EPFL

Thèse n° 7174

Functional Inverse Problems on Spheres: Theory, Algorithms and Applications

Présentée le 28 février 2020

à la Faculté informatique et communications
Laboratoire de communications audiovisuelles
Programme doctoral en informatique et communications

pour l'obtention du grade de Docteur ès Sciences

par

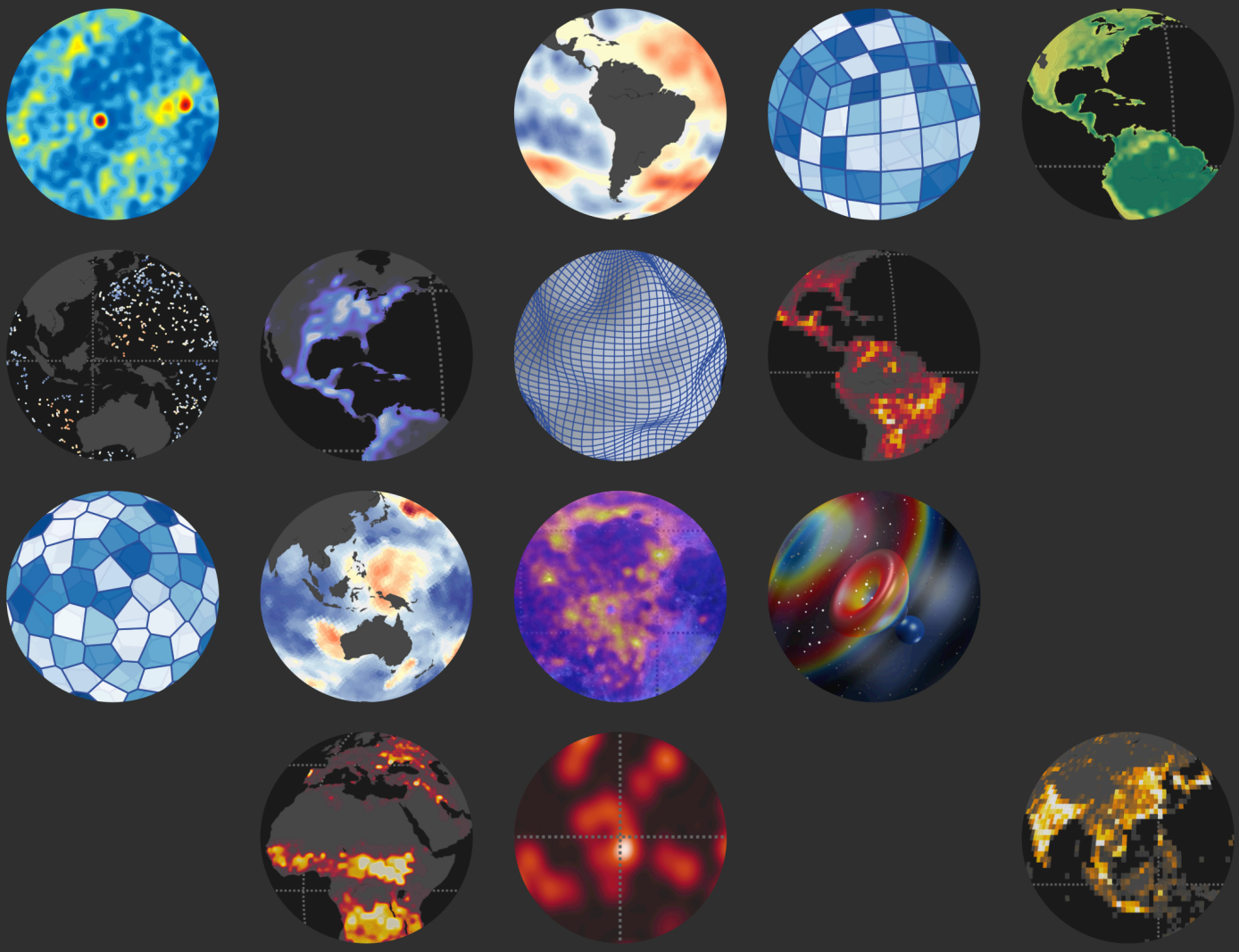
Matthieu Martin Jean-Andre SIMEONI

Acceptée sur proposition du jury

Prof. P. Vandergheynst, président du jury
Prof. M. Vetterli, Prof. V. Panaretos, directeurs de thèse
Prof. D. Marinucci, rapporteur
Prof. T. Blu, rapporteur
Prof. M. Unser, rapporteur

École polytechnique fédérale de Lausanne

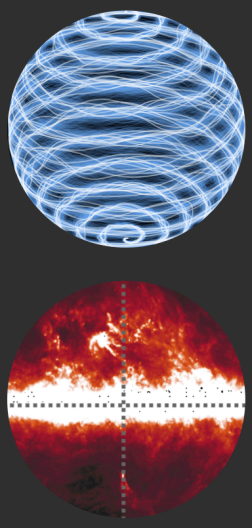
2020

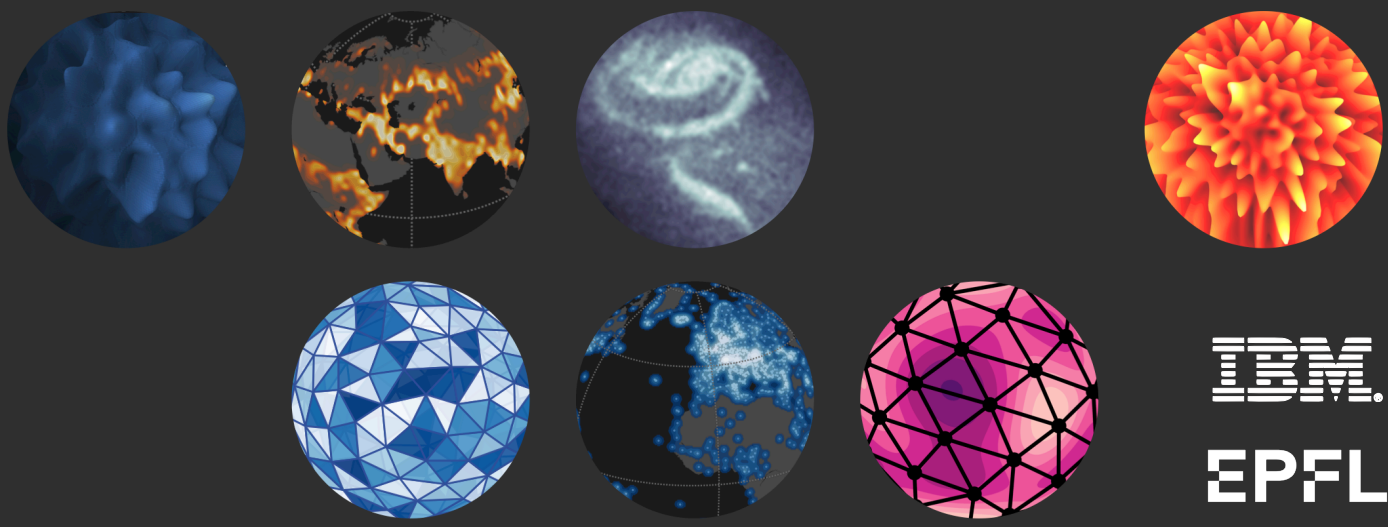


Functional Inverse Problems on Spheres

Theory, Algorithms and Applications

Matthieu Simeoni





IBM

EPFL

A thesis presented by Matthieu Simeoni to the

Doctoral School of Computer and Communication Sciences, Audiovisual Communications Laboratory (LCAV), École Polytechnique Fédérale de Lausanne (EPFL)

in partial requirements for the degree of Doctor *ès* Sciences.

Thesis committee: Pr. Pierre Vandergheynst (president), Pr. Martin Vetterli and Pr. Victor Panaretos (co-advisors), Pr. Michaël Unser (internal expert), Pr. Domenico Marinucci and Pr. Thierry Blu (external experts).

This thesis was funded by and carried out at IBM Research in Zurich, under the supervision of Pr. Paul Hurley in the group of Dr. Costas Bekas.

Lausanne, École Polytechnique Fédérale de Lausanne (EPFL), January 3, 2020

Acknowledgements

I would like to express my sincere gratitude to my academic advisors Pr. Martin Vetterli and Pr. Victor Panaretos, as well as my technical and industrial supervisor Pr. Paul Hurley. All three of them greatly contributed to the success of this thesis through countless scientific contributions and insightful pieces of advice:

- Martin's strong personal ethic, great humility and impressive scientific track record were truly inspiring for me and helped me grow as a young scientist and researcher. His personal engagement for open science and reproducible research had a lasting influence on my own work. I am deeply honoured and proud to have been his student and would like to thank him warmly for guiding me expertly throughout my PhD.
- I have known Victor for almost seven years now: first as a teacher during my undergraduate studies, then as a supervisor on multiple academic projects –including my Bachelor and Master thesis– and finally as my thesis co-advisor. He was during all these years –and continue to be today, a true role model and mentor for me, who influenced in many ways the researcher that I am today. His great pedagogical and presentation skills are notably at the origin of my current interest in statistics and applied mathematics at large. I feel truly blessed to have had the opportunity to work and interact with such a brilliant mind, and would like to thank him again for his help and support during all these years.
- Paul and I finally, have met during my work at IBM Research, where he was my technical supervisor for a little bit more than four years. I have learnt immensely at his side. He is a gifted researcher –driven by a passion for science and mathematics– and a great leader, who knows how to put his students before himself, listen to their ideas and help them grow in their careers. The work environment he managed to create at IBM as a technical lead was incredibly stimulating for me. I cherish in particular the memory of our endless scientific discussions, where we worked, as collaborators and friends, towards developing our shared vision for radio astronomy. I could never thank him enough for everything he did for me, during this thesis and before.

On December 4 2019, I defended during three intense and memorable hours the work of my thesis in front of a committee presided by Pr. Pierre Van-

dergheynst and composed of my two academic advisors and three world-class experts: Pr. Michael Unser, Pr. Domenico Marinucci and Pr. Thierry Blu. To be given the opportunity to discuss the results of my research with such esteemed members of the community of statistics and signal processing was for me a great honour and an incredible experience. I would like to thank them all again for their interest in my work, as well as the stimulating discussion and the very positive feedback they provided me with. I am particularly grateful to Pr. Michael Unser, who provided me with very useful feedback on earlier versions of this thesis, and whose own work has greatly influenced my research.

I must also acknowledge IBM Research Zurich for offering me an industrial PhD position and providing me with the freedom needed to carry out my own original research. I would like to thank in particular Dr. Costas Bekas for welcoming me in his group, as well as the many great minds I met there: Paul, Sepand, Lucien, Merve, Vesna, Hanjie, Patrick, Erwan and many more that I cannot list here. I have a particular thought for my dear friends Lucien and Sepand, with whom I worked very closely during my PhD. I will miss in particular the numerous passionated discussions I had with Lucien, which strongly influenced the direction of my thesis. He is notably the one who introduced me to the spline framework long before I got acquainted with the work of Michael Unser. I feel extremely lucky to count him among my friends. I am also indebted to Sepand in so many ways, who is an incredibly generous office mate. He notably introduced me to the wonders of Python and Numpy and expertly proofread most of my papers as well as the entirety of my thesis. I am truly grateful to him and could never thank him enough for his tremendous help. I hope I will get the opportunity to return him the favour in the future. I am also very glad to continue working with him as part of my next job.

During my time at EPFL, I also got to know, interact and collaborate with brilliant researchers: Adrien, Julien, Nathan, Miranda, Gilles, Robin, Adam, Benjamin, Hanjie, Frederike, Michalina, Golnoosh, Ivan, Luc, Mihailo, Eric and many more. All of them contributed to making my four years of PhD even more enjoyable. I am particularly grateful to Adrien Besson and Julien Fageot, who helped me on multiple occasions and with whom I worked very closely. Adrien and I met at a workshop three years ago where we had a very interesting discussion, revealing common research interests as well as a great complementarity and compatibility, both scientifically and humanely. This lead us to co-author three papers together. Among the many things he did for me, he notably introduced me to the wonders of convex optimisation and helped me during the writing of my thesis. I sincerely enjoyed interacting with him, and consider him a very good friend of mine. Julien entered in my life more recently, during the last year of my PhD. He contributed to my thesis in many important ways: first through his seminal work with his former thesis advisor Michael Unser, and then by providing me with expert bits of advice as well as thorough feedback on my work on sparse spline approximation on the sphere. More recently, we even initiated a long term collaboration on related topics. I am very excited about this prospect and look forward to the great work. He is a fantastic mathematician and a beautiful person, driven by his insatiable intellectual curiosity. I am incredibly grateful for the time he dedicated to my

thesis out of pure kindness. His help was extremely precious to me, and I owe him a lot.

Finally, I would like to thank all my fantastic friends, my dear family, and my amazing girlfriend Léa. To my friends first, I would like to say thank you for supporting me in indirect but equally important ways throughout this adventure. I must cite in particular Aschkan, Mathias, Adlen, Valentin, Damien, Paolo, Nathan, Benedikt, Yassine, Lucien, Camille, Yasmin, Lionel, Delphine and Esther, with whom I shared many precious moments during these four years. I consider myself extremely lucky to have you all in my life. To my parents, brother, sister and close relatives, thank you for your unconditional support and your love, who gave me the strength and confidence to achieve something as crazy as writing a thesis. I owe you most of who I am today and feel truly blessed to be surrounded by such a wonderful family. To Léa finally, who has been my partner in life for the past eleven years, I dedicate this thesis. She has always believed in me –even when I would not– and has been incredibly supportive and patient throughout this ordeal. Her trust and love gave me the strength to surpass myself. She is one of the kindest and empathetic person I know. I am extremely fortunate to have her along my side, and could never repay her enough for all the joy she brought and continues to bring every day in my life.

Matthieu Simeoni, January 3, 2020.

Foreword

"To finish the moment, to find the journey's end in every step of the road, to live the greatest number of good hours, is wisdom."

–Ralph Waldo Emerson.

A thesis is both a journey and a destination. You, the reader, are about to discover Matthieu's destination. I get to tell you a little bit about the journey.

Matthieu started an internship at IBM Research in Zurich, Switzerland, working on signal processing algorithms for radio astronomy. This lead to his masters thesis on a nascent imager, discrete in nature but whose geometry was already spherical. His PhD thesis journey began with a re-think, shaking up the unsatisfactory regular grid to embrace a "continuous" imager on the sphere that felt more graceful and deceptively simple. Its optimality is in the least-squares sense, and thus its derivation conceptually a form of Moore-Penrose pseudo-inverse. Thus we get closer to the document you will now read.

Some side-avenues lead to the philosophy of continuity carrying into a clever beamforming algorithm. The mathematical generality brought excursions into Positron Emission Tomography and ultrasound.

That none of these items found a place in this thesis is a testament to the cohesiveness, and ultimately significance, of the results that do and to the power of this journey of discovery.

This moment has now finished and it has been my pleasure to have accompanied and guided Matthieu.

Paul Hurley, October 2019.

Abstract

Many scientific inquiries in natural sciences involve approximating a *spherical field* –namely a scalar quantity defined over a *continuum* of directions– from generalised samples of the latter (e.g. directional samples, local averages, etc.). Such an approximation task is often carried out by means of a convex optimisation problem, assessing an optimal trade-off between a data-fidelity and regularisation term. To solve this problem numerically, scientists typically discretise the spherical domain by means of *quasi-uniform spherical point sets*. Finite-difference methods for approximating (pseudo-)differential operators on such discrete domains are however unavailable in general, making it difficult to work with generalised Tikhonov (gTikhonov) or generalised total variation (gTV) regularisers, favouring physically admissible spherical fields with smooth and sharp variations respectively.

To overcome such limitations, canonical spline-based discretisation schemes have been proposed. In the case of gTikhonov regularisation, the optimality of such schemes has been proven for spherical scattered data interpolation problems with quadratic cost functionals. This result is however too restrictive for most practical purposes, since it is restricted to directional samples and Gaussian noise models. Moreover, a similar optimality result for gTV regularisation is still lacking.

In this thesis, we propose a unified theoretical and practical spherical approximation framework for *functional inverse problems* on the hypersphere $\mathbb{S}^{d-1}$. More specifically, we consider recovering spherical fields directly in the continuous domain using penalised convex optimisation problems with gTikhonov or gTV regularisation terms. Our framework is compatible with various measurement types as well as non-differentiable convex cost functionals. Via novel representer theorems, we characterise the solutions of the reconstruction problem for both regularisation strategies. For gTikhonov regularisation, we show that the solution is unique and can be expressed as a linear combination of the sampling linear functionals –modelling the acquisition process– primitived twice with respect to the gTikhonov pseudo-differential operator. For gTV regularisation, we show that the solutions are convex combinations of spherical splines with less innovations than available measurements. We use both results to design *canonical* spline-based discretisation schemes, *exact* for gTikhonov regularisation and with vanishing approximation error for gTV regularisation.

We propose *efficient* and *provably convergent* proximal algorithms to solve the discrete optimisation problems resulting from both discretisation schemes. We illustrate the superiority of our continuous-domain spherical approximation framework over traditional methods on a variety of real and simulated datasets in the fields of *meteorology*, *forestry*, *radio astronomy* and *planetary sciences*. The sampling functionals, cost functions and regularisation strategies considered in each case are diverse, showing the versatility of both our theoretical framework and algorithmic solutions.

In the last part of this thesis finally, we design an efficient and locally convergent algorithm for recovering the spatial innovations of periodic Dirac streams with *finite rates of innovation*, and propose a recurrent neural-network for boosting spherical approximation methods in the context of real-time *acoustic imaging*.

Keywords: *functional inverse problem, spherical approximation, spherical splines, generalised Tikhonov regularisation, generalised total variation regularisation, representer theorems, continuous-domain recovery, canonical discretisation, proximal algorithms, geomathematics, environmental sciences, radio astronomy, planetary sciences, acoustic imaging*

Résumé

Dans le domaine des sciences naturelles, de nombreuses investigations scientifiques nécessitent l'approximation d'un *champ sphérique* –à savoir une quantité scalaire définie sur un *continuum* de directions– à partir d'échantillons généralisés de ce dernier (par exemple des échantillons directionnels, des moyennes locales, etc.). Une telle approximation s'effectue souvent en pratique par la résolution d'un problème d'optimisation convexe, visant à obtenir un compromis optimal entre un terme de fidélité aux données et un terme de régularisation. Pour résoudre numériquement ce problème d'optimisation, les scientifiques discrétisent souvent le domaine sphérique au moyen de collections finies de directions distribuées de manière *quasi uniforme*. En général, de tels domaines discrets ne permettent pas l'approximation d'opérateurs (pseudo-)différentiels par la méthode des différences finies. Cela rend donc les régularisations de type Tikhonov généralisée (gTikhonov) et variation totale généralisée (gTV) difficiles à mettre en oeuvre. Ces deux stratégies de régularisation ont tendance à produire des solutions aux propriétés physiques intéressantes, avec notamment des variations lisses pour gTikhonov et abrupte mais peu nombreuses pour gTV.

Pour pallier à ces limitations, des schémas de discrétisation canoniques utilisant des splines ont été proposés. L'optimalité de ces schémas a été montrée dans le cadre de la régularisation gTikhonov, pour des problèmes d'interpolation sphériques avec fonction de coût quadratique. Ce résultat est néanmoins trop restrictif dans la plupart des cas pratiques, car limité à des échantillons directionnels et des modèles de bruit Gaussiens. De plus, un résultat similaire pour la régularisation gTV reste à montrer, ce qui limite l'utilisation de cette dernière par les praticiens.

Dans cette thèse, nous proposons un cadre d'approximation sphérique unifié, à la fois théorique et pratique, pour les *problèmes fonctionnels inverses* définis sur l'hyperpshère $\mathbb{S}^{d-1}$. Plus spécifiquement, nous considérons la reconstruction de champs sphérique directement dans le domaine continu, faisant usage de problèmes d'optimisation convexes pénalisés au moyen de termes de régularisation gTikhonov et gTV. Notre formulation est compatible avec de nombreux types de mesures, ainsi qu'avec des fonctions de coût convexes et potentiellement non différentiables. Par l'intermédiaire de théorèmes de représentation inédits, nous caractérisons les solutions des problèmes de re-

construction pour les deux stratégies de régularisation. Pour la régularisation gTikhonov, nous montrons que la solution est unique et peut être écrite comme une combinaison linéaire des fonctionnelles d'échantillonnage –modélisant le processus d'acquisition– primitivées deux fois par rapport à l'opérateur pseudo-différentiel utilisé dans le terme de régularisation gTikhonov. Pour la régularisation gTV, nous montrons que les solutions peuvent être écrites comme combinaison convexes de splines sphériques dont les innovations sont en nombre inférieur au nombre de mesures disponibles.

Nous utilisons ces résultats pour construire deux schémas de discrétisation *canoniques* basés sur des splines, l'un exact pour la régularisation gTikhonov et l'autre approximatif mais avec erreur controlée pour la régularisation gTV. Nous proposons de plus des algorithmes proximaux *efficients* pour résoudre les problèmes d'optimisation discrets résultant de ces deux schémas de discrétisation. Pour illustrer la supériorité de ces méthodes d'approximation sphérique en domaine continu par rapport aux méthodes conventionnelles, nous les testons sur de nombreux jeux de données, à la fois simulés et réels, issus des domaines suivant: *météorologie*, *foresterie*, *radioastronomie* et *planétologie*. Les fonctionnelles d'échantillonnage, fonctions de coût et stratégies de régularisation considérés dans ces cas sont diverses et variés, montrant la versatilité de notre cadre théorique ainsi que de nos solutions algorithmiques.

Dans la dernière partie de cette thèse finalement, nous construisons un algorithme efficient et convergeant localement pour estimer les innovations spatiales de signaux périodiques à *taux d'innovation finis*, et proposons un réseau neuronal récurrent pour accélérer les méthodes d'approximation sphérique dans le cadre d'un problème d'imagerie acoustique en temps réel.

Mots clés: *problème inverse fonctionnel, approximation sphérique, splines sphériques, régularisation Tikhonov généralisée, régularisation par variation totale généralisée, théorèmes de représentation, reconstruction en domaine continu, discrétisation canonique, algorithmes proximaux, géographie mathématique, sciences de l'environ- nement, radioastronomie, planétologie, imagerie acoustique.*

Table of Contents

List of Figures, Tables & Algorithms

Figures

Tables

Algorithms

List of Definitions & Theorems

Definitions

Theorems

Propositions

Lemmas

Corollaries

Vocabulary

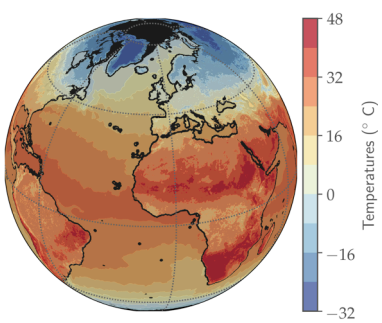
1 Introduction

1 Environmental Motivation

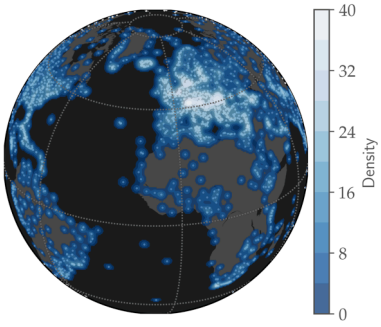
The human population has increased by 35% over the last 25 years, reaching an unprecedented size of 7 billion individuals. This ever-growing number exerts considerable stress on Earth's fragile ecosystem, pushed every year beyond capacity in an attempt to satisfy the needs of the population. In December 2017, W. Ripple et al. [149] published, with the support of more than 15'000 scientists from all over the world, an alarming manifesto declaring a state of ecological emergency and urging humanity to take a series of 13 concrete steps to ensure a sustainable future. Their manuscript studied the evolution of key environmental indicators over the past 25 years, and outlined troubling trends with potentially irreversible consequences on Earth's climate and biodiversity:

- Freshwater is becoming a scarce resource, with per capita availability reduced by 26%.
- Land use for agricultural and industrial production, cities, and transportation infrastructures has increased dramatically, with more than 1.2 million square kilometres of forest lost, a surface area equivalent to that of South Africa.
- Greenhouse gases emissions have doubled as a result of burning fossil fuels, hence accelerating the pace of global warming.
- The biosphere has entered the sixth mass extinction event in 540 million years, with a 29% drop in the population of vertebrate species, many of which are destined for extinction by the end of the century.

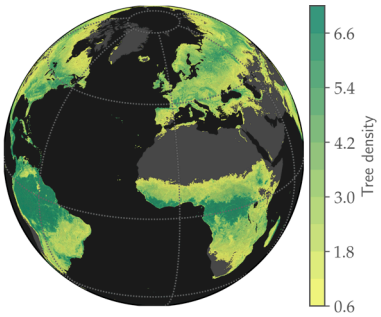
This unprecedented ecological crisis stresses the need for improved mathematical tools for Earth sciences –called *geomathematics* [60]– allowing scientists to monitor more accurately the planet's health. This task notably involves recording and inferring the temporal and spatial evolution of various natural phenomena occurring at a global scale, such as land and sea surface temperatures, natural hazards (droughts, floods...), or population dynamics of various life forms. Since the Earth's surface deviates from a sphere by less than 0.4% of its radius [60], such natural phenomena are often modelled as *spherical fields*, e.g. *functions*, *measures*, or *distributions* defined over a *continuum* of directions. A typical scientific inquiry consists then in approximating the spherical field modelling a particular phenomenon using finitely many *observations* of the



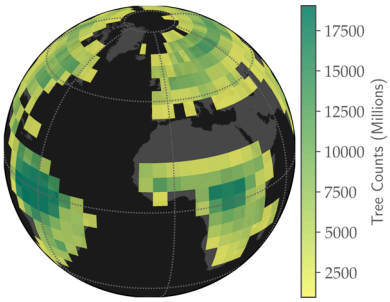
(a) Temperature distribution at the surface of the globe (illustration purposes only).



(b) Meteorological stations (dots), and corresponding density heatmap (shades of blue).



(c) Global tree density function at the surface of the globe (illustration purposes only).



(d) Tree counts in patches of size $5° \times 5°$ ($\simeq$ 300'000 km^2 at equator).

Figure 1.1: Examples of spherical approximation problems encountered in Earth sciences. Figs. 1.1a and 1.1b: the goal is to recover the unknown global temperature distribution Fig. 1.1a from temperature records of non-uniformly distributed weather stations in Fig. 1.1b. Figs. 1.1c and 1.1d: the goal is to recover the unknown tree density function Fig. 1.1c from low resolution tree counts in patches of size $5° \times 5°$ displayed in Fig. 1.1d. The illustrative temperature distribution in Fig. 1.1a was generated from average daytime surface temperatures for November 2016, obtained by thermal infrared measurements from NASA's satellites [171]. Similarly, the theoretical tree density distribution in Fig. 1.1c was constructed from the Leaf Area Index (LAI) map [169], estimating the density of the canopy from satellite observations. Finally, the weather station locations were obtained from OSCAR/Surface, WMO's official repository [196], and the tree counts simulated as independent Poisson realisations of an undersampled LAI map.

latter, collected by a linear *sensing device*.

For example, one may wish to recover the instantaneous temperature distribution at the surface of the Earth from various temperature records collected by weather stations scattered across the globe (see Figs. 1.1a and 1.1b). Mathematically speaking, this can be seen as an *interpolation* problem with *non-uniform spatial samples* – Fig. 1.1b shows that there is indeed a higher concentration of weather stations in western countries than in developing countries or over the oceans. Similarly, the distribution of trees at the surface of the globe is often modelled as a *spatial point process* [144, 161, 172]. Global tree density maps [45] –crucial in forestry to fight deforestation and illegal

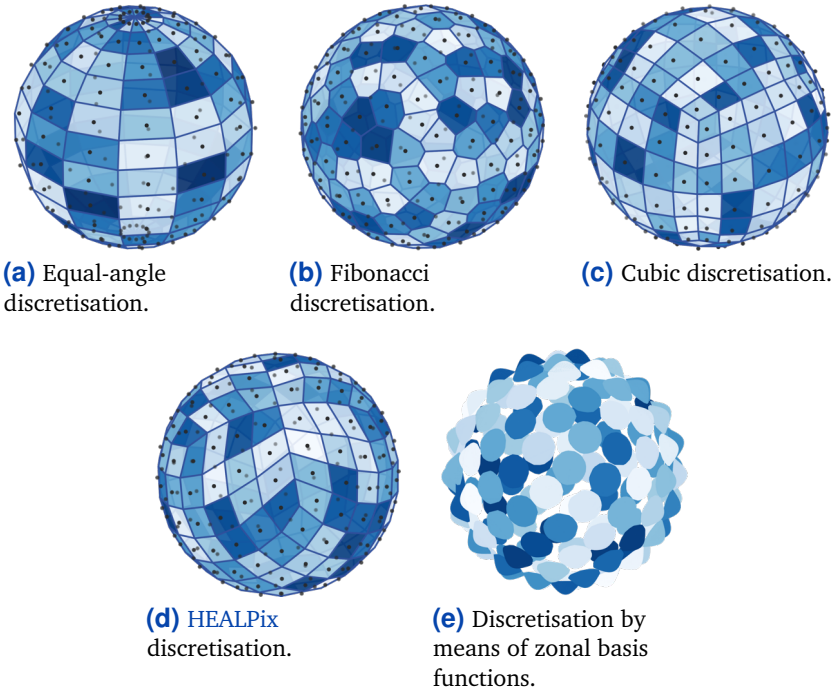
logging– are then obtained by estimating the *intensity function* of the point process (see Fig. 1.1c for an illustration). In this case however, the data no longer consists of spatial samples as in the previous meteorological example. Indeed, tree density maps are typically constructed from global surveys that count the number of trees in large geographical zones tiling the globe, either on the ground with human agents or remotely by analysing satellite images. These surveys are of course very expensive and tedious to conduct, and hence have necessarily limited accuracy (see Fig. 1.1d). Obtaining high-resolution density maps as in Fig. 1.1c from such datasets thus requires the use of advanced spherical approximation methods such as those proposed in this thesis.

2 Spherical Approximation: An Overview

Many scientific inquiries in natural sciences, such as environmental and planetary sciences [45, 101, 195], acoustics [142] or astronomy [120, 133, 163], involve approximating a *spherical field* –a scalar quantity such as a function or measure defined over a *continuum* of directions, from a finite number of measurements acquired by probing sensors. During the reconstruction task, the physical evidence is compared to some prior model of the unknown spherical field, reflecting the analyst's a priori beliefs about the latter. In practice, a trade-off between fidelity to the data and compliance with this prior is assessed via a composite convex optimisation problem, linear combination of a *cost functional* and a *regularisation* term. Popular regularisation strategies include generalised Tikhonov (gTikhonov) or generalised total variation (gTV) [72], which favour physically admissible spherical fields with *smooth* and *sharp* variations respectively. Since spherical fields encountered in nature are continuous and hence have infinitely many degrees of freedom, scientists often constrain the approximation problem using *discretisation schemes*, which help reducing the number of degrees of freedom to something more manageable, ideally comparable to the size of the available data. For Euclidean domains, it is for example common practice to approximate the continuum by means of discrete *uniformly distributed point sets*, typically forming regular rectangular grids.[1] The popularity of such domain discretisation schemes can be primarily explained by their simplicity and computational conveniency. Indeed, signals defined over rectangular grids admit a natural representation as multi-dimensional arrays, a data structure commonly used in computer science for computation, storage and visualisation purposes.

[1] This discretisation scheme is sometimes called *pixelisation* in visual computing.

Unfortunately, the sphere manifold structure makes it much more difficult to discretise by means of uniformly distributed point sets. For example, points gridded regularly on the azimuth-elevation domain $[0, 2\pi] \times [-\pi/2, \pi/2]$ are highly non-uniformly distributed at the surface of the sphere, with a much higher concentration of points near the poles (see Fig. 1.2a). As a matter of fact, uniform spherical point sets are only known [142, Chapter 3] for fixed numbers of points: 4, 6, 8, 12 and 20. They are respectively obtained from the vertices of the five Platonic solids: the *tetrahedron*, *cube*, *octahedron*, *dodecahedron* and *icosahedron*. For arbitrary numbers of points, spherical point sets with *quasi-uniform* distribution have been proposed [67, 78, 142]. The



(a) Equal-angle discretisation.

(b) Fibonacci discretisation.

(c) Cubic discretisation.

(d) HEALPix discretisation.

(e) Discretisation by means of zonal basis functions.

Figure 1.2: Examples of discretisation schemes on the sphere, with an approximate resolution of 200 for each scheme. Figs. 1.2a to 1.2d show examples of non-uniform (Fig. 1.2a) and quasi-uniform (Figs. 1.2b to 1.2d) spherical point sets (marked by black dots). Fig. 1.2e on the other hand, shows an example of parametric discretisation by means of bell-shaped zonal basis functions. The equal-angle point set in Fig. 1.2a is obtained by gridding the azimuth-elevation domain. The point set in Fig. 1.2b is called the Fibonacci lattice, and can be generated as explained in [66]. The point sets in Figs. 1.2c and 1.2d are obtained from the cell centroids of the cubic and HEALPix spherical tessellations respectively. The cubic tessellation is obtained by projecting the pixelated faces of a cube onto the sphere. The HEALPix tessellation, very popular in cosmology and astronomy, is constructed by hierarchically subdividing the faces of a dodecahedron [67].

spherical Voronoi diagrams of the latter typically tile the sphere with near-regular polygonal tiles,[2] see for example Fig. 1.2. Unfortunately, quasi-uniform spherical point sets are significantly more complicated to work with as they are not easily represented by array-like data-structures. Moreover, derivatives and more generally pseudo-differential operators are difficult to approximate on quasi-uniform spherical point sets [30, 48], making it cumbersome to work with gTikhonov and gTV priors.

The difficulty in designing domain discretisation schemes for the sphere has led scholars to consider alternative *parametric* discretisation schemes, where the unknown spherical field is constrained to a finite dimensional functional space, typically spanned by *zonal basis functions*[3] [61, 65, 79, 91, 126], i.e. functions with rotational invariance around a particular central direction on the sphere. The majority of zonal basis functions used in practice take the form of positive and *smooth* bell-shaped functions, sharply decaying to zero as the angular distance from their central direction increases (see Fig. 1.2e). They possess moreover many useful properties, particularly convenient for practical

[2] In practice, quasi-uniform spherical point sets are actually often constructed from centroids of spherical tessellation cells.

[3] Zonal basis functions are the spherical analogs of *radial basis functions* [201], used for scattered data interpolation in Euclidean settings.

purposes:

- They are identical, spatially localised and highly symmetric, and thus easy to evaluate and amenable to sparse, parallel computations.
- Their overlapping supports and strong regularity make them well-suited for approximating smooth natural phenomena.
- Their centres can be positioned arbitrarily at the surface of the sphere, permitting for example the concentration of more zonal basis functions in regions more susceptible of welcoming the signal.
- *Strictly positive definite* zonal basis functions [65] are all linearly independent, irrespective of the chosen centres [14]. This guarantees a non-redundant representation and limits the risk of numerical instability.
- They are particularly well-suited for *scattered data interpolation problems* [61, 79, 126] where the spatial samples to interpolate may be non-uniformly distributed.

This last fact is probably the main reason for their wide adoption in the literature. As a matter of fact, some zonal basis functions are not merely well-suited but *canonical* to spherical scattered data interpolation. This is notably the case for a specific type of zonal functions, called *spherical splines* [125], which arise naturally as solutions of interpolation problems on the sphere. As an illustration, consider the simplest interpolation problem where one wishes to find all *maximally smooth* functions with prescribed values $\{y_1, \dots, y_L\} \subset \mathbb{R}$ at directions $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^2$. The relevant notion of smoothness is of course application dependent, but is generally enforced by seeking an interpolant with minimal generalised Tikhonov (gTikhonov) norm, induced by some linear, self-adjoint and strictly positive definite *pseudo-differential operator* $\mathscr{D}$. In mathematical terms, the prototypical interpolation problem can be formulated as:

$$\mathcal{V} = \underset{f \in \mathscr{H}_{\mathscr{D}}}{\arg\min} \left\{ \|\mathscr{D} f\|_2 \quad \text{such that} \quad f(\boldsymbol{r}_i) = y_i,\ i = 1, \dots, L \right\}, \tag{1.1}$$

where the search space $\mathscr{H}_{\mathscr{D}}$ is an appropriately chosen reproducing kernel Hilbert space (RKHS) so that all the quantities involved in (1.1) are well-defined. It is then possible to show that there exists a *unique* maximally smooth interpolant $\mathcal{V} = \{f^\star\}$, and that the latter has exactly L degrees of freedom. Moreover, the maximally smooth interpolant $f^\star$ can be expressed [125, Section 6.3] as a spherical spline[4] with *knots* coinciding with the sampling directions $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^2$. The L spline weights can moreover be recovered by solving a square linear system [125, Section 6.3]. This result is quite remarkable, since it provides us with a *canonical* discretisation scheme operating in a *lossless* fashion: the infinite-dimensional optimisation problem (1.1) is transformed into an *equivalent* finite-dimensional optimisation problem, amenable to numerical optimisation. Theorem 6.40 in [125, Section 6.4.2] generalises this result to the *smoothing spline approximation* problem:

[4] With respect to the operator $\mathscr{D}^2$.

$$\mathcal{V} = \underset{f \in \mathscr{H}_{\mathscr{D}}}{\arg\min} \left\{ \sum_{i=1}^{L} |y_i - f(\boldsymbol{r}_i)|^2 + \lambda \|\mathscr{D} f\|_2^2 \right\}, \tag{1.2}$$

used as an alternative to (1.1) in the context of noisy spatial samples, since it is less prone to overfitting.

Unfortunately, both problems (1.1) and (1.2) are too restrictive for most spherical approximation tasks encountered in practice. This is for example the case when the measurements are corrupted by non Gaussian noise –hence requiring a more general cost functional than a quadratic one– or do not consist in directional samples of the spherical field, but rather in local averages or more generally filtrations of the latter. In addition, gTikhonov-regularised optimisation problems *à la* (1.1) and (1.2) suffer from two main drawbacks: they tend to produce *overly smooth* interpolants and are too sensitive to the sampling locations $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^2$. This is a general behaviour of smooth spline approximation even in the Euclidean setting [72], which can be explained in part by the fact that the gTikhonov regularising norm is a weighted $\mathscr{L}^2$ norm. The latter favours indeed functions with relatively *smooth* variations, rendering the spline interpolant incapable of adapting to rapid changes in the data. To overcome this limitation scholars have, motivated by empirical studies [185], advocated the use of generalised total variation (gTV) regularisation norms, promoting functions with *relatively sparse* but potentially *sharp* variations, as often encountered in natural phenomena. However, representer theorems [28, 29, 57, 72, 179] characterising the form of the solutions yielded by the use of such regularisation strategies are, to date, unadapted to spherical geometries (see Section 5 for a discussion on prior art). For this reason, and in contrast with most fields of signal processing, total variation based penalties are still very much unexplored in spherical setups. As explained in greater detail in Section 3, one of the goals of this thesis is to close this theoretical gap and promote the wider use of such recovery methods across the community of geomathematicians.

3 Contributions of this Thesis

A primary goal of this thesis is to offer a unified theoretical and practical approximation framework for gTikhonov and gTV regularised infinite dimensional inference on the hypersphere $\mathbb{S}^{d-1}$ of arbitrary dimension $d \geq 2$. Out of concern for making the content of this thesis accessible to a wider audience, care has been taken in thoroughly interpreting and analysing the stated results and their assumptions, with a particular focus on their practical implications. Multiple real-life examples from environmental sciences and radio astronomy are moreover considered. The main contributions of this thesis are listed hereafter and, in agreement with the subtitle of this thesis, classified into three categories: *theory*, *algorithms* and *applications*. A partial summary of the contributions is also available in Table 1.1. For simplicity of exposition, our theoretical results are stated here in the particular case of *spline-admissible*, *self-adjoint* and *invertible* pseudo-differential operators $\mathscr{D}$ (see Chapter 4 for details).

Regularisation		gTikhonov	gTV
Representer Theorems	Abstract	Theorem 5.3, Corollary 5.6	Theorem 2.12
	Continuous Domain	Theorem 5.3	Theorem 5.4
	Discrete Domain	Theorem 6.7	Theorem 6.8
Generalised Sampling		Proposition 5.1, Lemma 5.5	Propositions 5.2 and A.1, Lemma 5.7
Discretisation		Theorem 6.2 (exact)	Theorem 6.5 (approx.)
Algorithms	PDS	Algorithms 7.3 and 7.7	Algorithms 7.5, 7.9 and 7.10
	APGD	Algorithms 7.4 and 7.8	Algorithm 7.6
Applications (Chapter 9)		Meteorology 1, Planetary Sciences 4	Meteorology 1, Forestry 2, Radio Astronomy 3

Table 1.1: Summary of the various contributions of this thesis (Chapters 10 and 11 excluded).

3.1 Theory

3.1.1 Representer Theorems

The main theoretical contributions of this thesis are *representer theorems*. The latter characterise the forms and degrees of freedom of the solutions to convex optimisation problems formulated over various duality pairs of finite and infinite dimensional Hilbert or Banach spaces. The novel representer theorems established in this thesis are discussed hereafter in order of appearance.

(a) Abstract Representer Theorems

In Chapter 2, we establish abstract representer theorems pertaining to penalised convex optimisation in abstract Hilbert and Banach spaces. The latter serve as foundations to the various representer theorems of Chapter 5 for functional inverse problems on the sphere with gTikhonov and gTV regularisation.

First, Theorem 2.9 considers a generic penalised convex optimisation problem formulated over the topological dual $\mathscr{B}'$ of some Banach space $\mathscr{B}$. The penalised optimisation problem takes the form of a sum between a *convex* cost functional and a *strictly convex* penalty term, related to the (strictly convex) *dual norm* on $\mathscr{B}'$ via a (strictly) *convex and strictly increasing* function. The measurement process is moreover modelled by means of linear sampling functionals in the predual $\mathscr{B}$ of $\mathscr{B}'$. This penalised optimisation problem is shown to admit a *unique* solution, related to a certain linear combination of the sampling linear functionals via the isometric *duality map* between $\mathscr{B}$ and $\mathscr{B}'$. This result is based on [177, Theorem 5], which is established on a different set of assumptions,[5] slightly more restrictive for practical purposes.[6] In the Hilbert case, we choose $\mathscr{B} = \mathscr{H}''$ and $\mathscr{B}' = \mathscr{H}$ and obtain Corollary 2.10, which shows that the unique solution of the penalised optimisation problem is the image by the *linear* Riesz map of a certain linear combination of the sampling linear functionals. The proofs of the spherical representer theorems from Chapter 5 pertaining to gTikhonov regularisation are all based on Corollary 2.10, which can be seen as an abstract formulation of the latter.

[5] Theorem 5 of [177] assumes a *strictly convex* cost functional and *convex* penalty term.

[6] Indeed, some of the cost functionals commonly used in practice are *not* strictly convex.

In Theorem 2.12 then, we relax the assumption of strict convexity on the penalty term. More precisely, we consider a *convex* cost functional and penalty term as well as a *non strictly convex* search space $\mathscr{B}'$. In this case, we show that the solution set of the penalised optimisation problem is *nonempty* and the (weak*) *closed convex hull of extreme points*, taking the form of linear combinations of at most L extreme points of the regularisation ball, where L denotes the number of available linear measurements. The proof of this result is based on [177, Theorem 5], [72, Proposition 8] and [28, Theorem 3.1]. Note that, in the case where the cost functional is strictly convex, it is also possible to invoke [177, Theorem 5] and characterise the solution set in terms of the duality map as before. Such a characterisation is however much less practical, since the duality map for a *non strictly convex* search space $\mathscr{B}'$ is set-valued, and often unknown. The proofs of the spherical representer theorems from Chapter 5 pertaining to gTV regularisation are all based on Theorem 2.12, which can be seen as an abstract formulation of the latter.

(b) Continuous-Domain Representer Theorems for Spherical Approximation

In Chapter 5, we propose two representer theorems –Theorems 5.3 and 5.4, for functional penalised Tikhonov (FPT) and functional penalised basis pursuit (FPBP) problems respectively. These two classes of optimisation problems are envisioned for solving spherical approximation problems as in Section 1, formulated as *functional linear inverse problems* on the hypersphere $\mathbb{S}^{d-1}$. FPT and FPBP problems essentially seek spherical functions, measures or distributions minimising an optimal trade-off between a *convex* cost functional and a gTikhonov or gTV penalty term respectively. To accommodate various measurements types, both classes of optimisation problems are built upon a convenient *generalised sampling framework*, modelling the measurement process in terms of sampling linear functionals.

First, Theorem 5.3 shows that FPT problems admit a *unique* solution with finite degrees of freedom. This solution is moreover given by a linear combination of the sampling linear functionals primitived twice w.r.t. the invertible pseudo-differential operator $\mathscr{D}$ involved in the definition of the gTikhonov regularisation term. In the specific case of directional sampling, we show in Corollary 5.6 that the solution can be characterised as a spherical $\mathscr{D}^2$-spline, with knots coinciding with the sampling directions. Theorem 5.3 can hence be thought as a generalisation to arbitrary convex cost functionals and measurement types of the classical spherical approximation results [125, Theorems 6.30 and 6.40] discussed in Section 2.

Theorem 5.4 then, shows that the solution sets of FPBP problems can be characterised geometrically as the (weak*) *closed convex hull* of their extreme points. These extreme points take moreover the form of spherical $\mathscr{D}$-splines with *sparse* innovations –i.e. fewer degrees of freedom than the total number of measurements– and unknown knots. Corollary 5.8 specifies this result to the specific case of directional interpolation.

Note that Theorems 5.3 and 5.4 are both formulated under the assumption of an *invertible* pseudo-differential operator $\mathscr{D}$ in the gTikhonov and gTV regularisation terms respectively. With this assumption, Theorems 5.3 and 5.4

can indeed be obtained as corollaries of Corollary 2.10 and Theorem 2.12 respectively. In practice however, some pseudo-differential operators, such as the *Laplace-Beltrami operator* $\Delta_{\mathbb{S}^{d-1}}$, are non-invertible. In Remarks 5.5 and 5.8 we show how such operators can be brought into the scope of Theorems 5.3 and 5.4 respectively, if properly regularised on their nullspaces.

(c) Discrete-Domain Representer Theorems for Spherical Approximation
In Section 2 of Chapter 6, we propose two representer theorems –Theorems 6.7 and 6.8 – for FPT and FPBP problems over discrete domains taking the form of *equidistributed spherical point sets*. The conclusions of both theorems are analogous to those of their continuous-domain counterparts, Theorems 5.3 and 5.4 respectively.

3.1.2 Continuous Sampling Functionals

In order for FPT and FPBP problems to be well-posed, the sampling linear functionals modelling the acquisition system must be *continuous* w.r.t. the particular topology[7] chosen over the search space. In other words, the sampling linear functionals must belong to the *topological dual* of the search space.

In the case of gTikhonov regularisation, we characterise in Proposition 5.1 the topological dual of the search space in terms of the regularising pseudo-differential operator $\mathscr{D}$. In Lemma 5.5, we consider the special case of directional sampling and provide a sufficient condition on the spectral growth order of $\mathscr{D}$ so that all Dirac measures are included in this topological dual.

