{-1},0}} : f \mapsto (f, 0). \tag{5.10}$$

$$I_{C_{\mathcal{D}_s^{-1},0}} : C_{\mathcal{D}_s^{-1},0} \to C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}. \quad I_{C_{\mathcal{D}_s^{-1},0}} : g \mapsto (0, g). \tag{5.11}$$

One will remark that $H \circ I_{C_{\mathcal{D}_t^{-1},0}} \in \mathcal{M}_{\mathcal{D}_t}$ and $H \circ I_{C_{\mathcal{D}_s^{-1},0}} \in \mathcal{M}_{\mathcal{D}_s}$, thus $(H \circ I_{C_{\mathcal{D}_t^{-1},0}}, H \circ I_{C_{\mathcal{D}_s^{-1},0}}) \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$. Moreover, for $(f, g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ we have the chain:

$$(H \circ I_{C_{\mathcal{D}_t^{-1},0}}, H \circ I_{C_{\mathcal{D}_s^{-1},0}})(f, g) \underset{5.7.8}{=} H \circ I_{C_{\mathcal{D}_t^{-1},0}}(f) + H \circ I_{C_{\mathcal{D}_s^{-1},0}}(g) \underset{5.7.9}{=}$$

$$H \circ \left(I_{C_{\mathcal{D}_t^{-1},0}}(f) + I_{C_{\mathcal{D}_s^{-1},0}}(g) \right) \underset{5.7.10}{=} H(f,g).$$

Previously, (5.7.8) holds by definition of the action of the product operator; (5.7.9) holds by linearity of H; (5.7.10) holds by definition. This shows the second inclusion and concludes the proof. ■

6 Father, let the dawn rise: The representer theorem.

In this chapter we will establish the formalization of Equation (1.1). Then we will prove a representer theorem, the dawn, that will establish the shape of the solution set of the previous formalization. Finally, we will structure the general representer to fit an interesting exemple.

The solution set we just talked about will have a specific shape whose characterization will use extremal points. Let us quickly review this notion.

6.1 On extremal points

▶ **Definition 6.1.** For X a real vector space, possibly of infinite dimension, K a subspace of X. We say that for $\{x, y, z\} \in K$, x lies between y and z if there exits a $t \in (0, 1)$ such that $x = ty + (1 - t)z$. If for $x \in K$ there exits no couple $(y, z) \in K^2$ such that x lies between y and z, we say that x is an extremal point in K. We denote the set of extremal point of K by $\mathrm{Extremal}(K)$. ◀

The previous condition of extremality is defined in opposition to the existence of a segment $[y, z]$ where x would lie. The issue is that it is not very practical as one (probably) does not want to show that each segment $[y, z]$ does not contain x in order to prove that x is an extremal point. The following theorem gives a direct characterization of extremality.

▶ **Theorem 6.2.** Let X be a normed vector space and K a convex set in X. A point $x \in K$ is an extremal point if and only if any of the following conditions hold.

1. For $y \in X$, $x + y \in K$ and $x - y \in K$ implies $y = 0$.
2. If $y, z \in K$ are such that $x \in [y, z]$ then $y = x = z$.

◀

Proof. The theorem is proved in [Bec11] (page 279, theorem 9.2.2). ■

In order to state the representer theorem we need to understand a specific set of extremal points, the one of the unit ball in $\mathcal{M}(\mathbb{S}^2 \times \mathbb{R}) \times \mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$.

▶ **Definition 6.3.** Let $(X, \|\cdot\|_X)$ be a normed vector space, possibly of infinite dimension. For $r \in \mathbb{R}^+$ and $m \in X$, we will denote by $\mathbf{B}_X(r, m)$ the ball of radius r, centered around m,

$$\{x \in X| \quad \|x - m\|_X \leq r\}, \tag{6.1}$$

in $(X, \|\cdot\|_X)$. The detail of which norm is on X is omitted in the notation $\mathbf{B}_X(r, m)$. ◀

▶ **Lemma 6.4.** For $r \in \mathbb{R}^+$

$$\text{Extremal}(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}(r, 0)) = \{\tau\delta_{p,x}| \quad (p, x) \in \mathbb{S}^2 \times \mathbb{R}, \tau \in \mathbb{R}, |\tau| = r\}. \tag{6.2}$$

$$\text{Extremal}(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)) = \{\tau\delta_{p,c}| \quad (p, c) \in \mathbb{S}^2 \times \mathbb{S}^1, \tau \in \mathbb{R}, |\tau| = r\}. \tag{6.3}$$

◀

Proof of Lemma 6.4: Without proof. ■

The following lemma can be found as a consequence of lemma 1 in [Sha21], nevertheless we furnish this paper with our own proof.

▶ **Lemma 6.5.** Consider the space $\mathcal{M}(\mathbb{S}^2 \times \mathbb{R}) \times \mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$ equipped with the norm $\|(\cdot,\cdot)\|_{TV} = \alpha\|\cdot\|_{TV} + \beta\|\cdot\|_{TV}$, $\{\alpha, \beta\} \in \mathbb{R}^+_\star$. We have that for $r \in \mathbb{R}^+$,

$$\text{Extremal}(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)) =$$
$$\left(\text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{r}{\alpha}, 0\right)\right) \times 0\right) \cup \left(0 \times \text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{r}{\beta}, 0\right)\right)\right). \tag{6.4}$$

◀

Proof of Lemma 6.5: A ball in a normed vector space is always convex, so we can apply Theorem 6.2 with $K = \mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)$.

StepI $:\Leftarrow$.
We will first show, using characterization (1) of Theorem 6.2, that any points in

$$\left(\text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{r}{\alpha}, 0\right)\right) \times 0\right) \cup \left(0 \times \text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{r}{\beta}, 0\right)\right)\right)$$

is indeed an extremal point. Thus, suppose that

$$x \in \left(0 \times \text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{r}{\beta}, 0\right)\right)\right) \underset{Lemma\ 6.4}{\iff} x = \tau\delta_{p,c}, \quad |\tau| = \frac{r}{\beta}$$

and take $(y, y') \in \mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)$ such that $\left\|(\pm y, \tau\delta_{(p,c} \pm y')\right\|_{TV} \leq r$. This implies that $\left\|z\delta_{(p,t)} \pm y'\right\|_{TV} \leq \frac{r}{\beta}$. Then using Lemma 6.4 and the characterization (1) of Theorem 6.2 (in the other way around) show that $y' = 0$. Then we have $\alpha\|y\|_{TV} + \beta\left\|\tau\delta_{(p,t)}\right\|_{TV} \leq r$ with $\beta\left\|\tau\delta_{(p,t)}\right\|_{TV} = r$, implying that $y = 0$ and finally $(y, y') = (0, 0)$. The proof when

$$x \in \left(\text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{r}{\alpha}, 0\right)\right) \times 0\right)$$

is done in a symmetric manner.

<u>**StepII**</u> $:\Rightarrow$.
We now show the other inclusion using characterization (2) and (1) of Theorem 6.2. If

$$(F, G) \in \text{Extremal}(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)).$$

Then

$$(F, G) \in \left(\text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{r}{\alpha}, 0\right)\right) \times 0\right) \cup \left(0 \times \text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{r}{\beta}, 0\right)\right)\right).$$

Lets first suppose that either F or G is equal to 0. Suppose $F = 0$, by characterization (1) of Theorem 6.2 if $\alpha\|\pm y\|_{TV} + \beta\|G \pm y'\|_{TV} \leq r$ then $(y, y') = (0, 0)$. But by taking $y = 0$ we get that $\beta\|G \pm y'\|_{TV} \leq r$ implies $y' = 0$. Again by characterization (1) we get that G is an extremal point of the $\frac{r}{\beta}$ ball in $\mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$ and thus

$$(F, G) \in \left(0 \times \text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{r}{\beta}, 0\right)\right)\right).$$

Suppose $G = 0$, then with the same argument one gets that

$$(F, G) \in \left(\text{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{r}{\alpha}, 0\right)\right) \times 0\right).$$

We are now left to prove that either F or G is equal to 0. By contradiction, suppose that neither F nor G is equal to 0. By characterzation (2) of Theorem 6.2, it is easy to see that an extremal point in a closed ball has to be in the outermost layer, i.e here to be of norm r. Then because

$$\alpha\|F\|_{TV} + \beta\|G\|_{TV} = r. \qquad \alpha\|F\|_{TV} \neq 0. \qquad \beta\|G\|_{TV} \neq 0.$$

we have that

$$\exists\{\epsilon_1, \epsilon_2\} \in \mathbb{R}_\star^+ : \quad \alpha\|F\|_{TV} + \epsilon_1 < r, \quad \beta\|G\|_{TV} + \epsilon_2 < r.$$

Moreover, wlog. we can set $\epsilon = \epsilon_1 = \epsilon_2$ small enough, such that $0 < \frac{\epsilon}{\alpha\|F\|_{TV}} < 1$ and $0 < \frac{\epsilon}{\beta\|G\|_{TV}} < 1$. Then set $y = (\frac{\epsilon}{\alpha\|F\|_{TV}}F, -\frac{\epsilon}{\beta\|G\|_{TV}}G) \in \mathcal{M}(\mathbb{S}^2 \times \mathbb{R}) \times \mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$. Remark that y verifies the following chain

$$\|(F, G) + y\|_{TV} = \left\|\left(\left(1 + \frac{\epsilon}{\alpha\|F\|_{TV}}\right)F, \left(1 - (\frac{\epsilon}{\beta\|G\|_{TV}}\right)G\right)\right\|_{TV} =$$

$$\alpha\left(1 + \frac{\epsilon}{\alpha\|F\|_{TV}}\right)\|F\|_{TV} + \beta\left(1 - \frac{\epsilon}{\beta\|G\|_{TV}}\right)\|G\|_{TV} = \|(F, G)\|_{TV}.$$

We also have

$$\|(F, G) - y\|_{TV} = \|(F, G)\|_{TV}.$$

In consequence, $(F, G) \pm y \in \mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)$, and in regard of characterization (1) of Theorem 6.2, y should be 0. But it is not as F and G are both different from 0. Therefore, we have a contradiction with (F, G) being an extremal point of $\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}(r, 0)$ with both F and G different from 0. ■

The previous lemma provides us with characterization we wanted and we can now state the representer theorem.

6.2 The representer theorem

▶ **Theorem 6.6.** Suppose that the following assumptions hold true.

1. $\mathcal{D} = (\mathcal{D}_t, \mathcal{D}_s)$ is a STTS differential operator.

2. λ is a strictly positive regularisation constant and $\{\alpha, \beta\}$ are two strictly positive constants.

3. According to Lemma 5.6, $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}} = \alpha\|\cdot\|_{\mathcal{D}_t} + \beta\|\cdot\|_{\mathcal{D}_s}$ and its predual $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}^{-1}} = max(\frac{1}{\alpha}\|\cdot\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|\cdot\|_{\mathcal{D}_s^{-1}})$.

4. $\boldsymbol{\Phi}$ is the sampling operator defined by

$$\boldsymbol{\Phi}(\cdot,\cdot) : \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s} \to \mathbb{R}^{N\times M},$$

$$\boldsymbol{\Phi}(\cdot,\cdot) : H = (F, G) \mapsto \begin{bmatrix} H(f_{1,1}, g_{1,1}) & \dots & H(f_{1,M}, g_{1,M}) \\ & \dots & \\ H(f_{N,1}, g_{N,1}) & \dots & H(f_{N,M}, g_{N,M}) \end{bmatrix}.$$

Where $N \in \mathbb{N}, M \in \mathbb{N}$ and $\forall (i,j) \in [1,N]\times[1,M] \quad (f_{i,j}, g_{i,j}) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. Moreover $\{(f_{i,j}, g_{i,j})\}_{i,j=1,1}^{N,M}$ is a family of linearly independent vectors.

5. C is a cost functional defined by

$$C(\cdot,\cdot) : \mathbb{R}^{N\times M} \times \mathbb{R}^{N\times M} \to \mathbb{R}^{+} \cup \{\infty\}.$$

Such that $\forall \mathbf{Y} \in \mathbb{R}^{N\times M}$, the mapping

$$C(\mathbf{Y}, \cdot) : \mathbb{R}^{N\times M} \to \mathbb{R}^{+} \cup \{\infty\},$$

is proper, convex, lower semi-continuous.

Then $\forall \mathbf{Y} \in \mathbb{R}^{N\times M}$ the solution set of the infinite dimensional optimization problem:

$$\mathcal{V} = \underset{H \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}{\operatorname{argmin}} \{C(\mathbf{Y}, \boldsymbol{\Phi}(H)) + \lambda \|H\|_{\mathcal{D}}\} \tag{6.5}$$

is non-empty and representable as the *weak** closed convex hull of its extreme points. The latter are necessarily (but not exactly) of the form:

$$\begin{aligned} H &= \sum_{i=1}^{I_t} \gamma_{1,i} \left(\mathcal{D}_t^{-1}(\delta_{p_i^1, x_i}), 0 \right) + \sum_{i=1}^{I_s} \gamma_{2,i} \left(0, \mathcal{D}_s^{-1}(\delta_{p_i^2, c_i}) \right) \\ &= \sum_{i=1}^{I_t} \gamma_{1,i} \left(\psi^1_{p_i^1} \otimes \zeta_{x_i}, 0 \right) + \sum_{i=1}^{I_s} \gamma_{2,i} \left(0, \psi^2_{p_i^2} \otimes \phi_{c_i} \right) \end{aligned} \tag{6.6}$$

$$I_s + I_t \leq NM, \quad \{\{\gamma_{1,i}\}_{i=1}^{I_t}, \{\gamma_{2,i}\}_{i=1}^{I_s}\} \subset \mathbb{R}, \quad \{\{p_i^1\}_{i=1}^{I_t}\}, \{p_i^2\}_{i=1}^{I_s}\}\} \subset \mathbb{S}^2,$$
$$\{x_i\}_{i=1}^{I_t} \subset \mathbb{R}, \quad \{c_i\}_{i=1}^{I_s} \subset \mathbb{S}^1.$$

◀

Proof of Theorem 6.6: We can easily see that this theorem falls in the realm of the theorem 3.12 in [Jea20] as one can check that all the assumptions are verified. One direct consequence is that $\mathcal{V}$ is non-empty and representable as the *weak** closed convex hull of its extreme points. We still have to find their shape. Theorem 3.12 tells us that the extreme points are necessarily, but not always, of the form

$$H = \sum_{i=1}^{I} \gamma_i e_i, \quad I \leq NM, \quad \{\gamma_i\}_{i=1}^{I} \subset \mathbb{R}, \tag{6.7}$$

with e_i being extremal points of

$$\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}\left(\frac{1}{\lambda}, 0\right). \tag{6.8}$$

We are thus searching for the set $\mathrm{Extremal}(\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}(\frac{1}{\lambda}, 0))$. As $\mathcal{D}$ is a congruence between $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ and $\mathcal{M}(\mathbb{S}^2 \times \mathbb{R}) \times \mathcal{M}(\mathbb{S}^2 \times \mathbb{S}^1)$, we have that

$$\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}\left(\frac{1}{\lambda}, 0\right)\right) = \mathcal{D}^{-1}\left(\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})\times\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{1}{\lambda}, 0\right)\right)\right). \tag{6.9}$$

By using Lemma 6.5 and then Lemma 6.4 we get

$$\begin{aligned}
&\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}\left(\frac{1}{\lambda}, 0\right)\right) = \\
&\mathcal{D}^{-1}\left(\left(\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{1}{\alpha\lambda}, 0\right)\right) \times 0\right) \cup \left(0 \times \mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{1}{\beta\lambda}, 0\right)\right)\right)\right) = \\
&\left(\mathcal{D}_t^{-1}\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{R})}\left(\frac{1}{\alpha\lambda}, 0\right)\right) \times 0\right) \cup \left(0 \times \mathcal{D}_s^{-1}\mathrm{Extremal}\left(\mathbf{B}_{\mathcal{M}(\mathbb{S}^2\times\mathbb{S}^1)}\left(\frac{1}{\beta\lambda}, 0\right)\right)\right) = \\
&\left\{\tau\mathcal{D}_t^{-1}(\delta_{p,x}) \;\middle|\; (p, x) \in \mathbb{S}^2 \times \mathbb{R}, \tau \in \mathbb{R}, |\tau| = \frac{1}{\alpha\lambda}\right\} \cup \\
&\left\{\tau\mathcal{D}_s^{-1}(\delta_{p,c}) \;\middle|\; (p, c) \in \mathbb{S}^2 \times \mathbb{R}, \tau \in \mathbb{R}, |\tau| = \frac{1}{\beta\lambda}\right\}
\end{aligned} \tag{6.10}$$

Finally, putting together equation (6.7) and (6.10) prove equation (6.6) and thus conclude the proof. ■

▶ **Remark 6.7.** It is important to see that this version of the representer theorem also allows one the to have a vector of data and not a matrix, by just taking M=1. We used the matrix notation on purpose, as it is easier to specify the spatio-temporal dependency with it. ◀

6.3 Application of the representer theorem

We will see here how to use Theorem 6.6 to study Equation (1.1). In the following we will suppose that

$$\forall (p,x) \in \mathbb{S}^2 \times \mathbb{R}, \delta_{p,x} \in C_{\mathcal{D}_t^{-1},0}. \tag{6.11}$$

$$\forall (p,c) \in \mathbb{S}^2 \times \mathbb{S}^1, \delta_{p,c} \in C_{\mathcal{D}_s^{-1},0}. \tag{6.12}$$

In Theorem 6.2 we used a cost functional and a sampling functional. Lets specify the sampling functional first. We used the evaluation matrix

$$\begin{bmatrix} (f_{1,1}, g_{1,1}) & \dots & (f_{1,M}, g_{1,M}) \\ & \dots & \\ (f_{N,1}, g_{N,1}) & \dots & (f_{N,M}, g_{N,M}) \end{bmatrix},$$

which we now specify to be, thanks to Equation (6.11) and Equation (6.12),

$$\begin{bmatrix} (\delta_{p_1,x_1}, \delta_{p_1,\mathrm{P}_T(x_1)}) & \dots & (\delta_{p_1,x_M}, \delta_{p_1,\mathrm{P}_T(x_M)}) \\ & \dots & \\ (\delta_{p_N,x_1}, \delta_{p_N,\mathrm{P}_T(x_1)}) & \dots & (\delta_{p_N,x_M}, \delta_{p_N,\mathrm{P}_T(x_M)}) \end{bmatrix}. \tag{6.13}$$

Where

$$\{p_i\}_{i=1}^N \subset \mathbb{S}^2, \quad \{x_j\}_{j=1}^M \subset \mathbb{R}, \tag{6.14}$$

are considered to be the points on which the practitionner have queried data. Moreover, P_T is the projection of period T of the real line onto the circle, given by

$$\mathrm{P}_T(x) = \left(\cos\left(\frac{2\pi x}{T}\right), \sin\left(\frac{2\pi x}{T}\right)\right). \tag{6.15}$$

Finally the sampling functional Φ can be described as

$$\Phi(\cdot,\cdot) : H = (F,G) \mapsto \begin{bmatrix} H(\delta_{p_1,x_1}, \delta_{p_1,\mathrm{P}_T(x_1)}) & \dots & H(\delta_{p_1,x_M}, \delta_{p_1,\mathrm{P}_T(x_M)}) \\ & \dots & \\ H(\delta_{p_N,x_1}, \delta_{p_N,\mathrm{P}_T(x_1)}) & \dots & H(\delta_{p_N,x_M}, \delta_{p_N,\mathrm{P}_T(x_M)}) \end{bmatrix}. \tag{6.16}$$

$$\Phi(\cdot,\cdot) : H = (F,G) \mapsto \begin{bmatrix} F(\delta_{p_1,x_1}) + G(\delta_{p_1,\mathrm{P}_T(x_1)}) & \dots & F(\delta_{p_1,x_M}) + G(\delta_{p_1,\mathrm{P}_T(x_M)}) \\ & \dots & \\ F(\delta_{p_N,t_1}) + G(\delta_{p_N,\mathrm{P}_T(x_1)}) & \dots & F(\delta_{p_N,x_M}) + G(\delta_{p_N,\mathrm{P}_T(x_M)}) \end{bmatrix}. \tag{6.17}$$

Remark that in Theorem 6.6 we also ask the family of points $(f_{i,j}, g_{i,j})$ to be linearly independant. Here this family is given by $(\delta_{p_i,x_j}, \delta_{p_i,\mathrm{P}_T(x_j)})$ and it can be easily checked that as long as the x_j are all distinct and the p_i are all distinct, we do have linearly independance. It remains to specify the cost functional. In the following we will study a square cost, thus, C can be described as

$$C(\cdot,\cdot) : (\mathbf{Y}, \mathbf{\Phi}) \mapsto \sum_{i,j=1,1}^{N,M} (\mathbf{Y}_{i,j} - \mathbf{\Phi}_{i,j})^2 = \|\mathbf{Y} - \mathbf{\Phi}\|_2^2. \tag{6.18}$$

It is easy to see that this functional is, as require Theorem 6.6, proper and convex and lower semi-continous. We can now restate Theorem 6.6 with the previous specification of the sampling operator and the cost functional.

▶ **Remark 6.8.** It is to be remarked that the structure we just gave to the sampling functional implies at first glance that the data have to be structured in having the same spatial sampling points along time. We are giving data this format as it easier to apprehend and analyse with it. Nevertheless one can easily come back the "general" case where the spatial sampling points are not constant against time, by putting 0 at their corresponding location in the $\mathbf{Y}$ matrix and by considering a cost functional that does penalize with a 0 (not a quadratic term) in the corresponding location. In this case of ill-structure all the future results can be derived in the exact same manner as we are now doing and the maximal number of innovations here found to be $I_s + I_t$ will be reduced by the number of points where we lack information. Informatically speaking, one can deal with this inconvenience with the use of the so called mask matrix. ◀

▶ **Corollary 6.9 (Specification of the representer theorem).** Suppose that the following assumptions hold true.

1. $\mathcal{D} = (\mathcal{D}_t, \mathcal{D}_s)$ is a STTS differential operator.

2. λ is a strictly positive regularisation constant and $\{\alpha, \beta\}$ are two strictly positive constants.

3. According to Lemma 5.6, $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}} = \alpha\|\cdot\|_{\mathcal{D}_t} + \beta\|\cdot\|_{\mathcal{D}_s}$ and its predual $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}^{-1}} = max(\frac{1}{\alpha}\|\cdot\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|\cdot\|_{\mathcal{D}_s^{-1}})$.

4. $\forall (p,x) \in \mathbb{S}^2 \times \mathbb{R}, \delta_{p,x} \in C_{\mathcal{D}_t^{-1},0}$. $\quad \forall (p,c) \in \mathbb{S}^2 \times \mathbb{S}^1, \delta_{p,c} \in C_{\mathcal{D}_s^{-1},0}$.

5. N and M are two integers.$\{(p_i, x_j)\}_{i,j=1,1}^{N,M} \subset \mathbb{S}^2 \times \mathbb{R}$, are fixed, unformally seen as the points where data has been queried. $\mathbf{Y} \in \mathbb{R}^{N\times M}$, is fixed, unformally seen as the evaluation of a *phenomenon* on the previous set of points.

6. $\mathbf{\Phi}$ is the sampling operator defined above. C is the cost functional defined above.