In the case of gTV regularisation, the topological dual of the search space is more difficult to characterise entirely. Since the search space of an FPBP problem is itself a topological dual, it is however possible to characterise the predual of the search space, which can be embedded in the topological dual of the search space if the latter is equipped with the weak* topology.[8] We do so in Proposition 5.2, therefore characterising a subset of sampling linear functionals compatible with a particular gTV regularisation norm. Similarly to FPT problems, we moreover provide in Lemma 5.7 and Proposition A.1 sufficient conditions on the spectral growth order of the regularising pseudo-differential operator $\mathscr{D}$ such that all Dirac measures and square-integrable functions are included in the predual.

[7] For FPT problems, the search space is equipped with its canonical Hilbert topology, while for FPBP problems the search space is equipped with the weak* topology defined in Section 1.3 of Chapter 2.

[8] Again, see Section 1.3 of Chapter 2 for more details.

3.2 Practical Aspects & Algorithms

3.2.1 Canonical Search Space Discretisation Schemes

In Chapter 6, we leverage Theorems 5.3 and 5.4 from Chapter 5 in order to derive *canonical search space discretisation schemes* for gTikhonov and gTV regularisation respectively. Such schemes convert FPT and FPBP problems into simple finite dimensional optimisation problems –amenable to numerical optimisation– by restricting their search spaces to well chosen finite dimensional subspaces.

For FPT problems, we propose first to restrict the search space to the span of the sampling linear functionals primitived twice w.r.t. the regularising pseudo-differential operator. This discretisation, suggested by Theorem 5.3, is shown in Theorem 6.2 to be *exact*, in the sense that the resulting finite dimensional

optimisation problem is strictly equivalent to the original infinite dimensional FPT problem.

For FPBP problems, we then propose a search space discretisation scheme based on *quasi-uniform* $\mathscr{D}$-splines, i.e. splines whose knot sets form *quasi-uniform spherical point sets* [78]. This discretisation scheme is motivated by Proposition 6.4, which shows that, under mild conditions on the pseudo-differential operator $\mathscr{D}$, *quasi-uniform* $\mathscr{D}$-splines can approximate arbitrarily well most solutions of FPBP problems[9] when the number of knots tends to infinity. In Theorem 6.5 finally, we show that the finite dimensional problem resulting from this search space discretisation is a classical penalised basis pursuit (PBP) problem [178].

[9] Of course with gTV regularisation norms induced by $\mathscr{D}$.

3.2.2 Algorithms for Spherical Approximation

In Chapter 7, we propose to solve the various discrete optimisation problems from Chapter 6 by means of provably convergent *fully-split proximal iterative methods* [134], which only involve simple matrix-vector multiplications and *proximal steps*. We treat the most general case where the cost function is proximable but not necessarily differentiable with the *primal-dual splitting method (PDS)* introduced by Condat in his seminal work [43]. In the simpler (yet prevailing in practice) case where the cost functional is also *differentiable* and with *Lipschitz continuous* derivative, we leverage an optimal first-order method called *accelerated proximal gradient descent (APGD)* [16, 134], with faster convergence rate than the PDS method. Table 7.2 page 133 summarises the various algorithms presented in Chapter 7.

3.2.3 Practical Spherical Splines

Chapter 8 discusses suitable choices of pseudo-differential and spherical splines for practical purposes. We recommend the use of *Wendland* and *Matérn* pseudo-differential operators, whose Green kernels have simple closed-form expressions and good localisation in space. The latter have moreover spectra equivalent to those of *Sobolev operators* $\mathscr{D}_\beta = [\mathrm{Id} - \Delta_{\mathbb{S}^{d-1}}]^\beta$, $\beta > (d-1)/2$, often used in functional analysis.

3.2.4 Cadzow Plug-and-Play Gradient Descent

In Chapter 10,[10] we introduce a non-convex optimisation algorithm, baptised Cadzow plug-and-play gradient descent (CPGD), which estimates the spatial innovations of a *periodic Dirac stream* with *finite rate of innovation* [25] from generalised measurements of the latter. The algorithm is *extremely simple* and *very efficient*, outperforming the state-of-the-art algorithm proposed for this task in [130]. Unlike the latter, CPGD is moreover provably *locally convergent*. This algorithm could notably be used for the purpose of estimating extreme point solutions to FPBP problems formulated over the circle $\mathbb{S}^1$. It has however much wider applicability.

[10] The material presented in Chapter 10 is the result of joint work with A. Besson, P. Hurley and M. Vetterli, and is the topic of [162], currently under submission.

3.3 Applications

To demonstrate the versatility of our spherical approximation framework, we put it to the test in Chapter 9 on a mix of real and simulated data originating

from a variety of real-life spherical approximation problems encountered in *environmental sciences*, *radio astronomy* and *planetary sciences*.[11] A summary of all experiments investigated in this chapter is available in Table 9.1 page 146. Finally in Chapter 11 we show, in the context of *acoustic imaging*, how the convergence speed of the proximal algorithms from Chapter 7 can be "boosted" by means of recurrent neural-networks, notably for the purpose of real-time imaging.

[11] Interactive versions of the spherical maps produced in Chapter 9 are available at the following link: matthieumeo.github.io

3.3.1 Environmental Sciences

(a) Meteorology

In Section 1 of Chapter 9, we reconstruct a global map of sea surface temperature anomalies from recordings collected by drifting floats of the ARGO fleet [7, 98]. Such maps are used in environmental sciences to monitor global climate change as well as manage the population of marine species and ecosystems particularly sensitive to fluctuations in the water temperature. We compare in this context continuous-domain and discrete-domain recovery methods, for both choices of regularisation gTikhonov and gTV.

(b) Forestry

In Section 2.3 of Chapter 9, we build global density maps of trees and wildfires across the globe for the year 2016, using tree and fire counts recorded by NASA's Aqua and Terra satellites. Tree density maps are used in environmental sciences to monitor deforestation and illegal logging, as well as assess the amount of vegetal photosynthesis. Similarly, wildfire maps allow scientists to better understand atmospheric chemistry and its impact on climate. Because of the Poisson-like distribution of count data, we investigate the use of KL-*divergence* as a cost functional.

3.3.2 Radio Astronomy

In Section 3 of Chapter 9, we propose recovering full-sky intensity maps from the raw-measurements of radio telescopes such as Planck [2]. Such maps display the intensity (or equivalently the temperature) of every astronomical radio source across the celestial sphere. Using an FPBP problem with KL-*divergence* as cost functional, we obtain sky intensity maps with far greater resolution than the standard *dirty maps* outputted by radio telescopes.

3.3.3 Planetary Sciences

In Section 4 of Chapter 9, we build global distribution maps of radioactive elements at the surface of the Moon, using actual data collected by NASA's *Lunar Prospector (LP)* probe [101, 103]. Such maps, called *elemental abundance maps*, are used by scientists to retrace the Moon's geologic history [101]. The reconstruction is performed by means of an FPT problem with ℓ_2-ball cost function. For comparison purposes, we also provide the abundance map obtained with the state-of-the-art *Pixon method* [141], reproduced from [195].

3.3.4 Real-Time Acoustic Imaging

[12] The material presented in Chapter 11 is the result of joint work with S. Kashani, P. Hurley and M. Vetterli, and is the topic of [166].

In Chapter 11,[12] we propose a recurrent neural network for real-time reconstruction of acoustic camera spherical maps. The network, dubbed DeepWave, is both physically and algorithmically motivated: its recurrent architecture mimics the iterations from proximal algorithms, and its parsimonious parametrisation is based on the natural structure of acoustic imaging problems. Each network layer applies successive filtering, biasing and activation steps to its input, which can be interpreted as generalised *deblurring* and *sparsification* steps. Our real-data experiments show DeepWave has similar computational speed to the state-of-the-art delay and sum (DAS) imager with vastly superior resolution. While developed primarily for acoustic cameras, DeepWave could easily be adapted to other applications with real-time imaging requirements.

4 Organisation of this Thesis

This thesis is organised as follows. In Part I, we lay down the theoretical foundations of our functional approximation framework:

- In Section 1 of Chapter 2, we introduce the concept of *duality* in topological vector spaces, central to the generalised sampling framework used in Chapter 5 to model linear measurements of spherical fields.
- In Section 2 of Chapter 2, we establish abstract representer theorems characterising the form of the solutions to a variety of penalised convex optimisation problems defined over Hilbert or Banach spaces.
- In Chapter 3, we review important notions from Fourier analysis on the hypersphere, namely *spherical harmonics* and *spherical zonal functions*.
- In Chapter 4, we use these notions of harmonic analysis to define *hyperspherical splines*, which can be thought of as primitives –w.r.t. a certain pseudo-differential operator– of Dirac streams with finite innovations.

In Part II, we use the various concepts from Part I to perform infinite dimensional inference on the hypersphere:

- In Section 1 of Chapter 5, we introduce functional inverse problems on the hypersphere.
- In Section 2 of Chapter 5, we consider regularising functional inverse problems by means of generalised Tikhonov (gTikhonov) and generalised total variation (gTV) norms.
- In Section 3 of Chapter 5, we establish representer theorems for FPT and FPBP problems, which are functional inverse problems with gTikhonov and gTV regularisation respectively.
- In Chapter 6, we use Theorems 5.3 and 5.4 to design two canonical *search space discretisation schemes*, *exact* for gTikhonov regularisation and with vanishing approximation error for gTV regularisation. We moreover investigate and compare ourselves to alternative *domain discretisation schemes*, traditionally favoured by practitioners.

In Part III, we discuss practical aspects of the spherical approximation framework from Part II:

- In Chapter 7, we design efficient and provably convergent proximal algorithms for all discrete optimisation problems considered in Chapter 6.

We moreover propose rules of thumb for setting their various hyperparameters and provide the proximal operators of most common cost functionals.
- In Chapter 8, we discuss the use of Wendland and Matérn pseudo-differential operators when designing gTikhonov and gTV penalties, and their convenient properties for practical purposes.
- In Chapter 9, we test the spherical approximation framework from Part II and novel algorithms from Chapter 7 on a variety of real and simulated datasets, coming from the fields of meteorology, forestry, radio astronomy and planetary sciences. The sampling functionals, cost functionals and regularisation strategies considered in each case are very diverse, showing the versatility of both our theoretical framework and algorithmic solutions.

In Part IV, we discuss further topics and conclude:
- In Chapter 10, we design Cadzow plug-and-play gradient descent (CPGD), an efficient and locally convergent algorithm for recovering the spatial innovations of periodic Dirac streams with finite rates of innovation. This algorithm is envisioned as an alternative to the quasi-uniform spline discretisation scheme proposed in Chapter 6 for gTV regularised functional inverse problems.
- In Chapter 11, we design the DeepWave RNN for purposes of real-time acoustic imaging.
- In Chapter 12, we reflect back on the trajectory of this thesis and outline a few prospective research avenues building on top of the material of this work.

In Part V finally, we provide in Appendices A to C supplementary material to Chapters 9 to 11 respectively, including proofs, additional experiments and results.

5 Representer Theorems in the Literature

In this section, we review the most notable representer theorems proposed in the literature for gTikhonov and gTV regularisation, and discuss their limitations in the context of spherical approximation. A summary of this section is provided in Table 1.3 page 18.

5.1 gTikhonov Regularisation

Representer theorems for gTikhonov regularisation are better known in the context of statistical learning over reproducing kernel Hilbert spaces [93, 156]. Their origin can be traced back to the pioneering work of Kimeldorf and Wahba, who characterised in [93] the solutions of gTikhonov-penalised least-squares problems *à la* (1.2) in generic RKHSs. This result notably gave rise to smoothing spline representer theorems in non-parametric regression [69], such as [125, Theorems 6.30 and 6.40] in the spherical setup. Originally limited to quadratic cost functionals, the work of Kimeldorf and Wahba was later extended to arbitrary cost functionals and more general gTikhonov-based penal-

ties by Schölkopf, Herbrich, and Smola in [156]. More recently, Unser, Fageot et al. considered in a series of publications [10, 72] more general forms of measurements than spatial samples –not covered by the RKHS setup assumed in the previously cited works. More specifically, Gupta, Fageot and Unser considered in [72] gTikhonov-penalised convex optimisation problems for solving functional inverse problems over $\mathbb{R}^d$. Just as in Theorem 5.3 of this thesis, they considered arbitrary convex cost functionals and linear measurements. Unlike Theorem 5.3 however, they do not assume that the gTikhonov pseudo-differential operator is invertible, and allow it to have a finite dimensional nullspace. The same setup is investigated by Badoual, Fageot and Unser in [10], in the context of functional inverse problems involving *periodic* functions –i.e. with domains isomorphic to $\mathbb{S}^1$. In both cases, the conclusions of their representer theorems are very similar to those of Theorem 5.3 derived in this thesis. Unfortunately, both results were shown for the domains $\mathbb{R}^d$ and $\mathbb{S}^1$ *exclusively*, and do not cover hyperspherical domains $\mathbb{S}^{d-1}$ for $d \geq 3$. Note that the representer theorems proposed in [10, 72] were derived from the classical Hilbert result [191, Theorem 16.1]. As explained in Remark 2.3, Corollary 2.10 used in this thesis to derive Theorem 5.3 can be seen as a generalisation of [191, Theorem 16.1] to more general cost functionals than indicator functions of convex sets.

5.2 gTV Regularisation

Inspired by the pioneering work of Fisher and Jerome in [56], Unser et al. have investigated in a series of papers [46, 72, 179] gTV-regularised functional inverse problems over $\mathbb{R}^d$. As here, they consider a generalised sampling framework, compatible with a great variety of linear measurements. Similarly to the gTikhonov case, they moreover consider gTV pseudo-differential operators with finite dimensional nullspaces.

The conclusions of their representer theorem in [179] –derived for Euclidean domains only– are analogous to the ones of Theorem 5.4, proposed in this thesis for spherical domains. In subsequent publications [46, 72], the authors proposed canonical discretisation schemes as well as numerical algorithms for approximating extreme point solutions of gTV-penalised convex optimisation problems. In [72], they consider a discretisation based on cardinal splines of the gTV pseudo-differential operator, with uniform knots chosen over a dense grid. In [46], they propose a numerically stabler discretisation scheme, based this time on *multi-resolution B-splines* with refinable grid sizes. In both cases they solve the resulting discrete optimisation problem with a two-stage procedure leveraging *proximal gradient descent* and the *simplex* algorithm. In contrast with the proximal algorithms proposed in this thesis, this optimisation procedure is however limited to differentiable cost functionals with Lipschitz-continuous derivatives. While remarkably generic, their spline approximation framework is only valid for functions defined over $\mathbb{R}^d$, and cannot be used in the spherical setting.

In a subsequent work [57], Flinth et al. proposed an alternative proof of the representer theorem proposed in [179]. Their proof is based on a limit

argument, considering nested finite dimensional discretisations of the domain $\Omega \subset \mathbb{R}^d$ based on finer and finer uniform rectangular grids. They claim that such an approach, presented in the Euclidean case for the sake of simplicity, could easily be adapted to domains more general than $\mathbb{R}^d$, such as the torus or any separable, locally compact topological space. They specify however that such an extension would require modifying adequately the discretisation scheme to the specific geometry of the domain, without giving additional details on how this could be achieved canonically. Unfortunately, such a task may be very complex if even possible at all for geometries such as the sphere. Indeed, discretising the sphere by means of nested quasi-uniform point sets with finer and finer resolution is, as previously discussed, a nontrivial problem.

More recently, Boyer et al. [28] and Bredies et al. [29] have independently shown that the solutions to infinite dimensional optimisation problems with convex regularisers are convex combinations of extreme points of the regulariser level sets. This result applies notably to gTV regularisers with not only scalar but also *vector* pseudo-differential operators such as the gradient. This is in contrast with the previously cited works which were all limited to scalar pseudo-differential operators such as the Laplace-Beltrami operator. While theoretically applicable to spherical geometries, their result neither addresses existence conditions nor characterises the minimal search space (and its corresponding predual) associated to a certain gTV norm. This is problematic for practical purposes, where it is crucial to know if a given optimisation problem admits a solution or understand which sampling linear functionals are compatible with a specific choice of gTV penalty.

Finally, Unser [177] established a Banach representer theorem with very broad applicability. Unlike the previously cited results, this representer theorem relies on the notion of *duality map*, which generalises the Hilbert notion of *Riesz map* to Banach spaces. More precisely, it shows that the solutions of convex regularised inverse problems are contained in the image by a certain duality map of a linear combination of the sensing linear functionals. As acknowledged by the author, this result is however of limited use in the context of gTV regularisation, since the duality map is unknown, nonlinear and set-valued.

6 Notation and Terminology

Throughout the manuscript, we adopt the following conventions:

- We use the term *spherical field* to refer, depending on the context, to *functions*, *measures* or *generalised functions* [180] defined over the sphere $\mathbb{S}^{d-1}$ for any dimension[13] $d \geq 2$. In full generality, one shall think at a spherical field as an element of some *infinite-dimensional* Banach space $\nu \in \mathscr{B}$.
- It is traditional to call the 1-sphere $\mathbb{S}^1 \subset \mathbb{R}^2$ a *circle*, the 2-sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ a *sphere* and the $(d-1)$-sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$, $d \geq 2$ a *hypersphere*. For the sake of simplicity, we break with tradition and use the appellation "sphere" agnostic to the underlying dimension. Moreover, we denote by $\mathfrak{a}_d$ the *area* of the unit sphere $\mathbb{S}^{d-1}$, $d \geq 2$, given in general by: $\mathfrak{a}_d =$

[13] Of course the cases $d = 2, 3$ will be particularly prevailing in real-life applications.

Sets ($\mathcal{V}$ denotes a set)	
$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}_+, \mathbb{C}$	Positive integers, integers, real numbers, positive real numbers and complex numbers.
$\overline{\mathbb{R}}$	Extended real line: $\mathbb{R} \cup \{-\infty, +\infty\}$.
$[\![1, L]\!]$	$\{1, \dots, L\}$
$\emptyset, B, \mathscr{B}$	Empty set, Borel set, Borel algebra.
$\delta\mathcal{V}, \mathcal{V}^\circ$	Sets of extreme points and interior points of $\mathcal{V}$ convex.
$\mathscr{P}(\mathcal{V}), \overline{\mathcal{V}}^{\mathcal{T}}, \#\mathcal{V}$	Powerset, closure w.r.t. the topology $\mathcal{T}$, and cardinality of $\mathcal{V}$.
Functions and Operators	
Id, $\boldsymbol{\nabla}$	Identity and gradient operators.
$\exp, \ln$ (also $\log$)	Exponential and natural logarithm functions.
$\cos, \sin, \arccos, \arcsin$	Cosine, sine and their reciprocal functions.
sinc	Sinc function.
J_n, j_n	Bessel and spherical Bessel functions.
Probability	
$\mathbb{E}[X]$	Expectation of a random variable X.
$\mathcal{N}(\mu, \sigma^2), \mathrm{Poisson}(\lambda), \chi^2(N)$	Normal, Poisson and chi-square distributions.
$Q_F(\alpha)$	Quantile function of a distribution F.
Miscellaneous	
$o, \mathcal{O}, \omega, \Omega, \Theta$	Landau notations.
$\simeq, \cong, :=$	Approximate equality, isomorphism, definition.
δ_{nm}	Kronecker delta.
$\boldsymbol{\delta}_k \in \mathbb{R}^n$	k-th element of the canonical basis of $\mathbb{R}^n$.

Table 1.2: Miscellaneous notations used throughout this thesis.

$2\pi^{d/2}/\Gamma(d/2)$, with Γ denoting the *Gamma* function. We have notably $\mathfrak{a}_2 = 2\pi$ and $\mathfrak{a}_3 = 4\pi$.

- Vectors and matrices are written in bold face, in an attempt to make finite-dimensional quantities more apparent. The *adjoint*,[14] *Moore-Penrose pseudo-inverse*, *range* and *nullspace* of a linear operator Φ are denoted by Φ^*, $\Phi^\dagger$, $\mathcal{R}(\Phi)$ and $\mathcal{N}(\Phi)$ respectively. For scalars $z \in \mathbb{C}$ finally, we denote by $\bar{z}, |z|, \Re(z), \Im(z)$ the *conjugate*, *modulus*, *real part* and *imaginary part* of z respectively.
- A function $F : \mathbb{C}^N \to \mathbb{R} \cup \{-\infty, +\infty\}$ is called *convex* if

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N,\ \forall \theta \in [0,1] : \qquad F\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \leq \theta F(\boldsymbol{x}) + (1-\theta)F(\boldsymbol{y}), \tag{1.3}$$

[14] In the specific case of real and complex matrices $\boldsymbol{A}$, we denote the adjoint by $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$, respectively.

and *strictly convex* if

$$\forall \boldsymbol{x} \neq \boldsymbol{y} \in \mathbb{C}^N,\ \forall \theta \in [0,1]: \qquad F\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) < \theta F(\boldsymbol{x}) + (1-\theta) F(\boldsymbol{y}). \tag{1.4}$$

If moreover, $F(\boldsymbol{x}) > -\infty$ for all $\boldsymbol{x} \in \mathbb{C}^N$ and $D = \{\boldsymbol{x} \in \mathbb{C}^N : F(\boldsymbol{x}) < +\infty\} \neq \emptyset$, then F is called a *proper convex* function.[15]

- Let $(\mathcal{X}, \mathcal{T})$ be a *topological space*. A function $F : \mathcal{X} \to \mathbb{R} \cup \{-\infty, +\infty\}$ is said lower semi-continuous (lwsc)[16] at $x_0 \in \mathcal{X}$ if for every $y < F(x_0)$ there exists a neighborhood U of x_0 such that $F(x) > y$ for all $x \in U$. A function is lwsc i.f.f. all of its lower level sets $\left\{\{x \in \mathcal{X} : F(x) \leq y\}, y \in \overline{\mathbb{R}}\right\}$ are *closed* in $\mathcal{T}$. When $\mathcal{X}$ is a metric space, we assume the metric topology as underlying topology and do not specify it explicitly.
- Other miscellaneous notations used throughout this thesis are provided in Table 1.2.

[15] In short, a convex function is proper if its domain is nonempty and it never attains $-\infty$.

[16] w.r.t. the topology $\mathcal{T}$ on $\mathcal{X}$.

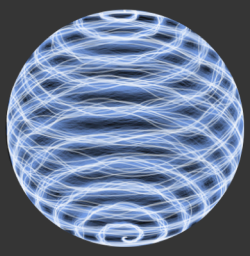
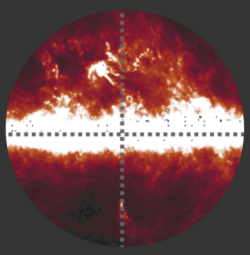
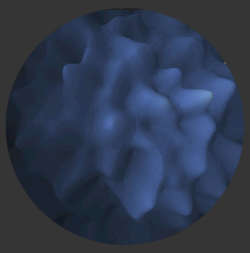
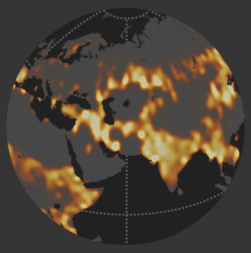
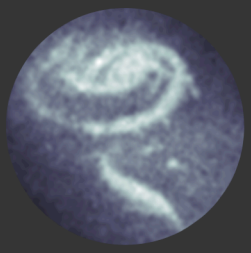
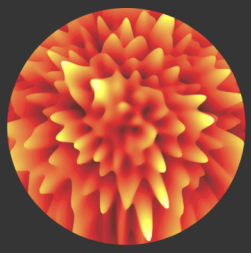
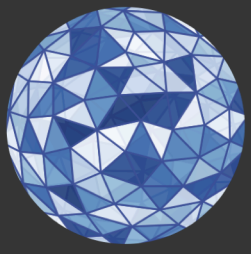
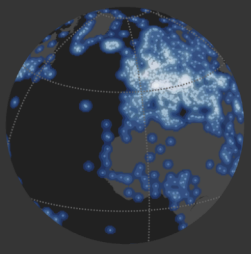
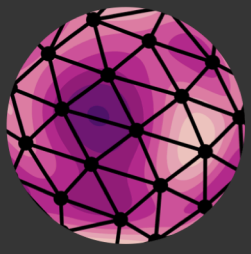
	Domain	Existence	Generalised Sampling	Cost Functional	gTikhonov	gTV	Charac. of Solutions	Search Space	Charac. of Predual
Kimeldorf et al. [93] Schölkopf et al. [156]	Agnostic	✓	✗	Convex (Non-Convex)	(✓)	✗	Parametric	RKHS, (Hilbert)	✓
Boyer et al. [28] Bredies et al. [29]	Agnostic	✗	✓	(quasi-) Convex	✗	(✓)	Geometric	Banach	✗
Unser [177]	Agnostic	✓	✓	Strictly Convex	(✓)	(✓)	Duality Map	Hilbert, Banach	✗
Unser, Fageot et al. [10, 72, 179]	$\mathbb{R}^d$ ($\mathbb{S}^1$)	✓	✓	Convex	✓	✓	Parametric, Geometric	Maximal	✓
This thesis	Agnostic & $\mathbb{S}^{d-1}$	✓	✓	Convex	✓	✓	Parametric, Geometric	Maximal	✓

Table 1.3: Comparison between existing representer theorems and the ones established in this thesis.

Theoretical Foundations

In this part, we lay down the theoretical foundations of the spherical functional approximation framework from Part II. Some of the highlights of this part are the following:

- *In Section 1 of Chapter 2, we introduce the concept of* duality *in topological vector spaces, central to the generalised sampling framework used in Chapter 5 to model linear measurements of spherical fields.*
- *In Section 1.5.3 of Chapter 2, we define and provide a dual characterisation of the* total variation norm *for regular Borel measures. The latter generalises the discrete ℓ_1 norm to continuous setups, and will be used in Chapter 5 as a sparsity-promoting regularisation norm.*
- *In Section 2 of Chapter 2, we establish abstract representer theorems characterising the form of the solutions to a variety of penalised convex optimisation problems defined over Hilbert or Banach spaces.*
- *In Chapter 3, we review important notions from Fourier analysis on the hypersphere, namely* spherical harmonics *and* spherical zonal functions*.*
- *In Chapter 4, we use these notions of harmonic analysis to define* hyperspherical splines*, which can be thought of as primitives –w.r.t. a certain pseudo-differential operator– of Dirac streams with finite innovations.*

2

Functional Analysis Primer

1 Duality in Topological Vector Spaces

The main reference for this section is Chapter 3 of [180].

In order to approximate a spherical field, modelled here as a generic element f of a vector space $\mathscr{B}$, one must first collect evidence of the latter, often by sensing it via a *linear acquisition device*. As we shall see in Chapter 5, these linear measurements can in general be modelled as the outcomes of a collection of device-specific *linear functionals* acting on the object f of interest. In this section, we investigate the structure of the space of all linear functionals associated with a given vector space $\mathscr{B}$, with a special focus on those that yield *well-defined* measurements when acting on any element $f \in \mathscr{B}$.

1.1 Schwartz Duality Product

For a vector space $\mathscr{B}$ over a scalar field $\mathbb{C}$, the space of all linear functionals $f : \mathscr{B} \to \mathbb{C}$ is a vector space called the *algebraic dual* and is denoted by $\mathscr{B}^*$. It is customary to write the action of a linear functional $f \in \mathscr{B}^*$ onto an element $h \in \mathscr{B}$ by means of a *bilinear map* $\langle \cdot | \cdot \rangle : \mathscr{B}^* \times \mathscr{B} \to \mathbb{C}$ called the *Schwartz duality product* defined as

$$\langle \cdot | \cdot \rangle : \begin{cases} \mathscr{B}^* \times \mathscr{B} \to \mathbb{C}, \\ (f, h) \mapsto \langle f | h \rangle := f(h). \end{cases} \tag{2.1}$$

The *bra-ket* notation used in (2.1) to denote the Schwartz duality product is common in quantum mechanics and was introduced by Paul Dirac in 1939. Its resemblance to the inner product is not fortuitous, and is motivated by Hilbert space theory. Indeed, for a Hilbert space $\mathscr{H}$, every linear functional in $\mathscr{H}^*$ can be written as an inner product with some unique element g of $\mathscr{H}$, i.e. $\forall f \in \mathscr{H}^*$, $\exists! g \in \mathscr{H}$ such that $\langle f | h \rangle = \langle h, g \rangle_{\mathscr{H}}$ (see Theorem 2.3).

1.2 Topological Dual

Any sensible acquisition system should react continuously to variations in its input. It seems hence reasonable to require that the linear functionals modelling it be *continuous* as well. The subset of continuous linear functionals in the algebraic dual is a linear subspace, called the *topological dual*[1] and de-

[1] In the manuscript, the shorthand expression "dual space" is sometimes used to refer to the topological dual space.

noted by $\mathscr{B}'$. In infinite dimensions, not all linear functionals are guaranteed to be continuous so we have in general $\mathscr{B}' \subset \mathscr{B}^*$. Moreover, since continuity is a topological notion, there can exist multiple topological duals of a given space $\mathscr{B}$ depending on the topology chosen on the latter. In the special case where $(\mathscr{B}, \|\cdot\|)$ is a Banach space equipped with its canonical normed topology, continuous linear functionals can be characterised as:

$$\mathscr{B}' = \left\{ f \in \mathscr{B}^* : |||f||| := \sup_{h \in \mathscr{B}, \|h\|=1} |\langle f|h\rangle| < \infty \right\}. \tag{2.2}$$

The norm $|||\cdot||| : \mathscr{B}' \to \mathbb{R}_+$ used in (2.2) makes $\mathscr{B}'$ a *Banach space* and is called the *dual norm* induced by the norm $\|\cdot\|$ on $\mathscr{B}$. In plain words, (2.2) states that continuous linear functionals are *bounded*[2] linear functionals in $\mathscr{B}^*$. The latter will hence produce bounded measurements, well-defined for any input $\nu \in \mathscr{B}$.

[2] Note that in finite dimensions things are much less complicated since every linear functional is bounded and all norms are topologically equivalent, hence $\mathscr{B}^* = \mathscr{B}'$.

Vocabulary 2.1 — *Predual and Duality Pair.* *A topological vector space $\mathscr{B}$ and its topological dual $\mathscr{B}'$ are said to form a* duality pair. *Moreover, $\mathscr{B}$ is called the* predual *of $\mathscr{B}'$.*

1.3 Weak* and Strong Topologies on the Topological Dual

Throughout, we will sometimes have cause to define a topology on the topological dual $\mathscr{B}'$. Two popular choices are the *strong topology* and the *weak* topology*. The strong topology is the Banach topology induced by the dual norm defined in (2.2). This is also the topology of *uniform convergence*. Indeed, a sequence $\{f_n, n \in \mathbb{N}\} \subset \mathscr{B}'$ converges towards a limit functional $f^* \in \mathscr{B}'$ with respect to the strong topology if and only if (i.f.f.):

$$\lim_{n \to +\infty} |||f^* - f_n||| = \lim_{n \to +\infty} \sup_{h \in \mathscr{B},\, \|h\|=1} |\langle f^* - f_n|h\rangle| = 0.$$

With this topology, the *closed balls are necessarily not compact* when $\mathscr{B}$ is infinite dimensional. As we shall see, this renders the strong topology cumbersome to work with. For this reason, we will prefer the weak* topology, or *topology of pointwise convergence*, which does not suffer from similar issues (see Theorem 2.1). It is the *coarsest* topology on $\mathscr{B}'$ such that elements $h \in \mathscr{B}$ are continuous functionals on $\mathscr{B}'$. It is induced by the family of seminorms:

The weak topology is the coarsest topology such that all elements of the predual are continuous linear functionals on the dual.*

$$\|\cdot\|_h : \begin{cases} \mathscr{B}' \to \mathbb{R}_+ \\ f \mapsto \|f\|_h = |\langle f|h\rangle| \end{cases} \qquad \forall h \in \mathscr{B}.$$

Convergence with respect to the weak* topology is *pointwise*. Indeed $\{f_n, n \in \mathbb{N}\} \subset \mathscr{B}'$ converges towards a limit functional $f^* \in \mathscr{B}'$ with respect to the weak* topology i.f.f.

$$\lim_{n \to \infty} |\langle f^* - f_n|h\rangle| = 0, \quad \forall h \in \mathscr{B}.$$

Clearly, uniform convergence is stronger than pointwise convergence, since:

$$|\langle f^* - f_n | h \rangle| \leq \|h\| |||f^* - f_n||| , \quad \forall h \in \mathscr{B},$$

from the very definition of the dual norm. Throughout, we will employ expressions such as "weak* compact", "weak* closed" or "weak* convergent" when it is important to make obvious the underlying topology with respect to which the topological notions should be understood.

1.3.1 Banach-Alaoglu Theorem

The *Banach-Alaoglu* theorem [153, p. 68] is one of the main reasons for our interest in the weak* topology. The latter provides us with an easy characterisation of weak* compact sets:

Theorem 2.1 — *Banach-Alaoglu [153].* Let $\mathscr{B}$ be a *normed topological vector space*. Then the *closed unit ball* of its *dual space* $\mathscr{B}'$ is *weak* compact*.

Weak compact sets are weak* bounded and* weak* closed.

The following proposition [179, Proposition 9] is a direct consequence of Theorem 2.1:

Proposition 2.2 — *Compactness in the Weak* Topology [179].* *Let $\mathscr{B}$ be a* normed topological vector space. *Then:*

- *Weak* compact sets of $\mathscr{B}'$ are* weak* bounded[3] *and* weak* closed. *In particular, for every $\epsilon > 0$, the closed ball $B_\epsilon = \{f \in \mathscr{B}' : |||f||| \leq \epsilon\}$ is weak* compact in $\mathscr{B}'$.*
- *If we assume further that $\mathscr{B}$ is separable, then any weak* closed and bounded set of $\mathscr{B}'$ is metrisable and hence* sequentially weak* compact. *In particular, if $\{f_n\}_{n\in\mathbb{N}} \subset \mathscr{B}'$ is a sequence bounded w.r.t. the operator norm $|||\cdot|||$ then we can extract a convergent subsequence from it with limit in $\mathscr{B}'$.*

[3] i.e. bounded w.r.t. the seminorms inducing the weak* topology.

1.4 Bidual and Reflexive Spaces

When equipped with the strong topology induced by the dual norm $|||\cdot|||$, the topological dual $\mathscr{B}'$ becomes a normed vector space of its own. It is then possible to consider its own topological dual, called *bidual* and denoted by $\mathscr{B}''$:

Definition 2.1 — *Bidual.* *Let $(\mathscr{B}, \|\cdot\|)$ be some normed vector space and $(\mathscr{B}', |||\cdot|||)$ its topological dual equipped with the strong topology. Then, the* bidual *$\mathscr{B}''$ is defined as the topological dual of $(\mathscr{B}', |||\cdot|||)$:*

$$\mathscr{B}'' = \left\{ g \in (\mathscr{B}')^* : |||g|||'' := \sup_{f \in \mathscr{B}', |||f|||=1} |\langle g | f \rangle| < \infty \right\}, \tag{2.3}$$

where the duality product is defined over $(\mathscr{B}')^* \times \mathscr{B}'$. The bidual norm *$|||\cdot|||''$ is the norm induced by the dual norm $|||\cdot|||$.*

When $\mathscr{B} \cong \mathscr{B}''$ (i.e. $\mathscr{B}$ and $\mathscr{B}''$ are isometrically isomorphic), we say that the space $\mathscr{B}$ is *reflexive*. For example, any Hilbert space $\mathscr{H}$ is reflexive. This is

due to the *Riesz-Fréchet representation theorem* [59] –one of the most famous results from functional analysis– which shows that every Hilbert space $\mathscr{H}$ is isometrically isomorphic to its topological dual:

Theorem 2.3 — *Riesz-Fréchet Representation Theorem [59].* Let $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ be a *Hilbert* space and $(\mathscr{H}', |||\cdot|||)$ its topological dual equipped with dual norm $|||\cdot|||$ induced by $\|\cdot\|_{\mathscr{H}}$ defined in (2.2). Then for every *continuous linear functional* $\nu \in \mathscr{H}'$ there exists a *unique* $\varphi_\nu \in \mathscr{H}$ such that

$$\nu(h) = \langle \nu | h \rangle = \langle h, \varphi_\nu \rangle_{\mathscr{H}}, \qquad \forall h \in \mathscr{H}. \tag{2.4}$$

We have moreover $|||\nu||| = \|\varphi_\nu\|_{\mathscr{H}}$.

Hilbert spaces are self-dual and reflexive.

The theorem yields the identification $\mathscr{H} \cong \mathscr{H}'$: we say that $\mathscr{H}$ is *self-dual*. Moreover, the duality pair $(\mathscr{H}, \mathscr{H}')$ is trivially reflexive since $\mathscr{H}'' \cong (\mathscr{H}')' \cong \mathscr{H}' \cong \mathscr{H}$. In general, Banach spaces are *not* always reflexive, as discussed in Sections 1.5.1 and 1.5.3 respectively.

1.5 Duality Pairs for Common Functional Spaces

In this section we provide well-known duality pairs for common functional spaces.

1.5.1 Duality Pairs of Lebesgue Spaces

The *Riesz representation theorem* [59] identifies duality pairs among the *Lebesgue spaces* $(\mathscr{L}^p(\mathcal{X}), \|\cdot\|_p)$, $1 \leq p < +\infty$:

Theorem 2.4 — *Riesz Representation Theorem [59].* Let $(\mathcal{X}, \mu)$ be a *measurable* space, $1 \leq p < +\infty$ and q the *conjugate* of p, i.e. $1/p + 1/q = 1$. Then for every *continuous linear functional* $f \in (\mathscr{L}^p(\mathcal{X}), \|\cdot\|_p)'$ there exists a *unique* $\varphi_f \in (\mathscr{L}^q(\mathcal{X}), \|\cdot\|_q)$ such that

$$f(h) = \langle f | h \rangle = \int_{\mathcal{X}} h(x)\overline{\varphi_f(x)}\mu(dx) = \langle h, \varphi_f \rangle_{\mathcal{X}}, \quad \forall h \in \mathscr{L}^p(\mathcal{X}).$$

We have moreover $|||f||| = \|\varphi_f\|_q$, where $|||\cdot|||$ is the dual norm induced by $\|\cdot\|_p$ on $(\mathscr{L}^p(\mathcal{X}), \|\cdot\|_p)'$.

When $\mathcal{X} = \mathbb{S}^{d-1}$ and μ is the Lebesgue measure on $\mathbb{S}^{d-1}$, one obtains in particular:

- $(\mathscr{L}^2(\mathbb{S}^{d-1}), \|\cdot\|_2)' \cong (\mathscr{L}^2(\mathbb{S}^{d-1}), \|\cdot\|_2)$ with dual norm $\|\cdot\|_2$. The space $\mathscr{L}^2(\mathbb{S}^{d-1})$ is *self-dual* and hence reflexive.
- $(\mathscr{L}^p(\mathbb{S}^{d-1}), \|\cdot\|_p)' \cong (\mathscr{L}^q(\mathbb{S}^{d-1}), \|\cdot\|_q)$ with dual norm $\|\cdot\|_q$, for $1 < p, q < +\infty$. These Lebesgue spaces are examples of reflexive Banach spaces.
- $(\mathscr{L}^1(\mathbb{S}^{d-1}), \|\cdot\|_1)' \cong (\mathscr{L}^\infty(\mathbb{S}^{d-1}), \|\cdot\|_\infty)$ with dual norm $\|\cdot\|_\infty$. Unlike the previous cases however, $\mathscr{L}^1(\mathbb{S}^{d-1})$ is *not* reflexive: $(\mathscr{L}^1(\mathbb{S}^{d-1}), \|\cdot\|_1)$ is *not*[4] the dual of $(\mathscr{L}^\infty(\mathbb{S}^{d-1}), \|\cdot\|_\infty)$.

[4] As a matter of fact, it can be shown that $\mathscr{L}^1(\mathbb{S}^{d-1})$ has *no predual*.

1.5.2 Schwartz Functions and Generalised Functions

The space of *generalised functions* or *distributions* is almost the largest functional space that can be defined on $\mathbb{S}^{d-1}$. It contains as subspaces the Lebesgue spaces as well as the spaces of continuous functions or regular Borel measures. It is denoted $\mathscr{S}'(\mathbb{S}^{d-1})$ and is defined as the topological dual of the space of *Schwartz functions*[5] $\mathscr{S}(\mathbb{S}^{d-1}) = \mathscr{C}^\infty(\mathbb{S}^{d-1})$, equipped with the *metric* topology generated by the family of norms:

$$\|\cdot\|_{n,\infty} : \begin{cases} \mathscr{C}^\infty(\mathbb{S}^{d-1}) \to \mathbb{R}_+ \\ h \mapsto \|h\|_{n,\infty} := \|(\mathrm{Id} - \Delta_{\mathbb{S}^{d-1}})^n h\|_\infty \end{cases} \qquad \forall n \in \mathbb{N},$$

where $\Delta_{\mathbb{S}^{d-1}}$ is the *Laplace-Beltrami* operator on $\mathbb{S}^{d-1}$ (see Section 1 of Chapter 3 for a definition of the Laplace-Beltrami operaor). The Schwartz space is a *locally convex Fréchet space*[6]. It is *not* normable and in particular not complete with the supremum norm (indeed, it is dense in the space of continuous functions). Note that since the metric topology on $\mathscr{S}(\mathbb{S}^{d-1})$ is not induced by a norm[7], it is not possible to define a strong topology on $\mathscr{S}'(\mathbb{S}^{d-1})$. We will hence always assume the weak* topology as canonical topology on $\mathscr{S}'(\mathbb{S}^{d-1})$.

[5] The hypersphere being compact and bounded, Schwartz functions simply reduce to infinitely smooth functions.

[6] Fréchet spaces generalise Banach spaces with metrics that do not originate from norms.

[7] As a matter of fact, it is possible to show that $\mathscr{S}(\mathbb{S}^{d-1})$ is not normable [180].

1.5.3 Continuous Functions, Measures and Total Variation Norm

The *Riesz-Markov theorem* in its most general form establishes a duality pair between *vanishing continuous functions* taking values in some *locally compact Hausdorff topological space* $\mathcal{X}$ and $\mathbb{C}$*-valued regular Borel measures* defined over the Borel sets of $\mathcal{X}$. When $\mathcal{X}$ is compact and bounded, as it is the case for the hypersphere $\mathbb{S}^{d-1}$, the theorem simplifies a little as vanishing continuous functions simply become continuous functions. In the specific case $\mathcal{X} = \mathbb{S}^{d-1}$, it reads [59, 68]:

Spherical continuous functions and spherical regular Borel measures form a non-reflexive duality pair.

Theorem 2.5 — *Riesz-Markov Representation Theorem – Spherical Setup.* For any *continuous linear functional* ν on $(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)'$, there exists a *unique* $\mathbb{C}$**-valued regular Borel measure** λ_ν on $\mathbb{S}^{d-1}$ such that

$$\nu(h) = \langle \nu | h \rangle = \int_{\mathbb{S}^{d-1}} h(\boldsymbol{r}) \lambda_\nu(d\boldsymbol{r}), \quad \forall h \in \mathscr{C}(\mathbb{S}^{d-1}).$$

We have moreover $|||\nu||| = \|\lambda_\nu\|_{TV} = |\lambda_\nu|(\mathcal{X})$ the *total variation* (see Definition 2.2) of λ_ν, where $|||\cdot|||$ is the dual norm induced by $\|\cdot\|_\infty$ on $(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)'$.

The theorem establishes hence the duality pair[8]

$$(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)' \cong (\mathcal{M}(\mathbb{S}^{d-1}), \|\cdot\|_{TV}),$$

where $\mathcal{M}(\mathbb{S}^{d-1})$ is the space of $\mathbb{C}$*-valued regular Borel measures* on $\mathbb{S}^{d-1}$, and provides us with a useful dual characterisation of $\mathcal{M}(\mathbb{S}^{d-1})$ (see (2.2)):

$$\mathcal{M}(\mathbb{S}^{d-1}) \cong \left\{ \nu \in \mathscr{C}(\mathbb{S}^{d-1})^* : \sup_{h \in \mathscr{C}(\mathbb{S}^{d-1}),\, \|h\|_\infty = 1} |\langle \nu | h \rangle| < +\infty \right\}$$

[8] It is a well-known fact however that $\mathscr{C}(\mathbb{S}^{d-1})$ is *not* reflexive [179].

$$\cong \left\{ \nu \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sup_{h \in \mathscr{S}(\mathbb{S}^{d-1}),\, \|h\|_\infty = 1} |\langle \nu | h \rangle| < +\infty \right\}, \tag{2.5}$$

where the second isomorphism results[9] from the density of Schwartz functions in the space of bounded continuous functions. Equation (2.5) permits us to see $\mathcal{M}(\mathbb{S}^{d-1})$ as the subspace of generalised functions with finite dual norm. The Riesz-Markov theorem tells us moreover that this dual norm is actually given by the total variation (TV) norm. For measures on the hypersphere, the latter is defined as follows:

[9] See Section 3 of [179] for more details on the matter, where the authors used an analogous characterisation for $\mathcal{M}(\mathbb{R}^d)$.

Definition 2.2 — *Total Variation Norm.* *Let ν be a complex-valued* measure *on $\mathbb{S}^{d-1}$ with* Borel algebra $\mathcal{B} \subset \mathscr{P}(\mathbb{S}^{d-1})$. *The* total variation *of ν over a Borel set $B \in \mathcal{B}$ is defined as*

$$|\nu|(B) := \sup \left\{ \sum_{n \in \mathbb{N}} |\nu(B_n)| : \{B_n, n \in \mathbb{N}\} \subset \mathcal{B},\ \cup_{n\in\mathbb{N}} B_n = B,\ B_n \cap B_m = \emptyset,\ n, m \in \mathbb{N} \right\}.$$

The total variation norm *of ν is obtained for $B = \mathbb{S}^{d-1}$:*

$$\|\nu\|_{TV} : \begin{cases} \mathcal{M}(\mathbb{S}^{d-1}) \to \mathbb{R}_+ \\ \nu \mapsto |\nu|\left(\mathbb{S}^{d-1}\right), \end{cases} \tag{2.6}$$

where $\mathcal{M}(\mathbb{S}^{d-1})$ is the space of $\mathbb{C}$-valued regular Borel measures *on $\mathbb{S}^{d-1}$ with finite total variation norm.*

Remark 2.1 — *TV, ℓ_1 and $\mathscr{L}^1$ norms.* *The total variation norm can be thought as an $\mathscr{L}^1$ norm for measures.*[10] *Indeed, consider for example discrete measures of the form:*

$$\nu = \sum_{k=1}^{n} \alpha_k \delta_{\boldsymbol{r}_k}, \quad n \in \mathbb{N},\ \{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{C},\ \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\} \subset \mathbb{S}^{d-1}.$$

Then it is easy to show that the total variation norm of ν is simply the discrete ℓ_1 norm of the Dirac amplitudes arranged as a vector:

$$|\nu|(\mathbb{S}^{d-1}) = \sum_{k=1}^{n} |\alpha_k| = \|\boldsymbol{\alpha}\|_1,$$

with $\boldsymbol{\alpha} := [\alpha_1, \cdots, \alpha_n] \in \mathbb{C}^n$. This behaviour extends also to absolutely continuous *Borel measures whose total variation is given by the continuous $\mathscr{L}^1$ norm of their "derivative".*[11] *Indeed, if ν is absolutely continuous with respect to the Lebesgue measure μ then, from the* Radon-Nikodym theorem*, ν admits a* Radon-Nikodym derivative $\rho = \frac{\partial \nu}{\partial \mu} \in \mathscr{L}^1(\mathbb{S}^{d-1})$. *We get hence from* Hölder's inequality*:*

$$|\langle \nu | h \rangle| = \left| \int_{\mathbb{S}^{d-1}} h(\boldsymbol{r}) \rho(\boldsymbol{r}) \mu(d\boldsymbol{r}) \right| \leq \|h\|_\infty \|\rho\|_1, \quad \forall h \in \mathscr{L}^\infty(\mathbb{S}^{d-1}),$$

which yields, from the dual characterisation of the TV norm provided by Theo-

[10] This motivates the use of the TV norm as sparsity-inducing regularisation norm in convex optimisation problems involving measures, such as the ones investigated in Chapter 5.

[11] We find back here the traditional notion of total variation in applied fields such as signal and image processing [185].

rem 2.5,

$$|\nu|\left(\mathbb{S}^{d-1}\right) = \sup_{h\in\mathscr{C}(\mathbb{S}^{d-1}),\|h\|_\infty=1} |\langle\nu|h\rangle| \leq \|\rho\|_1,$$

and finally $|\nu|(\mathbb{S}^{d-1}) = \left\|\frac{\partial\nu}{\partial\mu}\right\|_1 = \|\rho\|_1$ *since the bounded and continuous function identically equal to one saturates the inequality.*

1.6 Duality map

It is sometimes useful to map a Banach space $\mathscr{B}$ into its topological dual $\mathscr{B}'$, so as to interpret elements of $\mathscr{B}$ as continuous linear functionals in $\mathscr{B}'$. This is typically done by means of the *duality map* [177] which is defined as:

The duality map allows us to see functions in a given functional space as linear functionals in the dual.

Definition 2.3 — *Duality Map.* *Let* $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ *be a Banach space and* $\mathscr{B}'$ *its topological dual with dual norm* $|||\cdot|||$*. The* duality map*, is defined as:*

$$J_{\mathscr{B}} : \begin{cases} \mathscr{B} \to \mathscr{P}(\mathscr{B}'), \\ h \mapsto J_{\mathscr{B}}(h) := \left\{\nu \in \mathscr{B}' : \|h\|_{\mathscr{B}} = |||\nu|||,\ \langle\nu|h\rangle = \|h\|^2_{\mathscr{B}}\right\}. \end{cases} \tag{2.7}$$

The pairs $\{(h,\nu),\ \nu \in J_{\mathscr{B}}(h)\}$ *are called* dual pairs*, and elements of the set* $J_{\mathscr{B}}(h)$ *are called* Banach conjugates *of* $h \in \mathscr{B}$*.*

Note that the duality map is in general *set-valued*, i.e. it takes values in the powerset of $\mathscr{B}'$. The *Hahn-Banach theorem* guarantees that this set is nonempty for every $h \in \mathscr{B}$ [177]. When J is single-valued, we see the duality map as a function $J_{\mathscr{B}} : \mathscr{B} \to \mathscr{B}'$. When $\mathscr{B}$ is *reflexive*, and both $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ and $(\mathscr{B}', |||\cdot|||)$ have *strictly convex*[12] unit balls, then the duality map is *single-valued* and *bijective* [177, Theorem 3]. This is notably the case for Hilbert spaces.

[12] For brevity we say that a Banach space is strictly convex if its unit ball is strictly convex.

Vocabulary 2.2 — *Riesz Map.* *In the context of a Hilbert space* $\mathscr{H}$*, the inverse duality map* $J^{-1}_{\mathscr{H}} : \mathscr{H}' \to \mathscr{H}$ *is commonly called the* Riesz map *[177] which we denote by* $R_{\mathscr{H}} := J^{-1}_{\mathscr{H}}$*.*

In Propositions 2.6 and 2.7 below, we derive the duality map for an arbitrary Hilbert space and a strictly convex spherical Lebesgue space $\mathscr{L}^p(\mathbb{S}^{d-1})$ respectively. We will see that in the Hilbert case, the duality map is antilinear, while in the Lebesgue (Banach) case, the duality map is nonlinear. More generally, it can be shown that for a strictly convex and reflexive space with strictly convex dual, the duality map is antilinear i.f.f. the space is Hilbert [177, Proposition 4].

Proposition 2.6 — *Hilbert Duality Map.* *Let* $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ *denote an arbitrary Hilbert space. Then the duality map on* $\mathscr{H}$ *is single-valued, bijective and antilinear. It is given by*

$$J_{\mathscr{H}} : \begin{cases} \mathscr{H} \to \mathscr{H}', \\ \varphi \mapsto \langle J_{\mathscr{H}}(\varphi)|h\rangle := \langle h,\varphi\rangle_{\mathscr{H}}, \quad \forall h \in \mathscr{H}. \end{cases} \tag{2.8}$$

Proof. Let $\varphi \in \mathscr{H}$ and consider the linear functional defined as

$$\langle J_\varphi | h \rangle := \langle h, \varphi \rangle_{\mathscr{H}}, \qquad \forall h \in \mathscr{H}.$$

Then, we have from the equality case of the *Cauchy-Schwarz inequality*

$$|||J_\varphi||| = \sup_{h\in\mathscr{H}, \|h\|_{\mathscr{H}}=1} |\langle J_\varphi | h \rangle| = \sup_{h\in\mathscr{H}, \|h\|_{\mathscr{H}}=1} |\langle h, \varphi \rangle_{\mathscr{H}}| = \|\varphi\|_{\mathscr{H}}$$

and

$$\langle J_\varphi | \varphi \rangle = \langle \varphi, \varphi \rangle_{\mathscr{H}} = \|\varphi\|^2_{\mathscr{H}}.$$

From Definition 2.3, this hence yields $J_\varphi \in J_{\mathscr{H}}(\varphi)$. Moreover, since every Hilbert space is strictly convex [177] and reflexive, the duality map is actually single-valued and bijective, hence $J_\varphi = J_{\mathscr{H}}(\varphi)$. Note that, since the inner product is a *sesquilinear form*, the duality map is *antilinear*. We have indeed, for every $(\lambda, h) \in \mathbb{C} \times \mathscr{H}$:

$$\langle J_{\mathscr{H}}(\lambda\varphi) | h \rangle = \langle h, \lambda\varphi \rangle_{\mathscr{H}} = \bar{\lambda} \langle h, \varphi \rangle_{\mathscr{H}} = \bar{\lambda} \langle J_{\mathscr{H}}(\varphi) | h \rangle = \langle \bar{\lambda} J_{\mathscr{H}}(\varphi) | h \rangle,$$

and hence $J_{\mathscr{H}}(\lambda\varphi) = \bar{\lambda} J_{\mathscr{H}}(\varphi)$. ■

Proposition 2.7 — *Lebesgue Duality Map.* *Consider a Lebesgue space $\mathscr{L}^p(\mathbb{S}^{d-1})$, with $1 < p < +\infty$ and q its conjugate. Then, the duality map on $\mathscr{L}^p(\mathbb{S}^{d-1})$ is single-valued, bijective and nonlinear. It is given by*

$$J_{\mathscr{L}^p} : \begin{cases} \mathscr{L}^p(\mathbb{S}^{d-1}) \to (\mathscr{L}^p(\mathbb{S}^{d-1}))', \\ \varphi \mapsto \langle J_{\mathscr{L}^p}(\varphi) | f \rangle := \langle f, g_\varphi \rangle_{\mathbb{S}^{d-1}}, \quad \forall f \in \mathscr{L}^p(\mathbb{S}^{d-1}), \end{cases} \tag{2.9}$$

where $g_\varphi \in \mathscr{L}^q(\mathbb{S}^{d-1})$ is given by $g_\varphi = 0$ if $\varphi = 0$ and

$$g_\varphi = \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}}, \qquad \forall \varphi \neq 0.$$

The complex signum *function is moreover defined as* $\operatorname{sgn}(0) = 0$ *and* $\operatorname{sgn}(z) = \bar{z}/|z|, \ \forall z \neq 0$ *(we have* $\operatorname{sgn}(z)z = |z|, \ \forall z \in \mathbb{C}$*.)*

Proof. The case $\varphi = 0$ is trivial. Let hence $\varphi \in \mathscr{L}^p(\mathbb{S}^{d-1})\backslash\{0\}$ and consider the linear functional defined as

$$\langle J_\varphi | f \rangle := \left\langle f, \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\rangle_{\mathbb{S}^{d-1}}, \quad \forall f \in \mathscr{L}^p(\mathbb{S}^{d-1}).$$

Then, we have from *Hölder's inequality*:

$$|||J_\varphi||| = \sup_{f\in\mathscr{L}^p(\mathbb{S}^{d-1}), \|f\|_p=1} \left| \left\langle f, \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\rangle_{\mathbb{S}^{d-1}} \right| \leq \left\| \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\|_q,$$

where p and q are conjugate. Since $1/p + 1/q = 1 \Rightarrow ((p-1)q = p) \wedge (p - 1 =$

p/q), we have moreover

$$\left\| \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\|_q = \frac{1}{\|\varphi\|_p^{p-2}} \left(\int_{\mathbb{S}^{d-1}} |\varphi|^{(p-1)q} \right)^{1/q} = \frac{1}{\|\varphi\|_p^{p-2}} \left(\int_{\mathbb{S}^{d-1}} |\varphi|^p \right)^{1/q}$$
$$= \frac{\|\varphi\|_p^{\frac{p}{q}}}{\|\varphi\|_p^{p-2}} = \frac{\|\varphi\|_p^{p-1}}{\|\varphi\|_p^{p-2}} = \|\varphi\|_p.$$

The inequality is hence saturated for $f^\star = \varphi/\|\varphi\|_p$ since

$$\left\langle \frac{\varphi}{\|\varphi\|_p}, \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\rangle_{\mathbb{S}^{d-1}} = \frac{1}{\|\varphi\|_p^{p-1}} \int_{\mathbb{S}^{d-1}} |\varphi||\varphi|^{p-1} = \frac{\|\varphi\|_p^p}{\|\varphi\|_p^{p-1}} = \|\varphi\|_p, \tag{2.10}$$

and hence finally $|||J_\varphi||| = \|\varphi\|_p$ since $\|f^\star\|_p = 1$. From (2.10) we have moreover

$$\langle J_\varphi | \varphi \rangle = \left\langle \varphi, \frac{\operatorname{sgn}(\varphi)|\varphi|^{p-1}}{\|\varphi\|_p^{p-2}} \right\rangle_{\mathbb{S}^{d-1}} = \|\varphi\|_p^2.$$

From Definition 2.3, this hence yields $J_\varphi \in J_{\mathscr{L}^p}(\varphi)$. Moreover, for $1 < p < +\infty$ the Lebesgue space $\mathscr{L}^p(\mathbb{S}^{d-1})$ is strictly convex and reflexive (from Theorem 2.4) and hence the duality map is single-valued: $J_\varphi = J_{\mathscr{L}^p}(\varphi)$. Finally, $J_{\mathscr{L}^p}$ is clearly nonlinear since it involves an absolute value and a norm which are both nonlinear. ■

2 Abstract Representer Theorems

In Chapter 5, we consider optimisation problems formulated over infinite dimensional functional spaces. Solving such problems in practice is of course only feasible if they admit solutions with a *finite number* of degrees of freedom (df). In functional analysis, results characterising the form and degrees of freedom of the solutions to a particular (infinite-dimensional) optimisation problem are called *representer theorems*. In Sections 2.1 and 2.3, we present and introduce some important representer theorems pertaining to convex optimisation in abstract Banach spaces, which are leveraged in Chapter 5 to establish representer theorems for infinite-dimensional spherical approximation.

In infinite-dimensional optimisation, representer theorems are used to characterise the form of the solutions to a given problem, and hopefully re-express the optimisation task in terms of finitely many degrees of freedom.

2.1 Banach Representer Theorem

Recently, Unser has established in [177, Theorem 5] a very generic representer theorem, characterising the solutions of a broad class of unconstrained optimisation problems formulated over abstract Banach spaces. It is provided hereafter[13].

Theorem 2.8 — *Abstract Banach Representer Theorem [177].* Consider the following assumptions:

A1 $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ is a *Banach space*, with *topological dual* $(\mathscr{B}', |||\cdot|||)$ (where

[13] The notations from [177] have been slightly adapted to better align with those of the current document.

$|||\cdot|||$ is the *dual norm*), and $J_{\mathscr{B}} : \mathscr{B} \to \mathscr{B}'$ is the *duality map*;

A2 $\text{span}\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{B}$, with the φ_i being *linearly independent*;

A3 $\boldsymbol{\Phi} : \mathscr{B}' \to \mathbb{C}^L$ is a *sampling* operator, defined as

$$\boldsymbol{\Phi}(f) = (\langle f|\varphi_1\rangle, \cdots, \langle f|\varphi_L\rangle), \quad \forall f \in \mathscr{B}';$$

A4 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

is *proper*, *strictly convex* and lower semi-continuous;

A5 $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is some *arbitrary strictly increasing convex* function.

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the solution set of the optimisation problem

$$\mathcal{V} = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \; + \; \Lambda\left(|||f|||\right) \right\}, \tag{2.11}$$

is *non-empty*, *convex* and *weak* compact*. Moreover, we have

$$\mathcal{V} \subset J_{\mathscr{B}} \left(\sum_{i=1}^{L} \alpha_i \varphi_i \right), \tag{2.12}$$

for some weights $\{\alpha_1, \dots, \alpha_L\} \subset \mathbb{C}$.

Remark 2.2 — *Unicity Conditions.* *Notice that the solution set $\mathcal{V}$ degenerates to a single point when the functional $f \mapsto \Lambda(|||f|||)$ is strictly convex –see point (i) in the proof of [177, Theorem 5], which happens if at least one of the following conditions is verified:*

1. *$\mathscr{B}'$ is* strictly convex*;*
2. *Λ is* strictly convex*.*

Indeed, we have:

1. *$f \mapsto \Lambda(|||f|||)$ is strictly convex as a composition between a strictly convex function and a strictly increasing function.*
2. *$f \mapsto \Lambda(|||f|||)$ is strictly convex as a composition between a convex function and a strictly convex increasing function.*

In case 1, we moreover have from [177, Theorem 3, item 4] that the duality mapping is single-valued, and hence $\mathcal{V} = \{f^\star\} = J_{\mathscr{B}}\left(\sum_{i=1}^{L} \alpha_i \varphi_i\right)$. In case 2 however, the duality map may not necessarily be single-valued, and we only have $\mathcal{V} = \{f^\star\}$ with $f^\star \in J_{\mathscr{B}}\left(\sum_{i=1}^{L} \alpha_i \varphi_i\right)$.

Assumption A4 of Theorem 2.8 may in practice be slightly too restrictive, since many cost functionals are convex but not strictly convex (see examples in Section 5 of Chapter 7). Hopefully, an investigation of the proof of [177, Theorem 3] reveals that this assumption is only used in the nonunique case where $f \mapsto \Lambda(|||f|||)$ is not strictly convex, so as to guarantee the set $\boldsymbol{\Phi}(\mathcal{V}) \subset \mathbb{C}^L$ is a singleton –i.e. all solutions yield the same measurements. It is hence possible to relax A4, provided an additional assumption of strict convexity of the

regularisation functional. We do so in the next theorem:

Theorem 2.9 — *Representer Theorem (Non Strictly Convex Cost Functional).* Consider the following assumptions:

- B1 $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ is a *Banach space*, with *topological dual* $(\mathscr{B}', |||\cdot|||)$ (where $|||\cdot|||$ is the *dual norm*), and $J_{\mathscr{B}} : \mathscr{B} \to \mathscr{B}'$ is the *duality map*;
- B2 $\operatorname{span}\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{B}$, with the φ_i being *linearly independent*;
- B3 $\boldsymbol{\Phi} : \mathscr{B}' \to \mathbb{C}^L$ is a *sampling* operator, defined as

$$\boldsymbol{\Phi}(f) = (\langle f|\varphi_1\rangle, \cdots, \langle f|\varphi_L\rangle), \quad \forall f \in \mathscr{B}';$$

- B4 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

  is *proper*, *convex* and lower semi-continuous;
- B5 $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is some *arbitrary strictly increasing convex* function;
- B6 The functional $f \mapsto \Lambda(|||f|||)$ is *strictly convex*, i.e. $\mathscr{B}'$ is strictly convex or Λ is strictly convex (see Remark 2.2).

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the solution to the optimisation problem

$$f^\star = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \; + \; \Lambda\left(|||f|||\right) \right\}, \tag{2.13}$$

exists and is *unique*. Moreover, we have

$$f^\star \in J_{\mathscr{B}} \left(\sum_{i=1}^{L} \alpha_i \varphi_i \right),$$

for some weights $\{\alpha_1, \dots, \alpha_L\} \subset \mathbb{C}$.

Proof. The proof of [177, Theorem 5] remains valid under the assumptions of Theorem 2.9. Indeed, in this case Assumption A4 is unnecessary since Assumptions B4 and B6 guarantee that the minimising functional in (2.13) is strictly convex as a sum between a convex and a strictly convex function and hence admits a unique solution $f^\star \in \mathscr{B}'$, with unique measurement vector $\boldsymbol{\Phi}(f^\star) = \boldsymbol{z}^\star$. Problem (2.13) can hence indeed be transformed into a generalised interpolation problem of the form of (5) in the proof of [177, Theorem 5]. The rest of the proof remains unchanged. ∎

2.2 Hilbert Representer Theorem

In the Hilbert case, we can deduce the following corollary from Theorem 2.9:

Corollary 2.10 — *Hilbert Representer Theorem.* *Consider the following assumptions:*

C1 $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ *is a* Hilbert space*, with* topological dual $\mathscr{H}'$*, and* $R_{\mathscr{H}} : \mathscr{H}' \to \mathscr{H}$ *is the* Riesz map*;*

C2 $span\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{H}'$*, with the* φ_i *being* linearly independent*;*

C3 $\boldsymbol{\Phi} : \mathscr{H} \to \mathbb{C}^L$ *is a* sampling *operator, defined as*

$$\boldsymbol{\Phi}(h) = (\langle\varphi_1|h\rangle, \cdots, \langle\varphi_L|h\rangle), \quad \forall h \in \mathscr{H};$$

C4 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ *is a cost functional such that for all* $\boldsymbol{y} \in \mathbb{C}^L$*,*

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

is proper*,* convex *and lower semi-continuous;*

C5 $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ *is some* arbitrary strictly increasing convex *function;*

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$*, the solution to the optimisation problem*

$$h^\star = \underset{h \in \mathscr{H}}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(h)\right) \; + \; \Lambda\left(\|h\|_{\mathscr{H}}\right) \right\}, \tag{2.14}$$

exists *and is* unique*. Moreover, we have*

$$h^\star = \sum_{i=1}^{L} \beta_i R_{\mathscr{H}}(\varphi_i),$$

for some weights $\{\beta_1, \dots, \beta_L\} \subset \mathbb{C}$*.*

Proof. Every Hilbert space $\mathscr{H}$ is mapped isometrically on its bidual $\mathscr{H}''$ by the linear isomorphism

$$S_{\mathscr{H}} : \begin{cases} \mathscr{H} \to \mathscr{H}'' \\ h \mapsto g_h(\varphi) = \langle\varphi|h\rangle, \quad \forall \varphi \in \mathscr{H}, \end{cases}$$

where $\langle\cdot|\cdot\rangle : \mathscr{H}' \times \mathscr{H} \to \mathbb{C}$ is the duality product between $\mathscr{H}'$ and $\mathscr{H}$. We have hence

$$\boldsymbol{\Phi}(h) = \boldsymbol{\Psi}(S_{\mathscr{H}}(h)), \; \forall h \in \mathscr{H}, \quad \text{with} \quad \boldsymbol{\Psi}(g) = (g(\varphi_1), \dots, g(\varphi_L)), \; \forall g \in \mathscr{H}''.$$

Since $S_{\mathscr{H}}$ is an isometry, we have moreover $\|h\|_{\mathscr{H}} = |||S_{\mathscr{H}}(h)|||''$ where $|||\cdot|||''$ is the bidual norm. This allows us to rewrite (2.14) as the equivalent problem:

$$S_{\mathscr{H}}(\mathcal{V}) = \mathcal{V}'' = \underset{g \in \mathscr{H}''}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Psi}(g)\right) \; + \; \Lambda\left(|||g|||''\right) \right\}. \tag{2.15}$$

We then apply Theorem 2.9 to (2.15), choosing $\mathscr{B}' = \mathscr{H}''$ and $\mathscr{B} = \mathscr{H}'$. Since $\mathscr{H}''$ is a Hilbert space (since isomorphic to $\mathscr{H}$), it is strictly convex and Assumption B6 of Theorem 2.9 is verified. The solution is hence unique

$\mathcal{V}'' = \{g^\star\}$ and such that

$$g^\star \in J_{\mathscr{H}'}\left(\sum_{i=1}^{L} \alpha_i \varphi_i\right),$$

for some weights $\{\alpha_1, \ldots, \alpha_L\} \subset \mathbb{C}^L$, where $J_{\mathscr{H}'} : \mathscr{H}' \to \mathscr{H}''$ is the duality map on $\mathscr{H}'$. Since $\mathscr{H}''$ is a (strictly convex) Hilbert space, the duality map is moreover single valued and antilinear, yielding:

$$g^\star = \sum_{i=1}^{L} \bar{\alpha}_i J_{\mathscr{H}'}(\varphi_i).$$

The solution $h^\star$ to (2.14) is then obtained by applying the linear inverse map $S_{\mathscr{H}}^{-1} : \mathscr{H}'' \to \mathscr{H}$:

$$h^\star = S_{\mathscr{H}}^{-1}\left(\sum_{i=1}^{L} \bar{\alpha}_i J_{\mathscr{H}'}(\varphi_i)\right) = \sum_{i=1}^{L} \bar{\alpha}_i S_{\mathscr{H}}^{-1}\left(J_{\mathscr{H}'}(\varphi_i)\right).$$

Finally, notice that $S_{\mathscr{H}}^{-1} J_{\mathscr{H}'} : \mathscr{H}' \to \mathscr{H}$ is an isometric isomorphism between $\mathscr{H}'$ and $\mathscr{H}$, which is nothing else but the Riesz map $R_{\mathscr{H}} : \mathscr{H}' \to \mathscr{H}$. This finally yields the desired result

$$h^\star = \sum_{i=1}^{L} \beta_i R_{\mathscr{H}}(\varphi_i),$$

with $\beta_i = \bar{\alpha}_i$, $i = 1, \ldots, L$. ■

Remark 2.3 *The above theorem can be seen as a generalisation of the classical result [191, Theorem 16.1], which is obtained by choosing the cost function F as the indicator function of some convex and compact subset of $\mathbb{C}^L$.*

2.3 Extreme Point Representer Theorem

When the duality map is single-valued, i.e. $\mathscr{B}'$ is strictly convex, Theorem 2.8 tells us that the unique solution to (2.11) is the Banach conjugate of a certain linear combination of the sampling functionals. When the duality map is set-valued however, things are slightly more complicated. In this case, Theorem 2.8 tells us that the solutions to (2.11) are among the Banach conjugates of a common linear combination of the sampling functionals. This characterisation of the solution set is of course much weaker, since the theorem does not tell us which of the Banach conjugates are actually solutions of (2.11). Moreover, computing all candidate Banach conjugates can in practice be complicated if not impossible. In this section, we propose hence an alternative characterisation of $\mathcal{V}$ in the case where $\mathscr{B}'$ is not strictly convex. To this end, we introduce the concept of *extreme point* [178] of a convex set:

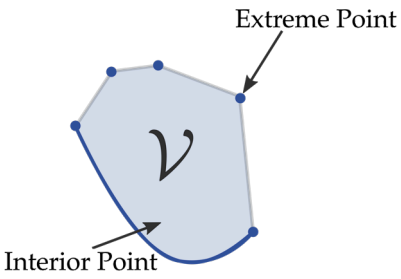


Figure 2.1: The convex set $\mathcal{V}$ is the closed convex hull (light blue) of its extreme points (dark blue).

Definition 2.4 — *Extreme Point.* *Let $\mathcal{V}$ be a* convex subset *of some topological vector space $\mathcal{X}$. An* extreme point *$v \in \mathcal{V}$ is a point such that*

$$\nexists (w, \nu) \in \mathcal{V}^2,\ \gamma \in]0,1[: \quad v = \gamma w + (1-\gamma)\nu.$$

In plain words, v is a point in $\mathcal{V}$ which does not lie in any open line segment joining two points of $\mathcal{V}$. We moreover call the set of all extreme points of $\mathcal{V}$ the extreme set*, denoted by $\delta\mathcal{V}$, and its complementary $\mathcal{V}^\circ = \mathcal{V} \backslash \delta\mathcal{V}$ the* interior set.

Extreme points are particularly convenient as any closed convex set $\mathcal{V}$ with extreme set $\delta\mathcal{V} \subset \mathcal{V}$ can be represented as the closed *convex hull* of its extreme points (see Fig. 2.1):

$$\mathcal{V} = \overline{\left\{ \sum_{k=1}^{n} \alpha_{i_k} v_{i_k} \;\middle|\; n \in \mathbb{N},\ i_k \in \mathbb{N},\ \sum_{k=1}^{n} \alpha_{i_k} = 1, \text{and } \ 0 \leq \alpha_{i_k} \leq 1, v_{i_k} \in \delta\mathcal{V} \right\}}. \tag{2.16}$$

[14] w.r.t. the topology on $\mathcal{X}$.

Note that $\mathcal{V}$ being closed,[14] it also contains infinite summations as limits of convergent sequences of finite summations. Finally, the *Krein-Milman theorem* [153, p. 75] tells us which sets in a *locally-convex space* admit extreme points and can be represented as the convex hull of these extreme points:

Theorem 2.11 — *Krein-Milman [153].* A *convex and compact* set $\mathcal{V}$ in a *locally convex topological space* is the *closed convex hull* of its *extreme points*. In particular, such a set has extreme points.

In the following theorem, we use Theorem 2.11 as well as results from [28, 72, 177] to characterise the solution set $\mathcal{V}$ of (2.11) as the weak* closed convex hull of extreme points with bounded df. Our result generalises [177, Theorem 6] to the case of non strictly convex cost functionals F, often encountered in practice.

Theorem 2.12 — *Extreme Point Representer Theorem.* Consider the following assumptions:

D1 $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ is a *Banach space*, with *topological dual* $(\mathscr{B}', |||\cdot|||)$;
D2 $\operatorname{span}\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{B}$, with the φ_i being *linearly independent*;
D3 $\boldsymbol{\Phi} : \mathscr{B}' \to \mathbb{C}^L$ is a *sampling* operator, defined as

$$\boldsymbol{\Phi}(f) = (\langle f|\varphi_1\rangle, \cdots, \langle f|\varphi_L\rangle), \quad \forall f \in \mathscr{B}';$$

D4 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

is *proper*, *convex* and lower semi-continuous;
D5 $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is some *arbitrary strictly increasing convex* function.
Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the solution set of the optimisation problem

$$\mathcal{V} = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \; + \; \Lambda\left(|||f|||\right) \right\}, \tag{2.17}$$

is *non-empty* and the *weak* closed convex hull* of its *extreme points*. The latter are moreover necessarily of the form:

$$f^\star = \sum_{m=1}^{M} \alpha_m e_m, \tag{2.18}$$

where $1 \leq M \leq L$, $\{\alpha_1, \dots, \alpha_M\} \subset \mathbb{C}$ and e_m are *extreme points* of the closed unit regularisation ball $\mathcal{B} := \{f \in \mathscr{B}' : \Lambda(|||f|||) \leq 1\}$.

Proof. Using the exact same arguments as in part i) of [177, Theorem 5] (which remain valid under the assumptions of Theorem 2.12), one can show that the functional $f \mapsto F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \; + \; \Lambda\left(|||f|||\right)$ is *proper, weak* lower semi-continuous, convex and coercive* on $\mathscr{B}'$. From [72, Proposition 8] the solution set $\mathcal{V}$ is hence *non-empty*, *convex* and *weak* compact*. Since $\mathscr{B}'$ equipped with the weak* topology is *locally convex and Hausdorff*, we can moreover invoke the *Krein-Milman theorem* to conclude that $\mathcal{V}$ is the *weak* closed convex hull of its extreme points*. In particular it has extreme points. Let $f_e \in \mathcal{V}$ be an arbitrary extreme point of $\mathcal{V}$ and let $\boldsymbol{z}_e := \boldsymbol{\Phi}(f_e) \in \mathbb{C}^L$. Then f_e is also in the solution set of the generalised interpolation problem

$$\mathcal{V}_e = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ \Lambda\left(|||f|||\right) \quad \text{s.t.} \quad \boldsymbol{\Phi}(f) = \boldsymbol{z}_e \right\}. \tag{2.19}$$

Using [28, Theorem 3.1] (with $j = 0$) we can moreover show that extreme points of $\mathcal{V}_e$ is of the form (2.18). Since $\mathcal{V}_e \subset \mathcal{V}$ and $f_e \in \mathcal{V}_e$, f_e is also an extreme point of $\mathcal{V}_e$ and hence is indeed of the form (2.18). This shows that every extreme point of $\mathcal{V}$ is of the form (2.18), which achieves the proof. ■

3

Fourier Analysis on the Hypersphere

The class of spherical pseudo-differential operators introduced in Chapter 4 are defined implicitly in the Fourier domain. In this chapter, we hence introduce the basic mathematical machinery needed for performing Fourier analysis on the hypersphere. The material presented in this chapter is based on the formalism adopted in [79, 125, 142].

1 Spherical Harmonics

One possible route towards defining the Fourier basis on the hypersphere is to proceed analogously to Fourier and study fundamental solutions of the *heat differential equation* on $\mathbb{S}^{d-1}$:

$$\frac{\partial u}{\partial t} - \alpha \Delta_{\mathbb{S}^{d-1}} u = 0, \quad \alpha > 0, \tag{3.1}$$

with $u : \mathbb{R}_+ \times \mathbb{S}^{d-1} \to \mathbb{C}$. The operator $\Delta_{\mathbb{S}^{d-1}}$ in (3.1) is called the *Laplace-Beltrami operator* on $\mathbb{S}^{d-1}$, and generalises the *Laplace operator* $\Delta_{\mathbb{R}^d}$ in $\mathbb{R}^d$ to the manifold setting. Both operators are linked by the relationship:

$$\Delta_{\mathbb{R}^d} = \frac{\partial^2}{\partial \rho^2} + \frac{d-1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \Delta_{\mathbb{S}^{d-1}}, \tag{3.2}$$

where, for every $\boldsymbol{x} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, we define $\boldsymbol{x} := \rho \boldsymbol{r}$ with $\rho := \|\boldsymbol{x}\|_{\mathbb{R}^d} \in \mathbb{R}_+$ and $\boldsymbol{r} \in \mathbb{S}^{d-1}$. The separation of variables technique reveals that the spherical component of such fundamental solutions are eigenfunctions of the *Laplace-Beltrami operator* on $\mathbb{S}^{d-1}$ (see [125, Chapter 5]). They are called *spherical harmonics*.

Spherical harmonics are orthonormal eigenfunctions of the Laplace-Beltrami operator on $\mathbb{S}^{d-1}$.

Definition 3.1 — *Spherical Harmonics.* *Let $\Delta_{\mathbb{S}^{d-1}}$ be the* Laplace-Beltrami operator *on $\mathbb{S}^{d-1}$ with* spectrum $\{\lambda_n = -n(n+d-2),\ n \in \mathbb{N}\}$. *We call* spherical harmonic *of order n any* eigenfunction Y *in the* eigenspace $Harm_n(\mathbb{S}^{d-1})$ *associated to the* eigenvalue λ_n*:*

$$Harm_n(\mathbb{S}^{d-1}) := \left\{ Y : \mathbb{S}^{d-1} \to \mathbb{C} \,|\, \Delta_{\mathbb{S}^{d-1}} Y = -n(n+d-2) Y \right\}.$$

Moreover, we denote by

$$\mathfrak{B}_n := \{Y_n^m,\ m = 1, \dots, N_d(n)\}$$

any orthonormal basis *of* $\text{Harm}_n(\mathbb{S}^{d-1})$*, where* $N_d(n)$ *is the* geometric multiplicity *of the eigenvalue* $\lambda_n = -n(n+d-2)$.

Remark 3.1 — *Geometric Multiplicity.* *The geometric multiplicity* $N_d(n)$ *of each eigenspace* $\text{Harm}_n(\mathbb{S}^{d-1})$ *can be computed explicitly [79, Chapter 2], and is given in general by:*

$$N_d(0) = 1, \qquad \& \qquad N_d(n) = \frac{2n+d-2}{n}\binom{n+d-3}{n-1}, \quad n \geq 1.$$

In particular, for $d = 2, 3$*, we get*

$$N_2(n) = 2, \qquad N_3(n) = 2n+1, \qquad n \geq 1.$$

We have moreover the asymptotic behaviour [79, Chapter 2]:

$$N_d(n) = \mathcal{O}\left(n^{d-2}\right). \tag{3.3}$$

The eigenspaces $\text{Harm}_n(\mathbb{S}^{d-1})$ are orthogonal for different eigenvalues λ_n, allowing us to represent the space of *spherical harmonics of order at most* n[1] as: $BL_n(\mathbb{S}^{d-1}) = \bigoplus_{k=0}^{n} \text{Harm}_k(\mathbb{S}^{d-1})$. It is possible to show[2] [125, Chapter 5] that this space becomes *asymptotically dense* in $\mathscr{L}^2(\mathbb{S}^{d-1})$:

$$\mathscr{L}^2(\mathbb{S}^{d-1}) = \bigoplus_{n=0}^{+\infty} \text{Harm}_n(\mathbb{S}^{d-1}).$$

[1] BL_n is the spherical analog to the space of *bandlimited* functions in traditional Fourier analysis.

[2] The proof relies on the equivalent characterisation of spherical harmonics as polynomials and the *Stone-Weierstrass theorem*.

This allows us to state the *Fourier expansion theorem* on the sphere:

Theorem 3.1 — *Spherical Fourier Expansion [125].* Let $d \geq 2$, $n \in \mathbb{N}$ and $\mathfrak{B}_n = \{Y_n^m, m = 1, \dots, N_d(n)\}$ be an *orthonormal basis* of $\text{Harm}_n(\mathbb{S}^{d-1})$. Then, every function $f \in \mathscr{L}^2(\mathbb{S}^{d-1})$ admits a *spherical Fourier expansion* given by

$$f \overset{\mathscr{L}^2}{=} \sum_{n=0}^{+\infty} \sum_{m=1}^{N_d(n)} \hat{f}_n^m\, Y_n^m, \tag{3.4}$$

where the *spherical Fourier coefficients* $\{\hat{f}_n^m\} \subset \mathbb{C}$ of f are given by the spherical harmonic transform (SHT):

$$\hat{f}_n^m = \langle f, Y_n^m \rangle_{\mathbb{S}^{d-1}} = \int_{\mathbb{S}^{d-1}} f(\boldsymbol{r})\overline{Y_n^m(\boldsymbol{r})}\, d\boldsymbol{r}, \quad n \in \mathbb{N},\ m = 1, \dots, N_d(n).$$

Remark 3.2 — $\mathscr{L}^2$ *convergence.* *Note that the equality (3.4) between* f *and its*

spherical Fourier expansion is in the $\mathscr{L}^2$ sense, i.e.

$$\lim_{N\to\infty} \left\| f - \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \hat{f}_n^m \, Y_n^m \right\|_2 = 0.$$

Uniform convergence can be achieved for sufficiently smooth *functions (see [125, Chapter 5]).*

Note that the spherical harmonics Y_n^m in Definition 3.1 and Theorem 3.1 are not uniquely specified, since there exists infinitely many orthonormal bases $\mathfrak{B}_n$ of $\mathrm{Harm}_n(\mathbb{S}^{d-1})$. In practice, the convention is to work with the system of so-called fully normalised spherical harmonics (FNSH),[3] obtained inductively by the method of separation of variables applied to the eigenvalue problem for the Laplace-Beltrami operator $\Delta_{\mathbb{S}^{d-1}}$. This construction is detailed in [125, Section 5.2] for the case $d = 2$ and in [167] for the general case. In all that follows, we will always assume the fully normalised spherical harmonics as canonical Fourier basis. The following example provides closed-form analytical expressions of the fully normalised spherical harmonics for $d = 2, 3$. Formulae for the more general case $d > 3$ are available in [167].

[3] The terminology "fully normalised" is slightly deceptive. Indeed, all Y_n^m are normalised, independently of the orthonormal system $\mathfrak{B}_n$ chosen on $\mathrm{Harm}_n(\mathbb{S}^{d-1})$.

Example 3.1 In this example, we detail the special cases of the *circle* $\mathbb{S}^1 \subset \mathbb{R}^2$ and the *sphere* $\mathbb{S}^2 \subset \mathbb{R}^3$.

- *Fourier expansion on the circle $\mathbb{S}^1$:* The eigenvalues of the Laplace-Beltrami operator are given by: $\lambda_n = -n^2,\ n \in \mathbb{N}$. They all have multiplicity $N_2(0) = 1$ and $N_2(n) = 2,\ n \geq 1$. The FNSH are given by

$$Y_0 = 1, \qquad Y_n^1 = e^{j2\pi\theta n}, \ Y_n^2 = e^{-j2\pi\theta n}, \quad n \geq 1.$$

  The Fourier expansion is then given for all $f \in \mathscr{L}^2(\mathbb{S}^1)$ by:

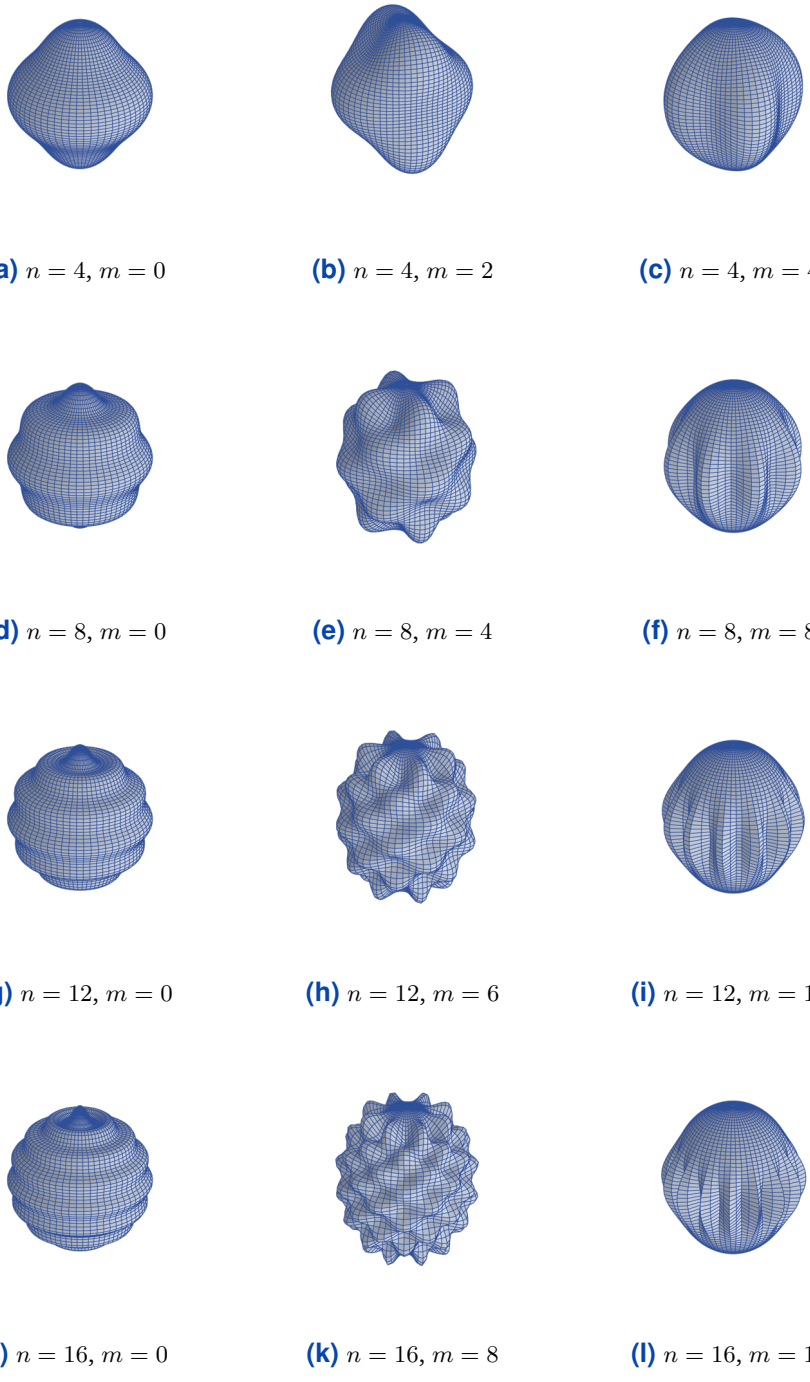
$$f(\theta) = \sum_{n=-\infty}^{+\infty} \hat{f}_n \, e^{j2\pi n\theta}, \quad \text{where} \quad \hat{f}_n = \int_{-\pi}^{\pi} f(\theta) e^{-j2\pi\theta n} \, d\theta,$$

  which corresponds to the traditional Fourier series expansion.
- *Fourier expansion on the sphere $\mathbb{S}^2$:* The eigenvalues of the Laplace-Beltrami operator are given by: $\lambda_n = -n(n+1),\ n \in \mathbb{N}$. They have multiplicity $N_3(n) = 2n + 1,\ n \geq 0$. In this case, it is customary to label the FNSH with m ranging from $-n$ to n, for each $n \in \mathbb{N}$. The latter are moreover given by [125, Section 5.2]

$$Y_n^m(\phi, \theta) := \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos(\theta)) e^{jm\phi}, \ \forall (\phi, \theta) \in [-\pi, \pi[\times[0, \pi],$$

  where $P_n^m : [-1, 1] \to \mathbb{R}$ denotes the *associated Legendre functions* [142, Chapter 1]. They can be classified into three groups:
  - *Zonal harmonics:* $\{Y_n^0,\ n \in \mathbb{N}\}$,
  - *Tesseral harmonics:* $\{Y_n^m,\ 0 < |m| < n,\ n \in \mathbb{N}\}$,

Zonal	Tesseral	Sectoral



(a) $n = 4, m = 0$ **(b)** $n = 4, m = 2$ **(c)** $n = 4, m = 4$

(d) $n = 8, m = 0$ **(e)** $n = 8, m = 4$ **(f)** $n = 8, m = 8$

(g) $n = 12, m = 0$ **(h)** $n = 12, m = 6$ **(i)** $n = 12, m = 12$

(j) $n = 16, m = 0$ **(k)** $n = 16, m = 8$ **(l)** $n = 16, m = 16$

Figure 3.1: Real part of selected fully normalised spherical harmonics Y_n^m for $\mathbb{S}^2$ and various n and m. As explained in Example 3.1, the spherical harmoncis are often classified into three types: *zonal*, *tesseral* and *sectoral*.

– *Sectoral harmonics:* $\{Y_n^m,\ m = \pm n,\ n \in \mathbb{N}\}$.

These designations are motivated by the patterns drawn by the zeros of the real and imaginary part of these functions on the sphere (see Fig. 3.1). Finally, the Fourier expansion is given for all $f \in \mathscr{L}^2(\mathbb{S}^2)$ by:

$$f \overset{\mathscr{L}^2}{=} \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} \hat{f}_n^m\, Y_n^m,$$

where

$$\hat{f}_n^m = \int_0^{\pi} \int_{-\pi}^{\pi} f(\phi,\theta)\overline{Y_n^m(\phi,\theta)}\, \sin(\theta) d\phi d\theta.$$

■

Remark 3.3 — *gSHT of Generalised Functions.* *It is possible to extend the SHT to generalised functions. Indeed, consider a generalised function*[4] $f \in \mathscr{S}'(\mathbb{S}^{d-1})$ *and a Schwartz function* $\varphi \in \mathscr{S}(\mathbb{S}^{d-1}) = \mathscr{C}^{\infty}(\mathbb{S}^{d-1})$*, with SHT*

[4] see Section 1.5.2 of Chapter 2 for a definition of generalised functions on the hypersphere.

$$\varphi = \sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m.$$

Then we have, from the bilinearity of the Schwartz duality product,

$$\begin{aligned}
\langle f|\varphi\rangle &= \lim_{N\to+\infty} \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m \overbrace{\langle f|Y_n^m\rangle}^{:=\hat{f}_n^m} \\
&= \lim_{N\to+\infty} \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \langle \varphi, Y_n^m\rangle_{\mathbb{S}^{d-1}}\, \hat{f}_n^m \\
&= \lim_{N\to+\infty} \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \langle Y_n^m|\varphi\rangle \hat{f}_n^m \\
&= \left\langle \lim_{N\to+\infty} \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \hat{f}_n^m Y_n^m \middle| \varphi \right\rangle, \qquad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}).
\end{aligned}$$

This observation motivates the definition of the generalised spherical harmonic transform (gSHT) of a generalised function $f \in \mathscr{S}'(\mathbb{S}^{d-1})$ *as*

$$f \overset{\text{weak}^*}{=} \sum_{n=0}^{+\infty} \sum_{m=1}^{N_d(n)} \hat{f}_n^m Y_n^m, \quad \text{where} \quad \hat{f}_n^m = \langle f|Y_n^m\rangle. \tag{3.5}$$

Notice that the convergence of the infinite series in (3.5) is w.r.t. to the weak topology and that the Fourier coefficients are well-defined. This is because the spherical harmonics are infinitely differentiable and hence in the predual* $\mathscr{S}(\mathbb{S}^{d-1})$ *of* $\mathscr{S}'(\mathbb{S}^{d-1})$.

2 Spherical Zonal Kernels

Another route towards defining the Fourier basis consists of looking at eigenfunctions of linear shift invariant systems, or *convolution operators* [183, Chapters 3 and 4]. As the hypersphere is a manifold, there is no intrinsically defined notion of convolution.[5] It is however possible to define a class of linear integral operators which "behave" as traditional convolution operators. In the Euclidean setting, convolution operators have *shift-invariant* kernels, whose value at a pair $(\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{R}^d$ depends only on the distance $\|\boldsymbol{r} - \boldsymbol{s}\|_{\mathbb{R}^d}$ between the two points. We can extend this notion to the hypersphere by noticing that the *chord distance* between two input directions $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}$ is given by:

$$\|\boldsymbol{r} - \boldsymbol{s}\|_{\mathbb{R}^d} = \sqrt{\|\boldsymbol{r}\|^2 + \|\boldsymbol{s}\|^2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle} = \sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}.$$

Notice that this quantity depends only on the *inner product* between the two directions $\boldsymbol{r}, \boldsymbol{s}$. This observation naturally leads to the notion of *zonal kernel*:

Definition 3.2 — *Spherical Zonal Kernel.* *A kernel $\Psi : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to \mathbb{C}$ is called a* spherical zonal kernel *if there exists a function $\psi : [-1, 1] \to \mathbb{C}$ such that:*

$$\Psi(\boldsymbol{r}, \boldsymbol{s}) = \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle), \quad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}.$$

For brevity, the function ψ is often abusively referred to as the zonal kernel *and no reference is made to Ψ.*

Vocabulary 3.1 — *Zonal Function.* *For every $\boldsymbol{s} \in \mathbb{S}^{d-1}$, we call the trace $\psi(\langle \cdot, \boldsymbol{s} \rangle) : \mathbb{S}^{d-1} \to \mathbb{C}$ of a zonal kernel a* zonal function.