Then $\forall \mathrm{Y} \in \mathbb{R}^{N\times M}$ the solution set of the infinite dimensional optimization problem:

$$\mathcal{V} = \underset{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{\operatorname{argmin}} \left\{ \sum_{i,j=1,1}^{N,M} (\mathbf{Y}_{i,j} - F(\delta_{p_i,x_j}) - G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\} \tag{6.19}$$

is non-empty and representable as the *weak** closed convex hull of its extreme points. The latter are necessarily (but not exactly) of the form:

$$\sum_{i=1}^{I_t} \gamma_{1,i}\left(\psi^1_{p^1_i} \otimes \zeta_{t_i}, 0\right) + \sum_{i=1}^{I_s} \gamma_{2,i}\left(0, \psi^2_{p^2_i} \otimes \phi_{c_i}\right) \tag{6.20}$$

$$I_s + I_t \leq NM, \quad \{\{\gamma_{1,i}\}_{i=1}^{I_t}, \{\gamma_{2,i}\}_{i=1}^{I_s}\} \subset \mathbb{R}, \quad \{\{p^1_i\}_{i=1}^{I_t}\}, \{p^2_i\}_{i=1}^{I_s}\}\} \subset \mathbb{S}^2,$$
$$\{t_i\}_{i=1}^{I_t} \subset \mathbb{R}, \quad \{c_i\}_{i=1}^{I_s} \subset \mathbb{S}^1.$$

◀

We are now interested to find a solution to the problem

$$\min_{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \left\{ \sum_{i,j=1,1}^{N,M} (\mathbf{Y}_{i,j} - F(\delta_{p_i,x_j}) - G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\}. \tag{6.21}$$

With this solution being a candidate to the representation of the *phenomenon*. The previous theorem tells us that the set of solution is a convex hull and characterize its extremal points. We mostly want one solution and do not necessarily care about finding all the solutions. Moreover we care about the simplicity of our solution. In this way Corollary 6.9 guarantees us that searching for an extremal points as a solution and not an interior point, is a good idea, as they have at most MN innovations. As a consequence we will study the problem of minimizing

$$\sum_{i,j=1,1}^{N,M} (\mathbf{Y}_{i,j} - F(\delta_{p_i,x_j}) - G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s}$$

subject to

$$(F,G) = \left(\sum_{i=1}^{NM} \gamma_{1,i} \psi^1_{p_i^1} \otimes \zeta_{t_i}, \sum_{i=1}^{NM} \gamma_{2,i} \psi^2_{p_i^2} \otimes \phi_{c_i} \right)$$

$\{\{\gamma_{1,i}\}_{i=1}^{NM}, \{\gamma_{2,i}\}_{i=1}^{NM}\} \subset \mathbb{R}, \quad \{\{p_i^1\}_{i=1}^{NM}\}, \{p_i^2\}_{i=1}^{NM}\}\} \subset \mathbb{S}^2, \quad \{t_i\}_{i=1}^{NM} \subset \mathbb{R}, \quad \{c_i\}_{i=1}^{NM} \subset \mathbb{S}^1.$

Where we have replaced both I_t and I_s by MN as we have no idea about their division. One could argue that we are loosing sharpness by doing so and enforcing $I_t + I_s = NM$ is also a good idea. Our idea is to enlarge our search space voluntarily in order to simplify its representation, to simply the latter algorithm. We will say that (F,G) verifies the previous conditions if and only if $(F,G) \in \Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})$. We now have to make two important remarks.

1. For $(F,G) \in \Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})$

$$\|(F,G)\|_{\mathcal{D}} = \alpha \left\| \sum_{i=1}^{NM} \gamma_{1,i} \psi^1_{p_i^1} \otimes \zeta_{t_i} \right\|_{\mathcal{D}_t} + \beta \left\| \sum_{i=1}^{NM} \gamma_{2,i} \psi^2_{p_i^2} \otimes \phi_{c_i} \right\|_{\mathcal{D}_s} =$$

$$\alpha \left\| \sum_{i=1}^{NM} \gamma_{1,i} \mathcal{D}_t(\psi^1_{p_i^1} \otimes \zeta_{t_i}) \right\|_{TV} + \beta \left\| \sum_{i=1}^{NM} \gamma_{2,i} \mathcal{D}_s(\psi^2_{p_i^2} \otimes \phi_{c_i}) \right\|_{TV} =$$

$$\alpha \left\| \sum_{i=1}^{NM} \gamma_{1,i} \mathcal{D}_t(\mathcal{D}_t^{-1}(\delta_{p_i^1,t_i})) \right\|_{TV} + \beta \left\| \sum_{i=1}^{NM} \gamma_{2,i} \mathcal{D}_s(\mathcal{D}_s^{-1}(\delta_{p_i^2,c_i})) \right\|_{TV} =$$

$$\alpha \sum_{i=1}^{NM} |\gamma_{1,i}| + \beta \sum_{i=1}^{NM} |\gamma_{2,i}| = \alpha \|\gamma_1\|_1 + \beta \|\gamma_2\|_1.$$

With the notation

$$\gamma_1 \in \mathbb{R}^{NM}, \gamma_{1,i} = \gamma_{1,i}. \quad \gamma_2 \in \mathbb{R}^{NM}, \gamma_{2,i} = \gamma_{2,i}.$$

In the last line of the derivation we made use of the TV norm of a sum of Dirac masses. To sum up

$$\|(F,G)\|_{\mathcal{D}} = \alpha \|\gamma_1\|_1 + \beta \|\gamma_2\|_1 \tag{6.22}$$

2. We know by Corollary 6.9 that some elements in $\Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})$ are solutions (minimizer) of the problem evoked Corollary 6.9. Therefore we have that

$$\min_{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}\left\{\sum_{i,j=1,1}^{N,M}(\mathbf{Y}_{i,j}-F(\delta_{p_i,x_j})-G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\} =$$
$$\min_{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}\left\{\sum_{i,j=1,1}^{N,M}(\mathbf{Y}_{i,j}-F(\delta_{p_i,x_j})-G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\} \tag{6.23}$$

and

$$\operatorname*{argmin}_{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}\left\{\sum_{i,j=1,1}^{N,M}(\mathbf{Y}_{i,j}-F(\delta_{p_i,x_j})-G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\} \supset$$
$$\operatorname*{argmin}_{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}\left\{\sum_{i,j=1,1}^{N,M}(\mathbf{Y}_{i,j}-F(\delta_{p_i,x_j})-G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\}. \tag{6.24}$$

Thanks to the two previous remark we can now state the last result of this chapter.

▶ **Corollary 6.10 (Specification 2 of the representer theorem).** Suppose that the following assumptions hold true.

1. $\mathcal{D} = (\mathcal{D}_t, \mathcal{D}_s)$ is a STTS differential operator.
2. λ is a strictly positive regularisation constant and $\{\alpha, \beta\}$ are two strictly positive constants.
3. According to Lemma 5.6, $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}} = \alpha\|\cdot\|_{\mathcal{D}_t} + \beta\|\cdot\|_{\mathcal{D}_s}$ and its predual $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is equipped with the norm $\|(\cdot,\cdot)\|_{\mathcal{D}^{-1}} = max(\frac{1}{\alpha}\|\cdot\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|\cdot\|_{\mathcal{D}_s^{-1}})$.
4. $\forall(p,x) \in \mathbb{S}^2 \times \mathbb{R}, \delta_{p,x} \in C_{\mathcal{D}_t^{-1},0}$. $\quad \forall(p,c) \in \mathbb{S}^2 \times \mathbb{S}^1, \delta_{p,c} \in C_{\mathcal{D}_s^{-1},0}$.
5. N and M are two integers.$\{(p_i, x_j)\}_{i,j=1,1}^{N,M} \subset \mathbb{S}^2\times\mathbb{R}$, are fixed, unformally seen as the points where data has been queried. $\mathbf{Y} \in \mathbb{R}^{N\times M}$, is fixed, unformally seen as the evaluation of a *phenomenon* on the previous set of points.
6. $\mathbf{\Phi}$ is the sampling operator defined above. C is the cost functional defined above.

Then $\forall \mathrm{Y} \in \mathbb{R}^{N\times M}$ the solution set of the infinite dimensional optimization problem:

$$\mathcal{V}' = \underset{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}} \left\{ \sum_{i,j=1,1}^{N,M} (\mathrm{Y}_{i,j} - F(\delta_{p_i,x_j}) - G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\} \tag{6.25}$$

is not empty. Moreover if $H \in \mathcal{V}'$ then H also belongs to

$$\underset{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{\operatorname{argmin}} \left\{ \sum_{i,j=1,1}^{N,M} (\mathrm{Y}_{i,j} - F(\delta_{p_i,x_j}) - G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\}.$$

◀

As suggested by Corollary 6.10 we can study the optimization problem restricted to $\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$ in order to find a solution to the optimization on $\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}$. By going in this direction, one question that quickly arise is; what is the shape of $\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$? Which degrees of freedom does it have ? Recall that $(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$ if and only if

$$(F,G) = \left(\sum_{i=1}^{NM} \gamma_{1,i}\psi^1_{p^1_i} \otimes \zeta_{t_i}, \sum_{i=1}^{NM} \gamma_{2,i}\psi^2_{p^2_i} \otimes \phi_{c_i}) \right)$$

$$\{\{\gamma_{1,i}\}_{i=1}^{NM}, \{\gamma_{2,i}\}_{i=1}^{NM}\} \subset \mathbb{R}, \quad \{\{p^1_i\}_{i=1}^{NM}\}, \{p^2_i\}_{i=1}^{NM}\}\} \subset \mathbb{S}^2, \quad \{t_i\}_{i=1}^{NM} \subset \mathbb{R}, \quad \{c_i\}_{i=1}^{NM} \subset \mathbb{S}^1.$$

We have the real amplitude parameters $\{\{\gamma_{1,i}\}_{i=1}^{NM}, \{\gamma_{2,i}\}_{i=1}^{NM}\}$ that are not a big problem as they can just be seen as multiplier of basis functions. Then we have the nodes $\{p^1_i\}_{i=1}^{NM}\}, \{p^2_i\}_{i=1}^{NM}\}\}, \{t_i\}_{i=1}^{NM}, \{c_i\}_{i=1}^{NM}$; the problem with them is thicker. Suppose we have a finite number these nodes, then the amplitudes are just scalar multipliers and $\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$ is just a vector space of dimension at most the number of nodes. But in our situation we have infinitely nodes, what suggests that $\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$ is a vector space of infinite dimension. Our goal is to structure the problem (5.25) of Corollary 6.10 and to solve it numerically. This will not be possible as of our current state of work, because of the possibly infinite dimensionality of $\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})$. This leads to the next chapter.

7 Discretization of the specification

The goal is still to furnish a solution to Equation (1.1). To do so we formulated an optimization problem in Theorem 6.6. Solving this problem, at least heuristically, would provide us with a solution to Equation (1.1). In Theorem 6.6 we characterized the set of solutions, with the main issue being its complexity. Therefore in Corollary 6.9 and Corollary 6.10 we restricted our research by fixing a sampling operator, fixing a cost functional and most important restricting the search space to a new one, embedded in the old one, that contains the extremal points of the original solution set. Our goal is now to solve this new restricted problem established Corollary 6.10 and as already discussed the biggest problem we face is its dimensionality. In this chapter we will work in the direction of the discretization of (5.25) from Corollary 6.10.

Recall the considered problem

$$\underset{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}}\left\{\sum_{i,j=1,1}^{N,M}(\mathbf{Y}_{i,j}-F(\delta_{p_i,x_j})-G(\delta_{p_i,\mathrm{P}_T(x_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\}$$

with

$$(F,G)=\left(\sum_{i=1}^{NM}\gamma_{1,i}\psi^1_{p_i^1}\otimes\zeta_{t_i},\sum_{i=1}^{NM}\gamma_{2,i}\psi^2_{p_i^2}\otimes\phi_{c_i})\right)$$

$$\{\{\gamma_{1,i}\}_{i=1}^{NM},\{\gamma_{2,i}\}_{i=1}^{NM}\}\subset\mathbb{R},\quad\{\{p_i^1\}_{i=1}^{NM}\},\{p_i^2\}_{i=1}^{NM}\}\}\subset\mathbb{S}^2,\quad\{t_i\}_{i=1}^{NM}\subset\mathbb{R},\quad\{c_i\}_{i=1}^{NM}\subset\mathbb{S}^1.$$

The notation might become heavy in the near future so we will try to be crystal clear. We shall keep in mind that $\{p_i,x_j\}_{i,j=1,1}^{N,M}$ and $\mathbf{Y}$ are fixed. In order to make this clear, in the following, we will replace $\{p_i,x_j\}_{i,j=1,1}^{N,M}$ by $\{\overline{p}_i,\overline{x}_j\}_{i,j=1,1}^{N,M}$ and $\mathbf{Y}$ by $\overline{\mathbf{Y}}$. The considered problem becomes

$$\underset{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}}\left\{\sum_{i,j=1,1}^{N,M}(\overline{\mathbf{Y}}_{i,j}-F(\delta_{\overline{p}_i,\overline{x}_j})-G(\delta_{\overline{p}_i,\mathrm{P}_T(\overline{x}_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\}.$$

We will also use $\tilde{}$ to indicate that a parameter is design dependant, in the sense that it is not absolutely fixed, the practitioner have to choose it prior to solving the optimization problem. We will use this notation for example in the next definition.

We will also use nothing, for example p, in contrast with $\tilde{p}$ or $\overline{p}$, to indicate that the parameter is free.

▶ **Definition 7.1.** For four integers $\{N_t, M_t, N_s, M_s\} \subset \mathbb{N}$. For four sets of points $\{\tilde{p}^1_i\}_{i=1}^{N_t} \subset \mathbb{S}^2$, $\{\tilde{x}_i\}_{i=1}^{M_t} \subset \mathbb{R}$, $\{\tilde{p}^2_i\}_{i=1}^{N_s} \subset \mathbb{S}^2$, $\{\tilde{c}_i\}_{i=1}^{M_s} \subset \mathbb{S}^1$ called the nodes of the discretization. We define

1. $\Pi(\mathcal{M}_{\mathcal{D}_t}) := \Pi(\mathcal{M}_{\mathcal{D}_t})\left[\{\tilde{p}^1_i\}_{i=1}^{N_t}, \{\tilde{x}_i\}_{i=1}^{M_t}\right] := \left\{\sum_{i,j=1,1}^{N_t,M_t} a_{i,j}\psi^1_{\tilde{p}^1_i} \otimes \zeta_{\tilde{x}_j} \middle| \quad \{a_{i,j}\}_{i,j=1,1}^{N_t,M_t} \subset \mathbb{R}\right\} \subset \mathcal{M}_{\mathcal{D}_t}$.

2. $\Pi(\mathcal{M}_{\mathcal{D}_s}) := \Pi(\mathcal{M}_{\mathcal{D}_s})\left[\{\tilde{p}^2_i\}_{i=1}^{N_s}, \{\tilde{c}_i\}_{i=1}^{M_s}\right] := \left\{\sum_{i,j=1,1}^{N_s,M_s} b_{i,j}\psi^2_{\tilde{p}^2_i} \otimes \phi_{\tilde{c}_j} \middle| \quad \{b_{i,j}\}_{i,j=1,1}^{N_s,M_s} \subset \mathbb{R}\right\} \subset \mathcal{M}_{\mathcal{D}_s}$.

3. $\Pi(\mathcal{M}_{\mathcal{D}_t})\times\Pi(\mathcal{M}_{\mathcal{D}_s}) := \Pi(\mathcal{M}_{\mathcal{D}_t})\left[\{\tilde{p}^1_i\}_{i=1}^{N_t}, \{\tilde{x}_i\}_{i=1}^{M_t}\right]\times\Pi(\mathcal{M}_{\mathcal{D}_s})\left[\{\tilde{p}^2_i\}_{i=1}^{N_s}, \{\tilde{c}_i\}_{i=1}^{M_s}\right]$.

On the interpretation, N_t has to be seen as the degrees of freedom we give to the spatial reconstruction of the trend part of our problem, M_t has to be seen as the degrees of freedom we give to the temporal reconstruction of the trend part of our problem. Same goes for N_s, M_s and the seasonal part. Moreover $\Pi(\mathcal{M}_{\mathcal{D}_t})$ has to be seen as the discretization with N_t spatial degrees of freedom and M_t temporal degrees of freedom, of the set

$$\left\{F = \sum_{i=1}^{NM} \gamma_{1,i}\psi^1_{p^1_i} \otimes \zeta_{x_i} \middle| \quad \{\gamma_{1,i}\}_{i=1}^{NM} \subset \mathbb{R}, \quad \{p^1_i\}_{i=1}^{NM} \subset \mathbb{S}^2, \quad \{x_i\}_{i=1}^{NM} \subset \mathbb{R}\right\}.$$

Also $\Pi(\mathcal{M}_{\mathcal{D}_t})$ is a vector space with at most N_tM_t dimensions. The same idea goes with $\Pi(\mathcal{M}_{\mathcal{D}_s})$, that is a discretized search space and a vector space of dimension at most N_sM_s. ◀

Our goal in this chapter is to replace the search space $\Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})$ by its discretized version $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$. We will here establish what the optimization problem constrainted on this space look like and in the next chapter we will look at its performance, asymptotic behavior. In order to understand the problem on $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$ we need to understand how the sampling operator works, what we will do first, and then to understand how the cost (penalization) can be formulated.

7.1 Adaptation of the sampling operator to the discretized search space $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$

We start this section by the definition of useful matrices

▶ **Definition 7.2.**

$$\Psi_{trend} := \begin{bmatrix} \psi^1_{\tilde{p}^1_1}(\delta_{\overline{p}_1}) & \dots & \psi^1_{\tilde{p}^1_{N_t}}(\delta_{\overline{p}_1}) \\ & \dots & \\ \psi^1_{\tilde{p}^1_1}(\delta_{\overline{p}_N}) & \dots & \psi^1_{\tilde{p}^1_{N_t}}(\delta_{\overline{p}_N}) \end{bmatrix} \tag{7.1}$$

$$\Psi_{seas} := \begin{bmatrix} \psi^2_{\tilde{p}^2_1}(\delta_{\overline{p}_1}) & \dots & \psi^2_{\tilde{p}^2_{N_t}}(\delta_{\overline{p}_1}) \\ & \dots & \\ \psi^2_{\tilde{p}^2_1}(\delta_{\overline{p}_N}) & \dots & \psi^2_{\tilde{p}^2_{N_t}}(\delta_{\overline{p}_N}) \end{bmatrix} \tag{7.2}$$

$$Z := \begin{bmatrix} \zeta_{\tilde{x}_1}(\delta_{\overline{x}_1}) & \dots & \zeta_{\tilde{x}_{M_t}}(\delta_{\overline{x}_1}) \\ & \dots & \\ \zeta_{\tilde{x}_1}(\delta_{\overline{x}_M}) & \dots & \zeta_{\tilde{x}_{M_t}}(\delta_{\overline{x}_M}) \end{bmatrix} \tag{7.3}$$

$$\Phi := \begin{bmatrix} \phi_{\tilde{c}_1}(\delta_{\mathrm{P}_T(\overline{x}_1)}) & \dots & \phi_{\tilde{c}_{M_s}}(\delta_{\mathrm{P}_T(\overline{x}_1)}) \\ & \dots & \\ \phi_{\tilde{c}_1}(\delta_{\mathrm{P}_T(\overline{x}_M)}) & \dots & \phi_{\tilde{c}_{M_s}}(\delta_{\mathrm{P}_T(\overline{x}_M)}) \end{bmatrix}. \tag{7.4}$$

◀

The previous matrices need to be thought as design matrices as they depend only on the pseudo differential operators we are using, the nodes of our discretization scheme, the points where we have information $(\overline{p}_i, \overline{x}_j)$. We also need to define the following weight matrices

▶ **Definition 7.3.** For $\{a_{i,j}\}_{i,j=1,1}^{N_t,M_t} \subset \mathbb{R}$ and $\{b_{i,j}\}_{i,j=1,1}^{N_s,M_s} \subset \mathbb{R}$ we define

$$A := \begin{bmatrix} a_{1,1} & \dots & a_{1,M_t} \\ & \dots & \\ a_{N_t,1} & \dots & a_{N_t,M_t} \end{bmatrix} \tag{7.5}$$

$$B := \begin{bmatrix} b_{1,1} & \dots & b_{1,M_s} \\ & \dots & \\ b_{N_s,1} & \dots & b_{N_s,M_s} \end{bmatrix} \tag{7.6}$$

◀

Using both Definition 7.2 and Definition 7.3 allow us to adapt the action of the sampling operator on $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$. This is summarized in the next lemma. Before it statement recall that

$$\Phi(F,G) = \begin{bmatrix} F(\delta_{\overline{p}_1,\overline{x}_1}) + G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_1)}) & \dots & F(\delta_{\overline{p}_1,\overline{x}_M}) + G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_M)}) \\ & \dots & \\ F(\delta_{\overline{p}_N,\overline{x}_1}) + G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_1)}) & \dots & F(\delta_{\overline{p}_N,\overline{x}_M}) + G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_M)}) \end{bmatrix} =$$

$$\begin{bmatrix} F(\delta_{\overline{p}_1,\overline{x}_1}) & \dots & F(\delta_{\overline{p}_1,\overline{x}_M}) \\ & \dots & \\ F(\delta_{\overline{p}_N,\overline{x}_1}) & \dots & F(\delta_{\overline{p}_N,\overline{x}_M}) \end{bmatrix} + \begin{bmatrix} G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_1)}) & \dots & G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_M)}) \\ & \dots & \\ G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_1)}) & \dots & G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_M)}) \end{bmatrix}.$$

▶ **Lemma 7.4.** For $(F,G) \in \Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$ as $(F,G) = \left(\sum_{i,j=1,1}^{N_t,M_t} a_{i,j}\psi^1_{\tilde{p}^1_i} \otimes \zeta_{\tilde{x}_j}, \sum_{i,j=1,1}^{N_s,M_s} b_{i,j}\psi^2_{\tilde{p}^2_i} \otimes \phi_{\tilde{c}_j}\right)$ we have that

$$\begin{bmatrix} F(\delta_{\overline{p}_1,\overline{x}_1}) & \dots & F(\delta_{\overline{p}_1,\overline{x}_M}) \\ & \dots & \\ F(\delta_{\overline{p}_N,\overline{x}_1}) & \dots & F(\delta_{\overline{p}_N,\overline{x}_M}) \end{bmatrix} = \Psi_{trend} A Z^{\star} \tag{7.7}$$

and

$$\begin{bmatrix} G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_1)}) & \dots & G(\delta_{\overline{p}_1,\mathrm{P}_T(\overline{x}_M)}) \\ & \dots & \\ G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_1)}) & \dots & G(\delta_{\overline{p}_N,\mathrm{P}_T(\overline{x}_M)}) \end{bmatrix}. = \Psi_{seas} B \Phi^{\star}. \tag{7.8}$$

Where $\Psi_{trend}, Z, \Psi_{seas}, \Phi$ are the design matrices from Definition 7.2 and A, B are the weight matrices from Definition 7.3. In consequence we have that

$$\Phi(F,G) = \Psi_{trend} A Z^{\star} + \Psi_{seas} B \Phi^{\star}. \tag{7.9}$$

◀

Proof of Lemma 7.4: We do not provide the proof as it is only a lengthy direct computation. ■

It is also possible to re-write the previous lemma in a different manner. Each of these manners will have their own advantages and will be later used in different contexts. To do so we need the following definition.