Fig. 3.2 shows various traces of example zonal kernels. The plots make obvious the rotational invariance of zonal kernels, analogous to the shift-invariant kernels of convolution operators in Euclidean settings. Zonal kernels hence seem like good candidates to construct *spherical convolution operators* [125]:

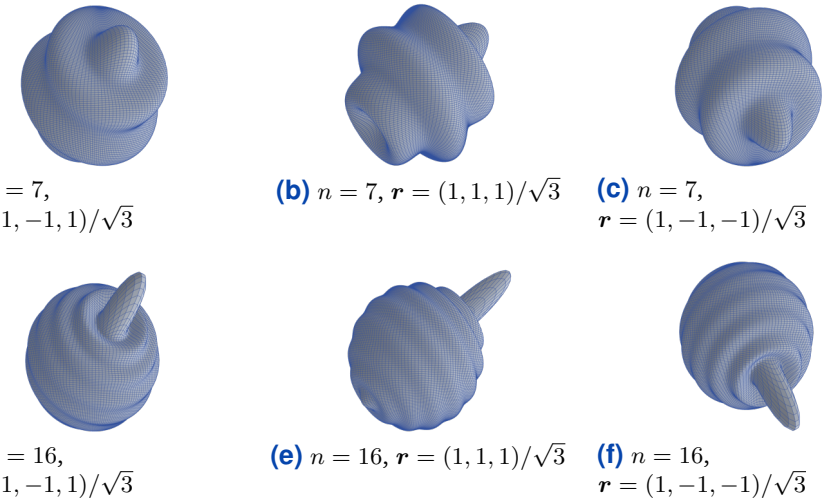
Definition 3.3 — *Spherical Convolution Operator.* *Let $\psi \in \mathscr{L}^2([-1, 1])$ be a* zonal kernel. *The* spherical convolution operator *$\mathscr{I}_\psi : \mathscr{L}^2(\mathbb{S}^{d-1}) \to \mathscr{L}^2(\mathbb{S}^{d-1})$ is defined as*

$$\mathscr{I}_\psi : \begin{cases} \mathscr{L}^2(\mathbb{S}^{d-1}) \to \mathscr{L}^2(\mathbb{S}^{d-1}) \\ f \mapsto \{\psi * f\}(\boldsymbol{r}) = \int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) f(\boldsymbol{s}) \, d\boldsymbol{s}, \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}. \end{cases} \tag{3.6}$$

Remark 3.4 *It is shown in [125, Theorem 7.2] that, under the assumptions of Definition 3.3, the image of the convolution operator (3.6) is indeed $\mathscr{L}^2(\mathbb{S}^{d-1})$.*

Remark 3.5 *Notice that the spherical convolution (3.6) is only defined between a zonal kernel and a function on the sphere but* not *between two arbitrary functions in $\mathscr{L}^2(\mathbb{S}^{d-1})$. This more general problem is addressed in [142, Chapter 1] for the case $d = 3$, where the authors propose to define the convolution between two*

[5] see Chapter 1 of [142] for an in-depth discussion on the topic.



(a) $n = 7$, $\boldsymbol{r} = (1, -1, 1)/\sqrt{3}$ **(b)** $n = 7$, $\boldsymbol{r} = (1, 1, 1)/\sqrt{3}$ **(c)** $n = 7$, $\boldsymbol{r} = (1, -1, -1)/\sqrt{3}$

(d) $n = 16$, $\boldsymbol{r} = (1, -1, 1)/\sqrt{3}$ **(e)** $n = 16$, $\boldsymbol{r} = (1, 1, 1)/\sqrt{3}$ **(f)** $n = 16$, $\boldsymbol{r} = (1, -1, -1)/\sqrt{3}$

Figure 3.2: Traces $\psi(\langle \cdot, \boldsymbol{r} \rangle)$ of the ultraspherical (a)-(c) and Shannon (d)-(f) zonal kernels for three focus directions $\boldsymbol{r} \in \mathbb{S}^2$. The ultraspherical kernel in (a)-(c) is the reproducing kernel of $\mathrm{Harm}_7(\mathbb{S}^2)$ (see Proposition 3.3). The Shannon kernel in (d)-(f) is the reproducing kernel of $BL_{16}(\mathbb{S}^2)$, the space of bandlimited functions on the sphere with bandwidth 16 (see Proposition 3.5). Observe that changing $\boldsymbol{r}$ simply rotates the kernel but does not change its shape.

functions on the sphere by looking at the correlation existing between rotated versions of themselves. Not surprisingly, eq. (3.6) *can also be understood in terms of rotations. Indeed, the effect of the term* $\langle \boldsymbol{r}, \boldsymbol{s} \rangle$ *in eq.* (3.6) *is to rotate and center the template function* ψ *around each direction* $\boldsymbol{r} \in \mathbb{S}^{d-1}$ *(see Fig. 3.2 for an example).*

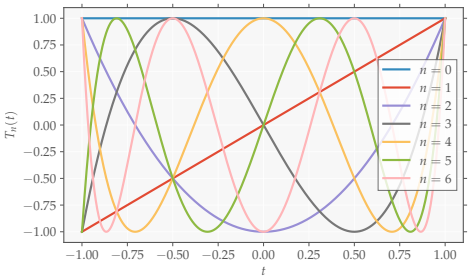
In the subsequent sections, we show that spherical convolution operators defined from zonal kernels are indeed diagonalised by the spherical harmonics. To this end, we first need to establish two important results: the *addition theorem* and the *Funk-Hecke formula*.

2.1 Ultraspherical Polynomials and Addition Theorem

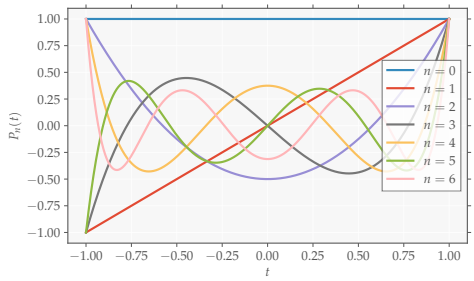
The *addition theorem* [79, 125] first, shows that the reproducing kernel of the RKHS $\mathrm{Harm}_n(\mathbb{S}^{d-1})$ is zonal and provides us with a closed-form expression for this kernel in terms of *ultraspherical polynomials* [79]. These polynomials are defined as follows:

Definition 3.4 — *Ultraspherical Polynomials.* *The* ultraspherical *or* Gegenbauer polynomials *are* polynomials $P_{n,d} : [-1, 1] \to \mathbb{R}$ *of degree* $n \in \mathbb{N}$*, defined via the Taylor expansion of generating functions:*

$$\frac{1 - ht}{1 + h^2 - 2ht} = \sum_{n=0}^{+\infty} P_{n,2}(t) h^n, \ h \in]0, 1[, \quad \text{for } d = 2,$$



(a) Chebyshev polynomials.



(b) Legendre polynomials.

Figure 3.3: Examples of Chebyshev and Legendre polynomials.

and:

$$\frac{1}{(1+h^2-2ht)^{(d-2)/2}} = \sum_{n=0}^{+\infty} P_{n,d}(t)h^n, \ h \in]0,1[, \quad \textit{for } d \geq 3.$$

Moreover, the ultraspherical polynomials are standardised *so that* $P_{n,d}(1) = 1$.

Remark 3.6 — *Orthogonality of Ultraspherical Polynomials.* *It is possible to show [79, Chapter 2] that the ultraspherical polynomials verify the orthogonality relationship:*

$$\int_{-1}^{1} P_{n,d}(t)P_{k,d}(t)(1-t^2)^{(d-3)/2}\,dt = \frac{\mathfrak{a}_d}{\mathfrak{a}_{d-1}N_d(n)}\delta_{nk}, \quad \forall n,k \in \mathbb{N},\ d \geq 2,$$

where δ_{nk} *is the Kronecker symbol and* $\mathfrak{a}_d$ *denotes the surface area of the hypersphere* $\mathbb{S}^{d-1}$.

Remark 3.7 — *Rodrigues' formula.* *The ultraspherical polynomials can also be defined via a recurrence relationship called* Rodrigues' formula *[79, Chapter 2].*

Example 3.2 We investigate here the special case of the *circle* $\mathbb{S}^1 \subset \mathbb{R}^2$ and the *sphere* $\mathbb{S}^2 \subset \mathbb{R}^3$.

- **Fourier expansion on the circle** $\mathbb{S}^1$**:** The ultraspherical polynomials

reduce to the *Chebyshev polynomials* (see Fig. 3.3a):

$$\begin{cases} P_{0,2}(t) & = 1 \\ P_{n,2}(t) & = T_n(t) = \cos(n \arccos(t)), \quad t \in [-1,1],\, n \geq 2. \end{cases}$$

They verify the recurrence relationship:

$$\begin{aligned} T_0(t) &= 1 \\ T_1(t) &= t \\ T_{n+1}(t) &= 2tT_n(t) - T_{n-1}(t), \quad n \geq 2. \end{aligned}$$

- **Fourier expansion on the sphere $\mathbb{S}^2$:** The ultraspherical polynomials reduce to the *Legendre polynomials* (see Fig. 3.3b), which verify *Bonnet's recurrence relationship*:

$$\begin{aligned} P_0(t) &= 1 \\ P_1(t) &= t \\ (n+1)P_{n+1}(t) &= (2n+1)tP_n(t) - nP_{n-1}(t), \quad n \geq 2. \end{aligned}$$

■

We are now ready to formulate the *addition theorem*:

Theorem 3.2 — *Addition Theorem [125].* Let $d \geq 2$, $n \in \mathbb{N}$, and $\mathfrak{B}_n = \{Y_n^m, m = 1, \ldots, N_d(n)\}$ be an *orthonormal basis* of $\mathrm{Harm}_n(\mathbb{S}^{d-1})$. Then, we have

$$\sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{r})\overline{Y_n^m(\boldsymbol{s})} = \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle), \quad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \tag{3.7}$$

where $P_{n,d} : [-1,1] \to \mathbb{R}$ are the *ultraspherical* polynomials of degree n, and $\mathfrak{a}_d > 0$ denote the surface area of $\mathbb{S}^{d-1}$.

Proof. See proof of [125, Theorem 5.11]. ■

Remark 3.8 *Notice that (3.7) holds for all choices of orthonormal systems $\mathfrak{B}_n$ on $\mathrm{Harm}_n(\mathbb{S}^{d-1})$, $n \in \mathbb{N}$, as is shown in [125, Theorem 5.11] and [79, Theorem 2.6].*

An immediate corollary of this theorem is

Proposition 3.3 — *Reproducing Kernel of $\mathrm{Harm}_n(\mathbb{S}^{d-1})$.* *The reproducing kernel of the RKHS $\mathrm{Harm}_n(\mathbb{S}^{d-1})$ is* zonal *and given by*

$$\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle), \quad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}.$$

It is called the ultraspherical kernel *of order $n \in \mathbb{N}$.*

Proof. From the Fourier expansion theorem we have, for every $f \in \mathrm{Harm}_n(\mathbb{S}^{d-1})$:

$$\begin{aligned} f(\boldsymbol{r}) &= \sum_{m=1}^{N_d(n)} \langle f, Y_n^m \rangle \, Y_n^m(\boldsymbol{r}) \\ &= \int_{\mathbb{S}^{d-1}} f(\boldsymbol{s}) \left[\sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{r}) \overline{Y_n^m(\boldsymbol{s})} \right] d\boldsymbol{s} \\ &= \int_{\mathbb{S}^{d-1}} f(\boldsymbol{s}) \left[\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) \right] d\boldsymbol{s}, \quad \forall \boldsymbol{r} \in \mathbb{S}^2, \end{aligned}$$

where the last equality results from the addition theorem. The kernel

$$\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle)$$

hence verifies the reproducing property for any $f \in \mathrm{Harm}_n(\mathbb{S}^{d-1})$:

$$f(\boldsymbol{r}) = \left\langle f, \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \cdot \rangle) \right\rangle \quad \forall \boldsymbol{r} \in \mathbb{S}^2.$$

■

Since the spaces $\{\mathrm{Harm}_n(\mathbb{S}^{d-1})\}_{n \in \mathbb{N}}$ are orthogonal, it follows that the ultraspherical kernels for different orders n are also orthogonal:

Proposition 3.4 — *Orthogonality of Ultraspherical Kernels.* *Let $d \geq 2$ and $m, n \in \mathbb{N}$. Then for every $(\boldsymbol{r}, \boldsymbol{\xi}) \in \mathbb{S}^{d-1}$, we have*

$$\int_{\mathbb{S}^{d-1}} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) P_{m,d}(\langle \boldsymbol{\xi}, \boldsymbol{s} \rangle) d\boldsymbol{s} = \frac{P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{\xi} \rangle) \mathfrak{a}_d}{N_d(n)} \delta_{mn}.$$

In particular, for $\boldsymbol{r} = \boldsymbol{\xi}$, we have

$$\langle P_{n,d}(\langle \boldsymbol{r}, \cdot \rangle), P_{m,d}(\langle \boldsymbol{r}, \cdot \rangle) \rangle_{\mathbb{S}^{d-1}} = \frac{\mathfrak{a}_d}{N_d(n)} \delta_{mn}.$$

Proof. First, note from the addition theorem that the function $P_{n,d}(\langle \boldsymbol{r}, \cdot \rangle)$ belongs to $\mathrm{Harm}_n(\mathbb{S}^{d-1})$ for every $\boldsymbol{r} \in \mathbb{S}^{d-1}$ and $n > 0$. Moreover, we have seen in Proposition 3.3 that the kernel $\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle)$ reproduces $\mathrm{Harm}_n(\mathbb{S}^{d-1})$. It is hence also the kernel of the orthogonal projection operator $\Pi_n : \mathscr{L}^2(\mathbb{S}^{d-1}) \to \mathrm{Harm}_n(\mathbb{S}^{d-1})$ with range $\mathrm{Harm}_n(\mathbb{S}^{d-1})$. Since the spaces $\mathrm{Harm}_n(\mathbb{S}^{d-1})$ are orthogonal, we have trivially for $m \neq n \in \mathbb{N}$:

$$\begin{aligned} \int_{\mathbb{S}^{d-1}} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) P_{m,d}(\langle \boldsymbol{\xi}, \boldsymbol{s} \rangle) d\boldsymbol{s} &= \frac{\mathfrak{a}_d}{N_d(n)} \int_{\mathbb{S}^{d-1}} P_{m,d}(\langle \boldsymbol{\xi}, \boldsymbol{s} \rangle) \left[\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) \right] d\boldsymbol{s} \\ &= \frac{\mathfrak{a}_d}{N_d(n)} \Pi_n \{ P_{m,d}(\langle \boldsymbol{\xi}, \cdot \rangle) \} = 0. \end{aligned}$$

When $m = n$, we leverage the reproducing property:

$$\begin{aligned}\int_{\mathbb{S}^{d-1}} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) P_{n,d}(\langle \boldsymbol{\xi}, \boldsymbol{s}\rangle) d\boldsymbol{s} &= \frac{\mathfrak{a}_d}{N_d(n)} \int_{\mathbb{S}^{d-1}} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) \left[\frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{\xi}, \boldsymbol{s}\rangle)\right] d\boldsymbol{s} \\ &= \frac{\mathfrak{a}_d}{N_d(n)} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{\xi}\rangle), \quad \forall (\boldsymbol{r}, \boldsymbol{\xi}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}.\end{aligned}$$

The last statement equality for $\boldsymbol{r} = \boldsymbol{\xi}$ follows trivially since $P_{n,d}(1) = 1$ (by definition). ■

We conclude this section by providing an expression for the reproducing kernel of the RKHS $BL_N(\mathbb{S}^{d-1})$ in terms of ultraspherical kernels:

Proposition 3.5 — *Reproducing Kernel of* $BL_N(\mathbb{S}^{d-1})$. *The reproducing kernel of the RKHS* $BL_N(\mathbb{S}^{d-1})$ *is* zonal *and given by*

$$\sum_{n=0}^{N} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle), \quad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}. \tag{3.8}$$

It is called the Shannon kernel *of order* $N \in \mathbb{N}$.

Proof. From the Fourier expansion theorem we have, for every $f \in BL_N(\mathbb{S}^{d-1})$:

$$\begin{aligned} f(\boldsymbol{r}) &= \sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} \langle f, Y_n^m \rangle Y_n^m(\boldsymbol{r}) \\ &= \int_{\mathbb{S}^{d-1}} f(\boldsymbol{s}) \left[\sum_{n=0}^{N} \sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{r}) \overline{Y_n^m(\boldsymbol{s})}\right] d\boldsymbol{s} \\ &= \int_{\mathbb{S}^{d-1}} f(\boldsymbol{s}) \left[\sum_{n=0}^{N} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle)\right] d\boldsymbol{s}, \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \end{aligned}$$

where the last equality results from the addition theorem. The kernel

$$\sum_{n=0}^{N} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle)$$

hence verifies the reproducing property for any $f \in BL_N(\mathbb{S}^{d-1})$:

$$f(\boldsymbol{r}) = \left\langle f, \sum_{n=0}^{N} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \cdot\rangle) \right\rangle \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}.$$

■

Example 3.3 We investigate here the special case of the *circle* $\mathbb{S}^1 \subset \mathbb{R}^2$ and the *sphere* $\mathbb{S}^2 \subset \mathbb{R}^3$.

- **Shannon kernel on the circle $\mathbb{S}^1$:** The Shannon kernel is called the *Dirichlet kernel* [183, Chapter 4]. Moreover, the summation in (3.8) simplifies to

$$D_N(\theta) = 1 + 2\sum_{n=1}^{N} T_n(\cos(\theta)) = \frac{\sin((N+1/2)\theta)}{\sin(\theta/2)}, \quad \theta \in [-\pi, \pi[,$$

  where $T_n : [-1,1] \to \mathbb{R}$ are the Chebyshev polynomials (see Example 3.2).
- **Shannon kernel on the sphere $\mathbb{S}^2$:** The *Shannon kernel* is given by [142, Christoffel formula], for $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}$,

$$\sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \begin{cases} (N+1)^2/(4\pi), & \text{if } \langle \boldsymbol{r}, \boldsymbol{s} \rangle = 1, \\ \frac{N+1}{4\pi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle - 1)} \left[P_{N+1}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) - P_N(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) \right], & \text{otherwise,} \end{cases}$$

  where $P_n : [-1,1] \to \mathbb{R}$ are the Legendre polynomials (see Example 3.2).

■

2.2 Funk-Hecke Formula

The next result needed to show that spherical harmonics diagonalise convolution operators is the *Funk-Hecke formula* [125, Chapter 7]. It shows that the zonal ultraspherical polynomials are eigenfunctions of convolution operators on $\mathbb{S}^{d-1}$:

Theorem 3.6 — *Funk-Hecke Formula.* Let $\psi \in \mathscr{L}^1([-1,1])$ and $n \in \mathbb{N}$. Then, we have $\forall n \in \mathbb{N}$:

$$\{\psi * P_{n,d}(\langle \boldsymbol{\eta}, \cdot \rangle)\}(\boldsymbol{r}) = \int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) P_{n,d}(\langle \boldsymbol{\eta}, \boldsymbol{s} \rangle) d\boldsymbol{s} = \hat{\psi}_n P_{n,d}(\langle \boldsymbol{\eta}, \boldsymbol{r} \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1},$$

where $\hat{\psi}_n$, $n \in \mathbb{N}$ are the *Fourier-Legendre coefficients* of ψ.

Proof. See proof of [125, Theorem 7.3]. ■

The Fourier-Legendre coefficients $\hat{\psi}_n$ above are obtained via a d-dimensional *Fourier-Legendre transform* [125, Chapter 3]:

Theorem 3.7 — *Fourier-Legendre Expansion.* Let $d \geq 2$ and $\{P_{n,d} : [-1,1] \to \mathbb{C},\ n \in \mathbb{N}\}$ the *ultraspherical polynomials*. Then, any function $\psi : [-1,1] \to \mathbb{C}$ such that

$$\int_{-1}^{1} |\psi(t)|^2 (1-t^2)^{(d-3)/2} dt < +\infty$$

admits a d-dimensional *Fourier-Legendre expansion* given by

$$\psi \overset{\omega-a.e.}{=} \sum_{n=0}^{+\infty} \hat{\psi}_n \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d},$$

with $\omega(t) = (1-t^2)^{(d-3)/2}$, $\mathfrak{a}_1 = 1$, and d-dimensional *Fourier-Legendre coefficients*

$$\hat{\psi}_n := \mathfrak{a}_{d-1} \int_{-1}^{1} \psi(t) P_{n,d}(t)(1-t^2)^{(d-3)/2} dt, \quad n \geq 0. \tag{3.9}$$

Remark 3.9 — *Computation of the Legendre Coefficients.* *In practice, the coefficients can be computed numerically using* Gaussian quadrature *with nodes and weights preserving the orthogonality property of the ultraspherical polynomials.*

We finally obtain from Theorem 3.6 and Proposition 3.3 that spherical harmonics diagonalise convolution operators:

Proposition 3.8 — *Spherical Harmonics & Spherical Convolution.* *Consider a function $\psi \in \mathscr{L}^1([-1,1])$ and a fixed $n \in \mathbb{N}$. Let moreover $\mathfrak{B}_n = \{Y_n^m, m = 1, \ldots, N_d(n)\}$ be an* orthonormal basis *of $\mathrm{Harm}_n(\mathbb{S}^{d-1})$. Then, we have:*

$$\{\psi * Y_n^m\}(\boldsymbol{r}) = \int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) Y_n^m(\boldsymbol{s}) d\boldsymbol{s} = \hat{\psi}_n Y_n^m(\boldsymbol{r}), \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1},$$

where $\{\hat{\psi}_n,\ n \in \mathbb{N}\} \subset \mathbb{C}$ are the d-dimensional Fourier-Legendre coefficients *of ψ (3.9).*

Proof. We have $Y_n^m \in \mathrm{Harm}_n(\mathbb{S}^{d-1})$. Using the reproducing kernel derived in Proposition 3.3, we get

$$\begin{aligned}
\{\psi * Y_n^m\}(\boldsymbol{r}) &= \int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) Y_n^m(\boldsymbol{s}) d\boldsymbol{s} \\
&= \int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) \left[\int_{\mathbb{S}^{d-1}} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{s}, \boldsymbol{\eta} \rangle) Y_n^m(\boldsymbol{\eta}) d\boldsymbol{\eta} \right] d\boldsymbol{s} \\
&= \frac{N_d(n)}{\mathfrak{a}_d} \int_{\mathbb{S}^{d-1}} \left[\int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) P_{n,d}(\langle \boldsymbol{s}, \boldsymbol{\eta} \rangle) d\boldsymbol{s} \right] Y_n^m(\boldsymbol{\eta}) d\boldsymbol{\eta}.
\end{aligned} \tag{3.10}$$

From Theorem 3.6 we have

$$\int_{\mathbb{S}^{d-1}} \psi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) P_{n,d}(\langle \boldsymbol{s}, \boldsymbol{\eta} \rangle) d\boldsymbol{s} = \hat{\psi}_n P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{\eta} \rangle). \tag{3.11}$$

Injecting (3.11) into (3.10) and using once more the reproducing property yields

$$\{\psi * Y_n^m\}(\boldsymbol{r}) = \hat{\psi}_n \int_{\mathbb{S}^{d-1}} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{\eta} \rangle) Y_n^m(\boldsymbol{\eta}) d\boldsymbol{\eta} = \hat{\psi}_n Y_n^m(\boldsymbol{r}).$$

■

4

Hyperspherical Splines

In this chapter we introduce *hyperspherical splines* –or *spherical splines* for short, which play a central role in spherical approximation theory [125, Chapter 6]. To this end, we extend the approach of [180, Chapter 6] to the spherical setting and construct spherical splines as "primitives" of finite Dirac streams w.r.t. a certain class of *pseudo-differential operators*, called *spline-admissible*. In short, spline-admissible operators are such that their fundamental solutions, called *Green functions*, are *ordinary* functions.[1] We derive a sufficient condition for spline-admissibility and provide examples of spline-admissible operators among the pseudo-differential operators most commonly used in practice.

[1] Ordinary functions are functions which are everywhere defined pointwise.

1 Spherical Pseudo-Differential Operators

By analogy with the Euclidean case, we define spherical pseudo-differential operators as *Fourier multipliers* with *slowly growing* spectra.

Definition 4.1 — *Spherical Pseudo-Differential Operator.* *We call* spherical pseudo-differential operator *any* linear *operator of the form*

$$\mathscr{D} : \begin{cases} \mathscr{S}(\mathbb{S}^{d-1}) \to \mathscr{S}(\mathbb{S}^{d-1}) \\ h \mapsto \mathscr{D}h := \sum_{n=0}^{+\infty} \hat{D}_n \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m \right], \end{cases} \tag{4.1}$$

where $\{\hat{h}_n^m,\ n \in \mathbb{N},\ m = 1, \dots, N_d(n)\}$ *are the* spherical Fourier coefficients *of* h *and* $\{\hat{D}_n\}_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ *is a sequence of* real numbers *such that the set*

$$\mathfrak{K}_{\mathscr{D}} := \left\{ n \in \mathbb{N} :\ |\hat{D}_n| = 0 \right\}, \tag{4.2}$$

is finite, *i.e.* $\#\mathfrak{K}_{\mathscr{D}} := N_0 < +\infty$, *and*

$$|\hat{D}_n| = \Theta\left(n^p\right), \tag{4.3}$$

for some real number $p \geq 0$, *called the* spectral growth order *of* $\mathscr{D}$.

Remark 4.1 *The functions Y_n^m in (4.1) denote the fully normalised spherical harmonics, and the convergence of the series is w.r.t. the canonical topology on $\mathscr{S}(\mathbb{S}^{d-1})$.*

Remark 4.2 — *Theta Notation.* *The condition $|\hat{D}_n| = \Theta(n^p)$ for some $p \geq 0$ means that $|\hat{D}_n| = \mathcal{O}(n^p)$ and $|\hat{D}_n| = \Omega(n^p)$, i.e. there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$ we have*

$$C_1 n^p \leq |\hat{D}_n| \leq C_2 n^p,$$

for some positive constants $C_1, C_2 \in \mathbb{R}_+$. In other words, the sequence $\{|\hat{D}_n|\}_{n\in\mathbb{N}}$ is asymptotically comparable to the polynomial n^p.

Pseudo-differential operators exhibit a "roughening" behaviour, similar to the derivative or Laplace-Beltrami operators.

Remark 4.3 — *Roughening Behaviour.* *Notice that a pseudo-differential operator $\mathscr{D}$ multiplies the Fourier coefficients of its argument h by a sequence $\{\hat{D}_n\}_{n\in\mathbb{N}}$ with polynomial growth order. This filtering operation effectively "boosts" the high frequency content of h, hence making it "rougher" (less regular). This behaviour is reminiscent of those of standard differential operators such as the prototypical* Laplace-Beltrami *operator.*

Vocabulary 4.1 — *Symbol of Pseudo-Differential Operator.* *In differential calculus, the sequence $\{\hat{D}_n\}_{n\in\mathbb{N}}$ is sometimes called the* symbol *of the pseudo-differential operator $\mathscr{D}$.*

As illustrated in the subsequent example, most of the pseudo-differential operators encountered in practice are constructed by transforming the Laplace-Beltrami operator $\Delta_{\mathbb{S}^{d-1}}$ by some polynomial or harmonic function:

Example 4.1 — *Common Pseudo-Differential Operators.* Consider the Laplace-Beltrami operator $\Delta_{\mathbb{S}^{d-1}}$ on $\mathbb{S}^{d-1}$, $d \geq 2$.

- **Iterated Laplace-Beltrami operators:** these operators are obtained as *integer powers* (i.e. successive compositions) of the Laplace-Beltrami operator $\mathscr{D} := \Delta_{\mathbb{S}^{d-1}}^k$, with $k \in \mathbb{N}$. They are indeed pseudo-differential operators (as a matter of fact they are even differential operators) since they can be written[a] as in (4.1) with

  $$\hat{D}_n = (-n(n+d-2))^k, \quad n \in \mathbb{N}.$$

  We have indeed $\hat{D}_n \in \mathbb{R}$ (since $k \in \mathbb{N}$), $|\hat{D}_n| = \Theta(n^{2k})$, and $\mathfrak{K}_{\mathscr{D}} = \{0\}$ is finite. Notice that $\Delta_{\mathbb{S}^{d-1}}^k$ is positive semi-definite for k even and negative semi-definite for k odd.
- **Fractional Laplace-Beltrami operators:** these operators are obtained as p-th *roots* of the negative Laplace-Beltrami operator

  $$\mathscr{D} := (-\Delta_{\mathbb{S}^{d-1}})^{1/p},$$

  with $p \in \mathbb{N}^*$. They are indeed pseudo-differential operators since they can be written as in (4.1) with

  $$\hat{D}_n = \sqrt[p]{n(n+d-2)}, \quad n \in \mathbb{N}.$$

Since $n(n+d-2) > 0, \forall n \geq 1, d \geq 2$, we have indeed $\hat{D}_n \in \mathbb{R}$. Moreover we also have $|\hat{D}_n| = \Theta(n^{2/p})$, and $\mathfrak{K}_{\mathscr{D}} = \{0\}$ is finite. Notice that $(-\Delta_{\mathbb{S}^{d-1}})^{1/p}$ is always positive semi-definite. The case $p = 2$ yields the square-root of the Laplace-Beltrami operator, which is intimately linked to the spherical *divergence* and *gradient* differential operators. The latter are however *vector-valued* differential operators, and hence do not belong to the class of pseudo-differential operators defined in Definition 4.1.

- **(Iterated) Beltrami operators:** these operators are defined as [60, Chapter 4]
$$\partial_k = \Delta_{\mathbb{S}^{d-1}} + k(k+d-2)\mathrm{Id},$$
where $k \in \mathbb{N}$ and Id denotes the identity operator. Such operators are indeed pseudo-differential operators, since their Fourier coefficients are given by
$$\hat{D}_n = k(k+d-2) - n(n+d-2), \quad n \in \mathbb{N},$$
and hence $\hat{D}_n \in \mathbb{R}$, $|\hat{D}_n| = \Theta(n^2)$, and $\mathfrak{K}_{\mathscr{D}} = \{k\}$ is finite. The *iterated Beltrami operators* are obtained by composing the Beltrami operators together:
$$\partial_{0\cdots k} = \partial_0 \partial_1 \cdots \partial_{k-1} \partial_k, \quad k \in \mathbb{N}.$$
They are also pseudo-differential operators since in this case, we have
$$\hat{D}_n = \prod_{j=0}^{k} (k(k+d-2) - n(n+d-2)), \quad n \in \mathbb{N},$$
and hence $\hat{D}_n \in \mathbb{R}$, $|\hat{D}_n| = \Theta(n^{2(k+1)})$, and $\mathfrak{K}_{\mathscr{D}} = \{0, \ldots, k\}$ is finite.
- **Sobolev operators:** these operators are defined as $\mathscr{D} := (\mathrm{Id} - \Delta_{\mathbb{S}^{d-1}})^{\beta}$, with $\beta > 0$. They are indeed pseudo-differential operators since their Fourier coefficients are given by
$$\hat{D}_n = (1 + n(n+d-2))^{\beta}, \quad n \in \mathbb{N},$$
and hence $\hat{D}_n \in \mathbb{R}$, $|\hat{D}_n| = \Theta(n^{2\beta})$, and $\mathfrak{K}_{\mathscr{D}} = \emptyset$. Notice that $(\mathrm{Id} - \Delta_{\mathbb{S}^{d-1}})^{\beta}$ is always positive definite.

■

[a] Recall that the spherical harmonics were defined as eigenfunctions of the Laplace-Beltrami operator: $\forall Y \in \mathrm{Harm}_n(\mathbb{S}^{d-1})$, $\Delta_{\mathbb{S}^{d-1}} Y = -n(n+d-2)Y$ (see Section 1 of Chapter 3).

In order to gain further insight on Definition 4.1 and the motivations behind it, it is helpful to look at some key properties of pseudo-differential operators.

Proposition 4.1 — *Properties of Pseudo-Differential Operators.* *Let $\mathscr{D}$ be a* spherical pseudo-differential operator *as in Definition 4.1. Then the following holds:*

1. *$\mathscr{D}$ is* self-adjoint, *i.e. $\mathscr{D}^* = \mathscr{D}$.*

2. $\mathscr{D}$ *is* isotropic, *i.e. any* $Y \in Harm_n(\mathbb{S}^{d-1})$ *is an* eigenfunction *of* $\mathscr{D}$*, with associated eigenvalue* $\lambda_n = \hat{D}_n$.
3. $\mathscr{D}$ *has* finite-dimensional *nullspace, given by*

$$\mathcal{N}(\mathscr{D}) = \left\{ h \in \mathscr{S}(\mathbb{S}^{d-1}) : \langle h, Y_n^m \rangle_{\mathbb{S}^{d-1}} = 0,\ n \in \mathbb{N} \backslash \mathfrak{K}_{\mathscr{D}},\ m = 1, \dots, N_d(n) \right\}$$
$$= \textit{span}\left\{ Y_n^m,\ n \in \mathfrak{K}_{\mathscr{D}},\ m = 1, \dots, N_d(n) \right\}.$$

4. $\mathscr{D}$ *is an* endomorphism[2] *on* $\mathscr{S}(\mathbb{S}^{d-1})$*, i.e. it maps infinitely differentiable functions onto infinitely differentiable functions.*

[2] Incidentally, this shows the well-posedness of $\mathscr{D}$ in Definition 4.1.

Proof. We prove below Items 1 to 4 of Proposition 4.1.

1. It trivially follows from the fact that the coefficients $\hat{D}_n$ are real.
2. We expand Y in the orthogonal basis $\mathfrak{B}_n = \{Y_n^m,\ m = 1, \dots, N_d(n)\}$ of $\text{Harm}_n(\mathbb{S}^{d-1})$: $Y = \sum_{m=1}^{N_d(n)} \langle Y, Y_n^m \rangle_{\mathbb{S}^{d-1}} Y_n^m$. From Definition 4.1, we have hence $\mathscr{D}Y = \hat{D}_n \sum_{m=1}^{N_d(n)} \langle Y, Y_n^m \rangle_{\mathbb{S}^{d-1}} Y_n^m = \hat{D}_n Y$. Y is therefore an eigenfunction of $\mathscr{D}$ with associated eigenvalue $\hat{D}_n$.
3. Let $h \in \mathscr{S}(\mathbb{S}^{d-1})$ be such that $\mathscr{D}h = 0$. Then, we have from Definition 4.1 $\sum_{n=0}^{+\infty} \hat{D}_n \sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m = 0$, where $\hat{h}_n^m = \langle h, Y_n^m \rangle_{\mathbb{S}^{d-1}}$. Since $\{Y_n^m, n \in \mathbb{N}, m = 1, \dots, N_d(n)\}$ forms a closed and complete orthonormal system of $\mathscr{L}^2(\mathbb{S}^{d-1})$, this implies that

$$(\hat{D}_n = 0) \vee (\hat{h}_n^m = 0) \quad \forall n \in \mathbb{N}, m = 1, \dots, N_d(n).$$

When $n \in \mathfrak{K}_{\mathscr{D}}$, we have $|\hat{D}_n| = 0$ and hence there is no constraint on $\hat{h}_n^m$. When $n \in \mathbb{N} \backslash \mathfrak{K}_{\mathscr{D}}$ however, we have $|\hat{D}_n| > 0$ and hence necessarily $\hat{h}_n^m = \langle h, Y_n^m \rangle_{\mathbb{S}^{d-1}} = 0$. This yields the following characterisation of the nullspace $\mathcal{N}(\mathscr{D})$ of $\mathscr{D}$:

$$\mathcal{N}(\mathscr{D}) = \left\{ h \in \mathscr{S}(\mathbb{S}^{d-1}) : \langle h, Y_n^m \rangle_{\mathbb{S}^{d-1}} = 0,\ n \in \mathbb{N} \backslash \mathfrak{K}_{\mathscr{D}},\ m = 1, \dots, N_d(n) \right\}$$

From the SHT of h, we have moreover that $h = \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \hat{h}_n^m Y_n^m$. This provides us with a second characterisation of $\mathcal{N}(\mathscr{D})$:

$$\mathcal{N}(\mathscr{D}) = \text{span}\left\{ Y_n^m,\ n \in \mathfrak{K}_{\mathscr{D}},\ m = 1, \dots, N_d(n) \right\}.$$

Since $\mathfrak{K}_{\mathscr{D}}$ is by assumption finite, the collection of spherical harmonics spanning $\mathcal{N}(\mathscr{D})$ is finite and hence the nullspace is indeed finite-dimensional.
4. We use the *Sobolev embedding theorem*, that tells us that the *angular power spectrum sequence* $\{S_n(h) := \sum_{m=1}^{N_d(n)} |\hat{h}_n^m|^2,\ n \in \mathbb{N}\}$ of a Schwartz function $h \in \mathscr{S}(\mathbb{S}^{d-1})$ decays faster than any rational function, i.e.

$$S_n(h) = \mathcal{O}\left(n^{-k}\right), \qquad \forall k \in \mathbb{N}.$$

Since the sequence $\{|\hat{D}_n|^2,\ n \in \mathbb{N}\}$ has at most polynomial growth of order $2p$ from (4.3), the angular power spectrum sequence of $\mathscr{D}h$ still

decays faster than any rational function –as multiplication between two sequences with supra-rational decay and polynomial growth, and is hence in $\mathscr{S}(\mathbb{S}^{d-1})$. $\mathscr{D}$ is hence indeed an endomorphism[3] on $\mathscr{S}(\mathbb{S}^{d-1})$. ■

[3] This fact is also discussed in [10] for the special case $d = 2$.

The proof of Proposition 4.1 reveals intimate links between the assumptions of Definition 4.1 and properties 1 to 4 of pseudo-differential operators:

- Properties 1 and 3 are direct consequences of the fact that the sequence $\{\hat{D}_n\}_{n\in\mathbb{N}}$ is respectively real and null for at most finitely many integers.
- The isotropy property 2 is implicitly assumed in Definition 4.1 since the spectrum of $\mathscr{D}$ was chosen in (4.1) to be constant for all $m = 1, \dots, N_d(n)$ in a given frequency level $n \in \mathbb{N}$. As shall be seen in Section 2, this construction guarantees that –when they exist– the spherical splines associated to a given pseudo-differential operator are sums of zonal functions, and hence fast to evaluate.
- Property 4 finally, results from the polynomial growth of the sequence $\{\hat{D}_n\}_{n\in\mathbb{N}}$. More specifically, it results from $\{\hat{D}_n\}_{n\in\mathbb{N}}$ being asymptotically bounded *from above* by a polynomial sequence $|\hat{D}_n| = \mathcal{O}(n^p)$, implied[4] by $|\hat{D}_n| = \Theta(n^p)$.

[4] As explained in Remark 4.2, $|\hat{D}_n| = \Theta(n^p)$ is equivalent to $|\hat{D}_n| = \mathcal{O}(n^p)$ and $|\hat{D}_n| = \Omega(n^p)$, i.e. $|\hat{D}_n|$ is asymptotically bounded from above and below by the polynomial sequence.

Surprisingly, the requirement that $\{\hat{D}_n\}_{n\in\mathbb{N}}$ is also asymptotically bounded *from below* by a polynomial sequence –i.e. $|\hat{D}_n| = \Omega(n^p)$– has not been used so far. This assumption comes into play when considering *primitives* w.r.t. a particular pseudo-differential operator $\mathscr{D}$, obtained via the *Moore-Penrose pseudo-inverse* $\mathscr{D}^\dagger$ of $\mathscr{D}$.

Proposition 4.2 — *Moore-Penrose Pseudo-Inverse of* $\mathscr{D}$. *Let $\mathscr{D}$ be a pseudo-differential operator as in Definition 4.1. The* Moore-Penrose pseudo-inverse $\mathscr{D}^\dagger$ *of $\mathscr{D}$ is given by*

$$\mathscr{D}^\dagger : \begin{cases} \mathscr{S}(\mathbb{S}^{d-1}) \to \mathscr{S}(\mathbb{S}^{d-1}) \\ h \mapsto \mathscr{D}^\dagger h := \displaystyle\sum_{n\notin\mathfrak{K}_\mathscr{D}} \frac{1}{\hat{D}_n} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m \right], \end{cases} \tag{4.4}$$

where $\{\hat{h}_n^m,\ n \in \mathbb{N},\ m = 1, \dots, N_d(n)\}$ are the spherical Fourier coefficients *of h.*

Proof. First, notice that since $|\hat{D}_n| = \Theta(n^p)$, we have in particular $|\hat{D}_n| = \Omega(n^p)$ and hence $\mathscr{D}^\dagger h \in \mathscr{S}(\mathbb{S}^{d-1})$ for all $h \in \mathscr{S}(\mathbb{S}^{d-1})$ (using similar arguments as for the proof of point 4 of Proposition 4.1). The compositions $\mathscr{D}\mathscr{D}^\dagger$ and $\mathscr{D}^\dagger\mathscr{D}$ are hence well-defined. Finally, we have from (4.1) and (4.4) that, for all $h \in \mathscr{S}(\mathbb{S}^{d-1})$

$$\mathscr{D}^\dagger\mathscr{D}\mathscr{D}^\dagger h = \sum_{n\notin\mathfrak{K}_\mathscr{D}} \frac{\hat{D}_n}{\hat{D}_n^2} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m \right] = \sum_{n\notin\mathfrak{K}_\mathscr{D}} \frac{1}{\hat{D}_n} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m \right] \mathscr{D}^\dagger h,$$

$$\mathscr{D}\mathscr{D}^\dagger\mathscr{D}h = \sum_{n\notin\mathfrak{K}_\mathscr{D}} \frac{\hat{D}_n^2}{\hat{D}_n}\left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m\right] = \sum_{n\in\mathbb{N}} \hat{D}_n\left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m\right] = \mathscr{D}h,$$

which shows that $\mathscr{D}^\dagger$ is a generalised inverse of $\mathscr{D}$. Moreover, we have that

$$\mathscr{D}\mathscr{D}^\dagger h = \sum_{n\notin\mathfrak{K}_\mathscr{D}} \frac{\hat{D}_n}{\hat{D}_n}\left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m\right] = \sum_{n\notin\mathfrak{K}_\mathscr{D}} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m\right] = \mathscr{D}\mathscr{D}^\dagger h.$$

Since both $\mathscr{D}$ and $\mathscr{D}^\dagger$ are self-adjoint, we have $(\mathscr{D}\mathscr{D}^\dagger)^* = \mathscr{D}^\dagger\mathscr{D} = \mathscr{D}\mathscr{D}^\dagger$ and $(\mathscr{D}^\dagger\mathscr{D})^* = \mathscr{D}\mathscr{D}^\dagger = \mathscr{D}^\dagger\mathscr{D}$, which shows that $\mathscr{D}^\dagger$ is actually the Moore-Penrose pseudo-inverse of $\mathscr{D}$ and concludes the proof. ■

As discussed in the proof of Proposition 4.2, the pseudo-inverse $\mathscr{D}^\dagger$ is well-defined since $|\hat{D}_n| = \Omega(n^p)$. Note that the primitive operator $\mathscr{D}^\dagger$ acts as an *integral operator* and smooths out high frequency content with a polynomially *decaying* sequence. In what follows, we will sometimes need to extend by duality the action of $\mathscr{D}$ (respectively $\mathscr{D}^\dagger$) to generalised functions $\nu \in \mathscr{S}'(\mathbb{S}^{d-1})$. Since $\mathscr{D}$ (respectively $\mathscr{D}^\dagger$) is self-adjoint, this can easily be achieved by understanding $\mathscr{D}\nu$ as the element of $\mathscr{S}'(\mathbb{S}^{d-1})$ with point-wise definition:

$$\langle \mathscr{D}\nu | \varphi\rangle := \langle \nu | \mathscr{D}\varphi\rangle, \quad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}). \tag{4.5}$$

Equation (4.5) is indeed well-defined since, from Item 4 of Proposition 4.1, $\mathscr{D}$ is an endomorphism on $\mathscr{S}(\mathbb{S}^{d-1})$.

2 Green Functions and Spline-Admissibility

The next important ingredient for the definition of spherical splines is the notion of *Green function* of a pseudo-differential operator $\mathscr{D}$. A Green function is a fundamental solution of $\mathscr{D}$, obtained by taking the primitive of some Dirac measure.

Definition 4.2 — *Green Function.* *Let $\mathscr{D}$ be a* pseudo-differential operator *as in Definition 4.1. Consider moreover the* Moore-Penrose pseudo-inverse *$\mathscr{D}^\dagger$ of $\mathscr{D}$, extended into an endomorphism on $\mathscr{S}'(\mathbb{S}^{d-1})$ with (4.5). Then, a generalised function $\Psi_{\boldsymbol{s}}^\mathscr{D} \in \mathscr{S}'(\mathbb{S}^{d-1})$ is said to be a* Green function *for $\mathscr{D}$ if:*

$$\Psi_{\boldsymbol{s}}^\mathscr{D} = \mathscr{D}^\dagger \delta_{\boldsymbol{s}}, \tag{4.6}$$

where $\delta_{\boldsymbol{s}} \in \mathcal{M}(\mathbb{S}^{d-1}) \subset \mathscr{S}'(\mathbb{S}^{d-1})$ is the Dirac measure *for some direction $\boldsymbol{s} \in \mathbb{S}^{d-1}$.*

Remark 4.4 *From (4.5), the pointwise definition of $\Psi_{\boldsymbol{s}}^\mathscr{D}$ in (4.6) is, for every $\boldsymbol{s} \in \mathbb{S}^{d-1}$,*

$$\left\langle \Psi_{\boldsymbol{s}}^\mathscr{D} \middle| \varphi \right\rangle = \left\langle \mathscr{D}^\dagger\delta_{\boldsymbol{s}} \middle| \varphi \right\rangle = \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^\dagger\varphi \right\rangle, \quad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}). \tag{4.7}$$

Remark 4.5 — *Unicity.* *Observe that for each $\boldsymbol{s}$, there exists a unique Green function. Indeed, assume $\Psi_{\boldsymbol{s}}^{\mathscr{D}}, \Phi_{\boldsymbol{s}}^{\mathscr{D}} \in \mathscr{S}'(\mathbb{S}^{d-1})$ verify Definition 4.2. Then, by linearity of the Schwartz duality product we have*

$$\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} - \Phi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle - \left\langle \Phi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^{\dagger}\varphi \right\rangle - \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^{\dagger}\varphi \right\rangle = 0,$$

for all $\varphi \in \mathscr{S}(\mathbb{S}^{d-1})$. Therefore $\Psi_{\boldsymbol{s}}^{\mathscr{D}} = \Phi_{\boldsymbol{s}}^{\mathscr{D}}$.

In physics, it is more common to define a Green function $\Psi_{\boldsymbol{s}}^{\mathscr{D}}$ of a pseudo-differential operator $\mathscr{D}$ via the relationship $\left\langle \mathscr{D}\Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \langle \delta_{\boldsymbol{s}} | \varphi \rangle$, for every test function $\varphi \in \mathscr{S}(\mathbb{S}^{d-1})$ verifying some specific *boundary conditions*. The next proposition makes the link between Definition 4.2 and the physicist's point of view.