▶ **Definition 7.5.**

- We define the Kronecker product $\otimes$ between matrices in the following way. For $\{p, q\} \in \mathbb{N}$ and $\{n, m\} \in \mathbb{N}$

$$\otimes : \mathbb{R}^{p\times q} \times \mathbb{R}^{n\times m} \longrightarrow \mathbb{R}^{pn\times qm},$$

$$\otimes : (A, B) \mapsto \begin{bmatrix} A_{1,1}B & \dots & A_{1,q}B \\ & \dots & \\ A_{p,1}B & \dots & A_{p,q}B \end{bmatrix}.$$

- We define the vectorization operator vec in the following way. For $A \in \mathbb{R}^{p\times q}$

$$\text{vec} : \mathbb{R}^{p\times q} \longrightarrow \mathbb{R}^{pq},$$

$$\text{vec} : A \mapsto \begin{bmatrix} A_{1,1} \\ A_{1,2} \\ \dots \\ A_{1,q} \\ A_{2,1} \\ \dots \\ A_{p-1,q} \\ A_{p,1} \\ \dots \\ A_{p,q} \end{bmatrix}.$$

  Unformally, vec can be seen as the operator that takes a matrix and returns a vector formed by the concatenated columns of the given matrix.

- We define the operator mat as the inverse of the operator vec.

◀

▶ **Lemma 7.6.** For the same setting as in the previous lemma, we have

$$vec(\Psi_{trend} A Z^{\star}) = Z \otimes \Psi_{trend}\text{vec}(A) \tag{7.10}$$

and

$$vec(\Psi_{seas} B \Phi^{\star}) = \Phi \otimes \Psi_{seas}\text{vec}(B). \tag{7.11}$$

In consequence we have

$$vec(\mathbf{\Phi}(F, G)) = Z \otimes \Psi_{trend}\text{vec}(A) + \Phi \otimes \Psi_{seas}\text{vec}(B). \tag{7.12}$$

◀

Proof of Lemma 7.6: Without entering the details we can say that the proof is a direct consequence of Lemma 7.4 and properties of the Kronecker product. ■

Continuing on adapting, one could say simplify, our optimization problem to the new search space. We tackle in the next section the penalization.

7.2 Adaptation of the penalization operator to the discretized search space $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$

For $(F, G) \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ we used to penalize our problem with

$$\lambda \|F, G\|_{\mathcal{D}}.$$

We see with the next lemma that this can be adapted or refined, provided that $(F, G) \in \Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$.

▶ **Lemma 7.7.** For $(F, G) \in \Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$ we have that

$$\|(F, G)\|_{\mathcal{D}} = \alpha \left\| \sum_{i,j=1,1}^{N_t, M_t} a_{i,j} \psi^1_{\tilde{p}^1_i} \otimes \zeta_{\tilde{x}_j} \right\|_{\mathcal{D}_t} + \beta \left\| \sum_{i,j=1,1}^{N_s, M_s} b_{i,j} \psi^2_{\tilde{p}^2_i} \otimes \phi_{\tilde{c}_j} \right\|_{\mathcal{D}_s} = \alpha \|A\|_1 + \beta \|B\|_1. \tag{7.13}$$

Where A and B are the weight matrices given in Definition 7.3. ◀

Proof of Lemma 7.7: The proof has already been made as the first point of a remark before the statement of Corollary 6.10. ■

A consequence of the previous lemma is that now the penalization can be re-written as

$$\lambda(\alpha \|A\|_1 + \beta \|B\|_1). \tag{7.14}$$

We are now at a point where, using section 6.2 and 6.1, we can formulate the optimization problem on the discretized search space $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$.

7.3 A new optimization problem on the discretized search space $\varPi(\mathcal{M}_{\mathcal{D}_t}) \times \varPi(\mathcal{M}_{\mathcal{D}_s})$

We recall that the original (as already discussed in the introduction of the chapter, it is restricted) problem was

$$\underset{(F,G)\in\varPi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}} \left\{ \sum_{i,j=1,1}^{N,M} \overline{\mathrm{Y}}_{i,j} - F(\delta_{\overline{p}_i,\overline{x}_j}) - G(\delta_{\overline{p}_i,\mathrm{P}_T(\overline{x}_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\}. \tag{7.15}$$

Using the discretization $\varPi(\mathcal{M}_{\mathcal{D}_t}) \times \varPi(\mathcal{M}_{\mathcal{D}_s})$ it gives

$$\underset{(F,G)\in\varPi(\mathcal{M}_{\mathcal{D}_t})\times\varPi(\mathcal{M}_{\mathcal{D}_s}))}{\operatorname{argmin}} \left\{ \sum_{i,j=1,1}^{N,M} \overline{\mathrm{Y}}_{i,j} - F(\delta_{\overline{p}_i,\overline{x}_j}) - G(\delta_{\overline{p}_i,\mathrm{P}_T(\overline{x}_j)}))^2 + \lambda\alpha\|F\|_{\mathcal{D}_t} + \lambda\beta\|G\|_{\mathcal{D}_s} \right\}. \tag{7.16}$$

Using Lemma 7.4 and Lemma 7.7 it gives

$$\underset{(F,G)\in\varPi(\mathcal{M}_{\mathcal{D}_t})\times\varPi(\mathcal{M}_{\mathcal{D}_s}))}{\operatorname{argmin}} \left\{ \left\|\overline{\mathrm{Y}} - \Psi_{trend} A Z^\star - \Psi_{seas} B \Phi^\star \right\|_2^2 + \lambda\alpha\|A\|_1 + \lambda\beta\|B\|_1 \right\}. \tag{7.17}$$

With equation (7.17) it is now easy to see that the only freedom we have to minimize our functional is through the matrices A and B. Therefore we can restate the problem into

$$\underset{(A,B)\in\mathbb{R}^{N_t\times M_t}\times\mathbb{R}^{N_s\times M_s}}{\operatorname{argmin}} \left\{ \left\|\overline{\mathrm{Y}} - \Psi_{trend} A Z^\star - \Psi_{seas} B \Phi^\star \right\|_2^2 + \lambda\alpha\|A\|_1 + \lambda\beta\|B\|_1 \right\}. \tag{7.18}$$

▶ **Corollary 7.8.** According to the previous derivation we have that

$$\underset{(A,B)\in\mathbb{R}^{N_t\times M_t}\times\mathbb{R}^{N_s\times M_s}}{\operatorname{argmin}} \left\{ \left\|\overline{\mathrm{Y}} - \Psi_{trend} A Z^\star - \Psi_{seas} B \Phi^\star \right\|_2^2 + \lambda\alpha\|A\|_1 + \lambda\beta\|B\|_1 \right\} =$$
$$\underset{(F,G)\in\varPi(\mathcal{M}_{\mathcal{D}_t})\times\varPi(\mathcal{M}_{\mathcal{D}_s}))}{\operatorname{argmin}} \left\{ \left\|\overline{\mathrm{Y}} - \Psi_{trend} A Z^\star - \Psi_{seas} B \Phi^\star \right\|_2^2 + \lambda\alpha\|A\|_1 + \lambda\beta\|B\|_1 \right\}. \tag{7.19}$$

Moreover $(A^\star, B^\star)$ is a minimizer of its respective problem if and only if

$$\left(\sum_{i,j=1,1}^{N_t,M_t} A^\star_{i,j} \psi^1_{\tilde{p}^1_i} \otimes \zeta_{\tilde{x}_j}, \sum_{i,j=1,1}^{N_s,M_s} B^\star_{i,j} \psi^2_{\tilde{p}^2_i} \otimes \phi_{\tilde{c}_j} \right)$$

is a minimizer of its respective problem. ◀

We shall keep in mind that the matrices $Z, \Psi_{trend}, \Psi_{seas}, \Phi$ are dependant on the discretization and therefore on the space $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$, even though it is not explicitly apparent in their written representation. Starting from now we will restrict our attention to Equation (7.18). As previously explained solving it provides us with a solution to the discretized problem and eventually a good approximation to the the original problem. It comes now to our mind the problem of how good this approximation is and the problem of studying the solution set of Equation (7.18). In a more abstract manner we want to study the structure of the solutions of Equation (7.18) and how it relates to the one of the original problem.

8 Discretization analysis of the specification

In this chapter we will study the solution set

$$\underset{(A,B)\in\mathbb{R}^{N_t\times M_t}\times\mathbb{R}^{N_s\times M_s}}{argmin}\left\{\left\|\overline{\mathbf{Y}}-\Psi_{trend}AZ^\star-\Psi_{seas}B\Phi^\star\right\|_2^2+\lambda\alpha\|A\|_1+\lambda\beta\|B\|_1\right\}. \quad (8.1)$$

and how it relates, through an asymptotic viewpoint, to its parent

$$\underset{(F,G)\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{argmin}\left\{\sum_{i,j=1,1}^{N,M}(\overline{\mathbf{Y}}_{i,j}-F(\delta_{\overline{p}_i,\overline{x}_j})-G(\delta_{\overline{p}_i,\mathrm{P}_T(\overline{x}_j)}))^2+\lambda\alpha\|F\|_{\mathcal{D}_t}+\lambda\beta\|G\|_{\mathcal{D}_s}\right\}. \quad (8.2)$$

To do so this chapter will be split in three sections, the analysis of the discretized problem,a warm-up for the asymptotic analysis, the asymptotic analysis. The goal will be to show that a solution of the discretized problem converges, in some sense, to a solution of the parent problem.

Before entering the details, we provide the time and space saving definition.

▶ **Definition 8.1.** For $H \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ we define

$$\mathcal{J}(H) = \left\|\overline{\mathbf{Y}} - \mathbf{\Phi}(H)\right\|_2^2 + \lambda\|H\|_{\mathcal{D}}, \quad (8.3)$$

$$\mathcal{J}(A,B) = \left\|\overline{\mathbf{Y}}-\Psi_{trend}AZ^\star-\Psi_{seas}B\Phi^\star\right\|_2^2+\lambda\alpha\|A\|_1+\lambda\beta\|B\|_1. \quad (8.4)$$

Where we assumed that $\mathbf{\Phi}$ is the same sampling operator (defined with $(\overline{p}_i, \overline{x}_j)$) that is used in Equation (8.1) and Equation (8.2). Therefore we can rewrite Equation (8.1) into

$$\underset{H\in\Pi(\mathcal{M}_{\mathcal{D}_t})\times\Pi(\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}} \mathcal{J}(H). \quad (8.5)$$

and Equation (8.2) into

$$\underset{H\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})}{\operatorname{argmin}} \mathcal{J}(H). \quad (8.6)$$

◀

8.1 Analysis of the finite discretization of the specification

In the following we will use the rk to denote the column rank of a matrix.

▶ **Lemma 8.2.** Suppose that the *rank* of $\Phi, \Psi_{trend}, Z, \Psi_{seas}$ is full. Then $\mathcal{J}$ admits a unique minimizer over $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$. ◀

Proof of Lemma 8.2: We start by noticing that, thanks to Corollary 7.8, showing that $\mathcal{J}(H)$ admits a unique minimizer over $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$ is equivalent to showing that $\mathcal{J}(A, B)$ admits a unique minimizer over $\mathbb{R}^{N_t \times M_t} \times \mathbb{R}^{N_s \times M_s}$. Moreover it is known that functionals that are strongly convex (with domain a product of real spaces) admit a unique minimizer. We will therefore try to show that the map

$$(A, B) \mapsto \mathcal{J}(A, B)$$

is strongly convex. It is first easy to see that the map

$$(A, B) \mapsto \lambda(\alpha\|A\|_1 + \beta\|B\|_1), \tag{8.7}$$

is convex. Thus it is enough to show that the map

$$(A, B) \mapsto \left\|\overline{\mathrm{Y}} - \Psi_{trend} A Z^\star - \Psi_{seas} B \Phi^\star\right\|_2^2 \tag{8.8}$$

is strongly convex, as the sum of a strongly convex and a convex map is strongly convex. In regard of Lemma 7.6 we can also rewrite Equation (8.8) into

$$(A, B) \mapsto \left\|vec(\overline{\mathrm{Y}}) - Z \otimes \Psi_{trend} vec(A) - \Phi \otimes \Psi_{seas} vec(B)\right\|_2^2 \tag{8.9}$$

It also does not change the problem to take $A \in \mathbb{R}^{N_t M_t}$ (not anymore in $\mathbb{R}^{N_t \times M_t}$) and $B \in \mathbb{R}^{N_s M_s}$ (not anymore in $\mathbb{R}^{N_s \times M_s}$). The map Equation (8.8) can then be re-written

$$C \mapsto \left\|vec(\overline{\mathrm{Y}}) - XC\right\|_2^2 \tag{8.10}$$

with $C := (A, B) \in \mathbb{R}^{N_t M_t + N_s M_s}$ and

$$X := \begin{bmatrix} Z \otimes \Psi_{trend} & 0 \\ 0 & \Phi \otimes \Psi_{seas} \end{bmatrix}. \tag{8.11}$$

It is again known that a sufficient condition for a twice differentiable functional to be strictly convex is having its Hessian to be positive definite. The Hessian of X can be found to be

$$2X^tX = 2\begin{bmatrix} (Z\otimes \Psi_{trend})^t Z\otimes \Psi_{trend} & 0 \\ 0 & (\Phi\otimes \Psi_{seas})^t \Phi\otimes \Psi_{seas} \end{bmatrix}.$$

It can easily be seen to be symmetric and semi-positive definite. Moreover it is positive-definite if and only if X has full *rank*. That event occurs if and only if both $Z\otimes \Psi_{trend}$ and $\Phi\otimes \Psi_{seas}$ have full *rank*. It is known that $rk(Z\otimes \Psi_{trend}) = rk(Z)rk(\Psi_{trend})$ and $rk(\Phi\otimes \Psi_{seas}) = rk(\Phi)rk(\Psi_{seas})$, as a property of the Kronecker product. Finally, we supposed that $\Phi, \Psi_{trend}, Z, \Psi_{seas}$ have full *rank*, this implies by the previous remark that both the *rank* of $Z\otimes \Psi_{trend}$ and $\Phi\otimes \Psi_{seas}$ are full and we have strong convexity. ■

It is worth noticing that the condition

$$\Phi, \Psi_{trend}, Z, \Psi_{seas} \text{ have all full rank,} \tag{8.12}$$

enforces the condition

$$N_t \le N, \quad N_s \le N, \quad M_t \le M, \quad M_s \le M. \tag{8.13}$$

8.2 A warm-up for the asymptotic analysis of the specification

In this section we will state and sometimes prove results that will be useful and necessary to conduct the asymptotic analysis. These results are rather theoretical and are here for the sake of rigour. Nevertheless not all definitions are provided as it is assumed that the reader is knowledgeable in topology and functional analysis. The goal will be to show that if X is a bounded subset of $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ then $weak^\star$ classical compactity of X implies $weak^\star$ sequential compactity of X.

▶ **Lemma 8.3.** The spaces $C_{\mathcal{D}_s^{-1},0}$ and $C_{\mathcal{D}_t^{-1},0}$ are separable. ◀

Proof of Lemma 8.3: These spaces are respectively congruent with $C_0(\mathbb{S}^2\times\mathbb{R})$ and $C_0(\mathbb{S}^2\times\mathbb{S}^1)$. Therefore it is sufficient to verify that $C_0(\mathbb{S}^2\times\mathbb{R})$ and $C_0(\mathbb{S}^2\times\mathbb{S}^1)$ are separable. It is known from classical functional analysis that the space of continuous real valued functions with for domain a compact space, equipped with the infinity norm, is separable. As $\mathbb{S}^2\times\mathbb{S}^1$ is compact, $C(\mathbb{S}^2\times\mathbb{S}^1)$ is separable and as $C_0(\mathbb{S}^2\times\mathbb{S}^1)\subset C(\mathbb{S}^2\times\mathbb{S}^1)$, $C_0(\mathbb{S}^2\times\mathbb{S}^1)$ is separable. We then study separability of $C_0(\mathbb{S}^2 \times \mathbb{R})$, because $\mathbb{S}^2 \times \mathbb{R}$ is not compact, we can not use the same argument. First define the (one sided stereographic projection) function:

$$
\begin{aligned}
&k : \mathbb{S}^2 \times (\mathbb{S}^1 - \{(0,1)\}) \longrightarrow \mathbb{S}^2 \times \mathbb{R}, \\
&\qquad k() : (p,(x,y)) \mapsto (p, \frac{x}{1-y}), \quad \forall p \in \mathbb{S}^2 \quad \forall (x,y) \in (\mathbb{S}^1 - \{(0,1)\}).
\end{aligned} \tag{8.14}
$$

Where $(\mathbb{S}^1 - \{(0,1)\})$ corresponds to the circle without the North pole. It is easy to see that this function is a bijection. We then consider the space $C := \{f \in C(\mathbb{S}^2 \times \mathbb{S}^1) | \quad \forall p \in \mathbb{S}^2, f(p,(0,1)) = 0\} \subset C(\mathbb{S}^2 \times \mathbb{S}^1)$, equipped with the infinity norm , it inherits its separability from $C(\mathbb{S}^2 \times \mathbb{S}^1)$. Finally define K as

$$
\begin{aligned}
&K : C_0(\mathbb{S}^2 \times \mathbb{R}) \longrightarrow C, \\
&\forall f \in: C_0(\mathbb{S}^2 \times \mathbb{R}) : \quad Kf(p,(x,y)) = \begin{cases} f(k(p,(x,y))), & (x,y) \neq (0,1), \\ 0, & (x,y) = (0,1), \end{cases}
\end{aligned} \tag{8.15}
$$

Because $\lim tf(p,t) = 0$, the function $K(f)$ is continuous in the North pole and $K(f) \in C$. We can also define the inverse function K^{-1} as

$$
\begin{aligned}
&K^{-1} : C \longrightarrow C_0(\mathbb{S}^2 \times \mathbb{R}), \\
&\qquad \forall f \in C : \quad \forall (p,t) \in \mathbb{S}^2 \times \mathbb{R} \quad K^{-1}f(p,t) = f(k^{-1}(p,t)).
\end{aligned} \tag{8.16}
$$

$K^{-1}f$ is obviously continuous and we have

$$
\lim_{\|(p,t)\| \to \infty} K^{-1}f(p,t) = \lim_{\|(p,t)\| \to \infty} f(k^{-1}(p,t)) = \lim_{(x,y) \to (0,1), p \to ?} f(p,(x,y)) = 0 \tag{8.17}
$$

In consequence $K^{-1}f \in C_0(\mathbb{S}^2 \times \mathbb{R})$. Finally it is easy to see that K^{-1} is the inverse of K and $\|f\|_\infty = \|K(f)\|_\infty$. The conclusions are that K is a congruence and the separability property of C gets transferred to $C_0(\mathbb{S}^2 \times \mathbb{R})$ through this congruence. ■

In order to make use of this separability result we need the general next lemma.

▶ **Lemma 8.4.** Let B be a separable normed vector space and $B^\star$ its dual equipped with the $weak^\star$ topology. For a bounded subspace X of $B^\star$ compactness is equivalent to sequential compactness. ◀

Proof of Lemma 8.4: It is known that in a metric space, compactness is equivalent to sequential compactness. We will then show that if $X \subset B^\star$ is bounded and equipped with the $weak^\star$ topology, then it is metrizable. As B is separable, we choose a family $\{b_n\}_{n=1}^\infty$ that is dense inside B. Inspired by "t.b." on math.stackexchange.com we put the following metric on X

$$d(f,g) = \sum_{n=1}^{\infty} 2^{-n} \frac{|f(b_n) - g(b_n)|}{1 + |f(b_n) - g(b_n)|}, \quad \forall \{f,g\} \in X. \tag{8.18}$$

First we check that d defines a metric. Symmetry is obvious. Identity of indiscernibles comes directly from the fact that f and g are continuous, thus if they agree on a dense subspace ($\{b_n\}_{n=1}^\infty$ here) they have to agree everywhere. Triangle inequality can be checked by some unwanted calculations. Secondly we have to verify that the topology generated by this metric agrees with the $weak^\star$ topology. To do so we will verify that $weak^\star$ convergence is equivalent to convergence in the metric d. Suppose that $weak^\star - \lim i f_i = f$. As $\frac{|f(b_n)-g(b_n)|}{1+|f(b_n)-g(b_n)|} \leq 1$ we can choose an integer N sufficiently big such that $\forall \{f,g\} \in X$ we have $\sum_{n=N}^\infty 2^{-n} \frac{|f(b_n)-g(b_n)|}{1+|f(b_n)-g(b_n)|} \leq \frac{\epsilon}{2}$. By supposition $\lim i f_i(b_n) = f(b_n)$ so that one can find, for a fixed N, I sufficiently big such that $\forall i > I \forall n < N$ we have $|f_i(b_n) - f(b_n)| \leq \frac{\epsilon}{2}$. This finally implies that $\forall i > I$, $\sum_{n=1}^\infty 2^{-n} \frac{|f(b_n)-f_i(b_n)|}{1+|f(b_n)-f_i(b_n)|} \leq \epsilon$ and $\lim i d(f, f_i) = 0$. Suppose now that $\lim i d(f, f_i) = 0$. We directly get that $\forall n \in \mathbb{N}$, $\underset{i\to\infty}{lim} f_i(b_n) = f(b_n)$. Moreover we get the chain

$$\begin{aligned} |f(b) - f_i(b)| &\leq |f(b) - f(b_n)| + |f(b_n) - f_i(b_n)| + |f_i(b_n) - f_i(b)| \\ &\leq 2\|b - b_n\|_B \underset{g\in X}{sup} \|g\|_{B^\star} + |f(b_n) - f_i(b_n)|. \end{aligned}$$

As X is bounded $\underset{g\in X}{sup}\|g\|_{B^\star} < \infty$ then the previous quantity can be made arbitrarily small by considering b_n close enough to b (density in separability) and i sufficiently big. We finally get that $weak^\star - \underset{i\to\infty}{lim} f_i = f$ and this concludes the proof. ■

Finally, in order to get sequential compactness, we need compactness. Which we will get with the help of the next theorem.

▶ **Theorem 8.5 (Banach-Alaoglu theorem).** For a Banach space B with dual space $B^\star$, the unit ball in the dual space is compact with respect to the $weak^\star$ topology. ◀

▶ **Lemma 8.6.** For $r \in \mathbb{R}_\star^+$, the ball $\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}(r, 0)$ is $weak^\star$ sequentially compact. ◀

Proof of Lemma 8.6: The Banach-Alaoglu theorem can easily be "upgraded" for the case when we have a ball of radius r. Then we know that $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is the dual space of the Banach space $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$. Therefore by the Banach-Alaoglu theorem, $\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}(r,0)$ is $weak^\star$ compact. Moreover we know by Lemma 8.3 that $C_{\mathcal{D}_t^{-1},0}$ and $C_{\mathcal{D}_s^{-1},0}$ are separable, then $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is also separable. Finally we can apply Lemma 8.4 with $B = C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ and the bounded subspace $X = \mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}(r,0)$, to get that for $\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}(r,0)$ sequentially compactness if equivalent to compactness. We know that this space is compact, therefore it is $weak^\star$ sequentially compact. ■

The warm up is finished and we can go picking the desired asymptotic results.

8.3 Asymptotic analysis of the specification

Take a discretization $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$ of $\Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})$ and recall that it depends on the sets of nodes $\left\{\tilde{p}_i^1, \tilde{x}_j\right\}_{i,j=1,1}^{N_t,M_t}$ and $\left\{\tilde{p}_i^2, \tilde{c}_j\right\}_{i,j=1,1}^{N_s,M_s}$ we used to discretize. The question we want to answer here is; What happens to the solutions of

$$\underset{H \in \Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})}{argmin} \quad \mathcal{J}(H)$$

when we refine more and more the search space $\Pi(\mathcal{M}_{\mathcal{D}_t}) \times \Pi(\mathcal{M}_{\mathcal{D}_s})$? Before answering this question we need to settle some notation. We will say that

$$\Xi^{1,1}(\mathbb{S}^2) = \left\{\tilde{p}_i^{1,1}\right\}_{i=1}^{N_{t,1}}, \quad \Xi^1(\mathbb{R}) = \left\{\tilde{x}_i^1\right\}_{i=1}^{M_{t,1}},$$
$$\Xi^{2,1}(\mathbb{S}^2) = \left\{\tilde{p}_i^{2,1}\right\}_{i=1}^{N_{s,1}}, \quad \Xi^1(\mathbb{S}^1) = \left\{\tilde{c}_i^1\right\}_{i=1}^{M_{s,1}}, \tag{8.19}$$

are the sets of point of our first discretization. The associated search space will be written

$$\Pi^1(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^1(\mathcal{M}_{\mathcal{D}_s}). \tag{8.20}$$

Iterating on these definitions, We will say

$$\Xi^{1,n}(\mathbb{S}^2) = \left\{\tilde{p}_i^{1,n}\right\}_{i=1}^{N_{t,n}}, \quad \Xi^n(\mathbb{R}) = \left\{\tilde{x}_i^n\right\}_{i=1}^{M_{t,n}},$$
$$\Xi^{2,n}(\mathbb{S}^2) = \left\{\tilde{p}_i^{2,n}\right\}_{i=1}^{N_{s,n}}, \quad \Xi^n(\mathbb{S}^n) = \left\{\tilde{c}_i^n\right\}_{i=1}^{M_{s,n}}, \tag{8.21}$$

are the sets of point of the n^{th} discretization. The associated search space will be written

$$\Pi^n(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^n(\mathcal{M}_{\mathcal{D}_s}). \tag{8.22}$$

It is straightforward to see that

$$\Pi^n(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^n(\mathcal{M}_{\mathcal{D}_s}) \subset \Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}), \tag{8.23}$$

and in consequence we have that

$$\underset{H \in \mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}}{min} \mathcal{J}(H) = \underset{H \in \Pi(\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s})}{min} \mathcal{J}(H) \leq \underset{H \in \Pi^n(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^n(\mathcal{M}_{\mathcal{D}_s})}{min} \mathcal{J}(H). \tag{8.24}$$

We want to know how tight is the last inequality. More precisely if it can become an equality when n tends to infinity. Before stating the result we need the following definition.

▶ **Definition 8.7.** For $X \in \{\mathbb{S}^2, \mathbb{S}^1, \mathbb{R}, [a, b]\}$ and $\Xi \subset X$ we define the nodal width mapping:

$$\Theta(X, \cdot) : \mathcal{P}(X) \longrightarrow \mathbb{R}^+ \cup \infty,$$

$$\Theta(X, \cdot) : \Xi \mapsto \underset{x \in X}{sup} \underset{y \in \Xi}{inf} \|x - y\|_2.$$

◀

▶ **Theorem 8.8 (Inspired by ?[Deb+19]).** Suppose that the following, now classical, assumptions hold:

1. $\mathcal{D} = (\mathcal{D}_t, \mathcal{D}_s)$ is a STTS differential operator.
2. λ is a strictly positive regularisation constant and $\{\alpha, \beta\}$ are two strictly positive constants.
3. According to Lemma 5.6, $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$ is equipped with the norm $\|(\cdot, \cdot)\|_{\mathcal{D}} = \alpha\|\cdot\|_{\mathcal{D}_t} + \beta\|\cdot\|_{\mathcal{D}_s}$ and its predual $C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ is equipped with the norm $\|(\cdot, \cdot)\|_{\mathcal{D}^{-1}} = max(\frac{1}{\alpha}\|\cdot\|_{\mathcal{D}_t^{-1}}, \frac{1}{\beta}\|\cdot\|_{\mathcal{D}_s^{-1}})$.
4. $\forall (p, x) \in \mathbb{S}^2 \times \mathbb{R}, \delta_{p,x} \in C_{\mathcal{D}_t^{-1},0}$. $\quad \forall (p, c) \in \mathbb{S}^2 \times \mathbb{S}^1, \delta_{p,c} \in C_{\mathcal{D}_s^{-1},0}$.
5. N and M are two integers.$\{(\overline{p}_i, \overline{x}_j)\}_{i,j=1,1}^{N,M} \subset \mathbb{S}^2 \times \mathbb{R}$, are fixed, unformally seen as the points where data has been queried. $\overline{\mathbf{Y}} \in \mathbb{R}^{N \times M}$, is fixed, unformally seen as the evaluation of a *phenomenon* on the previous set of points.

6. $\boldsymbol{\Phi}$ is the usual sampling operator defined with $(\overline{p}_i, \overline{x}_j)$. C is the usual square loss.
7. The function ψ_p^1 is continuous in p, the function ζ_x is continuous in x, the function ψ_p^2 is continuous in p and the function ϕ_c is continuous in c.

Moreover, suppose that we have at our disposition a sequence, indexed on n, of sets of nodes

$$\Xi^{1,n}(\mathbb{S}^2) = \left\{\tilde{p}_i^{1,n}\right\}_{i=1}^{N_{t,n}}, \quad \Xi^n(\mathbb{R}) = \left\{\tilde{x}_i^n\right\}_{i=1}^{M_{t,n}}, \\ \Xi^{2,n}(\mathbb{S}^2) = \left\{\tilde{p}_i^{2,n}\right\}_{i=1}^{N_{s,n}}, \quad \Xi^n(\mathbb{S}^n) = \left\{\tilde{c}_i^n\right\}_{i=1}^{M_{s,n}}, \tag{8.25}$$

with associated search spaces

$$\Pi^n(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^n(\mathcal{M}_{\mathcal{D}_s}), \tag{8.26}$$

such that for a fixed, but unknown, real interval $[a, b]$ we have

$$\lim_{n\to\infty} \Theta(\mathbb{S}^2, \Xi^{1,n}(\mathbb{S}^2)) = 0, \quad \lim_{n\to\infty} \Theta(\mathbb{S}^2, \Xi^{2,n}(\mathbb{S}^2)) = 0, \\ \lim_{n\to\infty} \Theta([a,b], \Xi^n(\mathbb{R})) = 0, \quad \lim_{n\to\infty} \Theta(\mathbb{S}^1, \Xi^n(\mathbb{S}^1)) = 0. \tag{8.27}$$

It results that, in this setting,

$$\lim_{n\to\infty} \min_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) = \min_{H\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) = \min_{H\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \mathcal{J}(H) \tag{8.28}$$

◀

Proof. Proof of Theorem 8.8:

<u>**Step I**: Construction of the candidate.</u>
The underlying idea of this proof is that a function

$$H^{\star,n} \in \operatorname*{argmin}_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) \tag{8.29}$$

does nearly as good as

$$H^\star \in \operatorname*{argmin}_{H\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \mathcal{J}(H). \tag{8.30}$$

The problem is that, in the current state of this work, we do not have much control on the solutions $H^{\star,n}$ and thus directly using them is odd. We will do the following.

Choose your favorite $H^\star$, thanks to Corollary 6.10 we know it will have the form

$$H^\star = \left(\sum_{i=1}^{I_{trend}} \gamma_{1,i}\psi^1_{p_i^{1,\star}} \otimes \zeta_{t_i^\star}, \sum_{i=1}^{I_{seas}} \gamma_{2,i}\psi^2_{p_i^{2,\star}} \otimes \phi_{c_i^\star})\right) \quad I_{seas} + I_{trend} \le 2NM. \tag{8.31}$$

Moreover we will make use of the mysterious "a fixed, but unknown, real interval $[a,b]$" with the supposition that $\{t_i^\star\}_{i=1}^{I_{trend}} \subset [a,b]$. This information might not be extracted from Corollary 6.10. We then define the projections

$$\forall i \le N_{trend} \quad \tilde{p}_i^{1,n,\star} = \underset{p\in\Xi^{1,n}(\mathbb{S}^2)}{argmin} \left\|p_i^{1,\star} - p\right\|_2. \quad \forall i \le N_{seas} \quad \tilde{p}_i^{2,n,\star} = \underset{p\in\Xi^{2,n}(\mathbb{S}^2)}{argmin} \left\|p_i^{2,\star} - p\right\|_2.$$

$$\forall i \le N_{trend} \quad \tilde{x}_i^{n,\star} = \underset{x\in\Xi^{n}(\mathbb{R})}{argmin}\left\|x_i^\star - x\right\|_2. \quad \forall i \le N_{seas} \quad \tilde{c}_i^{n,\star} = \underset{c\in\Xi^{n}(\mathbb{S}^1)}{argmin}\left\|c_i^\star - c\right\|_2.$$

The following bounds also apply

$$\forall i \le N_{trend} \quad \left\|p_i^{1,\star} - \tilde{p}_i^{1,n,\star}\right\|_2 \le \Theta(\mathbb{S}^2, \Xi^{1,n}(\mathbb{S}^2)). \qquad \forall i \le N_{seas} \quad \left\|p_i^{2,\star} - \tilde{p}_i^{2,n,\star}\right\|_2 \le \Theta(\mathbb{S}^2, \Xi^{2,n}(\mathbb{S}^2)).$$

$$\forall i \le N_{trend} \quad \left\|\tilde{x}_i^{\star} - \tilde{x}_i^{n,\star}\right\|_2 \le \Theta(a,b], \Xi^{n}(\mathbb{R})). \qquad \forall i \le N_{seas} \quad \left\|c_i^{\star} - \tilde{c}_i^{n,\star}\right\|_2 \le \Theta(\mathbb{S}^1, \Xi^{n}(\mathbb{S}^1)).$$

It is now possible to define a candidate H^n with good control

$$H^n = \left(\sum_{i=1}^{I_{trend}} \gamma_{1,i}\psi^1_{\tilde{p}_i^{1,n,\star}} \otimes \zeta_{\tilde{x}_i^{n,\star}}, \sum_{i=1}^{I_{seas}} \gamma_{2,i}\psi^2_{\tilde{p}_i^{2,n,\star}} \otimes \phi_{\tilde{c}_i^{n,\star}}\right) \in \Pi^n(\mathcal{M}_{\mathcal{D}_t}) \times \Pi^n(\mathcal{M}_{\mathcal{D}_s}). \tag{8.32}$$

Step II: $weak^\star$ convergence.

We prove here that $weak^\star - \lim_{n\to\infty} H^n = H^\star$. To do so let us take an arbitrary $(f,g) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ and verify that

$$\lim_{n\to\infty} H^n(f,g) - H^\star(f,g) = 0 \tag{8.33}$$

$$H^\star - H^n(f,g) =$$

$$\left(\sum_{i=1}^{I_{trend}} \gamma_{1,i}\mathcal{D}_t^{-1}\left(\delta_{(\tilde{p}_i^{1,n,\star},\tilde{x}_i^{n,\star})} - \delta_{(p_i^{1,\star},x_i^\star)}\right), \sum_{i=1}^{I_{seas}} \gamma_{2,i}\mathcal{D}_s^{-1}\left(\delta_{(\tilde{p}_i^{2,n,\star},\tilde{c}_i^{n,\star})} - \delta_{(p_i^{2,\star},c_i^\star)}\right)\right)(f,g) =$$

$$\sum_{i=1}^{I_{trend}} \gamma_{1,i}\left(\delta_{(\tilde{p}_i^{1,n,\star},\tilde{x}_i^{n,\star})} - \delta_{(p_i^{1,\star},x_i^\star)}\right)\mathcal{D}_t^{-1}f + \sum_{i=1}^{I_{seas}} \gamma_{2,i}\left(\delta_{(\tilde{p}_i^{2,n,\star},\tilde{c}_i^{n,\star})} - \delta_{(p_i^{2,\star},c_i^\star)}\right)\mathcal{D}_s^{-1}g =$$

$$\sum_{i=1}^{I_{trend}} \gamma_{1,i}\Big(\big[\mathcal{D}_{trend}^{-1} f\big](\tilde{p}_i^{1,n,\star}, \tilde{x}_i^{n,\star}) - \big[\mathcal{D}_t^{-1} f\big](p_i^{1,\star}, x_i^{\star})\Big) + \\ \sum_{i=1}^{I_{seas}} \gamma_{2,i}\Big(\big[\mathcal{D}_{seas}^{-1} f\big](\tilde{p}_i^{2,n,\star}, \tilde{c}_i^{n,\star}) - \big[\mathcal{D}_s^{-1} f\big](p_i^{2,\star}, c_i^{\star})\Big) \quad (8.34)$$

The $\lim_{n\to\infty}$ of Equation (8.34) is 0. One can see this as a direct consequence of

1. $I_{trend} + I_{seas} \leq NM < \infty$,
2. $\sum_{i=1}^{I_{trend}} |\gamma_{1,i}| + \sum_{i=1}^{I_{seas}} |\gamma_{2,i}| < \infty$,
3. $f \in C_{\mathcal{D}_t^{-1},0} \implies \mathcal{D}_t^{-1} f \in C_0(\mathbb{S}^2 \times \mathbb{R})$
4. $g \in C_{\mathcal{D}_s^{-1},0} \implies \mathcal{D}_s^{-1} g \in C_0(\mathbb{S}^2 \times \mathbb{S}^1)$
5. We supposed Equation (8.27) to hold true.

Therefore H^n is $weak^\star$ convergent to $H^\star$. We also notice that $\|H^n\|_{\mathcal{D}} = \|H^\star\|_{\mathcal{D}}$.

Step III: Conclusion.
From the $weak^\star$ convergence of H^n to $H^\star$ we get that

$$\lim_{n\to\infty} \Phi(H^n) = \Phi(H^\star). \quad (8.35)$$

Because of the classical continuity of $\left\|\overline{\mathrm{Y}} - \cdot\right\|_2^2$ we get that

$$\lim_{n\to\infty} \left\|\overline{\mathrm{Y}} - \Phi(H^n)\right\|_2^2 = \left\|\overline{\mathrm{Y}} - \Phi(H^\star)\right\|_2^2. \quad (8.36)$$

Finally, with using the fact $\|H^n\|_{\mathcal{D}} = \|H^\star\|_{\mathcal{D}}$ we get that

$$\lim_{n\to\infty} \left\|\overline{\mathrm{Y}} - \Phi(H^n)\right\|_2^2 + \lambda\|H^n\|_{\mathcal{D}} = \left\|\overline{\mathrm{Y}} - \Phi(H^\star)\right\|_2^2 + \lambda\|H^\star\|_{\mathcal{D}}, \quad (8.37)$$

and

$$\lim_{n\to\infty} \mathcal{J}(H^n) = \mathcal{J}(H^\star). \quad (8.38)$$

In consequence

$$\lim_{n\to\infty} \min_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) \leq \lim_{n\to\infty} \mathcal{J}(H^n) = \mathcal{J}(H^\star) = \min_{H\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H).$$

but the inequality

$$\lim_{n\to\infty} \min_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) \leq \min_{H\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) \tag{8.39}$$

has to be strict. Therefore we have

$$\lim_{n\to\infty} \min_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H) = \min_{H\in\Pi(\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H). \tag{8.40}$$

What concludes the proof. ■

The last theorem tells us that if our discretized search space is refined enough, we should be able to find a solution to the discretized problem

$$\min_{H\in\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H)$$

that performs nearly as well, with respect to the functional that is to be minimized, as a solution of the original problem

$$\min_{H\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \mathcal{J}(H).$$

Nevertheless there is a wound, here, that we are in the obligation to explore. Indeed, the last theorem does not tell us that if we take a sequence of solutions to the discretized problems, this sequence will converge to a solution to the original problem. That is because the H^n that we used in the proof need not to be a solution of their corresponding discretized problems. Is all convergence hope lost? No. The following theorem will establish, in a first version, a positive answer to that question.

▶ **Theorem 8.9.** Let us be in the setting of Theorem 8.8. Then, a sequence of solutions $\{H^{n,\star}\}_{n=1}^{\infty}$ for the discretized problems

$$H^{n,\star} \in \operatorname*{argmin}_{\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})} \mathcal{J}(H),$$

has a sub-sequence which is pointwise convergent to a solution in the sense that

$$\exists H^\star \in \operatorname*{argmin}_{\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \mathcal{J}(H) : \quad \forall (p,t,p',c) \in \mathbb{S}^2\times\mathbb{R}\times\mathbb{S}^2\times\mathbb{S}^1$$
$$\lim_{k\to\infty} H^{n_k,\star}(\delta_{(p,t)}\otimes\delta_{(p',c)}) = H^\star(\delta_{(p,t)}\otimes\delta_{(p',c)}). \tag{8.41}$$

More generally, $\{H^{n,\star}\}_{n=1}^{\infty}$ can be in only two disjoints regimes:

1. $H^{n,\star}$ is a pointwise convergent sequence, in the sense of **??**, to some $H^{\star}$.

2. $H^{n,\star}$ is composed of different sub-sequences which are all converging to a different $H^{\star}$.

◀

Proof of Theorem 8.9: It is worth noticing now that, because of Theorem 8.8, we have that $\lim\limits_{n\to\infty} \mathcal{J}(H^{n,\star}) = \min\limits_{H\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}} \mathcal{J}(H)$, which implies that $\hat{\mathcal{J}} := \max\limits_{l\in\mathbb{N}} \mathcal{J}(H^{l,\star})$ is a finite positive number. Independently, we have that

$$\forall n \in \mathbb{N} \quad \left\|H^{n,\star}\right\|_{\mathcal{D}} \leq \frac{\hat{\mathcal{J}}}{\lambda} \implies \{H^{n,\star}\}_{n=1}^{\infty} \subset \mathbf{B}_{\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}\left(\frac{\hat{\mathcal{J}}}{\lambda}, 0\right).$$

Lemma 8.6 tells us that

$$\mathbf{B}_{\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}\left(\frac{\hat{\mathcal{J}}}{\lambda}, 0\right)$$

is $weak^{\star}$ sequentially compact and we find that $\{H^{n,\star}\}_{n=1}^{\infty}$ admits a sub-sequence $\{H^{n_k,\star}\}_{k=1}^{\infty}$ that is $weak^{\star}$ convergent to some

$$H \in B_{\frac{\hat{\mathcal{J}}}{\lambda}}(0).$$

It is left to show that

$$H \in \underset{\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{argmin}\ \mathcal{J}(H). \tag{8.42}$$

First and with the help of [Bre11], proposition 3.13 page 63, we establish that

$$\|H\|_{\mathcal{D}} \leq \liminf_{k\to\infty} \left\|H^{n_k,\star}\right\|_{\mathcal{D}}. \tag{8.43}$$

Secondly, remark that because of the $weak^{\star}$ convergence of $H^{n_k,\star}$ to H, we have that

$$\liminf_{k\to\infty} \left\|\overline{\mathrm{Y}} - \Phi(H^{n_k,\star})\right\|_2^2 = \lim_{k\to\infty} \left\|\overline{\mathrm{Y}} - \Phi(H^{n_k,\star})\right\|_2^2 = \left\|\overline{\mathrm{Y}} - \Phi(H)\right\|_2^2.$$

Then, for

$$H^{\star} \in \underset{H\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{argmin}\ \mathcal{J}(H) \tag{8.44}$$

we have the chain

$$\mathcal{J}(H^\star) = \lim_{k\to\infty} \mathcal{J}(H^{n_k,\star}) = \liminf_{k\to\infty} \mathcal{J}(H^{n_k,\star}) = \|\mathbf{Y} - \Phi(H)\|_2^2 + \lambda \liminf_{k\to\infty} \left\|H^{n_k,\star}\right\|_{\mathcal{D}}.$$

But if

$$\|H\|_{\mathcal{D}} < \liminf_{k\to\infty} \left\|H^{n_k,\star}\right\|_{\mathcal{D}}$$

we then have $\mathcal{J}(H) < \mathcal{J}(H^\star)$, what is not possible as $H^\star$ is a minimizer. This implies that

$$\liminf_{k\to\infty} \|H^{n_k}\|_{\mathcal{D}} = \|H\|_{\mathcal{D}},$$

which implies that

$$\lim_{k\to\infty} \|H^{n_k}\|_{\mathcal{D}} = \|H\|_{\mathcal{D}},$$

and $\mathcal{J}(H) = \mathcal{J}(H^\star)$. Said differently, H, the *weak** and norm limit of H^{n_k} is a minimizer of $\mathcal{J}$ over $\mathcal{M}_{\mathcal{D}_t} \times \mathcal{M}_{\mathcal{D}_s}$. Finally one can prove that H^n follows two disjoints regimes in the following way. Suppose H^n does not follow the first one, then we can find a first sub-sequence, H^{n_k}, converging to a solution. Then by compacity the complementary sub-sequence of H^{n_k} in H^n, if non-empty, will have a converging (to a minimizer) sub-sequence. By proceeding like this, iteratively, we can exhaust the sequence H^n be finding that it is the superposition of (possibly countably many) sub-sequences that are converging to different minimizers of $\mathcal{J}$. ■

We can now, with the help of Theorem 8.9, go further in our questioning of the convergence. Can it be that the seasonal part of the solution of the discretized problem converges to the seasonal part of the solution of the original problem? The same question holds for the trending part. The answer to both of this question is, yes plus. The next theorem shows that and even more.

► **Theorem 8.10.** Let us be in the setting of Theorem 8.8. Then, a sequence of solutions $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ for the discretized problems

$$(F^{n,\star}, G^{n,\star}) \in \underset{\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})}{argmin} \mathcal{J}(H),$$

has a sub-sequence $\{(F^{n_k,\star}, G^{n_k,\star})\}_{k=1}^{\infty}$ that verifies the following

1. $\{(F^{n_k,\star}, G^{n_k,\star})\}_{k=1}^{\infty}$ is *weak** convergent to a solution $(F^\star, G^\star)$ of the original problem. Thus it is pointwise convergent to $(F^\star, G^\star)$.
2. $\lim_{k\to\infty} \left\|(F^{n_k,\star}, G^{n_k,\star})\right\|_{\mathcal{D}} = \left\|(F^\star, G^\star)\right\|_{\mathcal{D}}.$

3. $\{F^{n_k,\star}\}_{k=1}^{\infty}$ is $weak^{\star}$ convergent to $F^{\star}$. Thus it is pointwise convergent to $F^{\star}$.

4. $\{G^{n_k,\star}\}_{k=1}^{\infty}$ is $weak^{\star}$ convergent to $G^{\star}$. Thus it is pointwise convergent to $G^{\star}$.

5. $\lim_{k\to\infty} \|F^{n_k,\star}\|_{\mathcal{D}_t} = \|F^{\star}\|_{\mathcal{D}_t}$.

6. $\lim_{k\to\infty} \|G^{n_k,\star}\|_{\mathcal{D}_t} = \|G^{\star}\|_{\mathcal{D}_t}$.

More generally, $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ can be in only two disjoints regimes:

1. $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ verifies all the previous properties.

2. $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ is composed of different, disjoint, sub-sequences that all verify the previous properties, for a different limiting solution $(F^{\star}, G^{\star})$.

◀

Proof of Theorem 8.10: Consider a sequence of solutions $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$. By exactly the same arguments as in Theorem 8.9, we get that there exits a subsequence that verify

1. $\{(F^{n_k,\star}, G^{n_k,\star})\}_{n=1}^{\infty}$ is $weak^{\star}$ convergent to a solution $(F^{\star}, G^{\star})$ of the original problem. Thus it is pointwise convergent to $(F^{\star}, G^{\star})$.

2. $\lim_{k\to\infty} \|(F^{n_k,\star}, G^{n_k,\star})\|_{\mathcal{D}} = \|(F^{\star}, G^{\star})\|_{\mathcal{D}}$.