Proposition 4.3 — *Green Function (Physicist's Point of View).* *Let $\{\Psi_{\boldsymbol{s}}^{\mathscr{D}}, \boldsymbol{s} \in \mathbb{S}^{d-1}\} \subset \mathscr{S}'(\mathbb{S}^{d-1})$ be Green functions for a spline-admissible pseudo-differential operator $\mathscr{D}$. We have then, for each $\boldsymbol{s} \in \mathbb{S}^{d-1}$:*

$$\left\langle \mathscr{D}\Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \langle \delta_{\boldsymbol{s}} | \varphi \rangle, \qquad \forall \varphi \in \mathcal{N}(\mathscr{D})^{\perp}, \tag{4.8}$$

and

$$\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| Y_n^m \right\rangle = 0, \qquad \forall n \in \mathfrak{K}_{\mathscr{D}},\ m = 1, \ldots, N_d(n), \tag{4.9}$$

where $\mathcal{N}(\mathscr{D})^{\perp}$ is the orthogonal complement of $\mathcal{N}(\mathscr{D})$ in $\mathscr{S}(\mathbb{S}^{d-1})$, given by

$$\mathcal{N}(\mathscr{D})^{\perp} = \left\{ \varphi \in \mathscr{S}(\mathbb{S}^{d-1}) : \langle \varphi, Y_n^m \rangle_{\mathbb{S}^{d-1}} = 0,\ n \in \mathfrak{K}_{\mathscr{D}},\ m = 1, \ldots, N_d(n) \right\}.$$

Proof. From the pointwise definition (4.7) of $\Psi_{\boldsymbol{s}}$ we have, for each $\boldsymbol{s} \in \mathbb{S}^{d-1}$

$$\left\langle \mathscr{D}\Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \mathscr{D}\varphi \right\rangle = \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^{\dagger}\mathscr{D}\varphi \right\rangle, \qquad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}).$$

Moreover, for $\varphi \in \mathcal{N}(\mathscr{D})^{\perp}$ we have

$$\mathscr{D}^{\dagger}\mathscr{D}\varphi = \sum_{n \in \mathbb{N} \setminus \mathfrak{K}_{\mathscr{D}}} \frac{\hat{D}_n}{\hat{D}_n} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m = \sum_{n \in \mathbb{N}} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m = \varphi,$$

since by definition of $\mathcal{N}(\mathscr{D})^{\perp}$, $\hat{\varphi}_n^m = 0\ \forall n \in \mathfrak{K}(\mathscr{D})$. We have hence indeed

$$\left\langle \mathscr{D}\Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \langle \delta_{\boldsymbol{s}} | \varphi \rangle, \qquad \forall \varphi \in \mathcal{N}(\mathscr{D})^{\perp},$$

which proves (4.8). Finally we have, $\forall n \in \mathfrak{K}_{\mathscr{D}},\ m = 1, \ldots, N_d(n)$,

$$\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| Y_n^m \right\rangle = \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^{\dagger} Y_n^m \right\rangle = \sum_{n' \notin \mathfrak{K}_{\mathscr{D}}} \frac{1}{\hat{D}_{n'}} \left[\sum_{m'=1}^{N_d(n')} \delta_{n'n}\delta_{m'm} Y_n^m(\boldsymbol{s}) \right] = 0,$$

where $\delta_{n'n}$ is the Kronecker delta symbol, proving (4.9). ∎

The Green functions of an operator $\mathscr{D}$ can be expressed as traces of a certain zonal kernel, called the *zonal Green kernel*:

Proposition 4.4 — *Zonal Green Kernel.* *Let $\{\Psi_{\boldsymbol{s}}^{\mathscr{D}},\ \boldsymbol{s}\in\mathbb{S}^{d-1}\}\subset\mathscr{S}'(\mathbb{S}^{d-1})$ be Green functions for a spline-admissible pseudo-differential operator $\mathscr{D}$. We have then, for each $\boldsymbol{s}\in\mathbb{S}^{d-1}$:*

$$\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle = \int_{\mathbb{S}^{d-1}} \psi_{\mathscr{D}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle)\varphi(\boldsymbol{r})\,d\boldsymbol{r}, \quad \forall \varphi\in\mathscr{S}(\mathbb{S}^{d-1}). \tag{4.10}$$

The zonal Green kernel $\psi_{\mathscr{D}}$ *is moreover such that* $\{\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{s}\rangle),\ \boldsymbol{s}\in\mathbb{S}^{d-1}\}\subset\mathscr{S}'(\mathbb{S}^{d-1})$*, and is defined as*

$$\psi_{\mathscr{D}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) := \sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n} P_{n,d}(\langle \boldsymbol{r},\boldsymbol{s}\rangle), \quad \boldsymbol{r}\in\mathbb{S}^{d-1}, \tag{4.11}$$

where $\mathfrak{a}_d$ is the area of the unit sphere $\mathbb{S}^{d-1}$ and $P_{n,d}:[-1,1]\to\mathbb{R}$ denotes the d-dimensional ultraspherical polynomial *of degree $n\in\mathbb{N}$ (see Chapter 3).*

Proof. Let $\boldsymbol{s}\in\mathbb{S}^{d-1}$. From the generalised spherical harmonic transform (see Remark 3.3 in Chapter 3) applied to $\Psi_{\boldsymbol{s}}^{\mathscr{D}}$ we have:

$$\Psi_{\boldsymbol{s}}^{\mathscr{D}} = \sum_{n\in\mathbb{N}}\sum_{m=1}^{N_d(n)} \left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| Y_n^m \right\rangle Y_n^m = \sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}}\sum_{m=1}^{N_d(n)} \frac{1}{\hat{D}_n} Y_n^m(\boldsymbol{s}) Y_n^m,$$

since from (4.7), we have $\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| Y_n^m \right\rangle = \left\langle \delta_{\boldsymbol{s}} \middle| \mathscr{D}^\dagger Y_n^m \right\rangle = \hat{D}_n^{-1} Y_n^m(\boldsymbol{s})$ if $n\notin\mathfrak{K}_{\mathscr{D}}$ and zero otherwise. We have hence, from the bilinearity of the Schwartz duality product and the addition theorem 3.2:

$$\begin{aligned}
\left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle &= \sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{s}) \langle Y_n^m | \varphi\rangle \\
&= \sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{s})\, \langle \varphi, Y_n^m\rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{s}) \int_{\mathbb{S}^{d-1}} \varphi(\boldsymbol{r})\overline{Y_n^m}(\boldsymbol{r})\,d\boldsymbol{r} \\
&= \int_{\mathbb{S}^{d-1}} \varphi(\boldsymbol{r}) \left[\sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} Y_n^m(\boldsymbol{s})\overline{Y_n^m}(\boldsymbol{r}) \right] d\boldsymbol{r} \\
&= \int_{\mathbb{S}^{d-1}} \varphi(\boldsymbol{r}) \left[\sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}} \frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n} P_{n,d}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) \right] d\boldsymbol{r} \\
&= \int_{\mathbb{S}^{d-1}} \varphi(\boldsymbol{r})\psi_{\mathscr{D}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle)\,d\boldsymbol{r}, \qquad \forall\varphi\in\mathscr{S}(\mathbb{S}^{d-1}).
\end{aligned}$$

■

Observe that the traces of the zonal Green kernel (4.11) are generalised functions, which make sense when integrated against a Schwartz function but which may not admit a pointwise interpretation. When they do admit a pointwise interpretation, we say that the operator $\mathscr{D}$ is *spline-admissible*:

Definition 4.3 — *Spline-Admissible Pseudo-Differential Operator.* *Let $\mathscr{D}$ be a* pseudo-differential operator *with* zonal Green kernel $\psi_{\mathscr{D}}$. *We say that $\mathscr{D}$ is* spline admissible *if all traces $\{\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{s}\rangle),\ \boldsymbol{s}\in\mathbb{S}^{d-1}\}\subset\mathscr{S}'(\mathbb{S}^{d-1})$ of $\psi_{\mathscr{D}}$ are* ordinary functions*, i.e. they are* pointwise defined.

The following result provides us with a sufficient condition for a pseudo-differential operator to be spline-admissible:

Proposition 4.5 — *Sufficient Condition for Spline-Admissibility.* *Let $\mathscr{D}$ be a pseudo-differential operator, with* spectral growth order $p > d-1$ *and* zonal Green kernel $\psi_{\mathscr{D}}$. *Then we have*

$$\{\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{s}\rangle):\mathbb{S}^{d-1}\to\mathbb{R},\ \boldsymbol{s}\in\mathbb{S}^{d-1}\}\subset\mathscr{C}(\mathbb{S}^{d-1}),$$

and hence $\mathscr{D}$ is spline-admissible.

Pseudo-differential operators with spectral growth order $p > d-1$ are spline-admissible.

Proof. Let $\boldsymbol{s}\in\mathbb{S}^{d-1}$ be fixed but arbitrary. We show that, under the assumptions of Proposition 4.5, the series

$$\psi_{\mathscr{D}}(\langle\boldsymbol{r},\boldsymbol{s}\rangle)=\sum_{n\in\mathbb{N}\setminus\mathfrak{K}_{\mathscr{D}}}\frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n}P_{n,d}(\langle\boldsymbol{r},\boldsymbol{s}\rangle),\quad \boldsymbol{r}\in\mathbb{S}^{d-1},\tag{4.12}$$

converges uniformly (w.r.t. the variable $\boldsymbol{r}$). Since every summand is continuous, we can then conclude that the limit $\psi_{\mathscr{D}}(\langle\boldsymbol{r},\boldsymbol{s}\rangle)$ is continuous (see [125, Theorem 2.14]) and hence in particular pointwise defined –i.e. $\mathscr{D}$ is spline-admissible. To show that (4.12) is uniformly convergent, we consider its remainder for some $N>\max(\mathfrak{K}_{\mathscr{D}})$. Then, from the addition theorem 3.2 and the Cauchy-Schwarz inequality we get, for each $\boldsymbol{r}\in\mathbb{S}^{d-1}$:

$$\begin{aligned}
\left|\sum_{n=N}^{+\infty}\frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n}P_{n,d}(\langle\boldsymbol{r},\boldsymbol{s}\rangle)\right| &= \left|\sum_{n=N}^{+\infty}\frac{1}{\hat{D}_n}\sum_{m=1}^{N_d(n)}Y_n^m(\boldsymbol{s})\overline{Y_n^m}(\boldsymbol{r})\right| \\
&= \left|\sum_{n=N}^{+\infty}\sum_{m=1}^{N_d(n)}\left(\frac{Y_n^m(\boldsymbol{s})}{\operatorname{sgn}(\hat{D}_n)\sqrt{|\hat{D}_n|}}\right)\overline{\left(\frac{Y_n^m(\boldsymbol{r})}{\sqrt{|\hat{D}_n|}}\right)}\right| \\
&\le \left|\sum_{n=N}^{+\infty}\frac{\sum_{m=1}^{N_d(n)}|Y_n^m(\boldsymbol{s})|^2}{|\hat{D}_n|}\right|\left|\sum_{n=N}^{+\infty}\frac{\sum_{m=1}^{N_d(n)}|Y_n^m(\boldsymbol{r})|^2}{|\hat{D}_n|}\right| \\
&= \left|\sum_{n=N}^{+\infty}\frac{N_d(n)}{\mathfrak{a}_d|\hat{D}_n|}\right|^2.
\end{aligned}$$

Moreover, since $|\hat{D}_n| = \Theta(n^p)$ we have from (3.3) $N_d(n)|\hat{D}_n|^{-1} = \mathcal{O}(n^{d-2-p})$. Since $p > d-1 \Rightarrow d-2-p < -1$, the series $\sum_{n\in\mathbb{N}\setminus\hat{\mathfrak{K}}_{\mathscr{D}}} \frac{N_d(n)}{\mathfrak{a}_d|\hat{D}_n|}$ is convergent and hence its remainder tends to zero. Therefore

$$\left|\psi_{\mathscr{D}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) - \sum_{\substack{n=0\\ n\notin\hat{\mathfrak{K}}_{\mathscr{D}}}}^{N-1} \frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n} P_{n,d}(\langle \boldsymbol{r},\boldsymbol{s}\rangle)\right| = \left|\sum_{n=N}^{+\infty} \frac{N_d(n)}{\mathfrak{a}_d\hat{D}_n} P_{n,d}(\langle \boldsymbol{r},\boldsymbol{s}\rangle)\right|$$
$$\leq \left|\sum_{n=N}^{+\infty} \frac{N_d(n)}{\mathfrak{a}_d|\hat{D}_n|}\right|^2 \overset{N\to+\infty}{\longrightarrow} 0.$$

Moreover, since the bound is independent of $\boldsymbol{r}$ (and $\boldsymbol{s}$ as a matter of fact) the convergence is uniform, which achieves the proof. ■

We conclude this section by providing, for the specific case of $\mathbb{S}^2$, some examples (and non-examples) of spline-admissible pseudo-differential operators among the ones introduced in Example 4.1.

Example 4.2 — *Common Spline-Admissible Operators on* $\mathbb{S}^2$. Consider the specific case $d = 3$ and the pseudo-differential operators introduced in Example 4.1.

- **Laplace-Beltrami operator:** $\Delta_{\mathbb{S}^2}$ is *not* spline-admissible. Indeed, its zonal Green kernel is given by [60, Lemma 4.3],

$$\psi_{\Delta_{\mathbb{S}^2}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) = \frac{1}{4\pi}\ln(1-\langle \boldsymbol{r},\boldsymbol{s}\rangle) + \frac{1}{4\pi} - \frac{1}{4\pi}\ln 2, \quad \forall \boldsymbol{r},\boldsymbol{s}\in\mathbb{S}^{d-1},$$

  which is not defined for $\boldsymbol{r} = \boldsymbol{s}$. Note that we have $p = 2 = d-1$ which shows that the bound on the spectral growth order in Proposition 4.5 is tight.
- **Squared Laplace-Beltrami operator:** $\Delta^2_{\mathbb{S}^2}$ is, from Proposition 4.5, spline-admissible. Indeed, its spectral order p is such that $p = 4 > 2 = d-1$. Moreover, its zonal Green kernel is given by [60, Corollary 4.24],

$$\psi_{\Delta^2_{\mathbb{S}^2}}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) = \begin{cases} \frac{1}{4\pi}, & \text{for } \langle \boldsymbol{r},\boldsymbol{s}\rangle = 1 \\ \frac{1}{4\pi} - \frac{\pi}{24}, & \text{for } \langle \boldsymbol{r},\boldsymbol{s}\rangle = -1 \\ -\frac{1}{4\pi}\ln(1-\langle \boldsymbol{r},\boldsymbol{s}\rangle)\ln(1+\langle \boldsymbol{r},\boldsymbol{s}\rangle) & \\ +\frac{\ln 2}{4\pi}\ln(1-\langle \boldsymbol{r},\boldsymbol{s}\rangle^2) - \frac{1}{4\pi}\mathcal{L}_2\left(\frac{1-\langle \boldsymbol{r},\boldsymbol{s}\rangle}{2}\right) & \\ +\frac{1}{4\pi}(1-(\ln 2)^2) & \text{otherwise,} \end{cases}$$

where $\mathcal{L}_2$ denotes the so-called *Spence's functions* or *dilogarithm*:

$$\mathcal{L}_2(t) = -\int_0^t \frac{\ln(1-u)}{u} du, \quad t \in \mathbb{R}.$$

See Fig. 4.1 for a graphical representation of the zonal Green kernel.

- **Iterated Beltrami operator:** $\partial_{0\cdots k}$ is spline-admissible for every $k \geq 1$. Indeed, in such cases, we have $p = 2(k+1) > 2 = d-1$ and hence the spline-admissibility results from Proposition 4.5. For the specific case $k = 1$, the zonal Green kernel is moreover given by [60, Lemma 4.25]

$$\psi_{\partial_{0\cdots 1}}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \frac{1}{8\pi}(1 - \langle \boldsymbol{r}, \boldsymbol{s} \rangle)\ln(1 - \langle \boldsymbol{r}, \boldsymbol{s} \rangle) + \left(\frac{1}{12} + \frac{\ln 2}{2}\right)\frac{\langle \boldsymbol{r}, \boldsymbol{s} \rangle}{4\pi} + \frac{1}{4\pi}\left(\frac{1}{4} - \frac{\ln 2}{2}\right), \qquad \forall \boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}.$$

See Fig. 4.1 for a graphical representation of the zonal Green kernel.

- **Sobolev operators:** from Proposition 4.5, the Sobolev operators $(\mathrm{Id} - \Delta_{\mathbb{S}^2})^\beta$ are spline-admissible whenever $\beta > 1$ (more generally whenever $\beta > (d-1)/2$). There exists no known closed-form expression for their zonal Green kernel, which is given by

$$\psi_\beta(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \sum_{n=0}^{+\infty} \frac{2n+1}{4\pi\,(1+n(n+1))^\beta} P_n(\langle \boldsymbol{r}, \boldsymbol{s} \rangle), \qquad \forall \boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}.$$

■

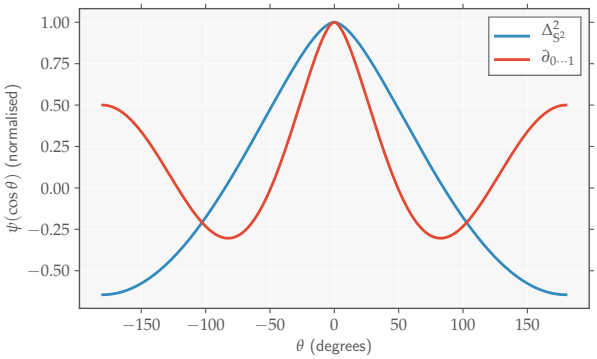
3 Spherical Splines

We are now in a position to introduce spherical splines. Roughly speaking, spherical splines are primitives (w.r.t. a particular spline-admissible pseudo-differential operator) of Dirac streams with finite innovations:

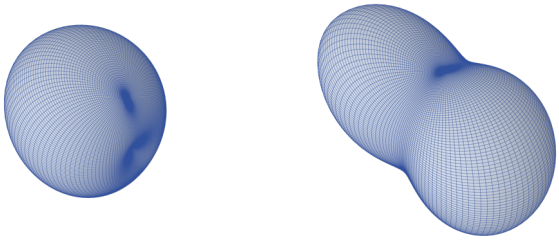
Definition 4.4 — $\mathscr{D}$-Spline. *Let $\Xi_M = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M\} \subset \mathbb{S}^{d-1}$ be a* set of points *on the hypersphere and $\mathscr{D}$ a* spline-admissible pseudo-differential operator. *Then, a* $\mathscr{D}$-spline *is a generalised function $\mathfrak{s} \in \mathscr{S}'(\mathbb{S}^{d-1})$ such that*

$$\mathscr{D}\mathfrak{s} = \sum_{i=1}^{M} \alpha_i \, \delta_{\boldsymbol{r}_i}, \tag{4.13}$$

where $\{\alpha_i,\ i = 1, \ldots, M\} \subset \mathbb{C}$ are called the amplitudes *of the spline, while the directions $\boldsymbol{r}_i$ in the* knot set *Ξ_M are called the* knots *of the spline. The pairs $(\alpha_i, \boldsymbol{r}_i)$ of amplitudes and knots are called the* innovations *of the spline, and their collection $\mathfrak{X}(\nu) = \{(\alpha_i, \boldsymbol{r}_i),\ i = 1, \ldots, M\}$ is called the* innovation set *of the spline.*



(a) Profile $\psi_{\mathscr{D}}(\cos(\theta))$ for $\Delta^2_{\mathbb{S}^2}$ and $\partial_{0\cdots1}$.



(b) Trace $\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_0\rangle)$ for $\Delta^2_{\mathbb{S}^2}$ and $\boldsymbol{r}_0=(-1,1,0)/\sqrt{2}$.

(c) Trace $\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_0\rangle)$ for $\partial_{0\cdots1}$ and $\boldsymbol{r}_0=(-1,1,0)/\sqrt{2}$.

Figure 4.1: Graphical representation (for $d=3$) of the squared Laplace-Beltrami and iterated Beltrami zonal Green kernels.

Finally, we denote by

$$\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1},\Xi_M):=\left\{\mathfrak{s}\in\mathscr{S}'(\mathbb{S}^{d-1}):\mathscr{D}\mathfrak{s}=\sum_{i=1}^{M}\alpha_i\delta_{\boldsymbol{r}_i},\ \alpha_i\in\mathbb{C},\ \boldsymbol{r}_i\in\Xi_M\right\}$$

the linear subspace of $\mathscr{D}$-splines associated with the knot set Ξ_M.

Remark 4.6 *Equation (4.13) tells us that the innovations of a $\mathscr{D}$-spline are revealed by applying $\mathscr{D}$ to it.*

Remark 4.7 — *Non-trivial Nullspace and Constrained Amplitudes.* *Notice that (4.13) implicitly constrains the spline amplitudes α_i when $\mathscr{D}$ has a nontrivial nullspace. Indeed, for a Schwartz function $\varphi\in\mathcal{N}(\mathscr{D})$, we have from the definition of $\mathscr{D}$ for generalised functions*

$$\langle\mathscr{D}\mathfrak{s}|\varphi\rangle=\langle\mathfrak{s}|\mathscr{D}\varphi\rangle=0.$$

However, we also have from the right hand-side of (4.13):

$$\langle\mathscr{D}\mathfrak{s}|\varphi\rangle=\sum_{i=1}^{M}\alpha_i\varphi(\boldsymbol{r}_i).$$

We have hence necessarily $\sum_{i=1}^{M}\alpha_i\varphi(\boldsymbol{r}_i)=0$ for all $\varphi\in\mathcal{N}(\mathscr{D})$, which holds if

and only if:

$$\sum_{i=1}^{M} \alpha_i Y_n^m(\boldsymbol{r}_i) = 0, \qquad \forall n \in \mathfrak{K}_{\mathscr{D}},\, m = 1, \ldots, N_d(n). \tag{4.14}$$

The following result characterises the splines associated to a spline-admissible operator in terms of its zonal Green kernel:

Proposition 4.6 — *Characterisation of $\mathscr{D}$-Splines.* *Let $\mathscr{D}$ be a* spline-admissible operator*, with* zonal Green kernel $\psi_{\mathscr{D}}$*. Let further $\mathfrak{s} \in \mathscr{S}'(\mathbb{S}^{d-1})$ be a $\mathscr{D}$-spline with knot set $\Xi_M = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M\} \subset \mathbb{S}^{d-1}$ and valid coefficients $\{\alpha_i\}_{i=1,\ldots,M} \subset \mathbb{C}$. Then, we have*

$$\mathfrak{s}(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle) \quad + \quad \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \hat{\beta}_n^m Y_n^m(\boldsymbol{r}), \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{4.15}$$

where $\hat{\beta}_n^m := \langle \mathfrak{s} | Y_n^m \rangle$, $\forall n \in \mathfrak{K}_{\mathscr{D}}, m = 1, \ldots, N_d(n)$. In particular, when $\mathfrak{K}_{\mathscr{D}} = \emptyset$ and $p > d - 1$, we have

$$\mathfrak{s}(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle), \qquad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{4.16}$$

and

$$\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_M) := \mathit{span}\, \{\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_i \rangle),\, \boldsymbol{r}_i \in \Xi_M,\, i = 1, \ldots, M\} \subset \mathscr{L}^2(\mathbb{S}^{d-1}). \tag{4.17}$$

Proof. Consider the function $\mathfrak{s}' : \mathbb{S}^{d-1} \to \mathbb{C}$ defined as

$$\mathfrak{s}'(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle) \quad + \quad \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle Y_n^m(\boldsymbol{r}), \quad \boldsymbol{r} \in \mathbb{S}^{d-1}.$$

Notice that since $\mathscr{D}$ is spline-admissible, the functions $\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_i \rangle)$ are ordinary functions which are hence bounded on $\mathbb{S}^{d-1}$ and hence $\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_i \rangle) \in \mathscr{L}^2(\mathbb{S}^{d-1})$, which implies in turn that $\mathfrak{s}' \in \mathscr{L}^2(\mathbb{S}^{d-1})$. We can hence interpret $\mathfrak{s}'$ as an element of $\mathscr{S}'(\mathbb{S}^{d-1})$ with pointwise definition:

$$\langle \mathfrak{s}' | \varphi \rangle = \langle \varphi, \mathfrak{s}' \rangle_{\mathbb{S}^{d-1}}, \quad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}).$$

We now show that $\mathfrak{s} = \mathfrak{s}'$, i.e.

$$\langle \mathfrak{s} | \varphi \rangle = \langle \mathfrak{s}' | \varphi \rangle, \qquad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}).$$

First, we write $\mathscr{S}(\mathbb{S}^{d-1}) = \mathcal{R}(\mathscr{D}) \oplus \mathcal{N}(\mathscr{D})$ such that every element φ of $\mathscr{S}(\mathbb{S}^{d-1})$

can be written as $\varphi = \mathscr{D}h + \eta$, with $(h, \eta) \in \mathcal{R}(\mathscr{D}) \times \mathcal{N}(\mathscr{D})$. We have hence

$$\begin{aligned}
\langle \mathfrak{s} | \varphi \rangle &= \langle \mathfrak{s} | \mathscr{D} h \rangle + \langle \mathfrak{s} | \eta \rangle \\
&= \langle \mathscr{D}\mathfrak{s} | h \rangle + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \hat{\eta}_n^m \langle \mathfrak{s} | Y_n^m \rangle \\
&= \left\langle \sum_{i=1}^{N} \alpha_i \delta_{\boldsymbol{r}_i} \middle| h \right\rangle + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \hat{\eta}_n^m \langle \mathfrak{s} | Y_n^m \rangle \\
&= \sum_{i=1}^{N} \alpha_i h(\boldsymbol{r}_i) + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \hat{\eta}_n^m .
\end{aligned}$$

Similarly we have from Proposition 4.4

$$\begin{aligned}
\langle \mathfrak{s}' | \varphi \rangle = \langle \varphi, \mathfrak{s}' \rangle_{\mathbb{S}^{d-1}} &= \sum_{i=1}^{M} \alpha_i \left\langle \varphi, \psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_i \rangle) \right\rangle_{\mathbb{S}^{d-1}} + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \langle \varphi, Y_n^m \rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{i=1}^{M} \alpha_i \left\langle \Psi_{\boldsymbol{s}}^{\mathscr{D}} \middle| \varphi \right\rangle + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \langle \eta, Y_n^m \rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{i=1}^{M} \alpha_i \left\langle \delta_{\boldsymbol{r}_i} \middle| \mathscr{D}^{\dagger} \mathscr{D} h \right\rangle + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \hat{\eta}_n^m \\
&= \sum_{i=1}^{M} \alpha_i \langle \delta_{\boldsymbol{r}_i} | h \rangle + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \hat{\eta}_n^m \\
&= \sum_{i=1}^{M} \alpha_i h(\boldsymbol{r}_i) + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \langle \mathfrak{s} | Y_n^m \rangle \hat{\eta}_n^m ,
\end{aligned}$$

and hence we have indeed $\langle \mathfrak{s} | \varphi \rangle = \langle \mathfrak{s}' | \varphi \rangle \ \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1})$ as claimed. Equation (4.16) then follows trivially from the fact that the summations involving $\mathfrak{K}_{\mathscr{D}}$ vanish when $\mathfrak{K}_{\mathscr{D}} = \emptyset$. Finally, (4.17) follows from the definition of $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_M)$ (see Definition 4.4) and the fact that when $\mathfrak{K}_{\mathscr{D}} = \emptyset$ the spline coefficients are unconstrained (see Remark 4.7). ■

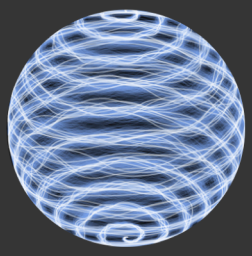
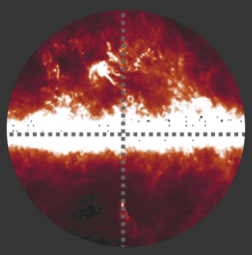
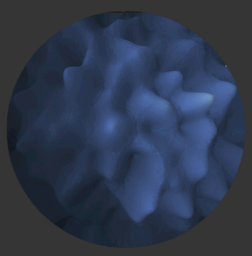
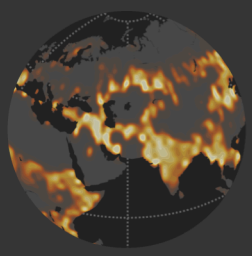
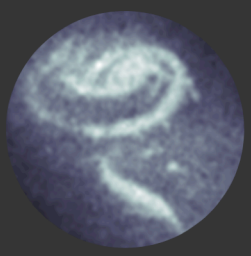
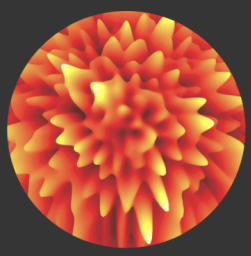
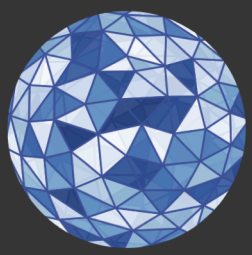
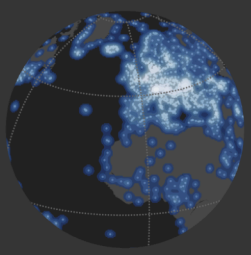
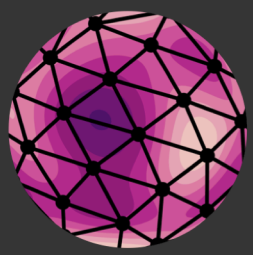
Remark 4.8 *Observe from (4.16) that when $\mathfrak{K}_{\mathscr{D}} = \emptyset$ the $\mathscr{D}$-splines are linear combinations of zonal functions, and hence very easy to evaluate. This nice feature is due to the fact that we restricted ourselves to* isotropic *pseudo-differential operators in Definition 4.1.*

II

Approximation on the Hypersphere

In this part, we use the various concepts from Part I to perform infinite dimensional inference on the hypersphere. Some of the highlights of this part are the following:

- *In Section 1 of Chapter 5, we introduce functional inverse problems on the hypersphere. Unlike the ad-hoc discrete methods traditionally favoured by practitioners, functional inverse problems present the advantage of being directly formulated in the continuous spherical domain, which is the natural domain of definition for the analog spherical signals encountered in nature.*
- *In Section 2 of Chapter 5, we consider regularising functional inverse problems by means of generalised Tikhonov (gTikhonov) and generalised total variation (gTV) norms.*
- *In Section 3 of Chapter 5, we show that, with gTikhonov and gTV regularisation, functional inverse problems admit* finite dimensional *solutions. For gTikhonov regularisation, we show in Theorem 5.3 that the solution is* unique *and can be expressed as a linear combination of the sampling linear functionals primitived twice w.r.t. the regularising pseudo-differential operator $\mathscr{D}$. For gTV regularisation, we show in Theorem 5.4 that the solutions are convex combinations of spherical $\mathscr{D}$-splines with less innovations than available measurements.*
- *In Chapter 6, we use Theorems 5.3 and 5.4 to design two canonical search space discretisation schemes,* exact *for gTikhonov regularisation and with vanishing approximation error for gTV regularisation. We moreover investigate alternative* domain discretisation schemes, *traditionally favoured by practitioners.*

5

Representer Theorems

In this chapter, we leverage the functional analysis tools introduced in Chapter 2 to formulate functional inverse problems in a common *generalised sampling* framework. Our formulation allows us to see most spherical approximation problems as specific instances of the penalised optimisation problem (2.11) of Chapter 2. We investigate the latter in the specific case of generalised Tikhonov and generalised total variation regularisation norms. In both cases, we define the search space of the optimisation problem, and characterise the predual in which the sampling linear functionals must be chosen. We moreover provide *representer theorems*, characterising the form of the solutions to both optimisation problems, and compare the effects of both regularisation strategies. Finally, we illustrate the representer theorems in the specific case of spherical interpolation.

1 Generalised Sampling & Functional Inverse Problems

Most real-life spherical approximation problems take the form of *functional inverse problems*. In a typical inverse problem formulation, an unknown *spherical field*[1] $f \in \mathscr{B}'$ is probed by some *sensing device*, resulting in a data vector $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{C}^L$ of L *measurements*. To account for potential inaccuracies in the measurement process, the data vector $\boldsymbol{y}$ is often modelled as the outcome of a *random vector* $\boldsymbol{Y} = [Y_1, \dots, Y_L] : \Omega \to \mathbb{C}^L$, fluctuating according to some application-dependent *noise distribution*.[2] When the measurement process is unbiased, entries of the expectation of $\boldsymbol{Y}$ can be thought of as the *ideal measurements* which would be obtained in a noise-free environment. In most cases, the ideal measurements are linked to the unknown spherical field by some *linear relationship*, called *generalised sampling* [179]:

$$\mathbb{E}[Y_i] = \langle f | \varphi_i \rangle, \qquad i = 1, \dots, L, \tag{5.1}$$

where $\langle \cdot | \cdot \rangle : \mathscr{B}' \times \mathscr{B} \to \mathbb{C}$ denotes the *Schwartz duality product* for some duality pair $(\mathscr{B}, \mathscr{B}')$ and $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{B}$ are *linear sampling functionals* modelling the action of the sensing device on the spherical field $f \in \mathscr{B}'$. Since most real-life acquisition systems react continuously to variations in their inputs, the dual space $\mathscr{B}'$ is generally equipped with the weak* topology, so that the linear

[1] The generic appellation "spherical field" is used here to designate any element of $\mathscr{S}'(\mathbb{S}^{d-1})$, such as a function or a measure defined over the sphere.

[2] In the absence of noise, $\boldsymbol{Y}$ can simply be chosen as a deterministic random vector.

functionals $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{B}$ modelling the instrument are all continuous (see Section 1.3 of Chapter 2). In such a formalism, the ideal measurements in (5.1) are often referred to as *generalised samples* [179] of f. This is because the Schwartz duality product is a generalised *evaluation* map $\langle f|\varphi_i\rangle = f(\varphi_i)$, allowing us to interpret the ideal measurements as samples of f evaluated at "points" $\varphi_1, \dots, \varphi_L \in \mathscr{B}$. For convenience, it is moreover customary to write the generalised sampling equations (5.1) in terms of a *sampling operator*[3] $\boldsymbol{\Phi} : \mathscr{B}' \to \mathbb{C}^L$ (see [183, Chapter 5]) defined as:

[3] In finite dimensions, the sampling operator is generally called *forward*, *design* or *sensing* matrix.

$$\boldsymbol{\Phi} : \begin{cases} \mathscr{B}' \to \mathbb{C}^L \\ f \mapsto [\langle f|\varphi_1\rangle, \cdots, \langle f|\varphi_L\rangle]\,. \end{cases} \tag{5.2}$$

Reformulating (5.1) in terms of $\boldsymbol{\Phi}$ yields:

$$\mathbb{E}[\boldsymbol{Y}] = \begin{bmatrix} \mathbb{E}[Y_1] \\ \vdots \\ \mathbb{E}[Y_L] \end{bmatrix} = \begin{bmatrix} \langle f|\varphi_1\rangle \\ \vdots \\ \langle f|\varphi_L\rangle \end{bmatrix} = \boldsymbol{\Phi}(f). \tag{5.3}$$

The goal of a functional inverse problem is then to recover a spherical field $f \in \mathscr{B}'$ which best explains the observed generalised samples $\boldsymbol{y}$, given a particular noise and functional data model (5.3). Since the search space $\mathscr{B}'$ is infinite-dimensional and the data finite-dimensional, this task is fundamentally *ill-posed* and will in general elicit *infinitely many* candidate solutions. To discriminate among such solutions, it is customary to resort to *regularisation*, which can be seen as implementing *Occam's razor principle*[4] by favouring solutions with *simple* behaviours. This is typically achieved by means of *penalised convex optimisation problems* of the form:

[4] Occam's razor principle is a philosophical principle also known as the "law of briefness" or in Latin *lex parsimoniae*. It was supposedly formulated by William of Ockham in the 14th century, who wrote in Latin *"Entia non sunt multiplicanda praeter necessitatem"*. In English, this translates to *"More things should not be used than are necessary"*. In essence, this principle states that when two equally good explanations for a given phenomenon are available, one should always favour the simplest, i.e. the one that introduces the least explanatory variables.

$$\mathcal{V} = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \;+\; \Lambda\left(|||f|||\right) \right\}, \tag{5.4}$$

where

- $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a *cost functional*, measuring the discrepancy between the *observed* and *predicted* generalised samples $\boldsymbol{y}$ and $\boldsymbol{\Phi}(f)$ respectively. Common choices of discrepancy measures are discussed in Remark 5.1. In what follows, we will assume that F is such that for all $\boldsymbol{y} \in \mathbb{C}^L$,
$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$
is *proper*, *convex* and lower semi-continuous.
- $|||\cdot||| : \mathscr{B}' \to \mathbb{R}_+$ is the *dual norm* on $\mathscr{B}'$, called *regularisation norm*, which implements Occam's razor principle. Intuitively, elements $f \in \mathscr{B}'$ with small regularisation norm are *simple* and *well-behaved*, typically with a finite number of degrees of freedom (df).
- $\Lambda : \mathbb{R} \to \mathbb{R}_+$ is some *convex regularisation function*, *strictly increasing* on $\mathbb{R}_+$. In practice, Λ often takes the form of a *monomial* $t \mapsto \lambda t^p$, where

$p \geq 1$ and $\lambda > 0$. The parameter λ is called *regularisation parameter* and controls the amount of regularisation by putting the regularisation norm and the cost functional on a similar scale.

Remark 5.1 — *Choosing the Cost Functional.* *In practice, the cost functional F is often chosen in one of the following two ways:*

- Noiseless case: *In a noiseless setup, one has full trust in the generalised samples. It is therefore natural to require that any solution of (5.4) be* consistent *[183, Chapter 5] with the samples at hand, i.e. $\boldsymbol{y} = \boldsymbol{\Phi}(f)$, $\forall f \in \mathcal{V}$. This can be achieved by choosing the cost functional as $F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) = \iota(\boldsymbol{y} - \boldsymbol{\Phi}(f))$, where $\iota : \mathbb{C}^L \to \{0, +\infty\}$ is the* indicator function

$$\iota(\boldsymbol{z}) = \begin{cases} 0 & \text{if } \boldsymbol{z} = \boldsymbol{0}, \\ +\infty & \text{otherwise.} \end{cases}$$

  Such cost functionals are for example used in interpolation problems *as discussed in Section 4.*
- Noisy case: *In a noisy setup, consistency is not desired anymore, as it almost always leads to* overfitting *the noisy data. In this case, one can use general ℓ_p cost functionals $F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) = \|\boldsymbol{y} - \boldsymbol{\Phi}(f)\|_p^p$, where $p \in [1, +\infty]$ is typically chosen according to the tail behaviour[5] of the noise distribution [146]. Another approach consists in using the* negative log-likelihood *of the data $\boldsymbol{y}$ as a measure of discrepancy, i.e. $F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) = -\ell(\boldsymbol{y}|\boldsymbol{\Phi}(f))$. This choice makes (5.4) resemble a* maximum a posteriori *problem with* improper prior. *In the case of centred Gaussian white noise, both discrepancy measures coincide, yielding the classical quadratic cost functional $F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) = \|\boldsymbol{y} - \boldsymbol{\Phi}(f)\|_2^2$.*

[5] The following rule of thumb is proposed in [146]: p should be close to 1 for heavy-tailed distributions, close to 2 for Gaussian-like distributions, and close to $+\infty$ for compactly supported distributions.

2 Regularisation Strategies

Notice that the regularisation norm $|||\cdot|||$ in (5.4) entirely determines the search space $\mathscr{B}'$. Candidate regularisation norms for spherical approximation problems can hence be constructed as follows:

1. Identify *interesting* functional spaces $\mathscr{B}' \subset \mathscr{S}'(\mathbb{S}^{d-1})$, whose elements are *regular* enough;
2. Find a norm $|||\cdot|||$ on $\mathscr{B}'$ such that $\mathscr{B}'$ admits a predual $\mathscr{B}$ and characterise this predual.

For example, one could consider choosing $\mathscr{B}'$ as a *generalised Sobolev space* of the form:

$$\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \mathscr{D} f \in \mathscr{L}^2(\mathbb{S}^{d-1}) \right\}, \tag{5.5}$$

where $\mathscr{D} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathscr{S}'(\mathbb{S}^{d-1})$ is some pseudo-differential operator as in Definition 4.1. This is the space of generalised functions regular enough so that their generalised derivatives w.r.t. $\mathscr{D}$ are *square-integrable*. While extensively used in the literature, this notion of regularity may however be considered too restrictive, since the Sobolev space (5.5) is notably not large enough[6] to

[6] $\mathscr{D}$-splines are indeed defined as $\mathscr{D}$-primitives of Dirac streams, i.e. $\mathscr{D}\mathfrak{s} = \sum_{i=1}^{M} \alpha_i \delta_{\boldsymbol{r}_i}$. Their $\mathscr{D}$-derivatives are hence not in $\mathscr{L}^2(\mathbb{S}^{d-1})$.

contain $\mathscr{D}$-splines in cases where $\mathscr{D}$ is *spline-admissible*. This is particularly cumbersome, since the latter are, by definition of spline-admissible operators, *ordinary functions* and hence relatively *well-behaved*. To include $\mathscr{D}$-splines, one must consider the larger space

$$\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \ \mathscr{D}f \in \mathcal{M}(\mathbb{S}^{d-1}) \right\}, \tag{5.6}$$

where $\mathcal{M}(\mathbb{S}^{d-1})$ denotes the space of *spherical regular Borel measures* introduced in Section 1.5.3 of Chapter 2. This is the space of generalised functions regular enough so that their generalised derivatives w.r.t. $\mathscr{D}$ are *Borel measures*.

In what follows, we investigate both regularisation strategies and derive their associated regularisation norms, called generalised Tikhonov (gTikhonov) and generalised total variation (gTV) norms respectively. For simplicity, we restrict our attention to pseudo-differential operators $\mathscr{D}$ with trivial nullspaces, i.e. $\mathfrak{K}_{\mathscr{D}} = \emptyset$.

2.1 Generalised Tikhonov Regularisation

Generalised Tikhonov regularisation is obtained by choosing $\mathscr{B}'$ in (5.4) as the generalised Sobolev space (5.5):

$$\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \ \mathscr{D}f \in \mathscr{L}^2(\mathbb{S}^{d-1}) \right\}.$$

[7] Notice that the sesquilinear form $\langle \cdot, \cdot \rangle_{\mathscr{D}}$ is indeed positive definite since $\mathscr{D}$ is assumed to have trivial nullspace.

It is easy to see that $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ can be equipped with the inner product[7]:

$$\langle f, g \rangle_{\mathscr{D}} := \langle \mathscr{D}f, \mathscr{D}g \rangle_2, \qquad \forall (f, g) \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \times \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

We denote by $\|f\|_{\mathscr{D},2} := \sqrt{\langle \mathscr{D}f, \mathscr{D}f \rangle_2}$ the inner product norm induced by $\langle \cdot, \cdot \rangle_{\mathscr{D}}$, and call it the generalised Tikhonov (gTikhonov) norm. If the quantity $\mathscr{D}f$ is understood as some generalised notion of *curvature*, then the gTikhonov norm can be interpreted as measuring the *bending energy* or *roughness* of f.

Since $\mathscr{D}$ has a trivial nullspace, it is bijective and its inverse $\mathscr{D}^{-1}$ defines an *isometric isomorphism* between the two inner product spaces $(\mathscr{L}^2(\mathbb{S}^{d-1}), \langle \cdot, \cdot \rangle_2)$ and $(\mathcal{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \langle \cdot, \cdot \rangle_{\mathscr{D}})$. Indeed, we can write uniquely any element $f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ as:

$$f = \mathscr{D}^{-1}h, \quad h \in \mathscr{L}^2(\mathbb{S}^{d-1}), \quad \text{with} \quad \|f\|_{\mathscr{D},2} = \|\mathscr{D}f\|_2 = \|h\|_2. \tag{5.7}$$

The space $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is hence a Hilbert space. Since Hilbert spaces are *reflexive*, we moreover have $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \cong \mathscr{H}''_{\mathscr{D}}(\mathbb{S}^{d-1})$, allowing us to identify the predual of $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ with its dual $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$. The latter can moreover be characterised as follows:

Proposition 5.1 — *Dual of* $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$. *The* dual *of the space* $(\mathcal{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \langle \cdot, \cdot \rangle_{\mathscr{D}})$ *is given by*

$$\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \ \mathscr{D}^{-1}g \in \mathscr{L}^2(\mathbb{S}^{d-1}) \right\}, \tag{5.8}$$

with dual norm

$$\|g\|_{\mathscr{D}^{-1},2} := \|\mathscr{D}^{-1}g\|_2, \qquad \forall g \in \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}). \tag{5.9}$$

Proof. From the definition of the dual space $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$, and the isometric isomorphism (5.7) we have

$$\begin{aligned}
\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}) &= \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sup_{f\in\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}),\|f\|_{\mathscr{D},2}=1} |\langle g|f\rangle| < +\infty \right\} \\
&= \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sup_{h\in\mathscr{L}^2(\mathbb{S}^{d-1}),\|h\|_2=1} |\langle g|\mathscr{D}^{-1}h\rangle| < +\infty \right\} \\
&= \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sup_{h\in\mathscr{L}^2(\mathbb{S}^{d-1}),\|h\|_2=1} |\langle \mathscr{D}^{-1}g|h\rangle| < +\infty \right\} \\
&= \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \|\mathscr{D}^{-1}g\|_2 < +\infty \right\}.
\end{aligned}$$

We have indeed

$$\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ g \in \mathscr{S}'(\mathbb{S}^{d-1}) : \mathscr{D}^{-1}g \in \mathscr{L}^2(\mathbb{S}^{d-1}) \right\},$$

and

$$|||g||| = \sup_{f\in\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}),\|f\|_{\mathscr{D},2}=1} |\langle g|f\rangle| = \|\mathscr{D}^{-1}g\|_2, \quad \forall g \in \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

■

Remark 5.2 — *Canonical Inner Product on* $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$. *Notice that the dual norm (5.9) is generated by the inner product on* $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$*, defined as*

$$\langle g, h\rangle_{\mathscr{H}'_{\mathscr{D}}} := \left\langle \mathscr{D}^{-1}g, \mathscr{D}^{-1}h \right\rangle, \qquad \forall g, h \in \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

We have hence established the duality pair

$$(\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}^{-1}\cdot\|_2)' \cong (\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},2}),$$

showing that the gTikhonov norm $\|\cdot\|_{\mathscr{D},2}$ is actually a dual norm which can hence be used as regularisation norm in (5.4). For such a choice of regularisation norm, it is customary to set the regularisation function to $\Lambda(t) = \lambda t^2$ with $\lambda > 0$. This yields the following optimisation problem:

$$\mathcal{V} = \underset{f\in\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) + \lambda\|\mathscr{D}f\|_2^2 \right\}, \tag{5.10}$$

where the sampling operator $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L, f \mapsto [\langle\varphi_1|f\rangle, \ldots, \langle\varphi_L|f\rangle]$ is such that $\{\varphi_1, \ldots, \varphi_L\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$. We call (5.10) a functional penalised Tikhonov (FPT) problem. Since the gTikhonov regularisation norm penalises

the roughness of f, solutions to an FPT problem are expected to be *smooth*. This intuition will be formalised in Section 3.1.

2.2 Generalised Total Variation Regularisation

Generalised total variation regularisation consists in choosing $\mathscr{B}'$ in (5.4) as the space (5.6)

$$\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \ \mathscr{D}f \in \mathcal{M}(\mathbb{S}^{d-1}) \right\},$$

where $\mathcal{M}(\mathbb{S}^{d-1})$ is the space of *spherical regular Borel measures* introduced in Section 1.5.3 of Chapter 2. This space can be equipped with the generalised total variation (gTV) norm, defined as[8]

$$\|f\|_{\mathscr{D},TV} := \|\mathscr{D}f\|_{TV} = \sup_{\varphi \in \mathscr{S}(\mathbb{S}^{d-1}),\, \|\varphi\|_\infty = 1} |\langle \mathscr{D}f | \varphi \rangle|, \quad \forall f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \tag{5.11}$$

where we have used the dual characterisation of the *total variation norm* $\|\cdot\|_{TV}$ defined in Definition 2.2 and the density of Schwartz functions into the space of bounded continuous functions (see (2.5) in Chapter 2). The gTV norm can be interpreted as measuring the *variations* of the generalised derivative $\mathscr{D}f$.

[8] Note that $\|\cdot\|_{\mathscr{D},TV}$ is indeed a norm since $\mathscr{D}$ is assumed to have trivial nullspace.

Again, since $\mathscr{D}$ is bijective we can consider its inverse which can be shown [179] to define an *isometric isomorphism* between the spaces $(\mathcal{M}(\mathbb{S}^{d-1}), \|\cdot\|_{TV})$ and $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV})$. Indeed, we can uniquely write any element f in $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ as:

$$f = \mathscr{D}^{-1}\mu, \quad \mu \in \mathcal{M}(\mathbb{S}^{d-1}), \quad \text{with} \quad \|f\|_{\mathscr{D},TV} = \|\mathscr{D}f\|_{TV} = \|\mu\|_{TV}. \tag{5.12}$$

This isometry implies that the metric space $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV})$ is actually a *Banach space*, and allows us to characterise its predual:

Proposition 5.2 — *Predual of* $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$. *The Banach space*

$$\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{ h \in \mathscr{S}'(\mathbb{S}^{d-1}) : \ h = \mathscr{D}\eta,\ \eta \in \mathscr{C}(\mathbb{S}^{d-1}) \right\}, \tag{5.13}$$

equipped with the norm

$$\|h\|_{\mathscr{D},\infty} = \|\mathscr{D}^{-1}h\|_\infty = \|\eta\|_\infty,$$

is the predual *of the Banach space* $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV})$*, i.e.*

$$(\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})' \cong (\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV}).$$

Proof. Notice first that $\mathscr{D}$ maps isometrically $(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)$ onto $(\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})$. Indeed, every element h of $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ can be uniquely written as

$$h = \mathscr{D}\eta, \quad \eta \in \mathscr{C}(\mathbb{S}^{d-1}), \quad \text{with} \quad \|h\|_{\mathscr{D},\infty} = \|\mathscr{D}^{-1}h\|_\infty = \|\eta\|_\infty.$$

We have hence the isometries $(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty) \cong (\mathscr{C}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})$ and $(\mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV}) \cong (\mathcal{M}(\mathbb{S}^{d-1}), \|\cdot\|_{TV})$. Moreover, we have from the *Riesz-Markov representation theorem* 2.5 the duality pair

$$(\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)' \cong (\mathcal{M}(\mathbb{S}^{d-1}), \|\cdot\|_{TV}),$$

which yields

$$(\mathscr{C}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})' \cong (\mathscr{C}(\mathbb{S}^{d-1}), \|\cdot\|_\infty)' \cong (\mathcal{M}(\mathbb{S}^{d-1}), \|\cdot\|_{TV}) \cong (\mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV}),$$

and hence $(\mathscr{C}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})' \cong (\mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV})$ as claimed. ■

We have hence established the duality pair

$$(\mathscr{C}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},\infty})' \cong (\mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathscr{D},TV}),$$

showing that the gTV norm $\|\cdot\|_{\mathscr{D},TV}$ is actually a dual norm which can hence be used as regularisation norm in (5.4). For such a choice of regularisation norm, it is customary to set the regularisation function to $\Lambda(t) = \lambda t$ with $\lambda > 0$. This yields the following optimisation problem:

$$\mathcal{V} = \underset{f \in \mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1})}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) + \lambda \|\mathscr{D} f\|_{TV} \right\}, \tag{5.14}$$

where the sampling operator $\boldsymbol{\Phi} : \mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1}) \to \mathbb{C}^L, f \mapsto [\langle f|\varphi_1\rangle, \ldots, \langle f|\varphi_L\rangle]$ is such that $\{\varphi_1, \ldots, \varphi_L\} \subset \mathscr{C}_\mathscr{D}(\mathbb{S}^{d-1})$. We call (5.14) a functional penalised basis pursuit (FPBP) problem. Because of the gTV regularisation norm, solutions to FPBP problems will tend to have *few* variations in their generalised derivatives. When $\mathscr{D}$ is *spline-admissible*, such functions are templated by the $\mathscr{D}$-splines, which, from Proposition 4.6, take the form

$$\mathfrak{s} = \sum_{i=1}^{M} \alpha_i \psi_\mathscr{D}(\langle \cdot, \boldsymbol{r}_i \rangle), \tag{5.15}$$

where $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M\} \subset \mathbb{S}^{d-1}$ and $\psi_\mathscr{D}$ is the zonal Green kernel of $\mathscr{D}$. For such functions, we have indeed (see Remark 2.1)

$$\|\mathscr{D}\mathfrak{s}\|_{TV} = \left\| \sum_{i=1}^{M} \alpha_i \mathscr{D}\psi_\mathscr{D}(\langle \cdot, \boldsymbol{r}_i \rangle) \right\|_{TV} = \left\| \sum_{i=1}^{M} \alpha_i \delta_{\boldsymbol{r}_i} \right\|_{TV} = \sum_{i=1}^{M} |\alpha_i| \underbrace{\|\delta_{\boldsymbol{r}_i}\|_{TV}}_{=1} = \|\boldsymbol{\alpha}\|_1.$$

Hence $\mathscr{D}$-splines with small ℓ_1 norm in their coefficients will also have small gTV norm. It is then expected for solutions $f \in \mathcal{V}$ to take the form of $\mathscr{D}$-splines (5.15) with *few* innovations M. In Section 3.2 we will show that extreme points of the solution set $\mathcal{V}$ indeed take such a form, with $M < L$.

Remark 5.3 *Note that generalised total variation regularisation can be considered less stringent than generalised Tikhonov regularisation since the space $\mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1})$ is strictly larger that the generalised Sobolev space $\mathscr{H}_\mathscr{D}(\mathbb{S}^{d-1})$. Indeed, we have $\mathscr{H}_\mathscr{D}(\mathbb{S}^{d-1}) \subset \mathcal{M}_\mathscr{D}(\mathbb{S}^{d-1})$ since every function in $\mathscr{L}^2(\mathbb{S}^{d-1})$ defines a*

regular Borel measure $\nu_f(B) = \int_B f(\boldsymbol{r})d\boldsymbol{r}$, for any Borel set B in a Borel algebra $\mathscr{B} \subset \mathscr{P}(\mathbb{S}^{d-1})$. However, we have $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) \not\subset \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ since Green functions of $\mathscr{D}$ are trivially[9] *in $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ but not in $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$.*

[9] Recall that for an injective pseudo-differential operator $\mathscr{D}$, Green functions $\{\Psi_{\boldsymbol{s}}^{\mathscr{D}}, \boldsymbol{s} \in \mathbb{S}^{d-1}\}$ are such that $\mathscr{D}\Psi_{\boldsymbol{s}}^{\mathscr{D}} = \delta_{\boldsymbol{s}}$.

3 Representer Theorems

We now make use of Corollary 2.10 and Theorem 2.12 from Chapter 2 to establish representer theorems characterising the solution sets of the FPT and FPBP problems (5.10) and (5.14) respectively. For simplicity, we state both theorems in the case where the pseudo-differential operator $\mathscr{D}$ used to define the gTikhonov or gTV regularisation norm has a *trivial null space*[10] ($\mathfrak{K}_{\mathscr{D}} = \emptyset$). The first representer theorem shows that an FPT problem (5.10) admits a *unique solution*, which can moreover be expressed as a linear combination of the *sampling linear functionals* φ_i primitived twice w.r.t. $\mathscr{D}$. The second representer theorem shows that the solution set of an FPBP problem is *nonempty* and the *weak* closed convex-hull* of extreme points taking the form of $\mathscr{D}$-primitives of Dirac streams[11] with less innovations than available measurements. Both results can be seen as extensions to the spherical setup of [72, Theorem 3] and [72, Theorem 4] respectively.

[10] See Remarks 5.5 and 5.8 for the case where $\mathscr{D}$ has a nontrivial nullspace.

[11] i.e. $\mathscr{D}$-splines when $\mathscr{D}$ is spline-admissible.

3.1 Representer Theorem for gTikhonov Regularisation

Our first representer theorem characterises the solutions of FPT problems:

Theorem 5.3 — *Representer Theorem for gTikhonov Regularisation.* Consider the following scenario:

- E1 $\mathscr{D} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathscr{S}'(\mathbb{S}^{d-1})$ is some *pseudo-differential operator* with *trivial nullspace*;
- E2 $(\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \langle \cdot, \cdot \rangle_{\mathscr{D}})$ is the *generalised Sobolev space* defined in (5.5), with *topological dual* $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ characterised in (5.8);
- E3 $\operatorname{span}\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$, with the φ_i being *linearly independent*;
- E4 $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is a *sampling* operator, defined as

$$\boldsymbol{\Phi}(f) = [\langle \varphi_1 | f \rangle, \cdots, \langle \varphi_L | f \rangle], \quad \forall f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1});$$

- E5 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

  is *proper*, *convex* and lower semi-continuous;
- E6 λ is a *positive* regularisation constant.

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the solution to the FPT optimisation problem

$$f^\star = \underset{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) + \lambda \|\mathscr{D} f\|_2^2 \right\}, \tag{5.16}$$

exists and is *unique*. Moreover, we have

$$f^\star = \sum_{i=1}^{L} \alpha_i \, \mathscr{D}^{-2} \varphi_i = \sum_{i=1}^{L} \alpha_i \left[\sum_{n \in \mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{\langle \varphi_i | Y_n^m \rangle}{|\hat{D}_n|^2} Y_n^m \right], \tag{5.17}$$

for some *weights* $\{\alpha_1, \ldots, \alpha_L\} \subset \mathbb{C}$. In particular, when $\{\varphi_i, i = 1, \ldots, L\} \subset \mathscr{L}^2(\mathbb{S}^{d-1})$ and $\mathscr{D}$ is *spline-admissible*, we have

$$f^\star = \sum_{i=1}^{L} \alpha_i \, \psi_{\mathscr{D}} * (\psi_{\mathscr{D}} * \varphi_i), \tag{5.18}$$

where $*$ is the *spherical convolution* and $\psi_{\mathscr{D}}$ is the *zonal Green kernel* of $\mathscr{D}$.

Proof. We apply Corollary 2.10 to (5.16), with $\mathscr{H} = \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$, $\mathscr{H}' = \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$, $\|\cdot\|_{\mathscr{H}} = \|\mathscr{D} \cdot\|_2$ and $\Lambda(t) = \lambda t^2$. Observe that the assumptions of the corollary are indeed verified since $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is a Hilbert space (see Section 2.1) and Λ is convex and strictly increasing. We have hence from Corollary 2.10 that the FPT problem (5.16) admits a unique solution, given by

$$f^\star = \sum_{i=1}^{L} \alpha_i \, R_{\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}(\varphi_i),$$

where $R_{\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}$ is the *Riesz map* for $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$, i.e. the isometric isomorphism from $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ to $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$. Since $\mathscr{D}$ is assumed to have a trivial nullspace, the Riesz map coincides moreover in this case with the bijective operator $\mathscr{D}^{-2}$. Indeed, the latter maps isometrically $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ onto $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ since every element $f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ can be uniquely written as

$$f = \mathscr{D}^{-2} g, \; g \in \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}), \text{ with } \|f\|_{\mathscr{D},2} = \|\mathscr{D}\mathscr{D}^{-2} g\|_2 = \|\mathscr{D}^{-1} g\|_2 = \|g\|_{\mathscr{D}^{-1},2},$$

where $\|\cdot\|_{\mathscr{D}^{-1},2}$ is the dual norm on $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ as shown in Proposition 5.1. We have hence $R_{\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})} = \mathscr{D}^{-2}$, yielding the first equality in (5.17):

$$f^\star = \sum_{i=1}^{L} \alpha_i \, \mathscr{D}^{-2} \varphi_i.$$

The second equality in (5.17) is obtained by considering the gSHT of $\mathscr{D}^{-2}\varphi_i$. We have indeed, for each $i = 1, \ldots, L$:

$$\left\langle \mathscr{D}^{-2} \varphi_i \middle| Y_n^m \right\rangle = \left\langle \varphi_i \middle| \mathscr{D}^{-2} Y_n^m \right\rangle = \frac{\langle \varphi_i | Y_n^m \rangle}{|\hat{D}_n|^2}, \quad \forall n \in \mathbb{N}, \; m = 1, \ldots, N_d(n),$$

where we have used the definition of pseudo-differential operators, the fact that $\mathscr{D}$ has a trivial nullspace and the bilinearity of the Schwartz duality product.

This yields the second equality in (5.17):

$$f^\star = \sum_{i=1}^{L} \alpha_i \left[\sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{\langle \varphi_i | Y_n^m \rangle}{|\hat{D}_n|^2} Y_n^m \right].$$

For (5.18) finally, notice that when $\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{L}^2(\mathbb{S}^{d-1})$ (which is possible since $\mathscr{L}^2(\mathbb{S}^{d-1}) \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ for any pseudo-differential operator), we have

$$f^\star = \sum_{i=1}^{L} \alpha_i \, \mathscr{D}^{-2} \varphi_i = \sum_{i=1}^{L} \alpha_i \left[\sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{(\hat{\varphi}_i)_n^m}{|\hat{D}_n|^2} Y_n^m \right], \tag{5.19}$$

where $(\hat{\varphi}_i)_n^m = \langle \varphi_i, Y_n^m \rangle_{\mathbb{S}^{d-1}}$ for $n \in \mathbb{N}$, $m = 1, \dots, N_d(n)$. From Propositions 3.8 and 4.4, we have moreover for any $\varphi \in \mathscr{L}^2(\mathbb{S}^{d-1})$,

$$\psi_{\mathscr{D}} * (\psi_{\mathscr{D}} * \varphi) = \psi_{\mathscr{D}} * \left[\sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{\hat{\varphi}_n^m}{\hat{D}_n} Y_n^m \right] = \sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{\hat{\varphi}_n^m}{|\hat{D}_n|^2} Y_n^m. \tag{5.20}$$

Plugging (5.20) into (5.19) hence finally yields

$$f^\star = \sum_{i=1}^{L} \alpha_i \, \psi_{\mathscr{D}} * (\psi_{\mathscr{D}} * \varphi_i).$$

Observe that the spherical convolutions in (5.20) are all well-defined, since $\mathscr{D}$ is spline-admissible. Indeed the kernel $\psi_{\mathscr{D}}$ of a spline-admissible operator $\mathscr{D}$ is (by definition of spline-admissibility) necessarily an ordinary function and hence in particular square-integrable. We can therefore consider the spherical convolution between $\psi_{\mathscr{D}}$ and any function in $\mathscr{L}^2(\mathbb{S}^{d-1})$, and the result of this convolution is again a function in $\mathscr{L}^2(\mathbb{S}^{d-1})$ (see Remark 3.4). ■

Remark 5.4 — *Case $\mathscr{D} = Id$.* *Observe that when $\mathscr{D}$ is chosen as the identity (which is indeed a pseudo-differential operator as per Definition 4.1), the solution $f^\star$ is included in the span of the sampling functionals. This fact is of course well-known for quadratic cost functionals, but much less so for arbitrary proper and convex cost functionals.*

It is important to note that Theorem 5.3 was established under the assumption that the pseudo-differential operator $\mathscr{D}$ has a *trivial nullspace*. While this may seem like a limiting assumption for practical purposes, we show in the next remark that pseudo-differential operators with nontrivial nullspaces (such as the Laplace-Beltrami operator $\Delta_{\mathbb{S}^{d-1}}$) can be brought into the scope of Theorem 5.3 if properly regularised on their nullspace.

Remark 5.5 — *gTikhonov Regularisation with Non-Injective Operators.* *Consider a pseudo-differential operator $\mathscr{D}$ with Fourier coefficients $\{\hat{D}_n\}_{n\in\mathbb{N}}$ such that $\mathfrak{K}_{\mathscr{D}} = \left\{ n \in \mathbb{N} : \ |\hat{D}_n| = 0 \right\} \neq \emptyset$. Then, we have from Proposition 4.1 that the*

nullspace $\mathcal{N}(\mathscr{D})$ of $\mathscr{D}$ is nontrivial *and given by*

$$\mathcal{N}(\mathscr{D}) = span\{Y_n^m,\, n \in \mathfrak{K}_{\mathscr{D}},\, m = 1, \dots, N_d(n)\}.$$

Because of this nontrivial nullspace, the operator $\mathscr{D}$ cannot be used to define a proper gTikhonov regularisation norm. To circumvent this issue, we propose to regularise $\mathscr{D}$ by changing its definition on its nullspace. To this end, we consider the orthogonal projection operator[12] $\Pi_{\mathcal{N}(\mathscr{D})} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathcal{N}(\mathscr{D})$. *Then, the regularised operator $\mathcal{L}_\gamma = \mathscr{D} + \gamma\Pi_{\mathcal{N}(\mathscr{D})}$ with $\gamma > 0$ is a* pseudo-differential operator with trivial nullspace. *We have indeed*

$$\mathcal{L}_\gamma\varphi = \sum_{n \notin \mathfrak{K}_{\mathscr{D}}} \hat{D}_n \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m \;+\; \gamma \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m, \qquad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}),$$

[12] The latter is easily shown to be given by $\Pi_{\mathcal{N}(\mathscr{D})} = \mathrm{Id} - \mathscr{D}\mathscr{D}^\dagger$, where $\mathscr{D}^\dagger$ is the *Moore-Penrose pseudo-inverse* of $\mathscr{D}$ provided in (4.4).

and hence the Fourier coefficients of $\mathcal{L}_\gamma$ are all nonzero and given by

$$\hat{L}_n^\gamma = \begin{cases} \gamma & \text{if } n \in \mathfrak{K}_{\mathscr{D}}, \\ \hat{D}_n & \text{if } n \notin \mathfrak{K}_{\mathscr{D}}. \end{cases}$$

Note that the operator $\mathcal{L}_\gamma$ coincides exactly with $\mathscr{D}$ when restricted to $\mathcal{N}(\mathscr{D})^\perp$, and behaves like an homothety when restricted to $\mathcal{N}(\mathscr{D})$. Being bijective, $\mathcal{L}_\gamma$ can moreover be used to define a gTikhonov regularisation norm. Since $\mathscr{D}$ and $\gamma\Pi_{\mathcal{N}(\mathscr{D})}$ have orthogonal ranges, *the* Pythagorean theorem *moreover gives us:*

$$\|\mathcal{L}_\gamma f\|_2^2 = \|\mathscr{D} f\|_2^2 + \gamma^2 \|\Pi_{\mathcal{N}(\mathscr{D})} f\|_2^2 = \|\mathscr{D} f\|_2^2 + \gamma^2 \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2, \;\; \forall f \in \mathscr{S}'(\mathbb{S}^{d-1}).$$

The FPT problem associated to this choice of gTikhonov regularisation norm is then:

$$\min_{f \in \mathscr{H}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1})} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \;+\; \lambda \left[\|\mathscr{D} f\|_2^2 + \gamma^2 \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2 \right] \right\}, \qquad (5.21)$$

where the sampling operator $\boldsymbol{\Phi} : \mathscr{H}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1}) \to \mathbb{C}^L, f \mapsto [\langle \varphi_1 | f \rangle, \dots, \langle \varphi_L | f \rangle]$ is such that $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{H}'_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1})$. Observe that the gTikhonov norm induced by $\mathcal{L}_\gamma$ penalises both the bending energy $\|\mathscr{D} f\|_2^2$ of an element $f \in \mathscr{H}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1})$ and the $\mathscr{L}^2$ norm of its projection onto the nullspace of $\mathscr{D}$. From the representer Theorem 5.3, the solution to (5.21) exists, is unique, and given by

$$\begin{aligned} f^\star &= \sum_{i=1}^{L} \alpha_i \left[\mathscr{D} + \gamma\Pi_{\mathcal{N}(\mathscr{D})}\right]^{-2} \varphi_i \\ &= \sum_{i=1}^{L} \alpha_i \left[\mathscr{D}^{2\dagger} + \frac{1}{\gamma^2}\Pi_{\mathcal{N}(\mathscr{D})}\right] \varphi_i \end{aligned}$$

$$= \sum_{i=1}^{L} \alpha_i \left[\sum_{n \notin \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \frac{\langle \varphi_i | Y_n^m \rangle}{|\hat{D}_n|^2} Y_n^m + \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \sum_{m=1}^{N_d(n)} \frac{\langle \varphi_i | Y_n^m \rangle}{\gamma^2} Y_n^m \right], \tag{5.22}$$

for some coefficients $\{\alpha_1, \dots, \alpha_L\} \in \mathbb{C}$. *Note that if* γ *is close to zero, the solution is mainly contained in the nullspace* $\mathcal{N}(\mathscr{D})$, *while if* γ *is very large, the solution is mainly contained in the orthogonal complement* $\mathcal{N}(\mathscr{D})^{\perp}$ *of the nullspace.*

Finally, when $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{L}^2(\mathbb{S}^{d-1})$ *and* $\mathscr{D}$ *is* spline-admissible, *it is moreover possible, using the addition theorem 3.2 and similar arguments as in the last part of the proof of Theorem 5.3, to rewrite (5.22) as*

$$f^\star = \sum_{i=1}^{L} \alpha_i \left[\psi_{\mathscr{D}} * (\psi_{\mathscr{D}} * \varphi_i) + \frac{1}{\gamma^2} \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d} * \varphi_i \right],$$

where $P_{n,d} : [-1, 1] \to \mathbb{R}$ *is the* ultraspherical polynomial *of degree* n *defined in Definition 3.4, and* $\mathfrak{a}_d > 0$ *denotes the surface area of* $\mathbb{S}^{d-1}$.

3.2 Representer Theorem for gTV Regularisation

Our second representer theorem characterises the solutions of FPBP problems:

Theorem 5.4 — *Representer Theorem for gTV Regularisation.* Consider the following assumptions:

- F1 $\mathscr{D} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathscr{S}'(\mathbb{S}^{d-1})$ is some *pseudo-differential operator* with *trivial nullspace* and *Green functions* $\{\Psi_{\boldsymbol{r}}^{\mathscr{D}}, \boldsymbol{r} \subset \mathbb{S}^{d-1}\}$;
- F2 $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_{TV})$ is the space defined in (5.6), with *topological dual* $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ characterised in (5.13);
- F3 $\operatorname{span}\{\varphi_i, i = 1, \dots, L\} \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$, with the φ_i being *linearly independent*;
- F4 $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is a *sampling* operator, defined as

$$\boldsymbol{\Phi}(f) = (\langle f | \varphi_1 \rangle, \cdots, \langle f | \varphi_L \rangle), \quad \forall f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1});$$

- F5 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

  is *proper*, *convex* and lower semi-continuous;
- F6 λ is a *positive* regularisation constant.

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the solution set of the FPBP problem

$$\mathcal{V} = \underset{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) + \lambda \|\mathscr{D} f\|_{TV} \right\}, \tag{5.23}$$

is *nonempty*, and the *weak* closed convex hull* of its *extreme points*. The latter

are moreover necessarily of the form:

$$f^\star = \sum_{i=1}^{M} \alpha_i \, \mathscr{D}^{-1}\delta_{\boldsymbol{r}_i} = \sum_{i=1}^{M} \alpha_i \, \Psi_{\boldsymbol{r}_i}^{\mathscr{D}} = \sum_{i=1}^{M} \alpha_i \left[\sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{Y_n^m(\boldsymbol{r}_i)}{\hat{D}_n} Y_n^m \right], \quad (5.24)$$

for some weights $\{\alpha_1, \ldots, \alpha_M\} \subset \mathbb{C}$, directions $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M\} \subset \mathbb{S}^{d-1}$, and where $1 \leq M \leq L$. In particular when $\mathscr{D}$ is *spline-admissible*, the extreme points of $\mathcal{V}$ are *ordinary functions* and take the form of *$\mathscr{D}$-splines*

$$f^\star(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \, \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle), \qquad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \quad (5.25)$$

where $\psi_{\mathscr{D}}$ is the *zonal Green kernel* of $\mathscr{D}$.

Proof. We apply Theorem 2.12 to (5.23). To this end, we set $(\mathscr{B}, \|\cdot\|_{\mathscr{B}}) = (\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}^{-1}\cdot\|_\infty)$, $(\mathscr{B}', |||\cdot|||) = (\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}\cdot\|_{TV})$ and $\Lambda(t) = \lambda t$. The assumptions of Theorem 2.12 are then indeed verified since, from the discussion in Section 2.2, $(\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}^{-1}\cdot\|_\infty)$ and $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}\cdot\|_{TV})$ form a duality pair of Banach spaces, and Λ is convex and strictly increasing. We get hence from Theorem 2.12 that the solution set to the FPBP problem (5.23) is *nonempty* and the *weak* closed convex hull* of its *extreme points*. The latter are moreover necessarily of the form:

$$f^\star = \sum_{i=1}^{M} \beta_i e_i, \quad (5.26)$$

where $1 \leq M \leq L$, $\{\beta_1, \ldots, \beta_M\} \subset \mathbb{C}$ and $e_i \in \mathscr{S}'(\mathbb{S}^{d-1})$ are *extreme points* of the closed regularisation ball

$$\mathcal{B}_{gTV,1/\lambda} = \{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) : \|\mathscr{D}f\|_{TV} \leq 1/\lambda\}.$$

We now compute the extreme points of $\mathcal{B}_{gTV,1/\lambda}$. Adopting the notations of Definition 2.4, we denote by $\delta\mathcal{V}$ the extreme points of an arbitrary convex set $\mathcal{V}$. We define moreover the gTV unit ball on $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$

$$\mathcal{B}_{gTV} = \{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) : \|\mathscr{D}f\|_{TV} \leq 1\},$$

as well as the TV unit ball on the space of $\mathbb{C}$-valued regular Borel measures $\mathcal{M}(\mathbb{S}^{d-1})$:

$$\mathcal{B}_{TV} = \{f \in \mathcal{M}(\mathbb{S}^{d-1}) : \|f\|_{TV} \leq 1\}.$$

First, we trivially have $\delta\mathcal{B}_{gTV,1/\lambda} = \lambda^{-1}\delta\mathcal{B}_{gTV}$. Second, we have shown in Section 2.2 that $\mathscr{D}^{-1}$ is an *isometric isomorphism* from $\mathcal{M}(\mathbb{S}^{d-1})$ to $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$, which yields $\delta\mathcal{B}_{gTV} = \mathscr{D}^{-1}(\delta\mathcal{B}_{TV})$ [28]. Finally, it is well known [177, Section 3.5] that extremes points of the total variation unit ball of complex regular Borel measures are of the form $z\delta_{\boldsymbol{r}}$ with $\boldsymbol{r} \in \mathbb{S}^{d-1}$ and $z \in \mathbb{C}$, $|z| = 1$. In

conclusion, we hence get

$$\delta\mathcal{B}_{gTV,1/\lambda} = \lambda^{-1}\delta\mathcal{B}_{gTV} = \lambda^{-1}\mathscr{D}^{-1}(\delta\mathcal{B}_{TV}) = \left\{ z\lambda^{-1}\mathscr{D}^{-1}\delta_{\boldsymbol{r}},\ \boldsymbol{r}\in\mathbb{S}^{d-1}, |z|=1 \right\}. \tag{5.27}$$

Plugging (5.27) into (5.26) allows us to write any extreme point of the solution set (5.23) as

$$f^\star = \sum_{i=1}^{M} \frac{\beta_i z_i}{\lambda}\mathscr{D}^{-1}\delta_{\boldsymbol{r}_i} = \sum_{i=1}^{M}\alpha_i\mathscr{D}^{-1}\delta_{\boldsymbol{r}_i},$$

for some constants $\{\alpha_1,\dots,\alpha_M\}\subset\mathbb{C}$ and directions $\{\boldsymbol{r}_1,\dots,\boldsymbol{r}_M\}\subset\mathbb{S}^{d-1}$, and where $1\leq M\leq L$. This provides us with the first equality in (5.24). The second equality in (5.24) follows trivially from the definition of Green functions (see Definition 4.2). The last equality in (5.24) is obtained by considering the gSHT of $\mathscr{D}^{-1}\delta_{\boldsymbol{r}}$, $\boldsymbol{r}\in\mathbb{S}^{d-1}$. We have indeed,

$$\left\langle \mathscr{D}^{-1}\delta_{\boldsymbol{r}} \middle| Y_n^m \right\rangle = \left\langle \delta_{\boldsymbol{r}} \middle| \mathscr{D}^{-1}Y_n^m \right\rangle = \frac{Y_n^m(\boldsymbol{r})}{\hat{D}_n}, \quad \forall n\in\mathbb{N},\ m=1,\dots,N_d(n),$$

and hence $\mathscr{D}^{-1}\delta_{\boldsymbol{r}} = \sum_{n\in\mathbb{N}}\sum_{m=1}^{N_d(n)}\hat{D}_n^{-1}Y_n^m(\boldsymbol{r})Y_n^m$. Finally, we have from Proposition 4.4 that

$$f^\star = \sum_{i=1}^{M}\alpha_i\Psi_{\boldsymbol{r}_i}^{\mathscr{D}} = \sum_{i=1}^{M}\alpha_i\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_i\rangle),$$

where $\psi_{\mathscr{D}}$ is the zonal Green kernel of $\mathscr{D}$ and where the equality is in the sense of (4.10). Moreover, when $\mathscr{D}$ is spline-admissible, all traces $\{\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}\rangle),\ \boldsymbol{r}\in\mathbb{S}^{d-1}\}$ of the zonal Green kernel are ordinary functions and hence $f^\star$ is also an ordinary function. This shows (5.25) and achieves the proof. ■

Remark 5.6 *Theorem 5.4 allows us to write the solution set $\mathcal{V}$ of (5.23) as the weak* closed convex-hull of a (potentially infinite) set of extreme points $\delta\mathcal{V}\subset\mathcal{V}$:*

$$\mathcal{V} = \overline{Hull\,(\delta\mathcal{V})}^{weak^*} = \overline{\left\{ \sum_{k=1}^{n}\alpha_{i_k}f_{i_k}^\star \,\middle|\, n\in\mathbb{N},\ \{i_1,\dots,i_n\}\subset\mathbb{N},\ \sum_{k=1}^{n}\alpha_{i_k}=1, and\ \ 0\leq\alpha_{i_k}\leq 1, f_{i_k}^\star\in\delta\mathcal{V} \right\}}^{weak^*}, \tag{5.28}$$

where the extreme points $f^\star\in\delta\mathcal{V}$ are, when $\mathscr{D}$ is spline-admissible, $\mathscr{D}$-splines of the form:

$$f^\star(\boldsymbol{r}) = \sum_{(\alpha_m,\boldsymbol{r}_m)\in\Xi(f^\star)}\alpha_m\psi_{\mathscr{D}}(\langle\boldsymbol{r},\boldsymbol{r}_m\rangle), \quad \boldsymbol{r}\in\mathbb{S}^{d-1}, \quad \#\Xi(f^\star)=M(f^\star)\leq L.$$

Note that the spline innovation sets $\{\Xi(f^\star),\ f^\star\in\delta\mathcal{V}\}\subset\mathscr{P}(\mathbb{C}\times\mathbb{S}^{d-1})$ are a priori unknown *and* different *for each extreme point. However, they all have* bounded *cardinality $\#\Xi(f^\star)\leq L$, where L corresponds to the dimension of the data vector $\boldsymbol{y}$. Extreme points are hence $\mathscr{D}$-splines with* sparse innovations*:* they have at most as many degrees of freedom as available data. *This remarkable*

result is reminiscent of a similar property of basis pursuit problems in discrete setups (see [178, Theorem 19]). Unfortunately, it is only valid for extreme points and does not hold for arbitrary interior points in $\mathcal{V}^\circ = \mathcal{V}\backslash\delta\mathcal{V}$. Indeed, $\mathcal{V}^\circ$ consists in general of all finite convex combinations of extreme points *in $\delta\mathcal{V}$, as well as* limits *of sequences of the latter under the weak* topology. Such limits may take the form of* infinite *summations and are hence not $\mathscr{D}$-splines anymore.*

Remark 5.7 — *Case $\mathscr{D} = \mathrm{Id}$.* *Observe that when $\mathscr{D}$ is chosen as the identity operator, the extreme points of $\mathcal{V}$ degenerate into Dirac measures and hence any solution of (5.23) is a convex combination of Dirac measures. Such functions are of course particularly cumbersome to work with in practice.*[13] *It is hence recommended to work with nontrivial pseudo-differential operators when using gTV regularisation strategies.*

[13] Indeed, they do not even have a pointwise interpretation.

Again, it is important to note that Theorem 5.3 was established under the assumption that the pseudo-differential operator $\mathscr{D}$ has a trivial nullspace. The case of pseudo-differential operators with nontrivial nullspaces can be treated similarly as for the gTikhonov regularisation by regularising the non-injective operator $\mathscr{D}$ (see Remark 5.5). In this case however, the gTV norm induced by the regularised operator is harder to interpret, as explained in the subsequent remark.

Remark 5.8 — *gTV Regularisation with Non-Injective Operators.* *Consider a pseudo-differential operator $\mathscr{D}$ with nontrivial nullspace $\mathcal{N}(\mathscr{D})$. Then, as explained in Remark 5.5, the regularised operator operator $\mathcal{L}_\gamma = \mathscr{D} + \gamma\Pi_{\mathcal{N}(\mathscr{D})}$ with $\gamma > 0$ is an injective* pseudo-differential operator. *The latter can hence be used to define a gTV regularisation norm. The FPBP problem associated to this choice of gTV regularisation norm is then:*

$$\min_{f\in\mathcal{M}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1})} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}(f)\right) \;+\; \lambda \|(\mathscr{D} + \gamma\Pi_{\mathcal{N}(\mathscr{D})})f\|_{TV} \right\}, \tag{5.29}$$

where the sampling operator $\boldsymbol{\Phi} : \mathcal{M}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1}) \to \mathbb{C}^L, f \mapsto [\langle f|\varphi_1\rangle, \dots, \langle f|\varphi_L\rangle]$ is such that $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{C}_{\mathcal{L}_\gamma}(\mathbb{S}^{d-1})$. Note that, in contrast with the gTikhonov norm in Remark 5.5, the gTV regularisation norm used in (5.29) cannot easily be related to the semi-norm $\|\mathscr{D}\cdot\|_{TV}$ and is hence harder to interpret physically. It is however still possible to invoke the representer theorem 5.4 to show that the solution set of (5.29) is nonempty *and the* weak* closed convex hull *of extreme points of the form*

$$\begin{aligned} f^\star &= \sum_{i=1}^M \alpha_i \left[\mathscr{D} + \gamma\Pi_{\mathcal{N}(\mathscr{D})}\right]^{-1} \delta_{\boldsymbol{r}_i} = \sum_{i=1}^M \alpha_i \left[\mathscr{D}^\dagger + \frac{1}{\gamma}\Pi_{\mathcal{N}(\mathscr{D})}\right] \delta_{\boldsymbol{r}_i} \\ &= \sum_{i=1}^M \alpha_i \left[\sum_{n\notin\mathfrak{K}_\mathscr{D}} \sum_{m=1}^{N_d(n)} \frac{Y_n^m(\boldsymbol{r}_i)}{\hat{D}_n} Y_n^m + \sum_{n\in\mathfrak{K}_\mathscr{D}} \sum_{m=1}^{N_d(n)} \frac{Y_n^m(\boldsymbol{r}_i)}{\gamma} Y_n^m \right], \end{aligned} \tag{5.30}$$

with $1 \leq M \leq L$, $\{\alpha_1, \dots, \alpha_M\} \in \mathbb{C}$ and $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_M\} \subset \mathbb{S}^{d-1}$. Note that if γ is close to zero, the solution is mainly contained in the nullspace $\mathcal{N}(\mathscr{D})$, while if γ is very large, the solution is mainly contained in the orthogonal complement

$\mathcal{N}(\mathscr{D})^\perp$ of the nullspace.

Finally, when $\mathscr{D}$ is spline-admissible*, it is moreover possible, using the addition theorem 3.2 and similar arguments as in the last part of the proof of Theorem 5.4, to rewrite (5.30) as an ordinary function, given by*

$$f^\star(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \left[\psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle) + \frac{1}{\gamma} \sum_{n \in \mathfrak{K}_{\mathscr{D}}} \frac{N_d(n)}{\mathfrak{a}_d} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle) \right], \quad \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{5.31}$$

where $P_{n,d} : [-1, 1] \to \mathbb{R}$ is the ultraspherical polynomial *of degree n defined in Definition 3.4, and $\mathfrak{a}_d > 0$ denotes the surface area of $\mathbb{S}^{d-1}$. Observe that although resembling it, (5.31) is* not *a $\mathscr{D}$-spline as the coefficients $\{\alpha_1, \ldots, \alpha_M\}$ have a priori no reason to verify $\sum_{i=1}^{M} \alpha_i Y_n^m(\boldsymbol{r}_i) = 0, \ \forall n \in \mathfrak{K}_{\mathscr{D}}, \ m = 1, \ldots, N_d(n)$, as requested for spherical splines associated to a pseudo-differential operator with nontrivial nullspace (see discussion in Remark 4.7).*

3.3 Comparison

Given the conclusions of Theorems 5.3 and 5.4, gTikhonov regularisation appears like a much more efficient regularisation strategy than gTV regularisation, at least in terms of making the functional inverse problem well-posed. Indeed, not only is the solution to the gTikhonov-regularised problem (5.16) unique, but it also has *exactly as many* degrees of freedom (df) as the number of independent measurements, making it easy to approximate numerically. In contrast, the gTV-regularised problem (5.23) is not guaranteed to have a unique solution, and –as explained in Remark 5.6– its non-extreme solutions may not even necessarily have finite df, making them potentially difficult to approximate numerically. In a sense, gTV regularisation can hence be considered as failing to properly regularise the functional inverse problem. Nevertheless, it still presents one main advantage over gTikhonov regularisation, which partly explains its huge adoption by practitioners: it produces much "nicer-looking" solutions. This is because, gTikhonov regularisation has a tendency of producing *overly-smooth* estimates. Indeed, (5.17) reveals that the solution $f^\star$ to the gTikhonov-regularised problem (5.16) is a linear combination of the sampling functionals *primitived twice* w.r.t. the pseudo-differential operator $\mathscr{D}$. Roughly speaking, this means that $f^\star$ will be *twice as smooth* as the sampling linear functionals φ_i. In the case where the sampling functionals are square-integrable and $\mathscr{D}$ spline-admissible, (5.18) makes even more obvious this smoothing behaviour, by showing that primitiving twice the sampling functionals is equivalent to convolving them twice with the zonal Green kernel $\psi_{\mathscr{D}}$ of $\mathscr{D}$. In contrast, the solutions to the gTV-regularised problem (5.23) are in general convex combinations of Green functions of $\mathscr{D}$, which are by definition $\mathscr{D}$-primitives of Dirac measures and hence much less smooth.[14] This makes gTV regularisation much more indicated for recovering spherical fields with *sharp variations*, often encountered in practice.

One last important difference between the gTikhonov and gTV regularised problems (5.16) and (5.23) is the dependency of their solutions on the sam-

[14] As a matter of fact, we showed in Remark 5.3 that Green functions are not regular enough to even be included in the search space $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ of the gTikhonov-regularised problem (5.16).

pling functionals. Indeed, the solution to the gTikhonov-regularised problem (5.16) depends *linearly* on the linear functionals, while solutions to the gTV-regularised problem (5.23) appear *independent*[15] from the sampling functionals. Theorem 2.9 tells us that in reality there still exists a link between the solution set of (5.23) and the sampling functionals. The latter is however much more subtle since expressed through the unknown and *nonlinear* set-valued *duality map* of the space $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ (see Definition 2.3). Nevertheless, solutions to the gTV-regularised problem (5.23) are not as constrained as the solution to the gTikhonov-regularised problem (5.16) by the specific shapes of the sampling functionals, and should hence have a stronger approximation power. This however, comes at the cost of additional unknowns, namely the knots $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_M\}$ intervening in the parametrisation of the extreme points of the solution set of the gTV-regularised problem.

[15] Indeed, the sampling functionals do not intervene explicitly in (5.24) characterising the form of the extreme points of the solution set of (5.23).

4 Spatial Interpolation

We specify here both representer theorems to the specific case of *spatial interpolation*, particularly relevant for practical purposes. The goal of interpolation is to find a function, called *interpolant*, with *minimal* gTikhonov or gTV norm respectively, and with prescribed values $\{y_i,\, i = 1, \dots, L\} \subset \mathbb{C}$ at sampling locations $\Theta_L = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$:

$$y_i = f(\boldsymbol{r}_i), \quad \forall i = 1, \dots, L.$$

Mathematically, this translates into an optimisation problem of the form

$$\mathcal{V} = \underset{f \in \mathscr{B}'}{\arg\min} \left\{ |||f||| \quad \text{subject to} \quad f(\boldsymbol{r}_i) = y_i,\, \forall i = 1, \dots, L \right\}, \tag{5.32}$$

where $(\mathscr{B}', |||\cdot|||)$ is either $(\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_2)$ or $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_{TV})$ depending on the chosen regularisation strategy. Let $\iota : \mathbb{C}^L \to \{0, +\infty\}$ be the convex and proper *indicator function* defined as

$$\iota(\boldsymbol{z}) = \begin{cases} 0 & \text{if } \boldsymbol{z} = \boldsymbol{0}, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.33}$$

Then, it is easy to see that the interpolation problem (5.32) can be seen as an instance of (5.4), with $\Lambda = \mathrm{Id}$, a cost function $F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) = \iota(\boldsymbol{y} - \boldsymbol{\Phi}(f))$, and a sampling operator $\boldsymbol{\Phi} : \mathscr{B}' \to \mathbb{C}^L$ of the form

$$\boldsymbol{\Phi}(f) = [\langle f | \delta_{\boldsymbol{r}_1} \rangle, \cdots, \langle f | \delta_{\boldsymbol{r}_L} \rangle], \quad \forall f \in \mathscr{B}', \tag{5.34}$$

where $\{\delta_{\boldsymbol{r}_1}, \dots, \delta_{\boldsymbol{r}_L}\} \subset \mathcal{M}(\mathbb{S}^{d-1})$ are *Dirac measures*. We can hence use Theorems 5.3 and 5.4 to characterise the solutions of the generic spherical interpolation problem (5.32) in the specific case of gTikhonov and gTV regularisation respectively.

4.1 gTikhonov Regularised Interpolation

Given a set of sample values $\{y_1, \dots, y_L\} \subset \mathbb{C}^L$ and locations $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$, consider the *gTikhonov-regularised interpolation problem*:

$$\underset{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \{\|\mathscr{D} f\|_2 \quad \text{subject to} \quad f(\boldsymbol{r}_i) = y_i,\ \forall i = 1, \dots, L\}, \tag{5.35}$$

where $\mathscr{D}$ is some pseudo-differential operator and $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is the generalised Sobolev space introduced in Section 2.1. As previously discussed, this problem is clearly equivalent to the following FPT problem:

$$\underset{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \left\{\iota(\boldsymbol{y} - \boldsymbol{\Phi}(f)) + \|\mathscr{D} f\|_2^2\right\}, \tag{5.36}$$

where $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{C}^L$, $\iota : \mathbb{C}^L \to \{0, +\infty\}$ is the indicator function (5.33) and the sampling operator $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is given by (5.34) with $\mathscr{B}' = \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$. Note that for $\boldsymbol{\Phi}$ to be well-defined, we need the Dirac sampling functionals to be included in the dual $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ of $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$, i.e. $\{\delta_{\boldsymbol{r}_1}, \dots, \delta_{\boldsymbol{r}_L}\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$. Using the characterisation of $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ provided in Proposition 5.1, we give a sufficient condition on the spectral growth order of $\mathscr{D}$ for this to be true.

Lemma 5.5 — *Spatial Sampling and gTikhonov Regularisation.* *Let $\mathscr{D}$ be a* pseudo-differential operator *as in Definition 4.1 with* spectral growth order *$p > (d-1)/2$, and $(\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_2)$ the* generalised Sobolev space *defined in (5.5) equipped with the gTikhonov norm. Then, all Dirac measures are included in the dual $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ of $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$, i.e.*

$$\{\delta_{\boldsymbol{r}},\ \boldsymbol{r} \in \mathbb{S}^{d-1}\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

In other words, $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is a reproducing kernel Hilbert space (RKHS).

Proof. Using Proposition 5.1, we can show that $\delta_{\boldsymbol{r}} \in \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ if $\mathscr{D}^{-1}\delta_{\boldsymbol{r}}$ is in $\mathscr{L}^2(\mathbb{S}^{d-1})$, $\forall \boldsymbol{r} \in \mathbb{S}^{d-1}$. From the addition theorem 3.2 and Parseval's theorem, we have for all $\boldsymbol{r} \in \mathbb{S}^{d-1}$

$$\|\mathscr{D}^{-1}\delta_{\boldsymbol{r}}\|_2^2 = \sum_{n \in \mathbb{N}} \frac{1}{|\hat{D}_n|^2} \sum_{m=1}^{N_d(n)} |Y_n^m(\boldsymbol{r})|^2 = \sum_{n \in \mathbb{N}} \frac{N_d(n)}{\mathfrak{a}_d |\hat{D}_n|^2} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{r} \rangle) = \sum_{n \in \mathbb{N}} \frac{N_d(n)}{\mathfrak{a}_d |\hat{D}_n|^2},$$

where we have used the fact that ultraspherical polynomials are (by Definition 3.4) such that $P_{n,d}(1) = 1$. For $p > (d-1)/2$, we have moreover from (3.3) $N_d(n)|\hat{D}_n|^{-2} = \mathcal{O}(n^{-\beta})$ with $\beta = 2p - d + 2 > 1$ and hence $\|\mathscr{D}^{-1}\delta_{\boldsymbol{r}}\|_2^2 = \sum_{n \in \mathbb{N}} \frac{N_d(n)}{\mathfrak{a}_d |\hat{D}_n|^2} < +\infty$, or equivalently $\mathscr{D}^{-1}\delta_{\boldsymbol{r}} \in \mathscr{L}^2(\mathbb{S}^{d-1})$. The native space $\mathscr{H}_{\mathscr{D}}$ is hence a RKHS[16], since its topological dual contains all Dirac measures. ∎

[16] See [18] for a definition of a reproducing kernel Hilbert space.

We are now ready to formulate the representer theorem for gTikhonov-regularised spherical interpolation problems. This a classical result from spherical

approximation theory (see [125, Theorem 6.30]), but we derive it here as a corollary of Theorem 5.3.

Corollary 5.6 — *Representer Theorem for gTikhonov Interpolation.* *Consider the following scenario:*

G1 $\mathscr{D} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathscr{S}'(\mathbb{S}^{d-1})$ *is some* pseudo-differential operator *with* trivial nullspace *and spectral growth order* $p > (d-1)/2$*;*

G2 $(\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}), \langle \cdot, \cdot \rangle_{\mathscr{D}})$ *is the* generalised Sobolev space *defined in* (5.5)*, with* topological dual $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ *characterised in* (5.8).