Then, by considering the element $(f, 0) \in C_{\mathcal{D}_t^{-1},0} \times C_{\mathcal{D}_s^{-1},0}$ with $f \in C_{\mathcal{D}_t^{-1},0}$ and the $weak^{\star}$ convergence of $\{(F^{n_k,\star}, G^{n_k,\star})\}_{k=1}^{\infty}$,we get that $\{F^{n_k,\star}\}_{k=1}^{\infty}$ is $weak^{\star}$ convergent to $F^{\star}$. Thus it is pointwise convergent to $F^{\star}$. The same is done to see that $\{G^{n_k,\star}\}_{k=1}^{\infty}$ is $weak^{\star}$ convergent to $G^{\star}$. Thus it is pointwise convergent to $G^{\star}$. We are now left to prove the convergence of the norms (5-6 in the statement). We can apply the same arguments as in Theorem 8.9 to get that

$$\|F^{\star}\|_{\mathcal{D}_t} \le \liminf_{k\to\infty} \|F^{n_k,\star}\|_{\mathcal{D}_t}, \tag{8.45}$$

$$\|G^{\star}\|_{\mathcal{D}_s} \le \liminf_{k\to\infty} \|G^{n_k,\star}\|_{\mathcal{D}_s}. \tag{8.46}$$

We know, again by Theorem 8.9, that

$$\|(F^{\star}, G^{\star})\|_{\mathcal{D}} = \liminf_{k\to\infty} \|(F^{n_k,\star}, G^{n_k,\star})\|_{\mathcal{D}}. \tag{8.47}$$

If we have in in Equation (8.45) or Equation (8.46) an inequality which is strict, it implies that we can not have the equality in Equation (8.47). In consequence we have

$$\left\|F^{\star}\right\|_{\mathcal{D}_t} = \liminf_{k\to\infty}\left\|F^{n_k,\star}\right\|_{\mathcal{D}_t}, \tag{8.48}$$

$$\left\|G^{\star}\right\|_{\mathcal{D}_s} = \liminf_{k\to\infty}\left\|G^{n_k,\star}\right\|_{\mathcal{D}_s}. \tag{8.49}$$

It is known that liminf≤limsup, thus

$$\left\|F^{\star}\right\|_{\mathcal{D}_t} \leq \limsup_{k\to\infty}\left\|F^{n_k,\star}\right\|_{\mathcal{D}_t}, \tag{8.50}$$

$$\left\|G^{\star}\right\|_{\mathcal{D}_s} \leq \limsup_{k\to\infty}\left\|G^{n_k,\star}\right\|_{\mathcal{D}_s}. \tag{8.51}$$

If we manage to show that, in Equation (8.50) and Equation (8.51) the inequality is actually an equality, we will be able to conclude with

$$\left\|F^{\star}\right\|_{\mathcal{D}_t} = \lim_{k\to\infty}\left\|F^{n_k,\star}\right\|_{\mathcal{D}_t}, \tag{8.52}$$

$$\left\|G^{\star}\right\|_{\mathcal{D}_s} = \lim_{k\to\infty}\left\|G^{n_k,\star}\right\|_{\mathcal{D}_s}. \tag{8.53}$$

Suppose by contradiction that in Equation (8.50) or Equation (8.51), the equality is strict and remark that because

$$\left\|(F^{\star}, G^{\star})\right\|_{\mathcal{D}} = \lim_{k\to\infty}\left\|(F^{n_k,\star}, G^{n_k,\star})\right\|_{\mathcal{D}}, \tag{8.54}$$

we have

$$\left\|(F^{\star}, G^{\star})\right\|_{\mathcal{D}} = \limsup_{k\to\infty}\left\|(F^{n_k,\star}, G^{n_k,\star})\right\|_{\mathcal{D}}. \tag{8.55}$$

It is that Equation (8.55) can not hold because of the previous supposition. Finally the fact that $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ can only be in two different regimes is proved in the same manner as in Theorem 8.9. ■

We are now in a very good position. We have that the different parts of a subsequence of a sequence of solutions converge pointwise to the different parts of an asymptotic solution. We have the the norms of the different parts of the subsequence converge to the norms of the different parts of an asymptotic solution, in a mass preservation like property. We can not hope, in the general case, to go from the subsequence to the whole sequence convergence, as the solutions are not unique in the discretized space and in the original space. We can still wonder if this pointwise convergence can be strengthen. We will study this question in the next section.

8.4 Convergence strengthening in the asymptotic analysis of the specification

The general direction of this section has been established on the ideas/advices of Julien Fageot. The goal will be to strengthen the pointwise convergence into a uniform convergence using the Ascoli-Arzela theorem.

▶ **Definition 8.11.** For $C(X)$, the space of real-valued continuous functions on a metric space X. A sequence $\{f_n\}_{n=1}^{\infty} \subset C(X)$ is said to be equicontinuous if:

$$\forall x \in X \forall \epsilon > 0 \exists \delta > 0 : \quad (x' \in X : d(x,x') < \delta, n \in \mathbb{N}) \implies |f_n(x') - f_n(x)| < \epsilon.$$

◀

▶ **Theorem 8.12 (Arzelà-Ascoli).** Let X be a compact Hausdorff space. Then, a subspace F of $C(X)$ is compact, with respect to the topology of the uniform convergence, if and only if

1. F is closed.
2. F is equicontinuous.
3. F is pointwise bounded in the sense that, $\forall x \in X \quad \underset{n\in\mathbb{N}}{sup}|f_n(x)| < \infty$.

◀

The reader might already see what is going to happen. We will want to apply the Ascoli-Arzelà theorem to both of the sequences $\{F^{n,\star}\}_{n=1}^{\infty}$ and $\{G^{n,\star}\}_{n=1}^{\infty}$. The main obstacle that is on our way is the equicontinuity and additional assumptions on our splines are to be made in order to get it.

▶ **Remark 8.13.** By inspection of the representation of a dirac mass in the spherical harmonics basis for $\mathbb{S}^2$ and $\mathbb{S}^1$ and its representation with the Fourier transform for $\mathbb{R}$. It is possible to see that

$$\psi^1_{p'}(p) = \psi^1_p(p'), \quad \forall \{p,p'\} \subset \mathbb{S}^2. \tag{8.56}$$

$$\psi^2_{p'}(p) = \psi^2_p(p'), \quad \forall \{p,p'\} \subset \mathbb{S}^2. \tag{8.57}$$

$$\zeta_{x'}(x) = \zeta_x(x'), \quad \forall \{x, x'\} \subset \mathbb{R}. \tag{8.58}$$

$$\phi_{p'}(p) = \phi_p(p'), \quad \forall \{c, c'\} \subset \mathbb{S}^1. \tag{8.59}$$

Moreover and by the previous points, the classical assumption that

$$\forall (p, x) \in \mathbb{S}^2 \times \mathbb{R}, \delta_{p,x} \in C_{\mathcal{D}_t^{-1},0}, \quad \forall (p, c) \in \mathbb{S}^2 \times \mathbb{S}^1, \delta_{p,c} \in C_{\mathcal{D}_s^{-1},0},$$

implies the following

$$\psi^1_{p'}(p) \text{ is continuous with respect to } p \text{ and } p', \tag{8.60}$$

$$\psi^2_{p'}(p) \text{ is continuous with respect to } p \text{ and } p', \tag{8.61}$$

$$\zeta_{x'}(x) \text{ is continuous with respect to } x \text{ and } x', \tag{8.62}$$

$$\phi_{c'}(c) \text{ is continuous with respect to } c \text{ and } c'. \tag{8.63}$$

◀

Nevertheless, continuity is not enough to get equicontinuity and we need the following definition.

▶ **Definition 8.14.** We will say that $\psi^1_\cdot(\cdot)$ is L-uniformly Lipschitz if

$$\forall p \in \mathbb{S}^2 \forall \{p', p''\} \subset \mathbb{S}^2 \quad |\psi^1_p(p') - \psi^1_p(p'')| \leq L\|p' - p''\|_2. \tag{8.64}$$

In the same way, $\psi^2_\cdot(\cdot)$ is L-uniformly Lipschitz if

$$\forall p \in \mathbb{S}^2 \forall \{p', p''\} \subset \mathbb{S}^2 \quad |\psi^2_p(p') - \psi^2_p(p'')| \leq L\|p' - p''\|_2. \tag{8.65}$$

The function $\phi_\cdot(\cdot)$ is L-uniformly Lipschitz if

$$\forall c \in \mathbb{S}^1 \forall \{c', c''\} \subset \mathbb{S}^1 \quad |\phi^1_c(c') - \phi^1_c(c'')| \leq L\|c' - c''\|_2. \tag{8.66}$$

The function $\zeta_\cdot(\cdot)$ is L-uniformly Lipschitz on $[a, b]$ if

$$\forall x \in \mathbb{R} \forall \{x', x''\} \subset [a, b] \quad |\phi^1_x(x') - \phi^1_x(x'')| \leq L\|x' - x''\|_2. \tag{8.67}$$

◀

▶ **Theorem 8.15.** Let us be in the setting of Theorem 8.8. Moreover suppose that

1. The function $\psi^1_\cdot(\cdot)$ is $L(\psi^1)$-uniformly Lipschitz.
2. The function $\psi^2_\cdot(\cdot)$ is $L(\psi^2)$-uniformly Lipschitz.

3. The function $\zeta_{\cdot}(\cdot)$ is $L(\zeta)$-uniformly Lipschitz.
4. The function $\phi_{\cdot}(\cdot)$ is $L(\phi)$-uniformly Lipschitz.

Then, a sequence of solutions $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ for the discretized problems

$$(F^{n,\star}, G^{n,\star}) \in \underset{\Pi^n(\mathcal{M}_{\mathcal{D}_t})\times\Pi^n(\mathcal{M}_{\mathcal{D}_s})}{argmin} \mathcal{J}(H),$$

has a subsequence $\{(F^{n_{k_l},\star}, G^{n_{k_l},\star})\}_{l=1}^{\infty}$ that verifies,

$$\exists (F^{\star}, G^{\star}) \in \underset{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{argmin} \mathcal{J}(F,G)$$

such that

1. $\forall \{a,b\} \subset \mathbb{R}: \quad \lim_{l\to\infty} \left\| F^{n_{k_l}} - F^{\star} \right\|_{\mathbb{S}^2\times[a,b],\infty} = 0.$
2. $\lim_{l\to\infty} \left\| G^{n_{k_l}} - G^{\star} \right\|_{\mathbb{S}^2\times\mathbb{S}^1,\infty} = 0.$

More generally, $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ can be in only two disjoints regimes:

1. $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ verifies all the previous properties.
2. $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$ is composed of different, disjoint, sub-sequences that all verify the previous properties, for a different limiting solution $(F^{\star}, G^{\star})$.

◀

Proof of Theorem 8.15: We start with a sequence $\{(F^{n,\star}, G^{n,\star})\}_{n=1}^{\infty}$, then we use Theorem 8.10 to get a subsequence $\{(F^{n_k,\star}, G^{n_k,\star})\}_{k=1}^{\infty}$ that verifies all the nice convergences given in Theorem 8.10. We then want to apply the Ascoli-Arzelà theorem to the subsequence $\{F^{n_k,\star}\}_{k=1}^{\infty}$ and to the subsequence $\{G^{n_k,\star}\}_{k=1}^{\infty}$. We consider here the $G^{n_k,\star}$ to be continuous functions with domain $\mathbb{S}^2\times\mathbb{S}^1$, which is compact and Hausdorff. We consider the $F^{n_k,\star}$ to be continous functions with domain $\mathbb{S}^2\times[a,b]$, which is compact and Hausdorff. We are now in the setting of the Ascoli-Arzelà theorem. It is first easy to see that both families $\{F^{n_k,\star}\}_{k=1}^{\infty}$ and $\{G^{n_k,\star}\}_{k=1}^{\infty}$ are closed. We secondly tackle the pointwise boundedness. Theorem 8.10 tells us that

$$\forall (p,x) \in \mathbb{S}^2\times[a,b] \quad \lim_{k\to\infty} F^{n_k,\star}(\delta_{(p,x)}) = F^{\star}(\delta_{(p,x)}), \tag{8.68}$$

$$\forall (p,c) \in \mathbb{S}^2\times\mathbb{S}^2 \quad \lim_{k\to\infty} G^{n_k,\star}(\delta_{(p,c)}) = G^{\star}(\delta_{(p,c)}). \tag{8.69}$$

Using the remark we made at the start of this section, it is possible to see that the two previous equations are equivalent to the two following

$$\forall (p,x) \in \mathbb{S}^2 \times [a,b] \quad \lim_{k\to\infty} F^{n_k,\star}(p,x) = F^{\star}(p,x), \tag{8.70}$$

$$\forall (p,c) \in \mathbb{S}^2 \times \mathbb{S}^2 \quad \lim_{k\to\infty} G^{n_k,\star}(p,c) = G^{\star}(p,c). \tag{8.71}$$

What implies

$$\forall (p,x) \in \mathbb{S}^2 \times [a,b] \quad \sup_{k\in\mathbb{N}} \left|F^{n_k,\star}(p,x)\right| < \infty, \tag{8.72}$$

$$\forall (p,c) \in \mathbb{S}^2 \times \mathbb{S}^2 \quad \sup_{k\in\mathbb{N}} \left|G^{n_k,\star}(p,c)\right| < \infty. \tag{8.73}$$

Therefore both families $\{F^{n_k,\star}\}_{k=1}^{\infty}$ and $\{G^{n_k,\star}\}_{k=1}^{\infty}$ are pointwise bounded. We are left to show equicontinuity. To do so we will make explicit the shape of our families with

$$F^{n_k,\star} = \sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \gamma_{i,j}^{1,n_k,\star} \psi^1_{\tilde{p}_i^{1,n_k}} \otimes \zeta_{\tilde{x}_j}^{n_k}. \tag{8.74}$$

$$G^{n_k,\star} = \sum_{i,j=1,1}^{N_{s,n_k},M_{s,n_k}} \gamma_{i,j}^{2,n_k,\star} \psi^2_{\tilde{p}_i^{2,n_k}} \otimes \phi_{\tilde{c}_j}^{n_k}. \tag{8.75}$$

From Equation (8.74), the uniform-Lipschitzness of ψ^1 and ζ, we get that $\forall\{(p,x),(p',x')\} \subset \mathbb{S}^2 \times [a,b]$

$$\left|F^{n_k,\star}(p,x) - F^{n_k,\star}(p',x')\right| = \left|\sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \gamma_{i,j}^{1,n_k,\star}\left[\psi^1_{\tilde{p}_i^{1,n_k}}(p)\zeta_{\tilde{x}_j}^{n_k}(x) - \psi^1_{\tilde{p}_i^{1,n_k}}(p')\zeta_{\tilde{x}_j}^{n_k}(x')\right]\right| \leq \tag{8.76}$$

$$\sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{1,n_k,\star}\right|\left|\psi^1_{\tilde{p}_i^{1,n_k}}(p)\zeta_{\tilde{x}_j}^{n_k}(x) - \psi^1_{\tilde{p}_i^{1,n_k}}(p')\zeta_{\tilde{x}_j}^{n_k}(x')\right| \leq \tag{8.77}$$

$$\sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{1,n_k,\star}\right|\left(\left|\zeta_{\tilde{x}_j}^{n_k}(x)\right|L(\psi^1)\|p-p'\|_2 + \left|\psi^1_{\tilde{p}_i^{1,n_k}}(p')\right|L(\zeta)\|x-x'\|_2\right) \leq \tag{8.78}$$

$$\left(\|\zeta_{\cdot}(\cdot)\|_{\mathbb{R}\times[a,b],\infty}L(\psi^1)\|p-p'\|_2 + \left\|\psi^1_{\cdot}(\cdot)\right\|_{\mathbb{S}^2\times\mathbb{S}^2,\infty}L(\zeta)\|x-x'\|_2\right)\sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}}\left|\gamma_{i,j}^{1,n_k,\star}\right|. \tag{8.79}$$

Moreover we know by Theorem 8.10 that

$$\lim_{k\to\infty}\left\|F^{n_k,\star}\right\|_{\mathcal{D}_t} = \left\|F^{\star}\right\|_{\mathcal{D}_t} \tag{8.80}$$

and remember that

$$\left\|F^{n_k,\star}\right\|_{\mathcal{D}_t} = \sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{1,n_k,\star}\right|. \tag{8.81}$$

In consequence the sequence $\sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{1,n_k,\star}\right|$ is convergent and we have

$$\Gamma^1 := \sup_{k\in\mathbb{N}} \sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{1,n_k,\star}\right| < \infty. \tag{8.82}$$

Finally, from the previous equation we rewrite Equation (8.79) in

$$\begin{aligned}&\left|F^{n_k,\star}(p,x) - F^{n_k,\star}(p',x')\right| \leq \\ &\quad \Gamma^1\|\zeta_\cdot(\cdot)\|_{\mathbb{R}\times[a,b],\infty}L(\psi^1)\|p-p'\|_2 + \Gamma^1\left\|\psi_\cdot^1(\cdot)\right\|_{\mathbb{S}^2\times\mathbb{S}^2,\infty}L(\zeta)\|x-x'\|_2.\end{aligned} \tag{8.83}$$

Using the very same technique and by defining

$$\Gamma^2 := \sup_{k\in\mathbb{N}} \sum_{i,j=1,1}^{N_{t,n_k},M_{t,n_k}} \left|\gamma_{i,j}^{2,n_k,\star}\right| < \infty, \tag{8.84}$$

we get the similar bound

$$\begin{aligned}&\left|G^{n_k,\star}(p,c) - G^{n_k,\star}(p',c')\right| \leq \\ &\quad \Gamma^2\|\phi_\cdot(\cdot)\|_{\mathbb{S}^1\times\mathbb{S}^1,\infty}L(\psi^2)\|p-p'\|_2 + \Gamma^2\left\|\psi_\cdot^2(\cdot)\right\|_{\mathbb{S}^2\times\mathbb{S}^2,\infty}L(\phi)\|c-c'\|_2.\end{aligned} \tag{8.85}$$

We are now very close to our goal as the upper-bounds in Equation (8.83) and Equation (8.85) do not depend on n_k. We just need to verify that the constants in the upper-bounds are finite.

- $\|\phi_\cdot(\cdot)\|_{\mathbb{S}^1\times\mathbb{S}^1,\infty}$ is finite because $\phi_\cdot(\cdot)$ is continuous and $\mathbb{S}^1\times\mathbb{S}^1$ is a compact domain.
- $\left\|\psi_\cdot^1(\cdot)\right\|_{\mathbb{S}^2\times\mathbb{S}^2,\infty}$ is finite because $\psi_\cdot^1(\cdot)$ is continuous and $\mathbb{S}^2\times\mathbb{S}^2$ is a compact domain.

- $\left\|\psi^2_\cdot(\cdot)\right\|_{\mathbb{S}^2\times\mathbb{S}^2,\infty}$ is finite because $\psi^2_\cdot(\cdot)$ is continuous and $\mathbb{S}^2\times\mathbb{S}^2$ is a compact domain.
- $\|\zeta_\cdot(\cdot)\|_{\mathbb{R}\times[a,b],\infty}$ is finite even though $\mathbb{R}\times[a,b]$ is not compact. To see this we need two ingredients. First it is possible to see, in the same idea as in the remark at the start of this section, that $\zeta_x(x')$ is a radial function, i.e. its value only depends on $|x-x'|$ (this is due to the early supposition that D_t is self-adjoint). The second ingredient is that $\forall x\ \zeta_x(\cdot)\in C_0(\mathbb{R})$. From this, it is possible to conclude that $\zeta_\cdot(\cdot)\in C_0(\mathbb{R}^2)$ and therefore

$$\|\zeta_\cdot(\cdot)\|_{\mathbb{R}\times[a,b],\infty}\leq\|\zeta_\cdot(\cdot)\|_{\mathbb{R}\times\mathbb{R},\infty}<\infty.$$

Now, Equation (8.83) and Equation (8.85) allow us to conclude that both $\{F^{n_k,\star}\}_{k=1}^{\infty}$ and $\{G^{n_k,\star}\}_{k=1}^{\infty}$ are equicontinuous and by the Ascoli-Arzelà theorem they are also both compact in the uniform convergence topology. Moreover $\{F^{n_k,\star}\}_{k=1}^{\infty}$ lives in $C_0(\mathbb{S}^2\times\mathbb{R})$ and $\{G^{n_k,\star}\}_{k=1}^{\infty}$ lives in $C_0(\mathbb{S}^2\times\mathbb{S}^1)$, where both of these spaces are metric spaces with respect to the ∞ norm. Then sequential compacity is equivalent to compacity and both of the sequences $\{F^{n_k,\star}\}_{k=1}^{\infty}$ and $\{G^{n_k,\star}\}_{k=1}^{\infty}$ are sequentially compact. In consequence it is possible to find subsequences $\{F^{n_{k_l},\star}\}_{l=1}^{\infty}$ and $\{G^{n_{k_q},\star}\}_{q=1}^{\infty}$ that are convergent in the ∞ norm and by a classical subsequence imbedding argument we can suppose that they are both the same, namely $\{F^{n_{k_l},\star}\}_{l=1}^{\infty}$ and $\{G^{n_{k_l},\star}\}_{l=1}^{\infty}$ are both convergent in the ∞ norm. Finally there limit have to coincide with $F^\star$ and $G^\star$ by unicity of the limit and the original pointwise convergence. We thus get that

$$\lim_{l\to\infty}\left\|F^{n_{k_l}}-F^\star\right\|_{\mathbb{S}^2\times[a,b],\infty}=0,\tag{8.86}$$

$$\lim_{l\to\infty}\left\|G^{n_{k_l}}-G^\star\right\|_{\mathbb{S}^2\times\mathbb{S}^1,\infty}=0.\tag{8.87}$$

Finally it possible to find, using a classical diagonal sub-sequences argument (like the one in the proof of the Ascoli-Arzelà theorem), two subsequences $\{F^{n_{k_l},\star}\}_{l=1}^{\infty}$ and $\{G^{n_{k_l},\star}\}_{l=1}^{\infty}$ that verify,

1. $\forall\{a,b\}\subset\mathbb{R}:\quad \lim_{l\to\infty}\left\|F^{n_{k_l}}-F^\star\right\|_{\mathbb{S}^2\times[a,b],\infty}=0.$
2. $\lim_{l\to\infty}\left\|G^{n_{k_l}}-G^\star\right\|_{\mathbb{S}^2\times\mathbb{S}^1,\infty}=0.$

The argumentation for the existence of the two disjoint regimes is the same as the one in Theorem 8.9. This concludes the proof. ■

It will now be interesting to see through the next example which kind of splines ψ^1,ψ^2,ζ,ϕ verify the condition of being L-uniform Lipschitz.

► **Example 8.16.** The start is done by analyzing the situation for the real spline ζ. We thus have $\forall\{x, x', x''\} \in \mathbb{R}$:

$$|\zeta_x(x') - \zeta_x(x'')| = \left|\int_{-\infty}^{+\infty} \frac{e^{-2\pi i w x}}{p(w)} \left(e^{2\pi i w x'} - e^{2\pi i w x''}\right) \mathrm{d}w\right| \leq \int_{-\infty}^{+\infty} \frac{1}{|p(w)|} \left|e^{2\pi i w x'} - e^{2\pi i w x''}\right| \mathrm{d}w$$

$$\leq |x' - x''| \int_{-\infty}^{+\infty} \frac{2\pi |w|}{|p(w)|} dw.$$

Where p is the oscillatory / frequency response of the differential operator associated to the spline ζ. By defining $L := \int_{-\infty}^{+\infty} \frac{2\pi|w|}{|p(w)|} dw$ we get that $L < \infty$ if and only if p grows strictly faster than a polynomial of degree 2 (spectral growth strictly superior to 2) and have no zeros (except in 0 where 0 can be reached with a polynomial speed of at most 1), which is as the previous inequalities are not tight only a sufficient condition for the spline ζ to be L-uniform Lipschitz. Among the splines / differential operators verifying this condition we have $\left(\mathrm{Id} + \frac{\partial^2}{\partial^2 x}\right)^n$ with $n \in \mathbb{R} \cap]1; +\infty[$ and the Matern's operators $\left(\mathrm{Id} + \frac{\epsilon^2}{\beta - 1} \frac{\partial^2}{\partial^2 x}\right)^\beta$ with $\beta \in \mathbb{R} \cap]1; +\infty[$. The cases of ψ^1, ψ^2, ψ are done in a very similar fashion as well as the outcome will be that the sequences weighting the frequencies and thus defining the corresponding differential operators will have to grow strictly faster than 2 for ϕ and strictly faster than 3 for ψ^1, ψ^2 also in general they can not have a term being equal to 0 but this is already assumed not to happen by their invertibility. ◄

Part II

Application of spatio-temporal field dismantling

9 Synthetic applications

This first embedded in the applications chapter is framing a first synthetic empirical analysis of the previously created theory. In this sense, a python code has been created that implement, solves approximately, analyse, numerically the now well-known problem

$$\underset{(F,G)\in\mathcal{M}_{\mathcal{D}_t}\times\mathcal{M}_{\mathcal{D}_s}}{argmin} \left\|\overline{\mathrm{Y}} - \Phi(F,G)\right\|_2^2 + \alpha\|F\|_1 + \beta\|G\|_1.$$

The code can be found at
https://github.com/7804567/SFD
and all the examples we will encounter have been made reproducible with the existence of a python file, of the same name, in the same repository. More precisely, the goal of this chapter is multiple. First we want to show that it is possible to implement numerically an approximating solver for this problem, without entering in any programmatic details but rather by showing the results, fits of the code. Secondly we want to show two things that we argue to be duals, solving the previous problem is a good fit of the reality for and by the numerical approximation of the problem is, in itself, all theoretical considerations aside, a good fit of the reality. We shall remember the following ordered chain.

Solving the original problem. $\underset{approximation}{\longrightarrow}$ *Solving the problem on a discretized search space.* $\underset{approximation}{\longrightarrow}$ *Solving the problem on a discretized search space, numerically.*

In consequence if we suppose the previous approximations to be perfect we indeed have the stated duality. The goal being to show that the use of the abstract optimization problem is justified by the empirical efficiency of its approximation and that the latter promote the usage of our algorithm as a tool to approximate real-life phenomenons. Thirdly we want to assess to robustness of our method by noising and sparsing synthetic data.

The attention put on the choice of hyper-parameters and on the discretization scheme is little, as it is not here our point to explain how to choose them as well as everything can be recovered in the Github repository.

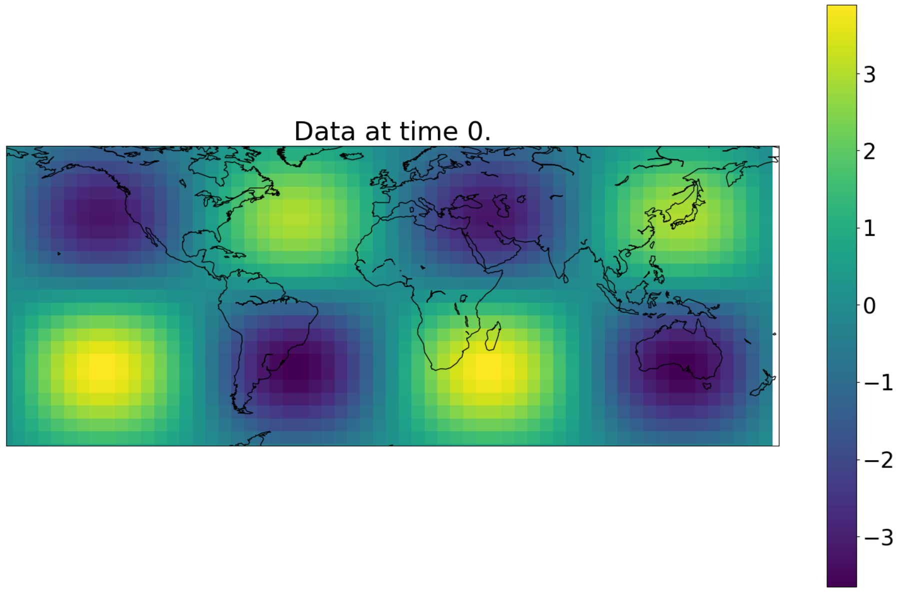


Figure 9.1: Data at time 0.

9.1 A first test

We start by placing ourself in an ideal case where we know f, which is the sum of a periodic f_{seas} part and a trending f_{trend} part defined by

$$f_{seas}(x, y, z, t) = 18 \cdot x \cdot y \cdot z \cdot \cos\left(\frac{2\pi t}{12}\right) \quad \forall (x, y, z) \in \mathbb{S}^2, t \in \mathbb{R}, \tag{9.1}$$

$$f_{trend}(x, y, z, t) = (x \cdot y + 0.2z) \cdot (1 + 0.07t) \quad \forall (x, y, z) \in \mathbb{S}^2, t \in \mathbb{R}. \tag{9.2}$$

In space we sampled this function on a transformed latitude-longitude lattice with a step size of 6 degrees. It is to be remarked that f_{seas} is 12 periodic, therefore, in time we sampled this function on an integer lattice ranging from 0 to 12*10 - 1, that is to be thought as 10 periods. We provide two spatial graphical representations of our sampled f through Figure 9.1 and Figure 9.2.

The goal here is to show that if we have no noise and a structured samples set, we can fit the samples with a continuous model and that continuous model is a

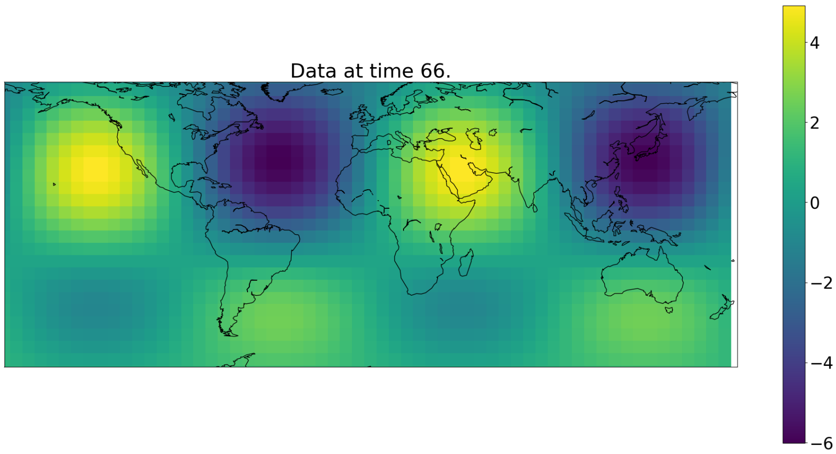


Figure 9.2: Data at time 66. 6.5 periods after the start at 0.

good approximation of both the underlying f_{seas} an f_{trend}. *I now have launched the algorithm and gotten the fit*, the present is ready to be open. As visualizing spatio-temporal data is not as easy as visualizing real data, we first propose the reader to look at a temporal section in position 12 (t=12), of the fields. We will also call $\tilde{f}_{seas}$ the fit to f_{seas} produced by the algorithm, $\tilde{f}_{trend}$ the fit to f_{trend}, $\tilde{f} = \tilde{f}_{seas} + \tilde{f}_{trend}$ the fit to f. Figure 9.3 suggests us that in the time section t=12, $\tilde{f}$ is a good

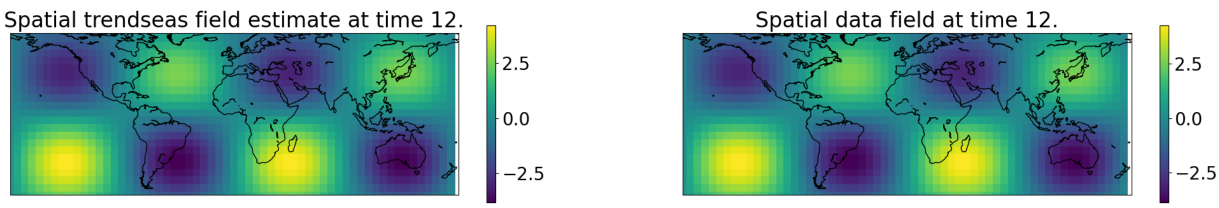


Figure 9.3: On the left is the fit (sum of the seasonal and trending fits) to the data at time 12. On the right is the same data at time 12.

approximation to f. In order to confirm this suggestion we can look at the residuals in Figure 9.4. But being in a spatio-temporal framework forces us to also check if

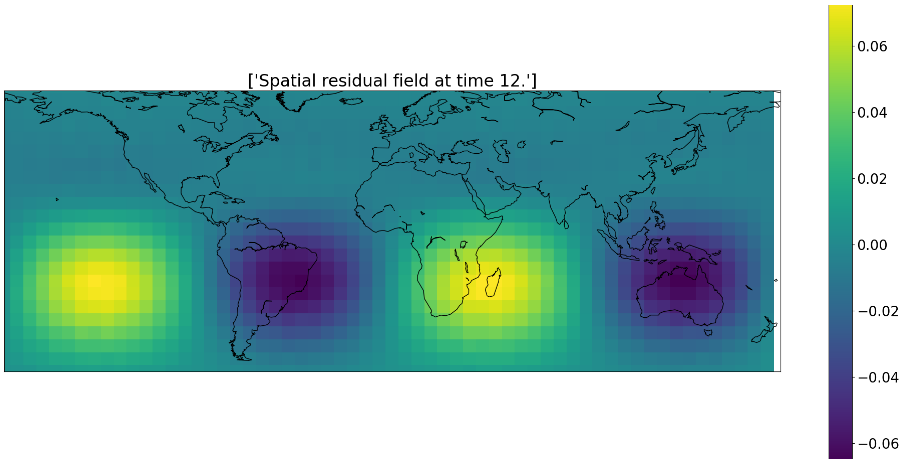


Figure 9.4: Presented are the residuals (data-fit) at the time section 12.

the residuals behaves in a good manner against time. Therefore it is, first, worth watching at a spatial section, i.e. fixed position on the sphere evolving against time, in order to convince ourself that the goodness of fit does not change as time runs and both Figure 9.5 and Figure 9.6 are here for this reason. Figure 9.5 will indubitably prove, at least on the point (longitude, latitude)=(-126, 36), that the residuals do evolve in a very good manner against time. Over and above we notice that there is nearly no issue at the queues. This behavior should nevertheless not be generalized as we saw that the fit can underestimate the data at their locations. While the queue behavior is not always as good as the one in Figure 9.5, we never observed symptomatic errors or bad oscillatory behaviors. Coming back to our main interest, we are still left wondering if the residuals do behave everywhere as good in time as in Figure 9.5. Before answering this question we shall pay attention to the Figure 9.6, that leverages another question; Are $\tilde{f}_{seas}$ and $\tilde{f}_{trend}$ good approximations of f_{seas} and f_{trend}. The previous question is ill-posed as $\tilde{f}_{seas}$ and $\tilde{f}_{trend}$ are a couple that is not uniquely defined in the sense that for a constant c, it is possible to get another solution $\tilde{f}_{seas} + c$ and $\tilde{f}_{trend} - c$, as a constant is periodic. The consequence of this inconsistency is the following, suppose that $\tilde{f}_{seas} = f_{seas} + c$ and $\tilde{f}_{trend} = f_{trend} - c$ is the solution found. Then this solution is perfect in the sense that it perfectly reconstructs f and we have no information to detect that the c does not fit the reality here, as we do not have any information on f_{seas} and f_{trend}. The take-away is multiple, first the solution $(\tilde{f}_{seas}, \tilde{f}_{trend})$ is always to be thought up to a constant, secondly if we want to estimate the seasonal residuals $(f_{seas} - \tilde{f}_{seas})$ and the trending residuals $(f_{trend} - \tilde{f}_{trend})$ we have to compensate for this c using, to be found, information on f_{trend}. In this example we are able to account for the c as we do know (f_{seas}, f_{trend}) and by doing this we are able to approximate both $\left\|f_{seas} - \tilde{f}_{seas}\right\|_\infty$ and $\left\|f_{trend} - \tilde{f}_{trend}\right\|_\infty$ using a Fibonacci lattice with 5000 points and a uniform discretization of $[0, 119]$ with 494 points, with the product lattice being called L.

$$\sup_{(p,t)\in L} \left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| = 0.468, \quad \frac{1}{|L|}\sum_{(p,t)\in L} \left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| = 0.083,$$

$$\frac{1}{|L|}\sum_{(p,t)\in L}\left(\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| - \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right|\right)^2 = 0.004.$$

$$\sup_{(p,t)\in L} \left|f_{seas}(p,t) - \tilde{f}_{seas}(p,t)\right| = 0.266, \quad \frac{1}{|L|}\sum_{(p,t)\in L} \left|f_{seas}(p,t) - \tilde{f}_{seas}(p,t)\right| = 0.084,$$

$$\frac{1}{|L|}\sum_{(p,t)\in L}\left(\left|f_{seas}(p,t) - \tilde{f}_{seas}(p,t)\right| - \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{seas}(p,t) - \tilde{f}_{seas}(p,t)\right|\right)^2 = 0.004.$$

The previous quantities are to be seen as the maximal absolute error, the mean absolute error and the variance of the absolute error. We can conclude that aside from some extreme values, after accounting for the c, $\tilde{f}_{seas}$ and $\tilde{f}_{trend}$ are both very good approximations of f_{seas} and f_{trend}, for this example.

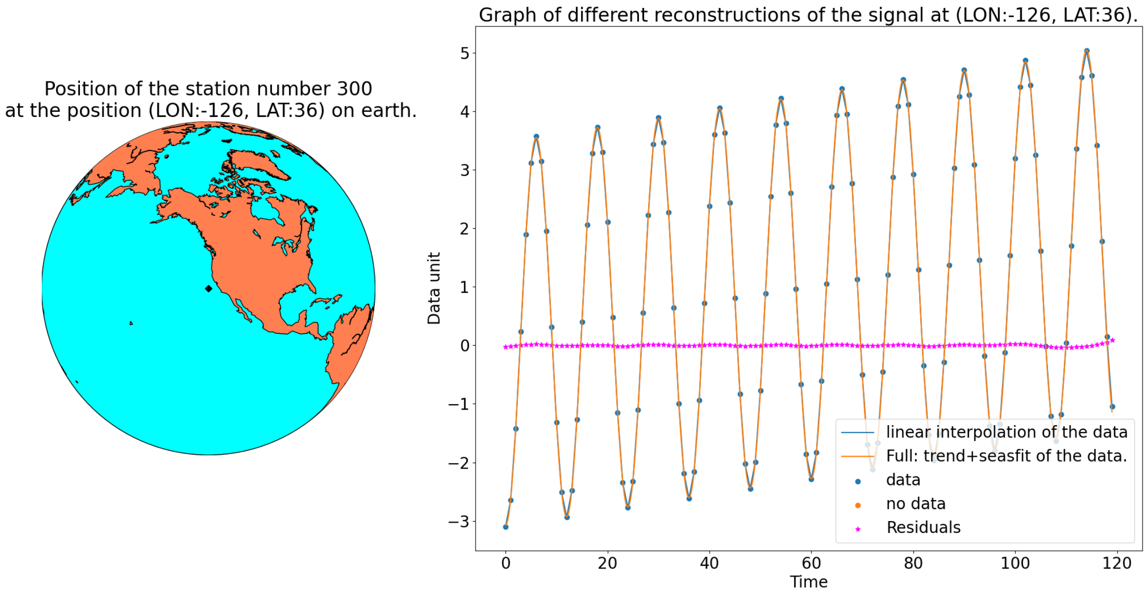


Figure 9.5: Data, fit, residuals evolving against time at a fixed spatial position.

We can now conclude this example by summing up where we are, what we have learned, where we are going. We have completed our first goal, implementing numerically an approximating solver for the original optimization problem. We have and will, found good arguments for our second goal, finding a good fit of the reality and arguing that the original problem is a good generator for such fits. In particular we saw that it is possible to estimate, up to a constant, both the periodic and trending part, if they exist, of a field provided with only samples of the same field. We still need to go towards our third goal, testing the robustness of the fitting procedure and testing the fitting procedure on non-synthetic data (more ill-structure and more chaotic behavior of the field).

9.2 A second test: sparsity robustness

In this second test we want to test the robustness of our algorithm against acquisition errors in the data collecting process. To do so we place ourself in the exact same setting as in section 9.1 and we suppose that 10% of the data we queried has been compromised, we can not use it. We did that by generating a random mask matrix with a fixed numbers of 0' s and the code in the Github repository is seeded

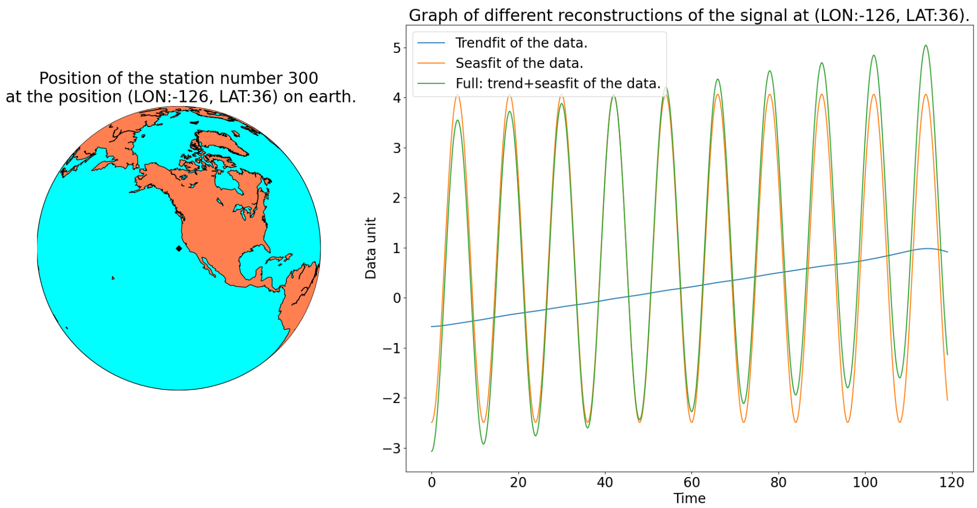


Figure 9.6: The full, seasonal, trending fit of the data are evolving against time at a fixed spatial position.

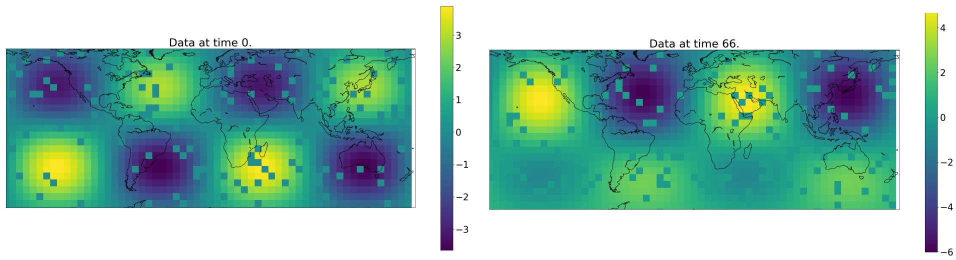


Figure 9.7: Corrupted data at time 0 on the left and at time 66 on the right. The green squares correspond to a place where data have been corrupted.

so that the reader can reproduce the exact same experiment. After the add of these synthetic errors we have a look on the new data in Figure 9.7.

We now wonder about two different things; Does the algorithm provides us with a couple $(\tilde{f}_{seas}, \tilde{f}_{trend})$ that fits the non-corrupted data well ? Does the couple $(\tilde{f}_{seas}, \tilde{f}_{trend})$ also fits well the data that we took out of the pool on purpose ? To answer this question we first look at the Figure 9.8 for a spatial interpretation and Figure 9.9 for a temporal interpretation.

Figure 9.8 suggests that the fit is spatially good where the data is not corrupted, the algorithm here does not seem to be affected by the lack of data on some points. Figure 9.9 suggests that the fit is temporally good where the data is not corrupted, again the algorithm here does not seem to be affected by the the lack of data at some

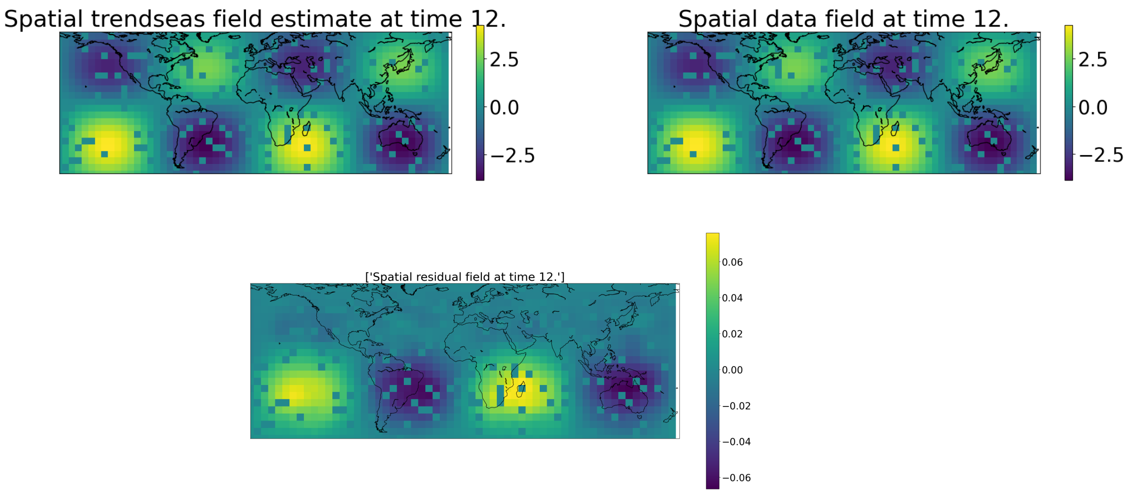


Figure 9.8: The top left image is a section of the estimated field at time 12. The top right image is a section of the data at time 12. The lower image is a section of the residuals at time 12. The greens squares still correspond to a place where data have been corrupted

points. We also notice that it seems, according to Figure 9.9, that the fit is good also where the data has been corrupted. Indeed one can see on the plot few orange dots, hidden in the purple ones. Those are where we lack information. In order to get a more accurate idea of how good it performs we will look at the following quantities. We start by defining S_1 as the set of spatio-temporal points where the data samples have not been corrupted and get

$$\sup_{(p,t)\in S_1}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.518, \quad \frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.066,$$

$$\frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left(\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| - \frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right|\right)^2 = 0.004.$$

$$\sup_{(p,t)\in S_1}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.269, \quad \frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.067,$$

$$\frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left(\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| - \frac{1}{|S_1|}\sum_{(p,t)\in S_1}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right|\right)^2 = 0.003.$$

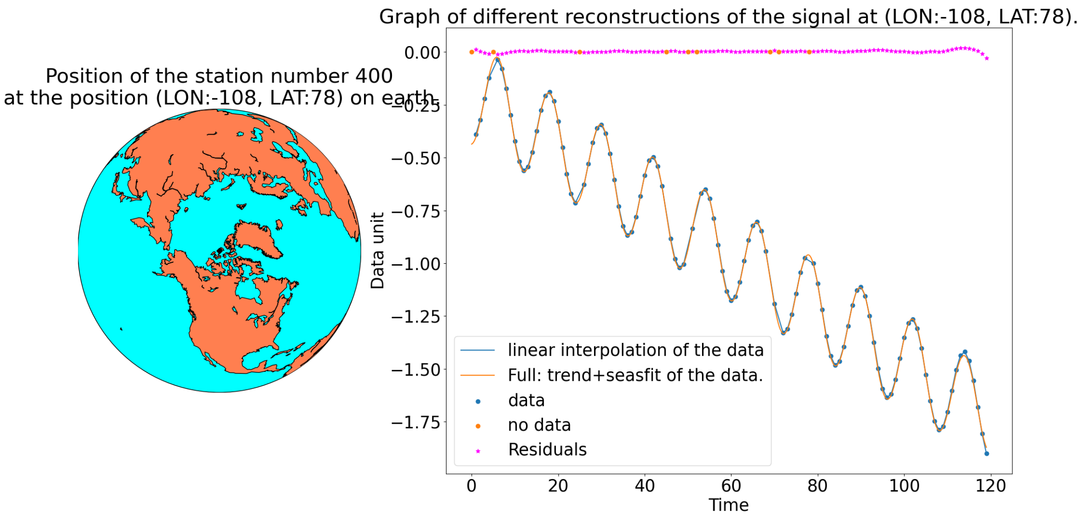


Figure 9.9: The full, seasonal, trending fit of the data are evolving against time at a fixed spatial position.