Then, for any sample values $\{y_1, \dots, y_L\} \subset \mathbb{C}^L$ *and* sample directions $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$*, the solution to the* gTikhonov-regularised interpolation problem*:*

$$f^\star = \underset{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \{ \|\mathscr{D} f\|_2 \quad \textit{subject to} \quad f(\boldsymbol{r}_i) = y_i, \, \forall i = 1, \dots, L \}, \tag{5.37}$$

exists*, is* unique *and is a* $\mathscr{D}^2$*-spline, given by*

$$f^\star(\boldsymbol{r}) = \sum_{i=1}^{L} \alpha_i \, \psi_{\mathscr{D}^2}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle), \qquad \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{5.38}$$

where $\psi_{\mathscr{D}^2}$ *is the* zonal Green kernel *of* $\mathscr{D}^2$*. The* weights $\boldsymbol{\alpha} := [\alpha_1, \dots, \alpha_L] \in \mathbb{C}^L$ *are moreover solutions to the square linear system*

$$\boldsymbol{G}\boldsymbol{\alpha} = \boldsymbol{y}, \quad \textit{where} \quad G_{ij} := \psi_{\mathscr{D}^2}(\langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle), \, \forall i, j = 1, \dots, L. \tag{5.39}$$

Proof. We formulate (5.37) equivalently as an unconstrained FPT problem

$$f^\star = \underset{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \left\{ \iota(\boldsymbol{y} - \boldsymbol{\Phi}(f)) + \|\mathscr{D} f\|_2^2 \right\}, \tag{5.40}$$

where $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{C}^L$, $\iota : \mathbb{C}^L \to \{0, +\infty\}$ is the indicator function (5.33), and the sampling operator $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is given by

$$\boldsymbol{\Phi}(f) = [\langle \delta_{\boldsymbol{r}_1} | f \rangle, \cdots, \langle \delta_{\boldsymbol{r}_L} | f \rangle], \quad \forall f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Notice that ι is *proper convex and lwsc* as indicator function of a convex set of $\mathbb{C}^L$ (here the singleton $\{\boldsymbol{y}\}$). Moreover, Lemma 5.5 tells us that, for a pseudo-differential operator $\mathscr{D}$ with spectral growth order $p > (d-1)/2$, all Dirac measures are included in the dual $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$, and hence in particular

$$\{\delta_{\boldsymbol{r}_1}, \dots, \delta_{\boldsymbol{r}_L}\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Finally, the Dirac measures associated to different directions are all linearly independent. We can hence apply Theorem 5.3 to the specific problem (5.40).

This gives us that $f^\star$ exists, is unique, and given by

$$f^\star = \sum_{i=1}^{L} \alpha_i \mathscr{D}^{-2} \delta_{\boldsymbol{r}_i} = \sum_{i=1}^{L} \alpha_i \Psi_{\boldsymbol{r}_i}^{\mathscr{D}^2},$$

where $\{\Psi_{\boldsymbol{r}_i}^{\mathscr{D}^2},\, i = 1, \dots, L\} \subset \mathscr{S}'(\mathbb{S}^{d-1})$ are Green functions of the operator $\mathscr{D}^2$ and $\{\alpha_1, \dots, \alpha_L\} \subset \mathbb{C}^L$. Proposition 4.4 allows us moreover to identify the Green functions of $\mathscr{D}^2$ as traces of its zonal Green kernel $\psi_{\mathscr{D}^2}$, yielding (5.38):

$$f^\star(\boldsymbol{r}) = \sum_{i=1}^{L} \alpha_i \, \psi_{\mathscr{D}^2}(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle), \qquad \boldsymbol{r} \in \mathbb{S}^{d-1}.$$

Note that $f^\star$ is indeed an ordinary function with a pointwise interpretation since it belongs to the search space $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ which is in this case an RKHS (see Lemma 5.5). As a matter of fact, it is even a spherical spline for the positive-definite pseudo-differential operator $\mathscr{D}^2$, which, from Proposition 4.5, is indeed spline-admissible under the assumption $p > (d-1)/2$.

The linear system (5.39) is finally obtained by noticing that, as solution of the constrained optimisation problem (5.40), $f^\star$ verifies the L sampling equations

$$y_i = f^\star(\boldsymbol{r}_i) = \sum_{j=1}^{L} \alpha_j \, \psi_{\mathscr{D}^2}(\langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle), \quad i = 1, \dots, L,$$

which can be rewritten as the system (5.40). ■

Remark 5.9 — *Invertibility of G.* *Notice that the unicity and existence of $f^\star$ in Corollary 5.6 implies that the* Gram *matrix $\boldsymbol{G} \in \mathbb{C}^{L \times L}$ defined in (5.39) is necessarily* invertible *for any set of L* sample directions *$\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$. We say that the zonal Green kernel $\psi_{\mathscr{D}^2}$ is* strictly positive definite *[125, Definition 6.25]. As shown in [125, Theorem 6.27], zonal Green kernels associated to positive definite pseudo-differential operators (like $\mathscr{D}^2$ here) are indeed strictly positive definite.*

4.2 gTV Regularised Interpolation

Assuming again a given set of sample values $\{y_1, \dots, y_L\} \subset \mathbb{C}^L$ and sample locations $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$, we consider this time the *gTV-regularised interpolation problem* given by:

$$\underset{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \; \{\|\mathscr{D} f\|_{TV} \quad \text{subject to} \quad f(\boldsymbol{r}_i) = y_i, \, \forall i = 1, \dots, L\}, \tag{5.41}$$

where $\mathscr{D}$ is some pseudo-differential operator and $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is the search space introduced in Section 2.2. As previously discussed, this problem is clearly equivalent to the following FPBP problem:

$$\underset{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})}{\arg\min} \; \{\iota(\boldsymbol{y} - \boldsymbol{\Phi}(f)) + \|\mathscr{D} f\|_{TV}\}, \tag{5.42}$$

where $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{C}^L$, $\iota : \mathbb{C}^L \to \{0, +\infty\}$ is the indicator function (5.33) and the sampling operator $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is given by (5.34) with $\mathscr{B}' = \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$. Again, for the sampling operator to be well-defined, we need

$$\{\delta_{\boldsymbol{r}_1}, \dots, \delta_{\boldsymbol{r}_L}\} \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Using the characterisation of $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ provided in Proposition 5.13, we give a sufficient condition on the spectral growth order of $\mathscr{D}$ for this to be true.

Lemma 5.7 — *Spatial Sampling and gTV Regularisation.* *Let $\mathscr{D}$ be a* pseudo-differential operator *as in Definition 4.1 with* spectral growth order $p > d-1$, *and $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_{TV})$ the space defined in (5.6) equipped with the gTV norm. Then, all Dirac measures are included in the predual $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ of $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$, i.e.*

$$\{\delta_{\boldsymbol{r}},\ \boldsymbol{r} \in \mathbb{S}^{d-1}\} \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Proof. According to Proposition 5.2, $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ can be characterised as

$$\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}) = \left\{h \in \mathscr{S}'(\mathbb{S}^{d-1}) :\ h = \mathscr{D}\eta,\ \eta \in \mathscr{C}(\mathbb{S}^{d-1})\right\}.$$

Thus, the Dirac measures $\{\delta_{\boldsymbol{r}},\ \boldsymbol{r} \in \mathbb{S}^{d-1}\}$ belong to $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ i.f.f. there exists, for every $\boldsymbol{r} \in \mathbb{S}^{d-1}$, a function $\Psi_{\boldsymbol{r}} \in \mathscr{C}(\mathbb{S}^{d-1})$ such that:

$$\mathscr{D}\Psi_{\boldsymbol{r}} = \delta_{\boldsymbol{r}}.$$

The functions $\Psi_{\boldsymbol{r}}$ satisfying the above equation are actually the Green functions of $\mathscr{D}$, which, from Proposition 4.4, can be identified with traces $\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}\rangle)$ of the zonal Green kernel of $\mathscr{D}$. Finally, we have shown in Proposition 4.5 that, for a pseudo-differential operator with spectral growth order $p > d-1$

$$\{\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}\rangle),\ \boldsymbol{r} \in \mathbb{S}^{d-1}\} \subset \mathscr{C}(\mathbb{S}^{d-1}),$$

and hence $\Psi_{\boldsymbol{r}}$ can indeed be identified with a continuous function and consequently all Dirac measures belong to the predual $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$. ■

We now state the representer theorem for gTV-regularised spherical interpolation problems.

Corollary 5.8 — *Representer Theorem for gTV Interpolation.* *Consider the following assumptions:*

- *H1* *$\mathscr{D} : \mathscr{S}'(\mathbb{S}^{d-1}) \to \mathscr{S}'(\mathbb{S}^{d-1})$ is some* pseudo-differential operator *with* trivial nullspace *and* spectral growth *order $p > d-1$;*
- *H2* *$(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D} \cdot\|_{TV})$ is the space defined in (5.6), with* topological dual *$\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ characterised in (5.13).*

Then, for any sample values *$\{y_1, \dots, y_L\} \subset \mathbb{C}^L$ and* sample directions

$\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^{d-1}$*, the solution set of the* gTV-regularised interpolation problem*:*

$$\mathcal{V} = \operatorname*{arg\,min}_{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})} \{\|\mathscr{D}f\|_{TV} \quad \textit{subject to} \quad f(\boldsymbol{r}_i) = y_i, \forall i = 1, \dots, L\}, \tag{5.43}$$

is nonempty*, and the* weak* closed convex hull *of its* extreme points*. The latter take moreover necessarily the form of a* $\mathscr{D}$-spline*:*

$$f^\star(\boldsymbol{r}) = \sum_{i=1}^{M} \alpha_i \, \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{\rho}_i \rangle), \qquad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{5.44}$$

for some weights $\{\alpha_1, \dots, \alpha_M\} \subset \mathbb{C}$*, directions* $\{\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_M\} \subset \mathbb{S}^{d-1}$*, and where* $1 \leq M \leq L$ *and* $\psi_{\mathscr{D}}$ *is the* zonal Green kernel *of* $\mathscr{D}$*.*

Proof. Similarly as in the proof of Corollary 5.6, we reformulate the interpolation problem (5.43) as an *unconstrained* FPBP optimisation problem:

$$\mathcal{V} = \operatorname*{arg\,min}_{f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})} \{\iota(\boldsymbol{y} - \boldsymbol{\Phi}(f)) + \|\mathscr{D}f\|_{TV}\}, \tag{5.45}$$

where $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{C}^L$, $\iota : \mathbb{C}^L \to \{0, +\infty\}$ is the indicator function (5.33) and the sampling operator $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is given by

$$\boldsymbol{\Phi}(f) = [\langle f | \delta_{\boldsymbol{r}_1} \rangle, \cdots, \langle f | \delta_{\boldsymbol{r}_L} \rangle], \quad \forall f \in \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Notice that ι is *proper convex and lwsc*, as indicator function of a convex set of $\mathbb{C}^L$ (here the singleton $\{\boldsymbol{y}\}$). Moreover, Lemma 5.7 tells us that, for a pseudo-differential operator $\mathscr{D}$ with spectral growth order $p > d-1$, all Dirac measures are included in the predual $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$, and hence in particular $\{\delta_{\boldsymbol{r}_1}, \dots, \delta_{\boldsymbol{r}_L}\} \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$. Finally, Dirac measures associated to different directions are all linearly independent. We can hence apply Theorem 5.4 to (5.45), which tells us that the solution set $\mathcal{V}$ is *nonempty* and the *weak* closed convex hull* of its extreme points. Since $p > d-1$, the *spline-admissibility* of $\mathscr{D}$ is guaranteed by Proposition 4.5. From (5.25), we obtain hence that the extreme points of $\mathcal{V}$ are indeed $\mathscr{D}$-splines of the form (5.44), which achieves the proof. ∎

6 Discretisation

In order to solve the FPT and FPBP problems (5.16) and (5.23) in practice, one must necessarily convert them into *finite-dimensional optimisation problems*. This process, called *discretisation*, can be performed in two ways. First, by *restricting the search space* to a finite-dimensional subspace, for example the span of some finite family of basis functions. Second, by *approximating the domain* $\mathbb{S}^{d-1}$ of the sought spherical fields by a *discrete manifold* [30], typically taking the form of a *tessellation graph*. We call these two strategies *search space discretisation* and *domain discretisation* respectively. Due to its appealing conceptual simplicity, domain discretisation is the prevailing approach in most applications. Unfortunately, it *necessarily* incurs some *approximation error*, sometimes even when the number of points composing the tessellation graph tends to infinity [30, 189]. This approximation error is moreover often very difficult to assess. In contrast, search space discretisation can in certain cases be *exact*, resulting in a finite-dimensional problem strictly equivalent to the original infinite-dimensional one. This is for example the case for the discretisation scheme proposed in Section 1.1 for FPT problems with gTikhonov regularisation. For non-exact search space discretisation schemes, it is moreover often possible to precisely assess the incurred approximation error, as we will demonstrate in Section 1.2 for the discretisation scheme proposed for FPBP problems with gTV regularisation. Finally, it is worth noting that the search space discretisation schemes proposed in Section 1.1 and Section 1.2 are both *canonical* to the gTikhonov and gTV regularisation norms respectively, in the sense that they transform the latter into their discrete counterparts, namely the ℓ_2 and ℓ_1 norm.

1 Canonical Search Space Discretisation

In this section, we use the representer theorems 5.3 and 5.4 to derive *canonical search space discretisation schemes* for the FPT and FPBP problems (5.16) and (5.23) respectively. The idea of search space discretisation schemes is to restrict the search space $\mathscr{B}'$ of an optimisation problem to a well-chosen finite-dimensional subspace of the form $\operatorname{span}\{\psi_1, \ldots, \psi_N\} \subset \mathscr{B}'$. To facilitate the formulation of our discretisation schemes, we need to introduce the concept of *syn-*

thesis operator associated to a family of (generalised functions) $\{\psi_1, \dots, \psi_N\}$ [183, Chapter 5]:

Definition 6.1 — *Synthesis Operator.* *Let $\mathfrak{F} = \{\psi_1, \dots, \psi_N\} \subset \mathscr{S}'(\mathbb{S}^{d-1})$ be a family of generalised functions. The* synthesis operator $\Psi : \mathbb{C}^N \to \mathit{span}\{\psi_1, \dots, \psi_N\}$ *associated with $\mathfrak{F}$ is defined as*

$$\Psi : \begin{cases} \mathbb{C}^N & \to \mathit{span}\{\psi_1, \dots, \psi_N\} \\ \boldsymbol{x} & \mapsto \Psi(\boldsymbol{x}) = \sum_{k=1}^{N} x_k \psi_k, \end{cases}$$

where x_k denotes the kth entry of the vector $\boldsymbol{x} = [x_1, \dots, x_N] \in \mathbb{C}^N$.

Since it transforms vectors into functions, the synthesis operator is often called an *interpolation* or *backprojection* operator. Observe that $\operatorname{span}\{\psi_1, \dots, \psi_N\}$ can be seen as the image of $\mathbb{C}^N$ by Ψ:

$$\operatorname{span}\{\psi_1, \dots, \psi_N\} = \left\{ \sum_{k=1}^{N} x_k \psi_k,\ x_1, \dots, x_N \in \mathbb{C} \right\} = \left\{ \Psi(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{C}^N \right\} = \Psi(\mathbb{C}^N).$$

When the functions $\{\psi_1, \dots, \psi_N\}$ are *linearly independent*, the synthesis operator Ψ is injective and hence defines an isomorphism from $\mathbb{C}^N$ to $\operatorname{span}\{\psi_1, \dots, \psi_N\}$. The inverse map from $\operatorname{span}\{\psi_1, \dots, \psi_N\}$ to $\mathbb{C}^N$ is moreover given by the (right) pseudo-inverse of Ψ, which can be expressed in terms of its adjoint:

Proposition 6.1 — *Pseudo-Inverse of Synthesis Operator.* *Let $\{\psi_1, \dots, \psi_N\} \subset \mathscr{S}'(\mathbb{S}^{d-1})$ be a family of* linearly independent *generalised functions, with associated synthesis operator $\Psi : \mathbb{C}^N \to \mathit{span}\{\psi_1, \dots, \psi_N\}$. Consider moreover some* arbitrary inner product[1] *$\langle \cdot, \cdot \rangle$ on the finite dimensional subspace $\mathit{span}\{\psi_1, \dots, \psi_N\}$, and the* canonical inner product *$\langle \cdot, \cdot \rangle_{\mathbb{C}^N}$ on $\mathbb{C}^N$. Then, the* (right) pseudo-inverse[2] *of Ψ is given by*

$$\boldsymbol{\Psi}^\dagger = (\boldsymbol{\Psi}^* \Psi)^{-1} \boldsymbol{\Psi}^*, \tag{6.1}$$

where $\boldsymbol{\Psi}^* : \mathit{span}\{\psi_1, \dots, \psi_N\} \to \mathbb{C}^N$ is the adjoint[3] *of Ψ w.r.t. the inner products on $\mathit{span}\{\psi_1, \dots, \psi_N\}$ and $\mathbb{C}^N$, given by*

$$\boldsymbol{\Psi}^*(f) = [\langle f, \psi_1 \rangle, \dots, \langle f, \psi_N \rangle], \qquad \forall f \in \mathit{span}\{\psi_1, \dots, \psi_N\}. \tag{6.2}$$

[1] Recall that any finite dimensional complex vector space can be made into an inner product. One possible inner product is indeed the sesquilinear form $\langle \cdot, \cdot \rangle$ s.t. $\langle \psi_k, \psi_j \rangle = \delta_{kj},\ \forall k, j$.

[2] A is a *right inverse* of B if $AB = \mathrm{Id}$.

[3] The adjoint of Ψ is defined by the relationship $\langle \Psi(\boldsymbol{x}), g \rangle = \langle \boldsymbol{x}, \boldsymbol{\Psi}^*(g) \rangle_{\mathbb{C}^N}$, for all $\boldsymbol{x} \in \mathbb{C}^N$ and $g \in \operatorname{span}\{\psi_1, \dots, \psi_N\}$.

Proof. First, we have from the sesquilinearity of the inner product,

$$\langle \Psi(\boldsymbol{x}), f \rangle = \left\langle \sum_{k=1}^{N} x_k \psi_k, f \right\rangle = \sum_{k=1}^{N} x_k \langle \psi_k, f \rangle = \sum_{k=1}^{N} x_k \overline{\langle f, \psi_k \rangle} = \langle \boldsymbol{x}, \boldsymbol{\Psi}^*(f) \rangle_{\mathbb{C}^N},$$

for all $\boldsymbol{x} \in \mathbb{C}^N$ and $f \in \operatorname{span}\{\psi_1, \dots, \psi_N\}$. Hence $\boldsymbol{\Psi}^*$ in (6.2) is indeed the adjoint of Ψ. Consider now the composition map $\boldsymbol{\Psi}^* \Psi : \mathbb{C}^N \to \mathbb{C}^N$. For all

$\boldsymbol{x} \in \mathbb{C}^N$, we have

$$\begin{aligned}\boldsymbol{\Psi}^*(\Psi(\boldsymbol{x})) = \boldsymbol{\Psi}^*\left(\sum_{k=1}^N x_k\psi_k\right) &= \sum_{k=1}^N x_k\boldsymbol{\Psi}^*(\psi_k)\\ &= \sum_{k=1}^N x_k\begin{bmatrix}\langle\psi_k,\psi_1\rangle\\ \vdots\\ \langle\psi_k,\psi_N\rangle\end{bmatrix} = \underbrace{\begin{bmatrix}\langle\psi_1,\psi_1\rangle & \cdots & \langle\psi_N,\psi_1\rangle\\ \vdots & \ddots & \vdots\\ \langle\psi_1,\psi_N\rangle & \cdots & \langle\psi_N,\psi_N\rangle\end{bmatrix}}_{:=\boldsymbol{H}\in\mathbb{C}^{N\times N}}\boldsymbol{x},\end{aligned}$$

and hence $\boldsymbol{\Psi}^*\Psi$ can be identified with the square matrix $\boldsymbol{H} \in \mathbb{C}^{N\times N}$. This matrix is moreover positive-definite since, from the linear independency of $\{\psi_1,\dots,\psi_N\}$, we have $\forall \boldsymbol{z} \in \mathbb{C}^N\backslash\{\boldsymbol{0}\}$

$$\boldsymbol{z}^H\boldsymbol{H}\boldsymbol{z} = \sum_{i,j=1}^N \overline{z_i}z_j\,\langle\psi_j,\psi_i\rangle = \left\langle\sum_{i=1}^N z_i\psi_i, \sum_{i=1}^N z_i\psi_i\right\rangle = \left\|\sum_{i=1}^N z_i\psi_i\right\|^2 > 0.$$

The matrix $\boldsymbol{H}$ admits hence an inverse and the operator $\boldsymbol{\Psi}^\dagger = (\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*$ is well-defined. We have moreover:

$$\boldsymbol{\Psi}^\dagger\Psi = (\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*\Psi = \boldsymbol{I}_N,$$

and hence $\boldsymbol{\Psi}^\dagger$ is a *right inverse* for Ψ, yielding in particular $\boldsymbol{\Psi}^\dagger\Psi\boldsymbol{\Psi}^\dagger = \boldsymbol{\Psi}^\dagger$, and $\Psi\boldsymbol{\Psi}^\dagger\Psi = \Psi$. Finally, we trivially have $(\boldsymbol{\Psi}^\dagger\Psi)^* = \boldsymbol{I}_N = \boldsymbol{\Psi}^\dagger\Psi$ and

$$(\Psi\boldsymbol{\Psi}^\dagger)^* = (\Psi(\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*)^* = \Psi(\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*,$$

which shows that $\boldsymbol{\Psi}^\dagger$ in (6.1) is indeed the pseudo-inverse of Ψ. ■

Vocabulary 6.1 — *Analysis Operator.* *The adjoint $\boldsymbol{\Psi}^*$ and pseudo-inverse $\boldsymbol{\Psi}^\dagger$ of the synthesis operator Ψ are both said to be* analysis operators *[183, Chapter 5]. The pseudo-inverse is moreover said to be the analysis operator* ideally matched *with Ψ, since $\boldsymbol{\Psi}^\dagger\Psi = \boldsymbol{I}_N$ [183, Chapter 5].*

1.1 Discretisation Scheme for gTikhonov Regularisation

The representer theorem 5.3 tells us that the solution to an FPT problem (5.16) is unique, and can be expressed as a linear combination of the primitives of the sampling functionals w.r.t. the pseudo-differential operator $\mathscr{D}$ used to define the gTikhonov regularisation norm. This parametric representation provides us with a natural search space discretisation scheme, which consists in restricting the search space of gTikhonov-regularised problems to the finite dimensional subspace

$$\text{span}\left\{\psi_k = \sum_{n\in\mathbb{N}}\sum_{m=1}^{N_d(n)} \frac{\langle\varphi_k|Y_n^m\rangle}{|\hat{D}_n|^2}Y_n^m, \quad k=1,\dots,L\right\} \subset \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}),$$

where $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ are the sampling functionals. As proved in the following result, this discretisation scheme is *exact*: the finite dimensional optimisation problem resulting from this search space restriction is *equivalent*[4] to the original FPT problem. Of course, for this scheme to be useful in practice, we need the functions $\{\psi_k,\, k = 1, \dots, L\}$ to be *ordinary* functions, with a pointwise interpretation. For this reason, we restrict our attention to pseudo-differential operators $\mathscr{D}$ with spectral growth order $p > (d-1)/2$. Indeed, for such operators the generalised Sobolev space $\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is an RKHS (see Lemma 5.5) and hence necessarily $\{\psi_k,\, k = 1, \dots, L\} \subset \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ are ordinary functions.

[4] Two optimisation problems are said *equivalent* if their solution sets are in bijection with one another.

Theorem 6.2 — *Canonical Discretisation of FPT Problems.* Consider the notations and assumptions E1 to E6 of Theorem 5.3. Consider additionally the following:

E7 $\mathscr{D}$ has spectral growth order $p > (d-1)/2$;

E8 $\{\psi_1, \dots, \psi_L\} \subset \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ are *interpolation functions*, defined as

$$\psi_k := \mathscr{D}^{-2}\varphi_k = \sum_{n\in\mathbb{N}} \sum_{m=1}^{N_d(n)} \frac{\langle \varphi_k | Y_n^m \rangle}{|\hat{D}_n|^2} Y_n^m, \quad \forall k = 1 \dots, L; \tag{6.3}$$

E9 $\Psi : \mathbb{C}^L \to \operatorname{span}\{\psi_1, \dots, \psi_L\} \subset \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is a *synthesis operator*, defined as

$$\Psi(\boldsymbol{x}) = \sum_{k=1}^{L} x_k \psi_k, \quad \forall \boldsymbol{x} \in \mathbb{C}^L.$$

Then, for any $\boldsymbol{y} \in \mathbb{C}^L$, the FPT problem

$$f^\star = \arg \min_{f \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})} \left\{ F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) \quad + \quad \lambda \|\mathscr{D} f\|_2^2 \right\}$$

and the following *discrete problem*

$$\boldsymbol{x}^\star = \arg \min_{\boldsymbol{x} \in \mathbb{C}^L} \left\{ F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \boldsymbol{x}^H \boldsymbol{G} \boldsymbol{x} \right\}, \tag{6.4}$$

are *equivalent*, in the sense that:

$$f^\star = \Psi(\boldsymbol{x}^\star) \qquad \text{and} \qquad \boldsymbol{x}^\star = \boldsymbol{\Psi}^\dagger(f^\star), \tag{6.5}$$

where $\boldsymbol{\Psi}^\dagger : \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \to \mathbb{C}^N$ is the *pseudo-inverse* (6.1) of Ψ. Moreover, the entries of the square matrix $\boldsymbol{G} \in \mathbb{C}^{L \times L}$ in (6.4) –called the *Gram matrix*, are given by

$$G_{mn} := \langle \varphi_n, \varphi_m \rangle_{\mathscr{H}'_{\mathscr{D}}} = \left\langle \mathscr{D}^{-1}\varphi_n, \mathscr{D}^{-1}\varphi_m \right\rangle_{\mathbb{S}^{d-1}}, \quad m, n = 1, \dots, L.$$

Proof. First, notice that the family of functions $\{\psi_1, \dots, \psi_L\}$ is linearly independent as image by the bijective map $\mathscr{D}^{-2}$ of a family $\{\varphi_1, \dots, \varphi_L\}$ of generalised functions assumed linearly independent. The synthesis operator Ψ hence defines a bijection from $\mathbb{C}^L$ to $\operatorname{span}\{\psi_1, \dots, \psi_L\}$. From this isomorphism, we

get

$$\begin{aligned} f^\star &= \arg\min_{f\in\mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1})} \left\{F(\boldsymbol{y},\boldsymbol{\Phi}(f)) \quad + \quad \lambda\|\mathscr{D}f\|_2^2\right\} \\ &= \Psi\left(\arg\min_{\boldsymbol{x}\in\mathbb{C}^L} \left\{F(\boldsymbol{y},\boldsymbol{\Phi}\Psi(\boldsymbol{x})) \quad + \quad \lambda\|\mathscr{D}\Psi(\boldsymbol{x})\|_2^2\right\}\right) \\ &= \Psi\left(\arg\min_{\boldsymbol{x}\in\mathbb{C}^L} \left\{F(\boldsymbol{y},\boldsymbol{\Phi}\Psi(\boldsymbol{x})) \quad + \quad \lambda\boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}\right\}\right) \end{aligned} \tag{6.6}$$

since

$$\begin{aligned} \|\mathscr{D}\Psi(\boldsymbol{x})\|_2^2 &= \sum_{i,j=1}^L \overline{x_i}x_j \left\langle\mathscr{D}\psi_j,\mathscr{D}\psi_i\right\rangle_{\mathbb{S}^{d-1}} = \sum_{i,j=1}^L \overline{x_i}x_j \left\langle\mathscr{D}\mathscr{D}^{-2}\varphi_j,\mathscr{D}\mathscr{D}^{-2}\varphi_i\right\rangle_{\mathbb{S}^{d-1}} \\ &= \sum_{i,j=1}^L \overline{x_i}x_j \left\langle\mathscr{D}^{-1}\varphi_j,\mathscr{D}^{-1}\varphi_i\right\rangle_{\mathbb{S}^{d-1}} = \sum_{i,j=1}^L \overline{x_i}x_j \left\langle\varphi_j,\varphi_i\right\rangle_{\mathscr{H}'_{\mathscr{D}}} \\ &= \sum_{i,j=1}^L \overline{x_i}x_j G_{ij} = \boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}, \qquad \forall\boldsymbol{x}\in\mathbb{C}^L. \end{aligned}$$

Notice that the linear operator $\boldsymbol{\Phi}\Psi : \mathbb{C}^L \to \mathbb{C}^L$ is finite dimensional, and can hence be represented as a matrix. From the bilinearity of the Schwartz duality product and the definition of the sampling operator $\boldsymbol{\Phi}$, we have indeed

$$\begin{aligned} (\boldsymbol{\Phi}\Psi\boldsymbol{x})_k &= \langle\varphi_k|\Psi(\boldsymbol{x})\rangle = \left\langle\varphi_k\middle|\sum_{j=1}^L x_j\psi_j\right\rangle = \left\langle\varphi_k\middle|\sum_{j=1}^L x_j\mathscr{D}^{-2}\varphi_j\right\rangle \\ &= \sum_{j=1}^L x_j\left\langle\varphi_k\middle|\mathscr{D}^{-2}\varphi_j\right\rangle = \sum_{j=1}^L x_j\left\langle\mathscr{D}^{-1}\varphi_k\middle|\mathscr{D}^{-1}\varphi_j\right\rangle, \quad k=1,\dots,L. \end{aligned}$$

Since $\mathscr{D}^{-1}$ maps $\mathscr{H}'_{\mathscr{D}}(\mathbb{S}^{d-1})$ onto $\mathscr{L}^2(\mathbb{S}^{d-1})$, we have moreover

$$\left\langle\mathscr{D}^{-1}\varphi_k\middle|\mathscr{D}^{-1}\varphi_j\right\rangle = \left\langle\mathscr{D}^{-1}\varphi_j,\mathscr{D}^{-1}\varphi_k\right\rangle_{\mathbb{S}^{d-1}} = \left\langle\varphi_j,\varphi_k\right\rangle_{\mathscr{H}'_{\mathscr{D}}} = G_{kj},$$

and hence

$$(\boldsymbol{\Phi}\Psi\boldsymbol{x})_k = \sum_{j=1}^L x_jG_{kj} = (\boldsymbol{G}\boldsymbol{x})_k, \quad k=1,\dots,L.$$

We can hence identify $\boldsymbol{\Phi}\Psi$ with the matrix $\boldsymbol{G}\in\mathbb{C}^{L\times L}$. Equation (6.6) hence reduces to

$$f^\star = \Psi\left(\arg\min_{\boldsymbol{x}\in\mathbb{C}^L} \left\{F(\boldsymbol{y},\boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda\boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}\right\}\right) = \Psi\left(\boldsymbol{x}^\star\right).$$

From the inverse map (6.1) we furthermore get $\boldsymbol{x}^\star = \boldsymbol{\Psi}^\dagger(f^\star) = (\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*(f^\star)$, which concludes the proof. ■

Remark 6.1 — *Canonical Discretisation Scheme.* *Note that the Gram matrix $\boldsymbol{G}$ is necessarily positive-definite since the sampling functionals are linearly independent (see Proposition 6.1). It therefore admits a Hermitian square root, allowing us to interpret (6.4) as a discrete Tikhonov problem of the form*

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{C}^L} \left\{ F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \|\boldsymbol{G}^{1/2}\boldsymbol{x}\|_2^2 \right\}.$$

The discretisation scheme is hence canonical to the gTikhonov regularisation norm, in the sense that it transforms the latter into a discrete weighted ℓ_2 norm, with weighting scheme described by $\boldsymbol{G}^{1/2}$.

Remark 6.2 — *Computing the Interpolation Functions.* *When the sampling functionals φ_i are in $\mathscr{L}^2(\mathbb{S}^{d-1})$, it is easy to show that the interpolation functions (6.3) can be computed via the formula:*

$$\psi_k = \psi_{\mathscr{D}^2} * \varphi_k, \qquad k = 1, \dots, L,$$

where $\psi_{\mathscr{D}^2}$ denotes the zonal Green kernel of the pseudo-differential operator $\mathscr{D}^2$, which from Proposition 4.5 is spline-admissible for $p > (d-1)/2$. Such convolutions can moreover be computed efficiently when $\psi_{\mathscr{D}^2}$ has small compact support, as explained in Section 3.3 of Chapter 8.

Remark 6.3 — *Practical Implementation.* *Theorem 6.2 provides us with a simple two-step procedure for computing a practical solution to FPT problems:*

1. *Find the unique solution $\boldsymbol{x}^\star \in \mathbb{C}^L$ to (6.4) using one of the algorithms described in Chapter 7.*
2. *Using the bijective synthesis operator $\Psi : \mathbb{C}^L \to \text{span}\{\psi_1, \dots, \psi_L\}$, map $\boldsymbol{x}^\star$ into the solution $f^\star = \Psi(\boldsymbol{x}^\star)$ of the original FPT problem:*

$$f^\star(\boldsymbol{r}) = (\Psi \boldsymbol{x}^\star)(\boldsymbol{r}) = \sum_{k=1}^{L} x_k^\star \psi_k, \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}.$$

1.2 Discretisation Scheme for gTV Regularisation

A discretisation scheme for the FPBP problem (5.23) is more complicated to obtain since the characterisation of its solution set $\mathcal{V}$ provided by the representer theorem 5.4 is *geometric* instead of *parametric*. Indeed, Theorem 5.4 tells us that the solution set $\mathcal{V}$ is the *closed convex-hull* of $\mathscr{D}$-splines[5] with *sparse*[6] innovation sets (see Remark 5.6). This essentially means that any non limit point[7] of $\mathcal{V}$ is itself a $\mathscr{D}$-spline, as *finite* convex combination of $\mathscr{D}$-splines. Our goal is therefore to find a finite family of functions capable of approximating well enough any arbitrary $\mathscr{D}$-spline. To this end, it will help to characterise the functional space in which $\mathscr{D}$-splines naturally live, called the *native space* [125, Chapter 6]. When $\mathscr{D}$ is a *positive-definite operator* with spectral growth order $p > d-1$, the native space is the generalised Sobolev space $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ associated to the Hermitian square-root $\mathscr{D}^{1/2}$ of $\mathscr{D}$:

Proposition 6.3 — *Native Space for $\mathscr{D}$-splines.* *Let $\mathscr{D}$ be a* positive-definite pseudo-differential operator *with* spectral growth order $p > d-1$. *Then the*

[5] Of course, under the assumption that the operator $\mathscr{D}$ used to define the gTV regularisation norm is spline-admissible.

[6] i.e. with cardinality bounded by the number of available measurements.

[7] As explained in Remark 5.6, the limit points are not necessarily splines. We are however not interested in approximating such limit points, since they may have infinitely many df.

Hermitian square-root $\mathscr{D}^{1/2}$ of $\mathscr{D}$ is a pseudo-differential operator with Fourier coefficients $\{\sqrt{\hat{D}_n},\, n \in \mathbb{N}\} \subset \mathbb{R}_+$. Moreover, the generalised Sobolev space $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ is an RKHS which contains all $\mathscr{D}$-splines.

Proof. It is easy to see that the Hermitian square-root of $\mathscr{D}$ is given by

$$\mathscr{D}^{1/2}\varphi = \sum_{n\in\mathbb{N}} \sqrt{\hat{D}_n} \sum_{m=1}^{N_d(n)} \hat{\varphi}_n^m Y_n^m, \qquad \forall \varphi \in \mathscr{S}(\mathbb{S}^{d-1}).$$

The latter is moreover a pseudo-differential operator since the Fourier coefficients $\hat{D}_n$ are all positive (from the assumption of positive-definiteness of $\mathscr{D}$) and hence $\{\sqrt{\hat{D}_n},\, n \in \mathbb{N}\} \subset \mathbb{R}_+$. The rest of the assumptions of Definition 4.1 trivially follow from $\mathscr{D}$ being a pseudo-differential operator. Moreover, from the assumption $p > d-1$ we get that the spectral growth order of $\mathscr{D}^{1/2}$, equal to $p/2$, is strictly larger than $(d-1)/2$. We can hence apply Lemma 5.5 to conclude that the generalised Sobolev space $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ is an RKHS, containing all Dirac measures in its dual. Moreover, using the same arguments as in the proof of Theorem 5.3, it is possible to show that the Riesz map $R_{\mathscr{H}_{\mathscr{D}^{1/2}}} : \mathscr{H}'_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1}) \to \mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ is $\mathscr{D}^{-1}$. Therefore, $\mathscr{D}$-splines[8] are all contained in $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ as images[9] by the Riesz map $\mathscr{D}^{-1}$ of elements of the dual, namely linear combinations of Dirac measures. ■

[8] $\mathscr{D}$-splines exist indeed since $p > d-1$ implies that $\mathscr{D}$ is spline-admissible from Proposition 4.5.

[9] From Definition 4.4, a $\mathscr{D}$-spline is such that $\mathscr{D}\mathfrak{s} = \sum_{i=1}^N \alpha_i \delta_{\boldsymbol{r}_i}$ which, for $\mathscr{D}$ invertible, is equivalent to $\mathfrak{s} = \mathscr{D}^{-1}(\sum_{i=1}^N \alpha_i \delta_{\boldsymbol{r}_i})$.

The next proposition, adapted from [125, Theorem 6.36], shows the error incurred by approximating elements of $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ –and hence in particular arbitrary $\mathscr{D}$-splines of interest here– with $\mathscr{D}$-splines with fixed knot set $\Xi_N \subset \mathbb{S}^{d-1}$ of size N.

Proposition 6.4 — *Approximation Error Analysis.* *Consider a* knot set $\Xi_N = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\} \subset \mathbb{S}^{d-1}$ *with* nodal width

$$\Theta_{\Xi_N} := \max_{\boldsymbol{r}\in\mathbb{S}^{d-1}} \min_{\boldsymbol{s}\in\Xi_N} \|\boldsymbol{r}-\boldsymbol{s}\|_{\mathbb{R}^d}. \tag{6.7}$$

Let further $\mathscr{D}$ denote a positive-definite, spline-admissible pseudo-differential operator *with* spectral growth order $p > \frac{d+1}{2}$ *and $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$ be the space of* spherical $\mathscr{D}$-splines *associated to the knot set Ξ_N. Then, for every function $h \in \mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ we have*

$$\frac{\|h - \mathfrak{s}_N^\perp\|_\infty}{\|h\|_{\mathscr{D}^{1/2}}} \leq 2^{3/2} L_{\mathscr{D}} \sqrt{\Theta_{\Xi_N}}, \tag{6.8}$$

where $\|h\|_{\mathscr{D}^{1/2}} := \sqrt{\langle \mathscr{D}^{1/2}h, \mathscr{D}^{1/2}h\rangle_{\mathbb{S}^{d-1}}}$, $L_{\mathscr{D}} > 0$ is a known[10] positive constant depending only on $\mathscr{D}$ and $\mathfrak{s}_N^\perp \in \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$ is a $\mathscr{D}$-spline verifying

$$\mathfrak{s}_N^\perp = \operatorname*{arg\,min}_{\mathfrak{s}\in\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1},\Xi_N)} \|h - \mathfrak{s}\|_{\mathscr{D}^{1/2}},$$

i.e. $\mathfrak{s}_N^\perp$ is the orthogonal projection *of h onto $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$.*

[10] As shown in the proof, $L_{\mathscr{D}}$ is the *uniform Lispschitz constant* of the zonal Green kernel $\psi_{\mathscr{D}}$.

Proof. Proposition 6.3 tells us that $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ is an RKHS. Therefore, any element $h \in \mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ is an ordinary function, and we have from Proposition 4.4

$$\begin{aligned}\langle h, \psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}\rangle)\rangle_{\mathscr{D}^{1/2}} &= \left\langle \mathscr{D}^{1/2}h, \mathscr{D}^{1/2}\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}\rangle)\right\rangle_{\mathbb{S}^{d-1}} \\ &= \left\langle \mathscr{D}^{1/2}\Psi_{\boldsymbol{r}}^{\mathscr{D}} \middle| \mathscr{D}^{1/2}h \right\rangle \\ &= \left\langle \mathscr{D}\Psi_{\boldsymbol{r}}^{\mathscr{D}} \middle| h \right\rangle \\ &= \langle \delta_{\boldsymbol{r}} | h\rangle \\ &= h(\boldsymbol{r}), \qquad \forall \boldsymbol{r} \in \mathbb{S}^{d-1},\end{aligned}$$

which shows that the zonal Green kernel $\psi_{\mathscr{D}}$ is the *reproducing kernel* [18] of $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$. Additionally, since $\mathscr{D}$ is positive-definite, it is in particular invertible, and we get from (4.17) that

$$\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N) = \operatorname{span}\{\psi_{\mathscr{D}}^n := \psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_n\rangle), \boldsymbol{r}_n \in \Xi_N\}.$$

The positive-definiteness of $\mathscr{D}$ implies moreover (see Remark 5.9 and [125, Theorem 6.27]) that the family of functions $\{\psi_{\mathscr{D}}^n, n = 1, \dots, N\}$ is *linearly independent* and hence forms a basis for $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$. Consequently, the orthogonal projection of h onto $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$ can be written as

$$\mathfrak{s}_N^{\perp} = \sum_{n=1}^{N} \langle h, \psi_{\mathscr{D}}^n\rangle_{\mathscr{D}^{1/2}} \widetilde{\psi}_{\mathscr{D}}^n = \sum_{n=1}^{N} h(\boldsymbol{r}_n)\widetilde{\psi}_{\mathscr{D}}^n,$$

where the second equality follows from the fact that $\psi_{\mathscr{D}}$ reproduces functions in $\mathscr{H}_{\mathscr{D}}^{1/2}(\mathbb{S}^{d-1})$ and $\{\widetilde{\psi}_{\mathscr{D}}^n, n = 1, \dots, N\} \subset \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$ is the *dual basis* [183, Chapter 2] of $\{\psi_{\mathscr{D}}^n, n = 1, \dots, N\}$, verifying the *biorthogonality property*

$$\left\langle \widetilde{\psi}_{\mathscr{D}}^m, \psi_{\mathscr{D}}^n \right\rangle_{\mathscr{D}^{1/2}} = \delta_{mn}, \quad \forall m, n = 1, \dots, N.$$

We have hence

$$\left\langle \mathfrak{s}_N^{\perp}, \psi_{\mathscr{D}}^n \right\rangle_{\mathscr{D}^{1/2}} = h(\boldsymbol{r}_n) = \langle h, \psi_{\mathscr{D}}^n\rangle_{\mathscr{D}^{1/2}}, \quad n = 1, \dots, N. \tag{6.9}$$

Moreover, [125, Lemma 6.34] tells us that, for spline-admissible pseudo-differential operators with growth order $p > \frac{d+1}{2}$, the zonal Green kernel $\psi_{\mathscr{D}}$ is *uniformly Lipschitz continuous*, i.e. there exists $L_{\mathscr{D}} > 0$ which only depends on the sequence $\{\hat{D}_n\}_{n\in\mathbb{N}}$ such that for any $\boldsymbol{\rho} \in \mathbb{S}^{d-1}$

$$|\psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{\rho}\rangle) - \psi_{\mathscr{D}}(\langle \boldsymbol{s}, \boldsymbol{\rho}\rangle)| \leq L_{\mathscr{D}}^2 \|\boldsymbol{r} - \boldsymbol{s}\|_2, \quad \forall \boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}. \tag{6.10}$$

With these two observations, we are now ready to prove the result. First, we get from (6.9) as well as the Cauchy-Schwarz and triangle inequalities

$$\left|h(\boldsymbol{r}) - \mathfrak{s}_N^{\perp}(\boldsymbol{r})\right| = |h(\boldsymbol{r}) - h(\boldsymbol{r}_n) + \mathfrak{s}_N^{\perp}(\boldsymbol{r}_n) - \mathfrak{s}_N^{\perp}(\boldsymbol{r})|$$

$$
\begin{aligned}
&= |\left\langle \psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n}, h - h_N^{\perp} \right\rangle_{\mathscr{D}^{1/2}}| \\
&\leq \|\psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n}\|_{\mathscr{D}^{1/2}} (\|h_N^{\perp}\|_{\mathscr{D}^{1/2}} + \|h\|_{\mathscr{D}^{1/2}}) \\
&\leq 2\|\psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n}\|_{\mathscr{D}^{1/2}} \|h\|_{\mathscr{D}^{1/2}}.
\end{aligned}
$$

Second, we obtain from the reproducing property, equation (6.10) and the definition (6.7) of the nodal width Θ_{Ξ_N}:

$$
\begin{aligned}
\|\psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n}\|_{\mathscr{D}^{1/2}}^2 &= \left\langle \psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n}, \psi_{\mathscr{D}}^{\boldsymbol{r}} - \psi_{\mathscr{D}}^{\boldsymbol{r}_n} \right\rangle_{\mathscr{D}^{1/2}} \\
&= \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r} \rangle) + \psi_{\mathscr{D}}(\langle \boldsymbol{r}_n, \boldsymbol{r}_n \rangle) - \psi_{\mathscr{D}}(\langle \boldsymbol{r}_n, \boldsymbol{r} \rangle) - \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle) \\
&\leq 2L_{\mathscr{D}}^2 \|\boldsymbol{r} - \boldsymbol{r}_n\|_{\mathbb{R}^d} \\
&\leq 2L_{\mathscr{D}}^2 \Theta_{\Xi_N}.
\end{aligned}
$$

In conclusion, this yields:

$$
\frac{\sup_{\boldsymbol{r} \in \mathbb{S}^{d-1}} |h(\boldsymbol{r}) - \mathfrak{s}_N^{\perp}(\boldsymbol{r})|}{\|h\|_{\mathscr{D}^{1/2}}} \leq 2^{3/2} L_{\mathscr{D}} \sqrt{\Theta_{\Xi_N}},
$$

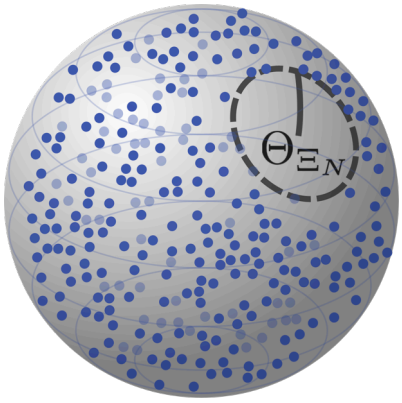
which achieves the proof. ■

Notice that the approximation error in Proposition 6.4 is bounded by the nodal width (6.7) of the knot set, which can be interpreted geometrically as the largest distance from an arbitrary point on the sphere to the knot set Ξ_N (see Fig. 6.1a). $\mathscr{D}$-splines with knot sets minimising this quantity for a fixed number of knots N will hence yield the smallest approximation error. From the geometric interpretation of the nodal width, it is easy to see that knot sets with minimal nodal width distribute their knots *uniformly* over the hypersphere. Unfortunately, distributing points uniformly over $\mathbb{S}^{d-1}$ is a notoriously hard problem for $d > 2$ [78], making uniform knot sets inpractical. For $d = 3$ however, it is possible to obtain *quasi-uniform* knot sets with *quasi-optimal* nodal widths [78]. An example of quasi-uniform knot set is the *Fibonacci lattice* [66, 78] described in the subsequent example. In [78], the authors provide a comprehensive list of quasi-uniform knot sets easy to generate in practice. For each knot set, the asymptotic behaviour of the nodal width is assessed, either numerically or theoretically.