Then we define the set S_2 as the set of spatio-temporal points where the data samples have been corrupted and get

$$\sup_{(p,t)\in S_2}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.492, \quad \frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.066,$$

$$\frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left(\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right|-\frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right|\right)^2 = 0.004.$$

$$\sup_{(p,t)\in S_2}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.269, \quad \frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.067,$$

$$\frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left(\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right|-\frac{1}{|S_2|}\sum_{(p,t)\in S_2}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right|\right)^2 = 0.003.$$

Finally we use a spatio-temporal lattice defined as the product between a (spherical) Fibonacci lattice with 5000 points and a real uniform lattice with 494 points, called

L. We get the results

$$\sup_{(p,t)\in L}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.501, \quad \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| = 0.083,$$

$$\frac{1}{|L|}\sum_{(p,t)\in L}\left(\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right| - \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{trend}(p,t)-\tilde{f}_{trend}(p,t)\right|\right)^2 = 0.004.$$

$$\sup_{(p,t)\in L}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.275, \quad \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.084,$$

$$\frac{1}{|L|}\sum_{(p,t)\in L}\left(\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| - \frac{1}{|L|}\sum_{(p,t)\in L}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right|\right)^2 = 0.004.$$

We extract the following points from the three previous results. First, the seasonal and trending fit (therefore also the full one) perform as good on the corrupted data as on the non-corrupted one: big maximal error but small expectation with even smaller variance. Secondly we can compare the results with the product lattice to the ones we had in section 9.1, where the fitting procedure had all data available. It is that there is a small difference for the maximal errors but absolutely no differences for the expectations and the variances. In conclusion we can say that with this example, with a known data corruption of 10%, the algorithm, fitting procedure, perform as good as if there is no corruption of the data, by mean of both the seasonal fit quality and the trending fit quality. That is, very good.

9.3 A third test: noisy measurements robustness

In this third test we want to test the robustness our fitting procedure against noise. To do so we again place ourself in the exact same setting as in section 9.1. The difference with section 9.2 is that we supposed the corrupted data to be unusable, one should imagine that we did like it never never was in our hand and therefore the fitting procedure was blind to it. Thus what we did suppose that the experimenter knows where and when the data have been corrupted. Now we suppose that all of the data is noisy, the experimenter does not know where and when the noise is, but the fitting procedure can see everything. We made the experiment reproducible by seeding it, which can be found in the Github. More precisely we used a normal noise with variance $\left(\frac{f}{15}\right)^2$, so that for example if one of our sample is $f((0.2, 0.2, 0.6), 10)$

then the noise is given by $\mathcal{N}\left(0, \left(\frac{f((0.2,0.2,0.6),10)}{15}\right)^2\right)$, and the algorithm will be fed with $f((0.2, 0.2, 0.6), 10) + \mathcal{N}\left(0, \left(\frac{f((0.2,0.2,0.6),10)}{15}\right)^2\right)$. This procedure is done on all samples which are then fed to the fitting procedure.

▶ **Remark 9.1.** This noise has not been chosen for a particular physical or mathematical meaning but rather to challenge the model. ◀

Before discussing the results the lecturer can have a look to the noised data in order to get a glimpse at how our new noisy measurements behave, with Figure 9.10 and Figure 9.11.

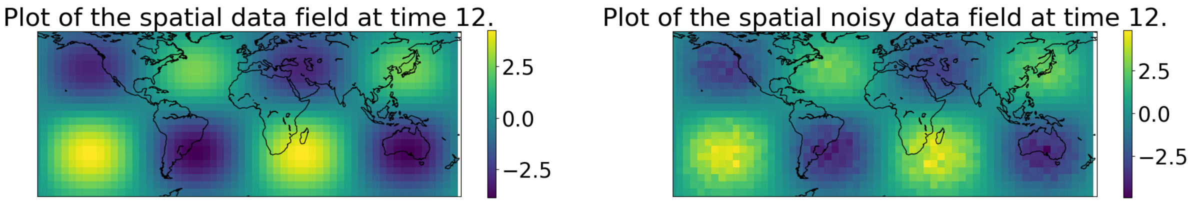


Figure 9.10: Comparison between the original data at time 12, on the left, and the noisy data at time 12, on the right.

We first want to know if the fit is a good estimation not of the noisy data but of the data and to do so we look at the Figure 9.12 for a spatial representation and at Figure 9.13 for a temporal representation.

It seems from Figure 9.12 that spatially the fitting procedures does a really good job de-noising. Figure 9.13 confirms this thought, temporally, and allows to think that the seasonal and trending part are not significantly affected by the noise (green and orange curve respectively). We confirm this intuition by defining S as the set of spatio-temporal points on which we have data, from which we get the following quantities:

$$\sup_{(p,t)\in S}\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| = 0.533, \qquad \frac{1}{|S|}\sum_{(p,t)\in S}\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| = 0.064,$$

$$\frac{1}{|S|}\sum_{(p,t)\in S}\left(\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right| - \frac{1}{|S|}\sum_{(p,t)\in S}\left|f_{trend}(p,t) - \tilde{f}_{trend}(p,t)\right|\right)^2 = 0.004.$$

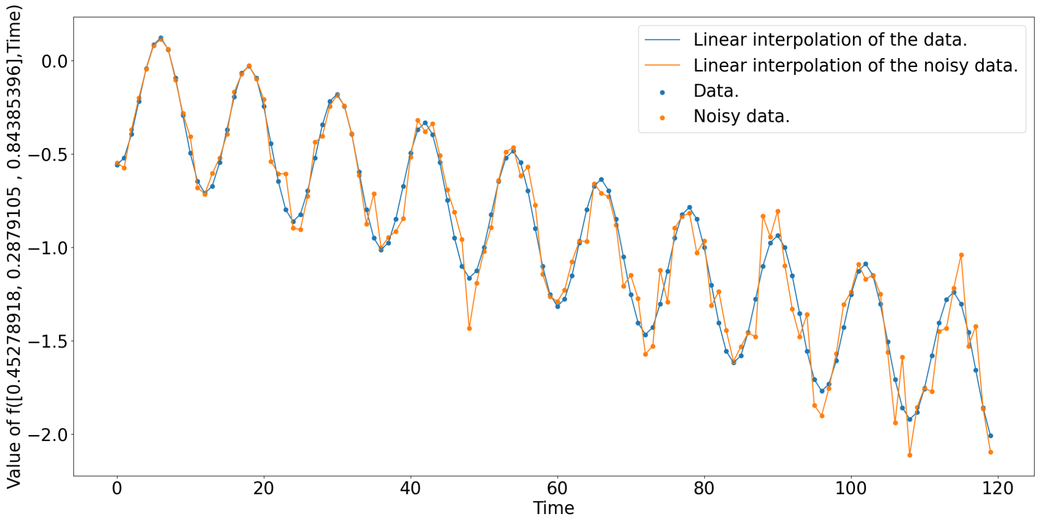
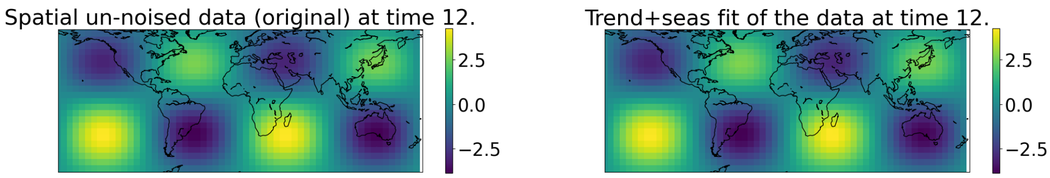


Figure 9.11: Comparison between original and noisy data against time, with fixed spatial position.

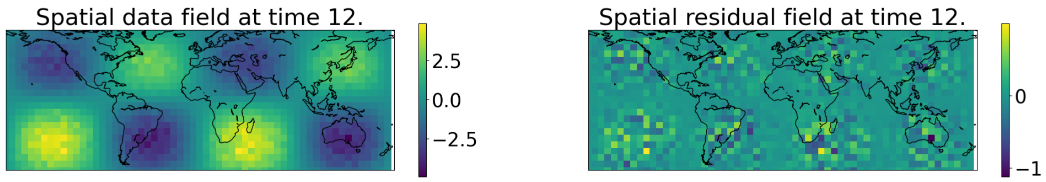
$$\sup_{(p,t)\in S}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.274, \quad \frac{1}{|S|}\sum_{(p,t)\in S}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| = 0.065,$$

$$\frac{1}{|S|}\sum_{(p,t)\in S}\left(\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right| - \frac{1}{|S|}\sum_{(p,t)\in S}\left|f_{seas}(p,t)-\tilde{f}_{seas}(p,t)\right|\right)^2 = 0.003.$$

Here, (f_{seas}, f_{trend}) correspond to the original de-noised functions and $(\tilde{f}_{seas}, \tilde{f}_{trend})$ correspond to the fitting functions returned by the algorithm, that are only based on the noised data. Comparing these results with the one given in section 9.1 yields the claim that, in this example, with this particular noise, the fitting procedure performs as good with noisy measurements as with normal measurements, by mean of both the fitting quality of $\tilde{f}_{seas}$ and $\tilde{f}_{trend}$. Remark that in the previous quantities the randomness has nothing to do with the randomness of the noise. Moreover, in order to get a more accurate estimation of the robustness of our fitting method we should first re-iterate the previous example with different generations of noise in order to get an expectation on the noise $\left(\underset{noise}{\mathbb{E}}\ \underset{data}{\mathbb{E}}\right)$ and secondly try all again with different noises. We skip this part as it is computationally intensive and not the main object of this paper.



(a) On the left is the plot of the original data at time 12. On the right is the fit the algorithm produced, provided with the noised data, at time 12.



(b) On the left is the plot of the noised data at time 12. On the right is the residual field, $f + \mathcal{N}\left(0, \left(\frac{f}{15}\right)^2\right) - \tilde{f}$, at time 12.

Figure 9.12

9.4 A fourth test: over-determined model robustness

In this section we want to investigate how the fitting procedure reacts when given a number of degrees of freedom far superior to the number of data. To do so we again place ourself in the setting of section 9.1. However we will restrict our attention, for computational efficiency, to a lattice between (-90,90) longitude and (-45, 45) latitude. The data from a spatial viewpoint now looks like.

Provided with this new data, we use in the fitting procedure 10 times more nodes than there is of data. Knowing that with a number of nodes equal to the number of data, we have a good fit, we would expect (in the best case) in this situation to have 90% of the weights matrices being equal to 0. Over 2000 iterations we used two algorithms, the primal dual splitting (PDS) and the accelerated proximal gradient descent (APGD). The results are summarized in the Figure 9.15 and Figure 9.16. From these plots we extract the following information. First, while the number of non-zero entries keep decreasing in the number of iterations with the APGD method, we do not observe this behavior with PDS method. Moreover, it seems that continuing

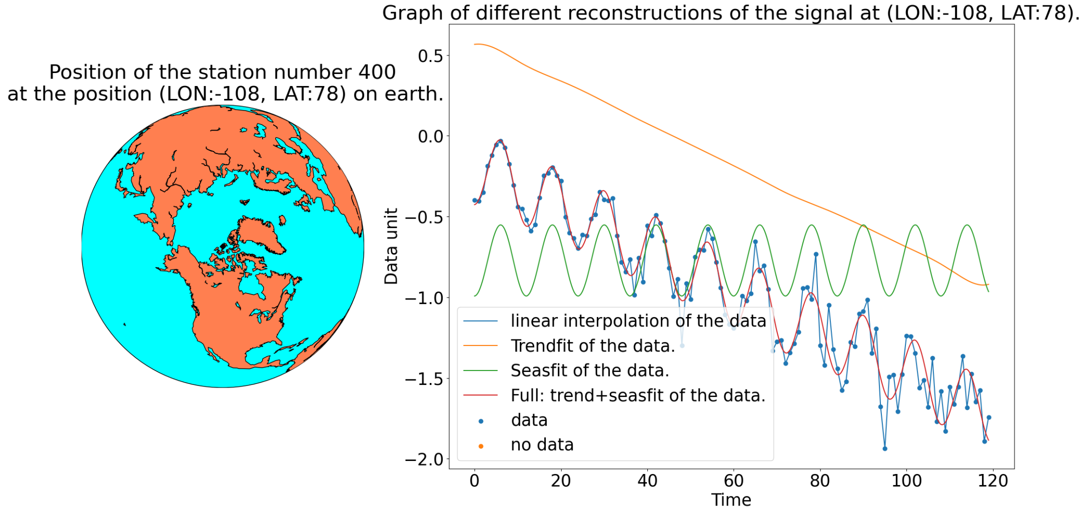
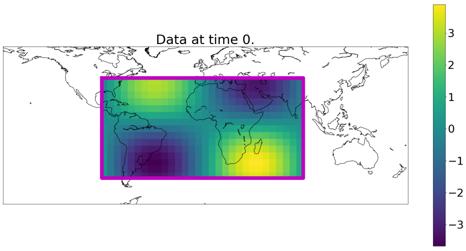
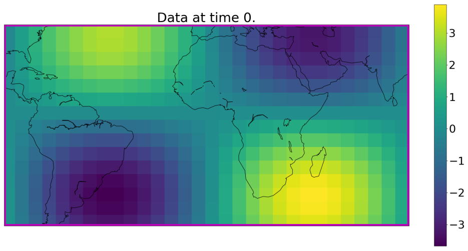


Figure 9.13: Time plot with different quantities involving the noised data at a fixed spatial location.

to run the APGD (over 2000 iteration) would continue decreasing the number of non-zeros entries in both the seasonal and trending weights matrices. Secondly, for the APGD, we get that there is in the end around 10% of non zeros entries in the seasonal matrix and 6% of non zeros entries in the trending matrix. This difference is easily explained by the fact that fitting a line for the trend requires less weights than fitting a non-linear in the period, periodic function. These numbers are fully satisfactory even though one could obtain better ones by running the APGD for a longer period of time. It is also worth noticing that for both the APGD and the PDS method, the fit is already extremely good at iteration 500. Therefore what happens between the iterations 500 and 2000, for the APGD, is an automatic parameters selection. Nevertheless further investigation are needed to find if the APGD fully takes advantage of the sparsity promoting behavior of the L1 norm or if one would need a sparsity promoting optimization scheme in order to take advantage of it. This last robustness test concludes chapter 9.

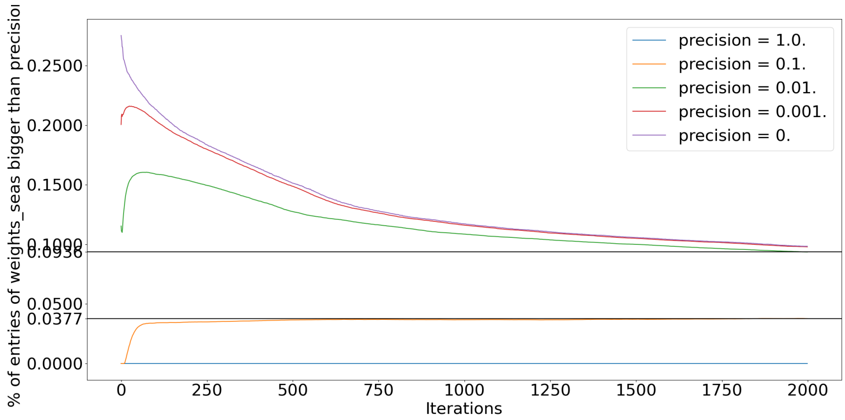


(a) Data on a section at time 0. Represented on the whole sphere.

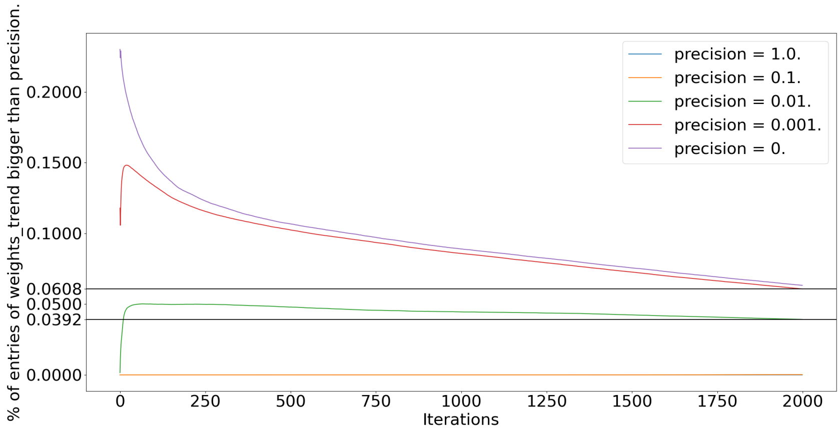


(b) Data on a section at time 0. Represented on a sphere section.

Figure 9.14

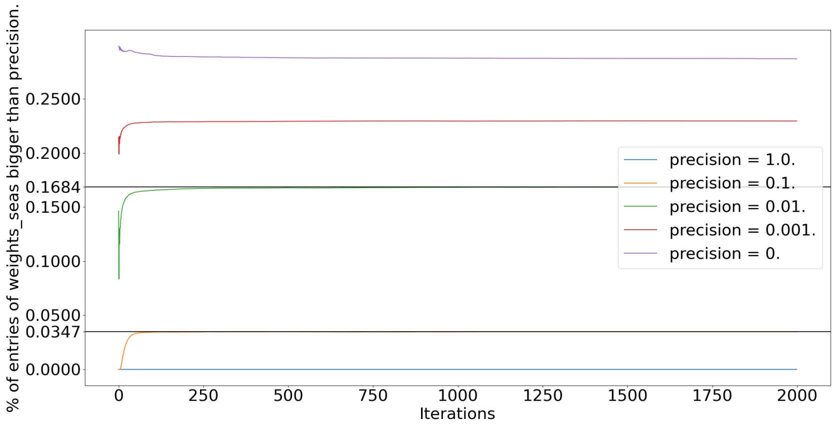


(a) Percentage of entries bigger than a certain precision in the weights matrix associated to the seasonal component.

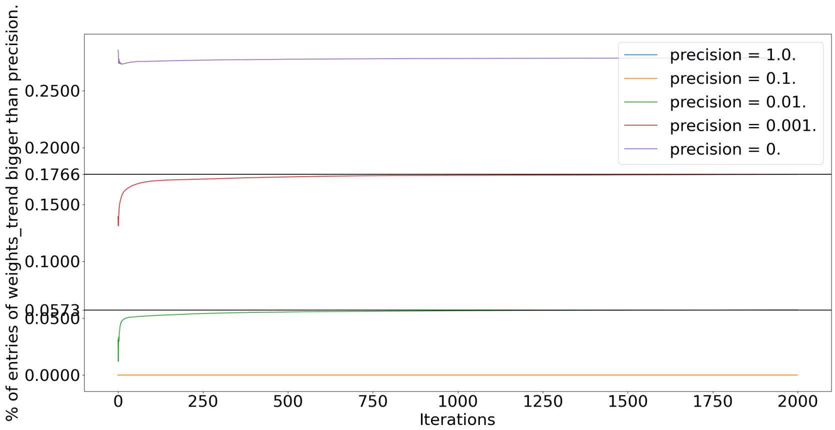


(b) Percentage of entries bigger than a certain precision in the weights matrix associated to the trending component.

Figure 9.15: Sparsity graphs for the fit done with APGD.



(a) Percentage of entries bigger than a certain precision in the weights matrix associated to the seasonal component.



(b) Percentage of entries bigger than a certain precision in the weights matrix associated to the trending component.

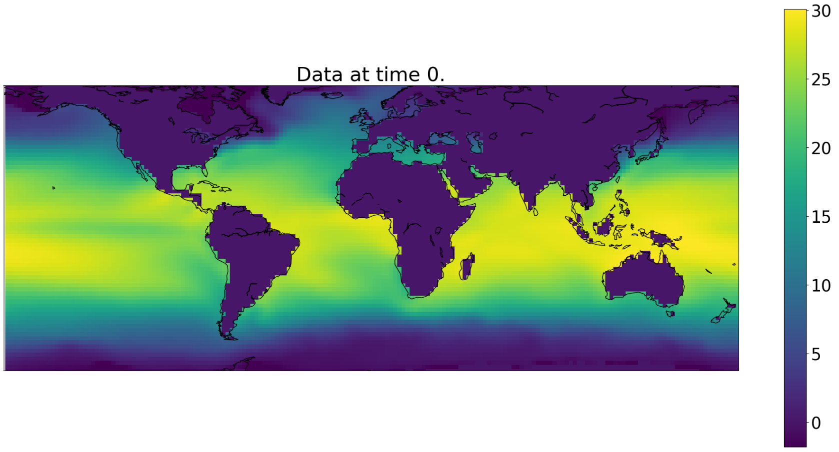
Figure 9.16: Sparsity graphs for the fit done with PDS.

10 Empirical applications

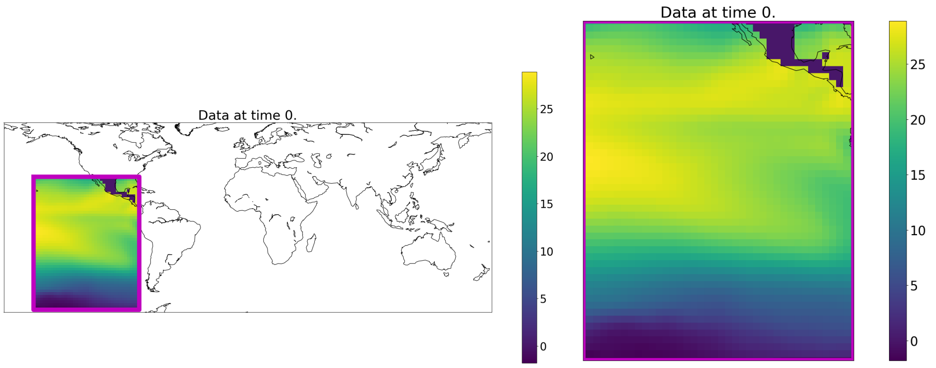
This second embedded in the applications chapter is framing an empirical analysis of the previously created theory. To do so we consider the dataset [-Z] and apply our fitting procedure on it. Before going further it is to be mentioned that NOAA-ERSST-V5 data has been provided by the NOAA-OAR-ESRL-PSL, Boulder, Colorado, USA, from their Web site. The goal of this chapter will be to show with a simple analysis that our framework does not only work in theory, on synthetic data, but also on a "real" dataset, with a more complex spatial field, seasonality, trend. By doing so we will see how far we can go with our method and thus what can, should be upgraded in a future work. In particular we will be able to bring to light non trivial seasonality (different from a simple cos-sin curve), short-time and long-time trending behavior. From this will arise the question; How to model non seasonal short-time and long-time trending behavior differently ?
This experiment is not available on the Github page, due to its size.

The data set in question is featuring a spatio-temporal map of temperatures on the oceans between 1854 and 2021. Because of computational issues we will not use the whole data set but a time section between 1930 and 2019, both extremes included, thus of 90 years in order to find out if we can get long-term trending behavior and eventually hints of a global (warming, freezing, oscillating ?). We will also restrict ourselves to a spatial window, framed by LONG [-158, -80] LAT [-68, 30], what can be appreciated in the Figure 10.1 and we will later have glance to the time series. The results of the experiment are displayed after Figure 10.1.

Just like before we start our analysis by having a look a global representation, through spacial sections of our fitted spatio-temporal field, displayed in Figure 10.2. We notice from (a), that generally the fit is really good at time 12, (b) provides us with a glance at how the seasonal and trending part share the mass. Here we adjusted the constant in the trend so that the seasonal fit has mean 0. The trend thus encapsulate the general behavior and the season encapsulate the yearly periodic variations, both described here in (b). Finally (c) gives us hints at how the residuals behave with respect to the original data. We can see that the residuals are dependant on the data in a way that when there is an important temperature gradient the residuals seem to be indicating either that the measurements are noisy when the gradient is important or that our method suffer from a lack of precision



(a) NOAA-ERRST-V5 data at time 0.



(b) Section of NOAA-ERRST-V5 data at time 0, not focused.

(c) Section of NOAA-ERRST-V5 data at time 0, focused.

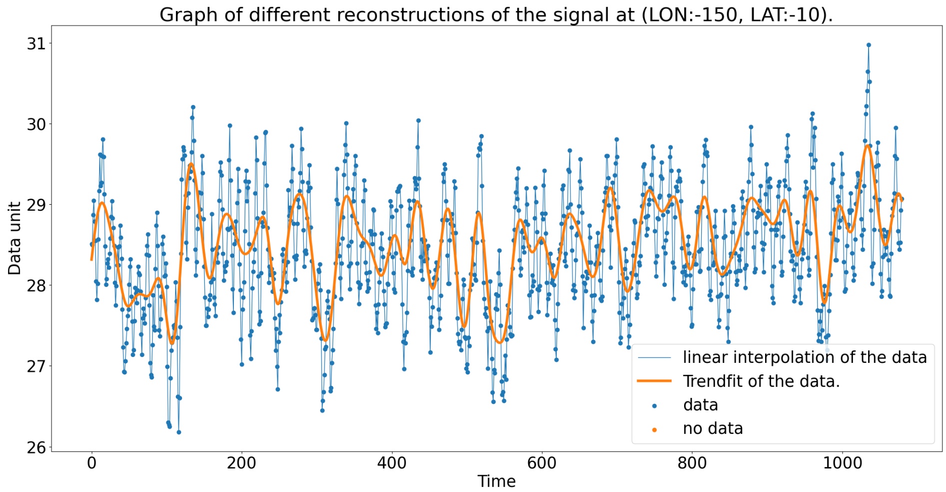
Figure 10.1: Different spatial representations of the NOAA-ERRST-V5 at time 0.

when the gradient is important. In the second case, the problem might be avoided by considering a lower regularization. Nevertheless it seems that the residuals are spatial Normal when the original temperature is spatially flat (close to constant). In order to continue the analysis further one would need to establish an histogram of the residuals, one for those which are linked to a small gradient and one for those which are linked to a big gradient. We do not conduct this experiment here but hope that in the flat case a normal curve would appear. After this experiment, if it was to seem that the residuals are normal when the region is flat it would be worth investing the distribution of the residuals normalized by the norm of the gradient. We should also wonder if the seasonal / trend splitting has been conducted successfully.

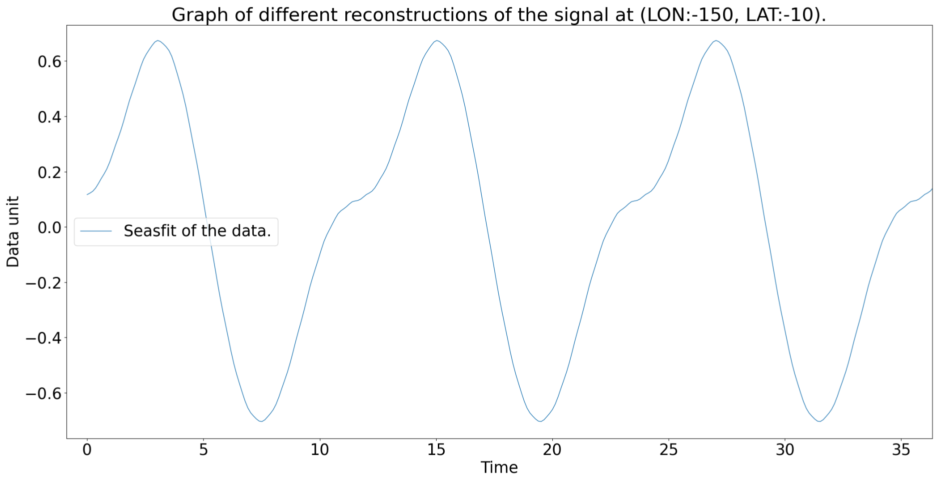
If there is no part of the season that has been included in the trend and if the trend does a good job in approximating the short-time and long-time, if they exit, variations. To do so we look at the temporal sections of our fitted fields. The results are displayed in Figure 10.2 and Figure 10.3.

If there is no part of the season that has been included in the trend and if the trend does a good job in approximating the short-time and long-time, if they exit, variations. To do so we look at the temporal sections of our fitted fields. The results are displayed in Figure 10.2 and Figure 10.3. We first observe that our procedure extracted non trivial seasonal parts in both cases as the first one features two regimes in the increasing part and the second one an extended minimum. A weird wiggling in the lower pick has to be seen in the first. This could be explained by to much splines used for the seasonal part and or splines that do not overlap enough. The last explanation is nevertheless less likely as we think that then this bad behavior would be general, the whole function would be wiggling. We then have a look at the trend, both are non-trivial, look-like they fit the general behavior well and are composed of, as previously hinted to the reader, a short-time trend and a long-time trend. The short-time trend is to be seen as those positive and negative yearly fluctuations that go along the natural temperature fluctuations. The reasons why the seasonal component has not digested these fluctuations is because several assumptions have been violated. First the amplitude of the season is not constant, secondly the time at which the season reaches its maximal temperature is not constant and last but not least, the period is not continuous in the sense that the value at the end of the period is not always the same as the value at the start of the period (ex. 0 degrees in january 2019, -2 degrees in january 2020) and if this process is severe and (on a short period of time) recurrent, the trend digests it. This is what we meant by short-time trend. Nevertheless we can also observe (semi) long-term trending behavior, better seen in the Figure 10.4. This suggests that the trend we have been considering in this paper can be philosophically be decomposed into different sub-trends at least in short and long time trend and that the approach we are now using is unstable in the number of splines we use to represent the trend. More spline and one has short time variations, less splines and one has long time variations. A better model would thus allow the experimenter to model both at the same time. We finish this discussion by considering the residuals plot, both of them in Figure 10.2 and Figure 10.3 feature residuals that are framed in a reasonable interval around 0, indicating that our model is efficient in fitting the corresponding time serie. We do not observe any pattern in the residuals, that would indicate for a missing term in our model. Finally the residuals look like they are spreading in a constant (heteroscedasticity) manner around -0.093 for Figure 10.2 and around -0.086 for Figure 10.3. This suggests first that we are doing a small overestimation in both cases, which might indicate a more general behaviour and secondly that for a fixed spatial position the residuals evolve against time without inflation or deflation and thus that there is no systematic dependencies of the residuals on the value of the underlying time serie, we have heteroscedasticity. Nevertheless only a deeper analysis would tell us if the auto-correlation function is trivial or if the different residuals are correlated. We would expect a short-time correlation. We conclude from this experimentation that our model is fitting efficient spatially and temporally (separately), as the residuals showed us. Also one of its main purpose was to break a spatio-temporal serie into a non-trivial seasonal and trending part, which we now know to be done with success. It remains the uncertainty about the different roles of the trend, that would require a more complex model, the uncertainty about the distribution of the residuals and especially from a spatio-temporal viewpoint. The spatio-temporal efficiency of our method remains to be proved, either by understanding better the distribution of the residuals or by performing a spatio-temporal cross-validation or a bootstrap in order to evaluate a (spatial? temporal? spatio-temporal?) generalized error.

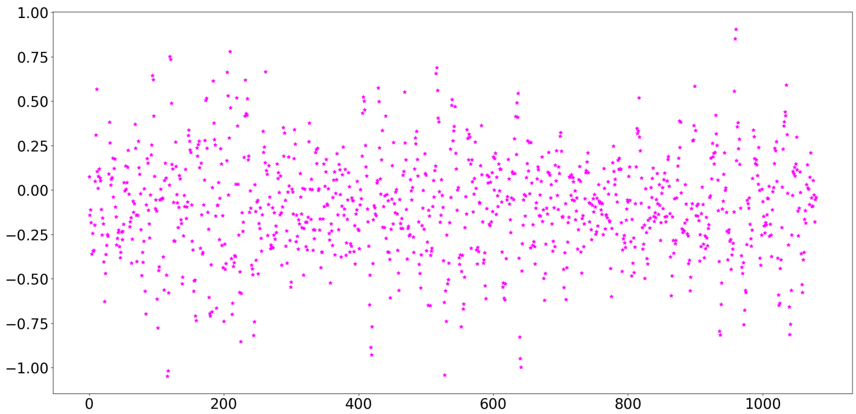
Last but not least, a video representing the spatio-temporal evolution of the seasonal field is available at https://youtu.be/6drC618Qu14 and a video representing the spatio-temporal evolution of the trend field is available at https://youtu.be/vE3xeA4iW9c.



(a) Between 1930-2019, at position (LONG, LAT)=(-150, -10), the time serie of NOAA-ERRST-V5 and the corresponding trend function fitted (the constant is corrected).

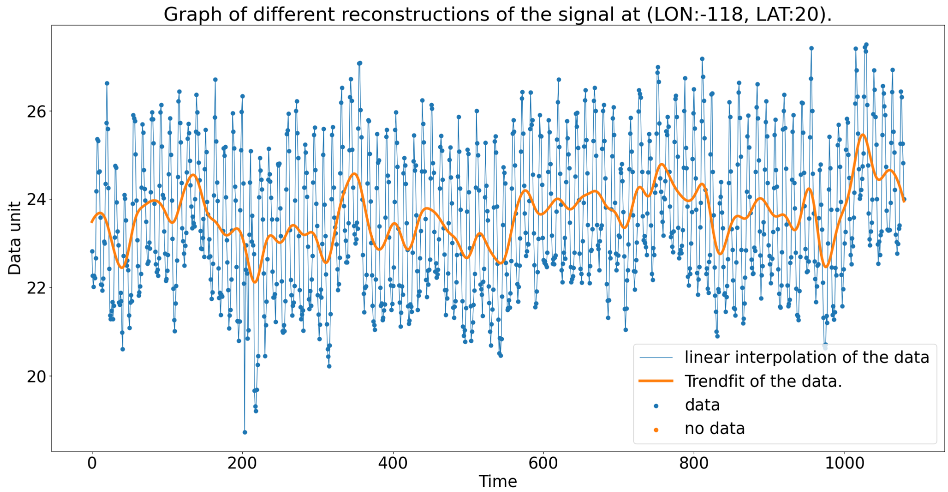


(b) At position (LONG, LAT)=(-150, -10), three periods of the seasonal function fitted to the time serie (the constant is corrected).

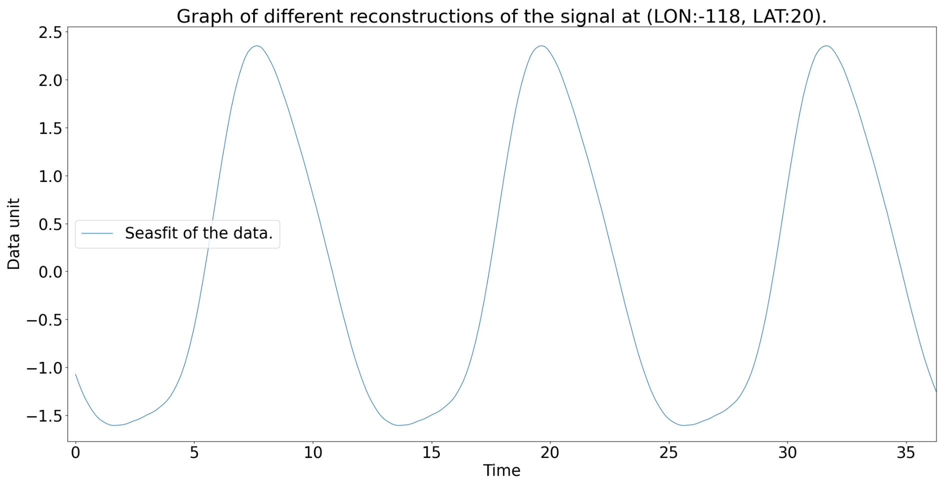


(c) Between 1930-2019, at position (LONG, LAT)=(-150, -10), the residual time serie.

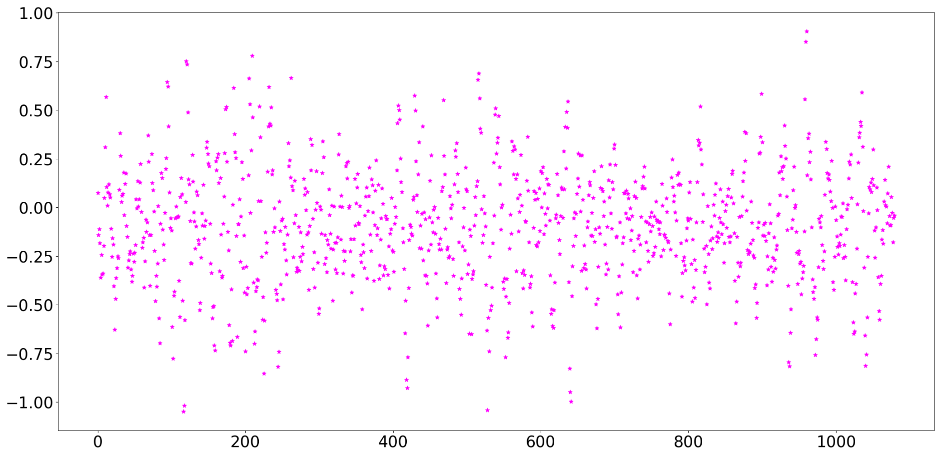
Figure 10.2: Different quantities corresponding to the fit of a time serie extracted from NOAA-ERRST-V5. Time is in month. Data unit is in degrees Celsius. The units for the residuals are the same.



(a) Between 1930-2019, at position (LONG, LAT)=(-118, -20), the time serie of NOAA-ERRST-V5 and the corresponding trend function fitted (the constant is corrected).

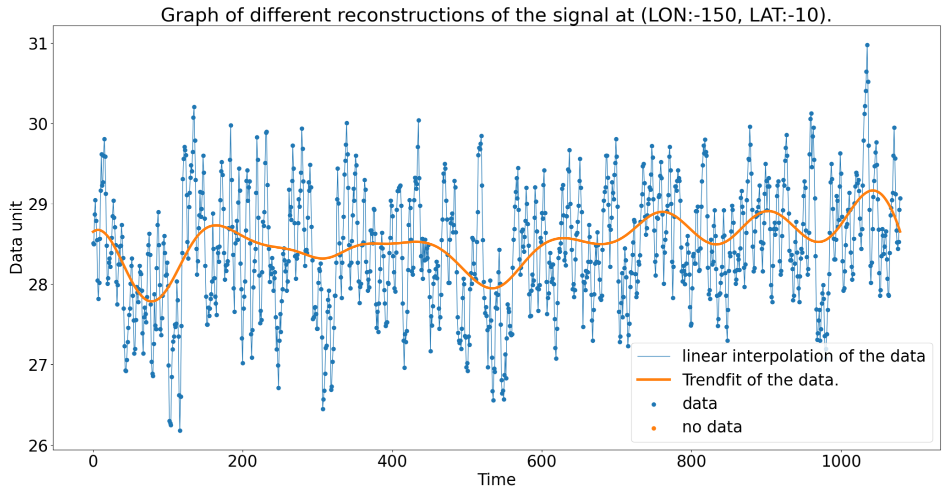


(b) At position (LONG, LAT)=(-118, -20), three periods of the seasonal function fitted to the time serie (the constant is corrected).

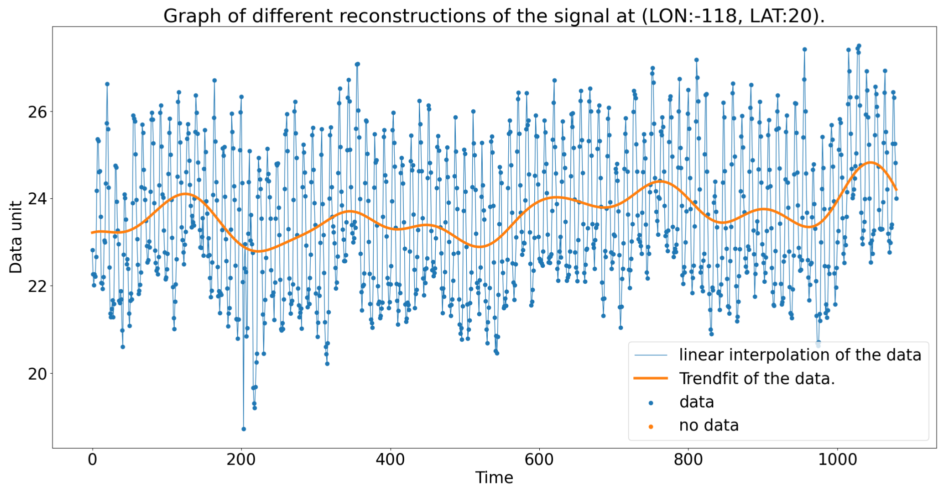


(c) Between 1930-2019, at position (LONG, LAT)=(-118, -20), the residual time serie.

Figure 10.3: Different quantities corresponding to the fit of a time serie extracted from NOAA-ERRST-V5. Time is in month. Data unit is in degrees Celsius. The units for the residuals are the same.



(a) Same time serie as in Figure 10.2 but with a trend composed of less splines (with bigger support).



(b) Same time serie as in Figure 10.3 but with a trend composed of less splines (with bigger support).

Figure 10.4: Two time series, the same as before but with a trend composed of less splines (with bigger support).

11 Conclusions & Outlook

The goal of this paper was to produce a framework, a method to allow one to dismantle a spatio-temporal field, from finitely many noisy measurements, into the superposition of a continuous seasonal and a continuous trending parts. The first chapter build the theory, giving life to it is the algorithm available on Github and the second chapter review basic applications of our framework and challenges it throughout different situations, in particular with a real data set. All of this form our, we think complete, first answer to this question and we hope that reading this paper was both instructive and a pleasure. While working on this paper, we thought about all the extensions, lemma, ideas we missed or simply did not have the time to work on, in the same idea as the cook who knows that there should be an additional spice in the recipe, that he forgot, which nobody noticed. We are now doing a non-exhaustive review of the fine spices we missed, hoping that we will have the time someday somehow to incorporate them into the recipe. These delicacies feature:

1. An extension of the theory to vector fields.
2. An extension of the theory to non-separable operators.
3. An extension of the theory to operators with an infinite-dimensional nullspace.
4. Understanding better the solution set of the discretized problem. Their cardinality, extreme points. The size of the sparser solutions.
5. Understanding better the solution set of the original problem. Can the solution be unique ? When?
6. An extension of the theory to other kinds of cost functional. (It should not be hard using our representer theorem.)
7. An extension of the theory to better dismantle the fields. Into smaller pieces, for examples with short time and long time trends. Short and long time season.
8. Understanding better the role of α, β in the penalization. How do they affect the solution set ?

9. Creating tools to understand the spatio-temporal residuals and their distributions.

10. Extend the algorithm to be GPU supported.

11. Extending our theory on other spaces than the sphere. For example if one has a shape (cylinder / plane / car / rubber band...) with scalar tensions, constraints on it, that can only be measured empirically. There is no deterministic theory or the equations are to complicated to be solved. Then one could measure these constraints and interpolate them with our technique to better understand how they evolve against time, in a certain environment.

12. A model that allows regression not only on a space of measures but also on a spaces of random measures.

Bibliography

[-Z] Boyin Huang - Peter W. Thorne - Viva F. Banzon - Tim Boyer - Gennady Chepurin - Jay H. Lawrimore - Matthew J. Menne - Thomas M. Smith - Russell S. Vose - and Huai-Min Zhang. **NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5**. *NOAA National Centers for Environmental Information* () (see page 87).

[Bec11] Lawrence Naciri - Edward Beckenstein. **Topological vector spaces**. Pure and applied mathematics. 2011 (see page 25).

[Bre11] Haim Brezis. **Functional Analysis, Sobolev Spaces and Partial Differential Equations**. 2011 (see page 56).

[Deb+19] Thomas Debarre, Julien Fageot, Harshit Gupta, and Michael Unser. **B-Spline-Based Exact Discretization of Continuous-Domain Inverse Problems With Generalized TV Regularization**. *IEEE Transactions on Information Theory* 65:7 (2019), 4457–4470. DOI: 10.1109/TIT.2019.2902926 (see page 51).

[Fra06] Trèves Francois. **Topological Vector Spaces, Distributions and Kernels**. 2006 (see page 11).

[Gui20] Vincent Guillemet. **A spline based approximation framework for spatio-temporal environmental monitoring**. 2020 (see pages 1, 2, 16).

[Jea20] Simeoni Matthieu Martin Jean-Andre. **Functional Inverse Problems on Spheres: Theory, Algorithms and Applications** (2020) (see pages 1, 13, 29).

[Mat20] Fageot Julien - Simeoni Matthieu. **TV-based reconstruction of periodic functions** (2020) (see page 1).

[Sha21] Unser Michael - Aziznejad Shayan. **Convex optimization in sums of Banach spaces** (2021) (see page 26).

[Uns18] Harshit Gupta - Julien Fageot - Michael Unser. **Continuous-Domain Solutions of Linear Inverse Problems With Tikhonov Versus Generalized TV Regularization**. *IEEE Transactions on Signal Processing* 66 (2018), 4670–4684 (see page 19).

[Wil10] V. Chandrasekaran - B. Recht - P. Parrilo - A. Willsky. **The Convex Geometry of Linear Inverse Problems** (2010) (see page 2).