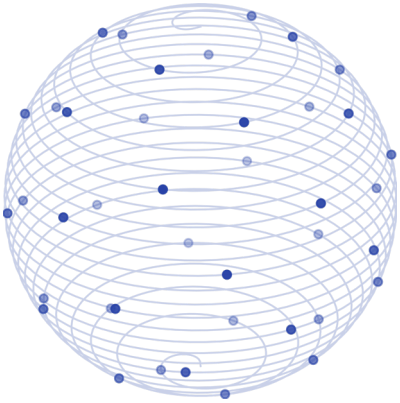
Example 6.1 — *Fibonacci Lattice.* In nature, many plant leaves are arranged according to *phyllotactic spiral patterns*, which are well modelled by the *Fibonacci lattice*. Points in the Fibonacci lattice are arranged uniformly along a spiral pattern on the sphere linking the two poles (see Fig. 6.1b). The lattice can very easily be generated from the following formula:

$$
\begin{cases}
\boldsymbol{r}_n = [\cos(\varphi_n)\sin(\theta_n), \sin(\varphi_n)\sin(\theta_n), \cos(\theta_n)]\,, \\
\text{where} \quad \varphi_n = 2\pi n \left(1 - \frac{2}{1+\sqrt{5}}\right) \quad \& \quad \theta_n = \arccos\left(1 - \frac{2n}{N}\right),
\end{cases}
\tag{6.11}
$$

with $n = 1, \ldots, N$. It can be shown [78] that if the knot set Ξ_N is constructed



(a) The nodal width of a knot set represents the largest distance from an arbitrary point on the sphere to the knot set.



(b) Fibonacci lattice with $N = 41$ points.

Figure 6.1: Visual representation of the nodal width **(a)** and the quasi-uniform Fibonacci lattice **(b)**.

according to the Fibonacci lattice (6.11), then the nodal width is *quasi-optimal* and approximately given by $\Theta_{\Xi_N} \simeq 2.728/\sqrt{N}$. ■

Notice that the nodal width of the Fibonacci lattice tends to zero as the number of knots N grows to infinity. This is a general behaviour of quasi-uniform knot sets [78]. Consequently, the uniform approximation error (6.8) in Proposition 6.4 tends to zero as the number of knots tends to infinity. In other words, any element of $\mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$ can be approximated *arbitrarily well* by $\mathscr{D}$-splines with *quasi-uniform* knot sets –called *quasi-uniform spherical splines*, provided a sufficient number of knots. In light of this discussion, we therefore propose to discretise FPBP problems by restricting their search spaces to subspaces spanned by quasi-uniform $\mathscr{D}$-splines. The following theorem shows that the solutions to FPBP problems restricted this way can be obtained by solving a *discrete penalised basis pursuit (PBP) problem*.

Theorem 6.5 — *Canonical Discretisation of FPBP Problems.* Consider the notations and assumptions F1 to F6 of Theorem 5.4. Consider additionally the following:

F7 $\mathscr{D}$ is *spline-admissible* and *positive-definite*;

F8 $\{\psi_1, \dots, \psi_N\} \subset \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ are *zonal functions* of the form

$$\psi_n := \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle), \quad \forall n = 1, \dots, N, \tag{6.12}$$

where $\psi_{\mathscr{D}}$ is the *zonal Green kernel* of $\mathscr{D}$ and $\Xi_N = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\} \subset \mathbb{S}^{d-1}$ for some $N \in \mathbb{N}$;

F9 $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N) = \operatorname{span}\{\psi_1, \dots, \psi_N\} \subset \mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is the space of $\mathscr{D}$-splines with knot set Ξ_N;

F10 $\Psi : \mathbb{C}^N \to \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)$ is a *synthesis operator*, defined as

$$\Psi(\boldsymbol{x}) = \sum_{n=1}^{N} x_n \psi_n, \quad \forall \boldsymbol{x} \in \mathbb{C}^N.$$

Then, for each $\boldsymbol{y} \in \mathbb{C}^L$, the *restricted FPBP problem*

$$\mathcal{V} = \arg \min_{f \in \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)} \{F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) \quad + \quad \lambda \|\mathscr{D} f\|_{TV}\} \tag{6.13}$$

and the following *discrete PBP problem*

$$\mathfrak{U} = \arg \min_{\boldsymbol{x} \in \mathbb{C}^N} \{F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \|\boldsymbol{x}\|_1\} \tag{6.14}$$

are *equivalent*, in the sense that their solution sets are in *bijection* with one another:

$$\mathcal{V} = \Psi(\mathfrak{U}) \qquad \text{and} \qquad \mathfrak{U} = \boldsymbol{\Psi}^{\dagger}(\mathcal{V}), \tag{6.15}$$

where $\boldsymbol{\Psi}^{\dagger} : \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N) \to \mathbb{C}^N$ is the *pseudo-inverse* (6.1) of Ψ. Moreover, the matrix $\boldsymbol{G} := \boldsymbol{\Phi}\Psi \in \mathbb{C}^{L \times N}$ in (6.14) is given by

$$G_{ln} := \langle \psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_n \rangle) | \varphi_l \rangle, \quad l = 1, \dots, L, \quad n = 1, \dots, N,$$

which simplifies to $G_{ln} = (\psi_{\mathscr{D}} * \varphi_l)(\boldsymbol{r}_n)$ when the sampling functionals $\{\varphi_l, \, l = 1, \dots, L\}$ are in $\mathscr{L}^2(\mathbb{S}^{d-1})$.

Proof. The spline-admissible pseudo-differential operator $\mathscr{D}$ being positive-definite, its Green kernel $\psi_{\mathscr{D}}$ is strictly positive-definite (see [125, Definition 6.25 and Theorem 6.27]) and hence according to [125, Lemma 6.26], the family of functions $\{\psi_n = \psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_n \rangle), \; n = 1, \dots, N\}$ is *linearly independent* for every set $\Xi_N = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_n\} \subset \mathbb{S}^{d-1}$ of N distinct points. The synthesis operator Ψ defines hence a bijection between $\mathbb{C}^N$ and $\mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N) = \text{span}\{\psi_n, \, n = 1, \dots, N\}$. From this isomorphism, we get notably

$$\begin{aligned} \mathcal{V} &= \arg \min_{f \in \mathfrak{S}_{\mathscr{D}}(\mathbb{S}^{d-1}, \Xi_N)} \{F(\boldsymbol{y}, \boldsymbol{\Phi}(f)) \quad + \quad \lambda \|\mathscr{D} f\|_{TV}\} \\ &= \Psi \left(\arg \min_{\boldsymbol{x} \in \mathbb{C}^N} \{F(\boldsymbol{y}, \boldsymbol{\Phi}\Psi(\boldsymbol{x})) \quad + \quad \lambda \|\mathscr{D}\Psi(\boldsymbol{x})\|_{TV}\} \right) \\ &= \Psi \left(\arg \min_{\boldsymbol{x} \in \mathbb{C}^N} \{F(\boldsymbol{y}, \boldsymbol{\Phi}\Psi(\boldsymbol{x})) \quad + \quad \lambda \|\boldsymbol{x}\|_1\} \right), \end{aligned} \tag{6.16}$$

since we have (see Section 2.2)

$$\|\mathscr{D}\Psi(\boldsymbol{x})\|_{TV} = \left\| \sum_{n=1}^{N} x_n \mathscr{D}\psi_{\mathscr{D}}(\langle \cdot, \boldsymbol{r}_n \rangle) \right\|_{TV} = \left\| \sum_{n=1}^{N} x_n \delta_{\boldsymbol{r}_n} \right\|_{TV} = \|\boldsymbol{x}\|_1.$$

Notice that the linear operator $\boldsymbol{\Phi}\Psi : \mathbb{C}^N \to \mathbb{C}^L$ is finite-dimensional, and can hence be represented as a matrix. From the bilinearity of the Schwartz duality

product, we have indeed

$$(\boldsymbol{\Phi}\Psi\boldsymbol{x})_l = \langle \Psi\boldsymbol{x}|\varphi_l\rangle = \left\langle \sum_{n=1}^{N} x_n \psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle) \middle| \varphi_l \right\rangle$$
$$= \sum_{n=1}^{N} x_n \langle \psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle)|\varphi_l\rangle = \sum_{n=1}^{N} x_n G_{ln}, \quad \forall l = 1,\dots,L.$$

We can hence identify $\boldsymbol{\Phi}\Psi$ with a matrix $\boldsymbol{G} \in \mathbb{C}^{L\times N}$, with entries given by

$$G_{ln} := \langle \psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle)|\varphi_l\rangle, \quad l = 1,\dots,L, \quad n = 1,\dots,N.$$

Since $\mathscr{D}$ is spline-admissible, the traces of the zonal Green kernel are ordinary functions and hence in particular square-integrable. When the sampling functionals $\{\varphi_1,\dots,\varphi_L\}$ are in $\mathscr{L}^2(\mathbb{S}^{d-1})$ we can hence obtain a simpler expression for the entries of $\boldsymbol{G}$:

$$G_{ln} = \langle \psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle)|\varphi_l\rangle = \langle \varphi_l, \psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle)\rangle_{\mathbb{S}^{d-1}}$$
$$= \int_{\mathbb{S}^{d-1}} \psi_{\mathscr{D}}(\langle\boldsymbol{r},\boldsymbol{r}_n\rangle)\varphi_l(\boldsymbol{r})d\boldsymbol{r}$$
$$= (\psi_{\mathscr{D}} * \varphi_l)(\boldsymbol{r}_n), \quad l = 1,\dots,L, \quad n = 1,\dots,N.$$

Equation (6.16) finally reduces to

$$\mathcal{V} = \Psi\left(\arg\min_{\boldsymbol{x}\in\mathbb{C}^N} \{F(\boldsymbol{y},\boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda\|\boldsymbol{x}\|_1\}\right) = \Psi\left(\mathfrak{U}\right),$$

as claimed. From the inverse map (6.1) we furthermore get $\mathfrak{U} = \boldsymbol{\Psi}^\dagger(\mathcal{V}) = (\boldsymbol{\Psi}^*\Psi)^{-1}\boldsymbol{\Psi}^*(\mathcal{V})$, which concludes the proof. ■

Remark 6.4 — *Canonical Discretisation Scheme.* *Notice that the discretisation scheme chosen in Theorem 6.5 is* canonical *w.r.t. the gTV norm induced by the pseudo-differential operator $\mathscr{D}$. Indeed, it conveniently transforms the gTV norm $\|\mathscr{D}\cdot\|_{TV}$ into a discrete ℓ_1 norm in (6.14). As detailed in the proof, this is because the basis functions $\{\psi_{\mathscr{D}}(\langle\cdot,\boldsymbol{r}_n\rangle),\ n = 1,\dots,N\}$ used in the discretisation are Green functions of the operator $\mathscr{D}$. Had the basis functions been chosen differently, such simplifications would not have been possible, hence making the discrete optimisation problem (6.14) considerably more difficult to solve in practice.*

Remark 6.5 — *Choice of N.* *The bound in (6.8) can be used in practice to set N. Indeed, one can choose N such that the relative approximation error falls below an acceptable accuracy threshold for any $h \in \mathscr{H}_{\mathscr{D}^{1/2}}(\mathbb{S}^{d-1})$, hence allowing us to approximate the solutions of the FPBP problem with controlled error.*

Remark 6.6 — *Practical Implementation.* *Again, Theorem 6.5 provides us with a simple two-step procedure for computing a practical solution to the restricted FPBP problem (6.13):*

1. *Minimise (6.14) using one of the algorithms described in Chapter 7 and obtain a solution $\boldsymbol{u} \in \mathfrak{U}$.*

2. *Using the synthesis operator* $\Psi : \mathbb{C}^N \to \mathfrak{S}(\mathbb{S}^{d-1}, \Xi_N)$ *and the fact that* $\mathcal{V} = \Psi(\mathfrak{U})$*, map* $\boldsymbol{u}$ *into a solution* $f = \Psi(\boldsymbol{u}) \in \mathcal{V}$ *of the restricted FPBP problem (6.13):*

$$f(\boldsymbol{r}) = (\Psi \boldsymbol{u})(\boldsymbol{r}) = \sum_{n=1}^{N} u_n \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}.$$

The latter step can in this case be interpreted as an interpolation *on* $\mathbb{S}^{d-1}$ *of samples* $\{u_n,\ n = 1, \dots, N\} \subset \mathbb{C}$ *with sampling locations* $\{\boldsymbol{r}_n,\ n = 1, \dots, N\} \subset \mathbb{S}^{d-1}$ *and interpolation kernel* $\psi_{\mathscr{D}}$*. Since the interpolating functions* $\psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle)$ *are* zonal*, such an interpolation can be carried out* very efficiently *in practice (and even more so when* $\psi_{\mathscr{D}}$ *has compact support, as explained in Section 3.1).*

Remark 6.7 — ***Form of the Solutions.*** *Applying [178, Theorem 6] to the discrete PBP problem (6.14) shows that* $\mathfrak{U}$ *is convex and compact with* L*-sparse extreme points. From the bijection (6.15), it implies in turn that* $\mathcal{V} = \Psi(\mathfrak{U})$ *is the closed convex-hull of extreme points taking the form of* sparse $\mathscr{D}$*-admissible spherical splines with at most* L *non-zero amplitudes:*

$$\forall f \in \delta\mathcal{V}, \quad f = \Psi \boldsymbol{u} \quad \textit{with} \quad \|\boldsymbol{u}\|_0 \leq L,$$

where $\|\cdot\|_0$ *denotes the "*ℓ_0 *norm", counting the number of non-zero elements in a vector. Solutions of the restricted FPBP problem (6.13) behave hence very similarly as the ones of the unrestricted FPBP problem (5.23) investigated in Theorem 5.4.*

2 Domain Discretisation

In this section, we discuss an alternative discretisation strategy, called *domain discretisation*. The latter, much appreciated by practitioners due to its simplicity, operates on functional inverse problems by approximating their continuous domain by some *discrete manifold* [30], hence implicitly converting them into finite dimensional inverse problems. In our case, this amounts to approximating the hypersphere $\mathbb{S}^{d-1}$ by some finite *spherical point set* $\Theta = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\} \subset \mathbb{S}^{d-1}$, often chosen *equidistributed* in practice (see Example 6.2 for more details). In this context, the aim of spherical approximation is then to recover "functions" $f : \Theta \to \mathbb{C}$ with discrete domain Θ and codomain $\mathbb{C}$. Of course, such objects do not naturally fall into the scope of the approximation framework introduced in Chapter 5, which was primarily designed for spherical fields, i.e. functions, measures or generalised functions defined over the *continuous* spherical domain. In the subsequent sections, we therefore adapt this approximation framework to the discrete manifold setting, discussing notably the discrete analogs of the $\mathscr{L}^2$ and TV norms as well as the sampling and pseudo-differential operators. As we shall see, although behaving similarly as their continuous counterparts, discrete pseudo-differential operators lack a canonical definition in the discrete manifold setting, complicating slightly the comparison with the continuous approximation framework.

(a) Cubic tessellation and its associated point set.

(b) HEALPix tessellation and its associated point set.

(c) Fibonacci Voronoi tessellation and its associated point set.

Figure 6.2: Examples of equidistributed point sets (marked by black dots) obtained from the centres of spherical tessellation cells. In this experiment, we chose an approximate resolution $N = 200$ for each point set. The cubic tessellation Fig. 6.2a is obtained by projecting the pixelated faces of a cube onto the sphere.The HEALPix tessellation Fig. 6.2b is constructed by hierarchical subdivision of the Voronoi cells of the *dodecahedron* vertices [67]. The Fibonacci Voronoi tessellation Fig. 6.2c is obtained by constructing the spherical Voronoi tessellation of the Fibonacci lattice (6.11).

Example 6.2 — *Equidistributed Point Sets.* Traditionally, the discrete set $\Theta = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\} \subset \mathbb{S}^{d-1}$ is chosen as an *equidistributed spherical point set* [78], with the property that the sequence of normalised *atomic measures*

$$\left\{ \nu_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{\boldsymbol{r}_n},\ N \in \mathbb{N} \right\} \subset \mathcal{M}(\mathbb{S}^{d-1})$$

converges in the weak* sense towards the Lebesgue measure μ on $\mathbb{S}^{d-1}$ when the number of points in the point set Θ grows to infinity. For $d = 3$, there exists many equidistributed point sets [78], among which the *Fibonacci lattice* discussed in Example 6.1 and the *Hierarchical Equal Area isoLatitude Pixelization (HEALPix) lattice* [67], developed by NASA for analysing the cosmic microwave radiation background (CMB). In most cases, equidistributed point sets are obtained from the barycentres of polygonal cells in *spherical tessellations*, which tile the sphere with near equal-area and near identical polygonal tiles (see [78] and Fig. 6.2). ■

Remark 6.8 — *Domain Discretisation and Spherical Pixelisation.* *The duality exhibited by Fig. 6.2 between equidistributed spherical point sets and spherical tessellations has mislead many scholars into considering domain discretisation as a form of* spherical pixelisation. *For example, signals defined over equidistributed spherical point sets are often represented visually by colouring the faces of the spherical tessellation associated to the specific point set configuration. Such a representation implicitly assumes that a function over Θ can be* interpolated *by the indicator functions of the tessellation cells, hence allowing us to interpret domain discretisation as some sort of search space discretisation. Unfortunately, the indicator functions of the tessellation cells are* discontinuous *step functions,*

and hence often not regular enough[11] *to be included in the search spaces of the FPT and FPBP problems. As such, they* cannot *be considered as basis functions of search space discretisation schemes for FPT and FPBP problems.*

[11] For example, taking the second derivative (in the sense of distributions) of the periodic box function yields a generalised function which is neither in $\mathscr{L}^2(\mathbb{S}^1)$ nor in $\mathcal{M}(\mathbb{S}^1)$.

2.1 Discrete $\mathscr{L}^2$ and TV Norms

The first ingredients we need for our discrete approximation framework are discrete analogues of the $\mathscr{L}^2$ and TV norms, which are central to the definitions of the gTikhonov and gTV regularisation norms. To this end, it helps to notice that, from the trivial bijection between Θ and $\{1,\dots,N\}$, one can identify the functional space $\mathbb{C}^\Theta = \{f : \Theta \to \mathbb{C}\}$ with $\mathbb{C}^N$. Indeed, any function $f \in \mathbb{C}^\Theta$ can be written uniquely as a *vector* of $\mathbb{C}^N$ given by

$$\boldsymbol{f} = [f(\boldsymbol{r}_1), \cdots, f(\boldsymbol{r}_N)]. \tag{6.17}$$

In the next result, we leverage this bijection to show that the $\mathscr{L}^2$ and TV norms of a function $f \in \mathbb{C}^\Theta$ are given by the ℓ_2 and ℓ_1 norm of its vector representation $\boldsymbol{f} \in \mathbb{C}^N$ respectively. We also characterise the dual of $\mathbb{C}^\Theta$ as well as the important functional spaces $\mathscr{L}^2(\Theta)$, $\mathscr{C}(\Theta)$ and $\mathcal{M}(\Theta)$.

Proposition 6.6 — *Functional Analysis over Discrete Domains.* *Consider a point set $\Theta = \{\boldsymbol{r}_1,\dots,\boldsymbol{r}_N\} \subset \mathbb{S}^{d-1}$ equipped with the* discrete topology *and the vector space $\mathbb{C}^\Theta = \{f : \Theta \to \mathbb{C}\}$ of functions with* domain Θ *and* co-domain $\mathbb{C}$. *Then the following holds:*

1. $(\mathscr{L}^2(\Theta), \|\cdot\|_2)$ *is isometrically isomorphic to* $(\mathbb{C}^N, \|\cdot\|_2)$.
2. $(\mathscr{C}(\Theta), \|\cdot\|_\infty)$ *is isometrically isomorphic to* $(\mathbb{C}^N, \|\cdot\|_\infty)$.
3. *The* algebraic dual $(\mathbb{C}^\Theta)^*$ *of $\mathbb{C}^\Theta$ is in bijection with $\mathbb{C}^N$. Moreover, the* Schwartz duality product $\langle\cdot|\cdot\rangle : (\mathbb{C}^\Theta)^* \times \mathbb{C}^\Theta \to \mathbb{C}$ *can be identified with the canonical inner product $\langle\cdot,\cdot\rangle_{\mathbb{C}^N}$ as follows*

$$\langle g|f\rangle = \langle \boldsymbol{f}, \overline{\boldsymbol{g}}\rangle_{\mathbb{C}^N}, \qquad \forall f, g \in \mathbb{C}^\Theta \times (\mathbb{C}^\Theta)^*,$$

   and where $\boldsymbol{f}, \boldsymbol{g} \in \mathbb{C}^N$ are the vector representations of f and g respectively.
4. $(\mathcal{M}(\Theta), \|\cdot\|_{TV})$ *is is isometrically isomorphic to* $(\mathbb{C}^N, \|\cdot\|_1)$.

Proof. Most of Items 1 to 4 are trivialities. For the sake of completeness, we provide here succinct derivations of the latter:

1. From bijection (6.17), we get

$$\|f\|_2^2 = \sum_{\boldsymbol{\rho}\in\Theta} |f(\boldsymbol{\rho})|^2 = \sum_{n=1}^N |f(\boldsymbol{r}_n)|^2 = \|\boldsymbol{f}\|_2^2, \quad \forall f \in \mathbb{C}^\Theta.$$

   This yields,

$$\mathscr{L}^2(\Theta) = \{f : \Theta \to \mathbb{C} : \|f\|_2 < +\infty\} \cong \{\boldsymbol{f} \in \mathbb{C}^N : \|\boldsymbol{f}\|_2 < +\infty\} = \mathbb{C}^N.$$

2. Since Θ is equipped with the discrete topology, any function in $\mathbb{C}^\Theta$ is continuous. We have hence $\mathscr{C}(\Theta) = \mathbb{C}^\Theta \cong \mathbb{C}^N$ from bijection (6.17).

Moreover, still from (6.17) we get

$$\|f\|_\infty = \sup_{\boldsymbol{\rho}\in\Theta} |f(\boldsymbol{\rho})| = \max_{n=1,\dots,N} |f(\boldsymbol{r}_n)| = \|\boldsymbol{f}\|_\infty, \quad \forall f \in \mathbb{C}^\Theta. \tag{6.18}$$

3. Using bijection (6.17), we can write any element f of $\mathbb{C}^\Theta$ uniquely as

$$f = \sum_{n=1}^{N} f(\boldsymbol{r}_n)\delta_n,$$

where δ_n is the function[12] in $\mathbb{C}^\Theta$ associated to the nth vector $\boldsymbol{\delta}_n \in \mathbb{C}^N$ of the canonical basis of $\mathbb{C}^N$, for $n = 1, \dots, N$. Consequently, we have, for any linear functional $g \in (\mathbb{C}^\Theta)^*$:

$$\langle g|f\rangle = \left\langle g \middle| \sum n = 1 f(\boldsymbol{r}_n)\delta_n \right\rangle = \sum_{n=1}^{N} f(\boldsymbol{r}_n)\langle g|\delta_n\rangle = \sum_{n=1}^{N} f(\boldsymbol{r}_n)\overline{g_n} = \langle \boldsymbol{f}, \overline{\boldsymbol{g}}\rangle_{\mathbb{C}^N}. \tag{6.19}$$

Hence any linear functional $g \in (\mathbb{C}^\Theta)^*$ can be written uniquely as a vector of $\mathbb{C}^N$ as:

$$\boldsymbol{g} = [\langle g|\delta_1\rangle, \dots, \langle g|\delta_N\rangle].$$

4. From (6.17), (6.18) and (6.19) we get that the total variation norm is given by:

$$\|g\|_{TV} = \sup_{f\in\mathbb{C}^\Theta, \|f\|_\infty=1} |\langle g|f\rangle| = \sup_{\boldsymbol{f}\in\mathbb{C}^N, \|\boldsymbol{f}\|_\infty=1} |\langle \boldsymbol{f}, \overline{\boldsymbol{g}}\rangle| = \|\boldsymbol{g}\|_1, \quad \forall g \in (\mathbb{C}^\Theta)^*.$$

Therefore we have

$$\mathcal{M}(\Theta) = \{g \in (\mathbb{C}^\Theta)^* : \|g\|_{TV} < +\infty\} \cong \{\boldsymbol{g} \in \mathbb{C}^N : \|\boldsymbol{g}\|_1 < +\infty\} = \mathbb{C}^N.$$

■

[12] This function is such that $\delta_n(\boldsymbol{r}_n) = 1$ and $\delta_n(\boldsymbol{r}_m) = 0$ for all $m \neq n$, $m = 1, \dots, N$.

2.2 Discrete Sampling Operators

In agreement with the generic definition (5.2) of a sampling operator, we define a *discrete sampling operator* $\boldsymbol{\Phi} : \mathbb{C}^\Theta \to \mathbb{C}^L$ over $\mathbb{C}^\Theta$ as

$$\boldsymbol{\Phi}(f) = [\langle g_1|f\rangle, \dots, \langle g_L|f\rangle] \in \mathbb{C}^L, \quad \forall f \in \mathbb{C}^\Theta,$$

where $\{g_1, \dots, g_L\}$ are linear functionals in the dual of $\mathbb{C}^\Theta$. Due to Item 3 of Proposition 6.6, discrete sampling operators can be identified with matrices in $\mathbb{C}^{L\times N}$. Indeed, we have, for all $f \in \mathbb{C}^\Theta$:

$$\boldsymbol{\Phi}(f) = \begin{bmatrix} \langle g_1|f\rangle \\ \vdots \\ \langle g_L|f\rangle \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{f}, \overline{\boldsymbol{g}_1}\rangle_{\mathbb{C}^N} \\ \vdots \\ \langle \boldsymbol{f}, \overline{\boldsymbol{g}_L}\rangle_{\mathbb{C}^N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1^H \\ \vdots \\ \boldsymbol{g}_L^H \end{bmatrix} \boldsymbol{f} = \boldsymbol{G}\boldsymbol{f}, \tag{6.20}$$

and hence $\boldsymbol{\Phi}$ can be identified with the matrix $\boldsymbol{G} \in \mathbb{C}^{L\times N}$, whose lth row is given by $\boldsymbol{g}_l^H \in \mathbb{C}^N$. We call $\boldsymbol{G}$ the *sensing matrix*.

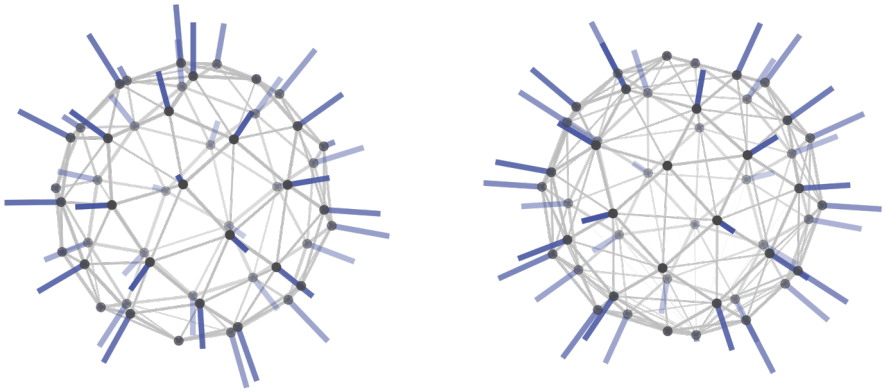
Remark 6.9 — *The Sensing Matrix in Practice.* *In practice, the matrix $\boldsymbol{G}$ is chosen so as to approximate well the acquisition system used to collect the data. Such an approximation may in some cases be particularly cumbersome to establish. Indeed, the physical world being analog, the operations performed by the acquisition system are inherently continuous, and may hence be very difficult to represent as discrete operations. This is for example the case for non-uniform spatial sampling scenarios, where the sampling directions may not coincide with the directions in Θ. In such cases, practitioners often have recourse to ad-hoc* gridding steps[13]*, so as to map the off-grid samples onto the point set Θ (see for example Section 1.4.2 of Chapter 9). Of course, the error incurred by such gridding steps are most often difficult, if not impossible, to assess precisely.*

[13] Such as nearest-neighbour interpolation.

2.3 Discrete Pseudo-Differential Operators

In Chapter 4, we defined pseudo-differential operators on the sphere as Fourier multipliers. We adapt here this construction to discrete domains and introduce *discrete pseudo-differential operators* as linear operators diagonalised by the Fourier basis and with specific constraints on their spectrum. To this end, we need to first define a Fourier basis over $\mathbb{C}^\Theta$. From bijection (6.17), it may seem tempting to choose to work with the canonical Fourier basis $e^{j2\pi k\boldsymbol{n}/N}$ on $\mathbb{C}^N$. Unfortunately, it turns out that the latter is particularly ill-suited for harmonic analysis purposes on $\mathbb{C}^\Theta$. Indeed, there is a fundamental *mismatch*[14] between *locality* in the discrete manifold Θ and *connectivity* in the vector representation $\boldsymbol{f} \in \mathbb{C}^N$ of a function $f \in \mathbb{C}^\Theta$: points which are close to one another in the discrete manifold can end up far apart from one another in the vector representation. Consequently, applying naively the discrete Fourier transform (DFT) to the vector representation $\boldsymbol{f}$ of a function f in $\mathbb{C}^\Theta$ would yield a fundamentally *flawed* spectrum, where some of the frequency content would be due to the unaccounted geometry of the underlying domain and not to the inherent fluctuations of the analysed signal. To properly account for the correlations in $\boldsymbol{f}$ arising from the underlying domain geometry, one possibility [47, 139] is to define an explicit *connectivity graph* $\mathcal{G} = (\Theta, \mathcal{E}, \boldsymbol{W})$, where $\mathcal{E} \subset \Theta^2$ is an *edge set* defining neighbouring vertices in Θ and $\boldsymbol{W} \in \mathbb{R}^{N\times N}$ is a *weighting matrix*, defining the *similarity* between two connected vertices. Given an arbitrary spherical point set Θ, the edge set can for example be defined as the *Delaunay triangulation* of Θ (which can be computed in practice efficiently via the *Quickhull* algorithm [11]). Such a construction can be thought as linking the points in Θ whose corresponding tessellation cells are adjacent (see Fig. 6.3). Consequently, we call the graph obtained this way a *tessellation graph*. The edge weights are moreover commonly defined as a function of the Euclidean distance separating two vertices in the lattice Θ. In [139], the authors recommend

[14] This mismatch is explained by the fact that manifolds are by definition not *homeomorphic* to Euclidean domains.



(a) Signal on the Fibonacci tessellation graph. **(b)** Signal on the HEALPix tessellation graph.

Figure 6.3: Examples of signals on spherical tessellation graphs ($N = 48$).

the following weighting scheme:

$$W_{nm} := \begin{cases} \exp\left(-\frac{\|\boldsymbol{r}_n - \boldsymbol{r}_m\|_2^2}{\rho^2}\right) & \text{if } (\boldsymbol{r}_n, \boldsymbol{r}_m) \in \mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\rho > 0$ is given by $\rho = \frac{1}{|\mathcal{E}|} \sum_{(\boldsymbol{r}_n, \boldsymbol{r}_m) \in \mathcal{E}} \|\boldsymbol{r}_n - \boldsymbol{r}_m\|_2$. With this additional structure, a function $f \in \mathbb{C}^\Theta$ can be seen as a *signal on a graph* (see Fig. 6.3), which can be processed by means of *graph signal processing* tools [159]. Similarly as in Chapter 3, the Fourier basis on a graph is typically defined as the eigenvectors of the *Laplacian* of the graph $\mathscr{G}$, which can be thought of as a discrete analog of the negative Laplace-Beltrami operator $-\Delta_{\mathbb{S}^{d-1}}$. As explained in [159], there exist many possible definitions of the graph Laplacian (see Remark 6.10). For example, one can consider the *normalised Laplacian* [139] given by:

$$\boldsymbol{L} := \boldsymbol{I} - \boldsymbol{\Lambda}^{-1/2} \boldsymbol{W} \boldsymbol{\Lambda}^{-1/2}, \tag{6.21}$$

where $\boldsymbol{I} \in \mathbb{R}^{N \times N}$ denotes the *identity matrix* and $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}$ is a diagonal matrix defined as:

$$\Lambda_{ii} = \sum_{n=1}^{N} W_{in}, \qquad i = 1, \ldots, N.$$

The Laplacian operator (6.21) has many useful properties in common with the negative Laplace-Beltrami operator [159, 189] (see however Remark 6.10). In particular, it is often extremely sparse[15], positive semi-definite and its induced

[15] At least in the context of the spherical tessellation graphs explored here.

semi-norm

$$\|\boldsymbol{f}\|_{\boldsymbol{L}} := \|\boldsymbol{L}^{1/2}\boldsymbol{f}\|_2 = \sqrt{\boldsymbol{f}^H \boldsymbol{L}\boldsymbol{f}} = \sqrt{\sum_{(i,j)\in\mathcal{E}} \frac{W_{ij}}{\sqrt{\Lambda_{ii}\Lambda_{jj}}}(f_i - f_j)^2}, \quad \forall \boldsymbol{f} \in \mathbb{C}^N, \tag{6.22}$$

can be thought of as a generalised finite-difference scheme for measuring the global *smoothness*[16] [159, Example 2] of a signal $\boldsymbol{f} \in \mathbb{C}^N$ defined on the graph of $\mathscr{G}$. This intimate link between the Laplacian and the Laplace-Beltrami operator has lead scholars to define the graph Fourier transform (GFT) from the eigenvectors $\boldsymbol{U} \in \mathbb{R}^{N\times N}$ of the Laplacian $\boldsymbol{L}$:

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{U}^T,$$

where $\boldsymbol{\Delta} \in \mathbb{R}_+^{N\times N}$ is the diagonal matrix of eigenvalues of $\boldsymbol{L}$ sorted in *ascending order,* and $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_N$. We have then the following *analysis* and *synthesis* formulae for the GFT:

$$\hat{\boldsymbol{f}} = \boldsymbol{U}^T\boldsymbol{f}, \qquad \& \qquad \boldsymbol{f} = \boldsymbol{U}\hat{\boldsymbol{f}}, \quad \forall \boldsymbol{f} \in \mathbb{C}^N.$$

It can moreover be shown [159] that the eigenvectors associated to the largest eigenvalues have more *oscillatory* behaviour than those associated to the smallest eigenvalues, hence allowing one to interpret large eigenvalues of $\boldsymbol{L}$ as *high frequencies*. With the availability of a suitable notion of Fourier transform on graphs, we can then define *discrete pseudo-differential operators* as linear operators $\boldsymbol{D} : \mathbb{C}^N \to \mathbb{C}^N$ "boosting" the high frequency content of a graph signal:

$$\boldsymbol{D}\boldsymbol{f} := \boldsymbol{U}\mathrm{diag}(\hat{D}_1, \ldots, \hat{D}_N)\boldsymbol{U}^T\boldsymbol{f} = \boldsymbol{U}\mathrm{diag}(\hat{D}_1, \ldots, \hat{D}_N)\hat{\boldsymbol{f}}, \quad \boldsymbol{f} \in \mathbb{C}^N, \tag{6.23}$$

where the sequence of coefficients $\{\hat{D}_n,\, n = 1, \ldots, N\} \subset \mathbb{R}$ is *non-decreasing*, *non identically null*, and such that

$$|\hat{D}_n| \in \{0\} \cup [1, +\infty[, \qquad \forall n = 1, \ldots, N.$$

Observe that this definition is reminiscent of Definition 4.1 for continuous spherical pseudo-differential operators. In practice, we will mostly consider discrete pseudo-differential operators that can be expressed as *real polynomials* $\boldsymbol{D} = \sum_{k=0}^K \theta_k \boldsymbol{L}^k$ of the graph Laplacian, such as for example $(\boldsymbol{L} + \boldsymbol{I}_N)^K$ for some $K \in \mathbb{N}$, which can be thought of as the discrete counterpart of the Sobolev operators from Example 4.1. Finally, we define *discrete $\boldsymbol{D}$-splines* as vectors $\boldsymbol{s} \in \mathbb{C}^N$, such that

$$\boldsymbol{D}\boldsymbol{s} = \sum_{k=1}^K \alpha_k \boldsymbol{\delta}_{n_k}, \tag{6.24}$$

where $1 \leq K \leq N$, $\{\alpha_k, k = 1, \ldots, K\} \subset \mathbb{C}$ and $\{n_k, k = 1, \ldots, K\} \subset \{1, \ldots, N\}$. Observe that (6.24) is the discrete analog to (4.13) in Definition 4.4.

[16] This explains why the Laplacian is considered a discrete analog of the continuous negative Laplace-Beltrami operator.

Remark 6.10 — *Discrete Laplacian vs. Laplace-Beltrami Operator.* *Although possessing many of the nice properties of the continuous (negative) Laplace-Beltrami operator*[17]*, discrete graph Laplacians fail to verify them all at the same time (see [189] for a detailed exposition of this theoretical limitation). This explains the multiple alternative definitions of graph Laplacians proposed in the literature. Moreover, discrete Laplacians generally do not converge [30, 189] towards the continuous (negative) Laplace-Beltrami operator as the number of points N in the discrete manifold Θ grows to infinity. Therefore, the notion of smoothness captured by (6.22) differs slightly from the traditional notion of smoothness in the continuous setup.*

[17] See for example [139] for the specific case of HEALPix tessellation graphs.

2.4 Discrete gTikhonov and gTV Regularisation and Representer Theorems

The developments of Sections 2.1 to 2.3 lead us to consider the following optimisation problems:

$$\mathcal{V} = \arg\min_{\boldsymbol{f}\in\mathbb{C}^N} \left\{ F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{f}) \quad + \quad \lambda \|\boldsymbol{D}\boldsymbol{f}\|_2^2 \right\}, \tag{6.25}$$

and

$$\mathcal{V} = \arg\min_{\boldsymbol{f}\in\mathbb{C}^N} \left\{ F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{f}) \quad + \quad \lambda \|\boldsymbol{D}\boldsymbol{f}\|_1 \right\}, \tag{6.26}$$

as natural generalisations of the FPT and FPBP problems (5.10) and (5.14) to discrete point set domains Θ with size N. For both problems (6.25) and (6.26), we assume moreover the following

- I1 $\boldsymbol{y} \in \mathbb{C}^L$ is an arbitrary *data vector*;
- I2 $\boldsymbol{G} = [\boldsymbol{g}_1, \dots \boldsymbol{g}_L]^H \in \mathbb{C}^{L\times N}$ is some *sensing matrix* as in (6.20) with *independent sampling vectors* $\{\boldsymbol{g}_i,\, i = 1, \dots, L\} \subset \mathbb{C}^N$;
- I3 $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is a cost functional such that for all $\boldsymbol{y} \in \mathbb{C}^L$,

$$F(\boldsymbol{y}, \cdot) : \begin{cases} \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\} \\ \boldsymbol{z} \mapsto F(\boldsymbol{y}, \boldsymbol{z}) \end{cases}$$

  is *proper*, *convex* and lower semi-continuous;
- I4 $\lambda > 0$ is some *regularisation parameter*;
- I5 $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ is some *invertible discrete pseudo-differential operator* defined in (6.23).

The assumption that the rows of $\boldsymbol{G}$ are linearly independent in I2 is the translation in discrete terms of the assumption of independent linear measurements in FPT and FPBP problems (see Assumptions E3 and F3 of Theorems 5.3 and 5.4 respectively). Note that a necessary condition for I1 to hold is $L \leq N$. The assumption that $\boldsymbol{D}$ is invertible in I5 guarantees that the discrete gTikhonov and gTV norms $\|\boldsymbol{D}\cdot\|_2$ and $\|\boldsymbol{D}\cdot\|_1$ are indeed *positive definite*. Using the abstract representer theorems Corollary 2.10 and Theorem 2.12 from Chapter 2, it is possible to characterise the solutions of problems (6.25) and (6.26):

Theorem 6.7 — *Representer Theorem for Discrete gTikhonov Regularisation.* Consider the assumptions I1 to I5 introduced above. Then, (6.25) admits a *unique* solution given by

$$\boldsymbol{f}^\star = \boldsymbol{D}^{-2}\boldsymbol{G}^H\boldsymbol{\alpha} = \sum_{i=1}^{L} \alpha_i \boldsymbol{D}^{-2}\boldsymbol{g}_i, \tag{6.27}$$

for some $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L] \in \mathbb{C}^L$.

Proof. We apply Corollary 2.10 to (6.25), with $\mathscr{H} = (\mathbb{C}^N, \langle \boldsymbol{D}\cdot, \boldsymbol{D}\cdot\rangle_{\mathbb{C}^N})$, $\mathscr{H}' = (\mathbb{C}^N, \langle \boldsymbol{D}^{-1}\cdot, \boldsymbol{D}^{-1}\cdot\rangle_{\mathbb{C}^N})$, $\Lambda(t) = \lambda t^2$ and sampling operator $\boldsymbol{\Phi} : \mathbb{C}^N \to \mathbb{C}^L$ given by

$$\boldsymbol{\Phi}(\boldsymbol{f}) = [\langle \boldsymbol{g}_1 | \boldsymbol{f}\rangle, \dots, \langle \boldsymbol{g}_L | \boldsymbol{f}\rangle] = [\langle \boldsymbol{f}, \overline{\boldsymbol{g}_1}\rangle_{\mathbb{C}^N}, \dots, \langle \boldsymbol{f}, \overline{\boldsymbol{g}_L}\rangle_{\mathbb{C}^N}] = \boldsymbol{G}\boldsymbol{f}, \quad \forall \boldsymbol{f} \in \mathbb{C}^N.$$

Note that the assumptions of the corollary are indeed verified since $\mathscr{H}$ is an Hilbert space, Λ is convex and strictly increasing and the sampling vectors $\boldsymbol{g}_i$ are assumed linearly independent. We deduce hence that (6.25) admits a unique solution given by

$$\boldsymbol{f}^\star = \sum_{i=1}^{L} \alpha_i \boldsymbol{R}\boldsymbol{g}_i,$$

for some coefficients $\alpha_1, \dots, \alpha_L \in \mathbb{C}$ where $\boldsymbol{R} : \mathscr{H}' \to \mathscr{H}$ denotes the isometric *Riesz map*. The latter is moreover given by $\boldsymbol{D}^{-2}$ since any $\boldsymbol{f} \in \mathscr{H}$ can be written uniquely as

$$\boldsymbol{f} = \boldsymbol{D}^{-2}\boldsymbol{g}, \quad \text{with} \quad \|\boldsymbol{f}\|_{\mathscr{H}} = \|\boldsymbol{D}\boldsymbol{f}\|_2 = \|\boldsymbol{D}\boldsymbol{D}^{-2}\boldsymbol{g}\|_2 = \|\boldsymbol{D}^{-1}\boldsymbol{g}\|_2 = \|\boldsymbol{g}\|_{\mathscr{H}'}.$$

This finally yields

$$\boldsymbol{f}^\star = \sum_{i=1}^{L} \alpha_i \boldsymbol{D}^{-2}\boldsymbol{g}_i,$$

which can also be written equivalently as

$$\boldsymbol{f}^\star = \boldsymbol{D}^{-2}\boldsymbol{G}^H\boldsymbol{\alpha}, \quad \text{where} \quad \boldsymbol{\alpha} := [\alpha_1, \dots, \alpha_L] \in \mathbb{C}^L.$$

■

Theorem 6.8 — *Representer Theorem for Discrete gTV Regularisation.* Consider the assumptions I1 to I5 introduced above. The solution set $\mathcal{V}$ to (6.26) is *nonempty* and the *closed convex-hull of sparse extreme points* taking the form of *discrete $\boldsymbol{D}$-splines*:

$$\boldsymbol{f}^\star = \sum_{i=1}^{M} \alpha_i \boldsymbol{D}^{-1}\boldsymbol{\delta}_{n_i}, \tag{6.28}$$

where $1 \leq M \leq L$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L] \in \mathbb{C}^L$, $\{n_i,\, i = 1, \dots, M\} \subset \{1, \dots, N\}$ and $\boldsymbol{\delta}_i \in \mathbb{C}^N$, denotes the ith element of the *canonical basis* of $\mathbb{C}^N$.

Proof. We apply Theorem 2.12 to (6.26). To this end, we set $(\mathscr{B}, \|\cdot\|_{\mathscr{B}}) = (\mathbb{C}^N, \|\boldsymbol{D}^{-1}\cdot\|_\infty)$, $(\mathscr{B}', |||\cdot|||) = (\mathbb{C}^N, \|\boldsymbol{D}\cdot\|_1)$, $\Lambda(t) = \lambda t$ and sampling operator $\boldsymbol{\Phi} : \mathbb{C}^N \to \mathbb{C}^L$ given by

$$\boldsymbol{\Phi}(\boldsymbol{f}) = [\langle \boldsymbol{g}_1 | \boldsymbol{f} \rangle, \dots, \langle \boldsymbol{g}_L | \boldsymbol{f} \rangle] = [\langle \boldsymbol{f}, \overline{\boldsymbol{g}_1} \rangle_{\mathbb{C}^N}, \dots, \langle \boldsymbol{f}, \overline{\boldsymbol{g}_L} \rangle_{\mathbb{C}^N}] = \boldsymbol{G}\boldsymbol{f}, \quad \forall \boldsymbol{f} \in \mathbb{C}^N.$$

The assumptions of Theorem 2.12 are then indeed verified since $(\mathbb{C}^N, \|\boldsymbol{D}^{-1}\cdot\|_\infty)$ and $(\mathbb{C}^N, \|\boldsymbol{D}\cdot\|_1)$ form indeed a duality pair of Banach spaces, Λ is convex and strictly increasing and the sampling vectors $\boldsymbol{g}_i$ are assumed linearly independent. We get hence from Theorem 2.12 that the solution set $\mathcal{V}$ to (6.26) is *nonempty* and the *weak* closed convex hull* of its *extreme points*. Since in finite dimension the weak* topology coincides with the strong topology, $\mathcal{V}$ is also closed w.r.t. the canonical Banach topology on $(\mathbb{C}^N, \|\boldsymbol{D}\cdot\|_1)$. From Theorem 2.12, we have that the extreme points of $\mathcal{V}$ are moreover necessarily of the form:

$$\boldsymbol{f}^\star = \sum_{i=1}^{M} \beta_i \boldsymbol{e}_i, \tag{6.29}$$

where $1 \leq M \leq L$, $\{\beta_1, \dots, \beta_M\} \subset \mathbb{C}$ and $\boldsymbol{e}_i \in \mathbb{C}^N$ are *extreme points* of the closed regularisation ball

$$\mathcal{B}_{gTV,1/\lambda} = \{\boldsymbol{f} \in \mathbb{C}^N : \|\boldsymbol{D}\boldsymbol{f}\|_1 \leq 1/\lambda\}.$$

Using similar arguments as in the proof of Theorem 2.12, it is possible to show that the latter are of the form

$$\delta\mathcal{B}_{gTV,1/\lambda} = \left\{ z\lambda^{-1}\boldsymbol{D}^{-1}\boldsymbol{\delta}_i,\, i = 1, \dots, N,\, |z| = 1 \right\}, \tag{6.30}$$

where $\{\boldsymbol{\delta}_i,\, i = 1, \dots, N\}$ is the canonical basis on $\mathbb{C}^N$. Plugging (6.30) into (6.29) therefore allows us to write any extreme points of the solution set $\mathcal{V}$ as

$$\boldsymbol{f}^\star = \sum_{i=1}^{M} \frac{\beta_i z_i}{\lambda} \boldsymbol{D}^{-1}\boldsymbol{\delta}_{n_i} = \sum_{i=1}^{M} \alpha_i \boldsymbol{D}^{-1}\boldsymbol{\delta}_{n_i},$$

for some constants $\{\alpha_1, \dots, \alpha_M\} \subset \mathbb{C}$ and indices $\{n_1, \dots, n_M\} \subset \{1, \dots, N\}$, and where $1 \leq M \leq L$. ∎

2.5 Comparison with Search Space Discretisation

Theorem 6.7 predicts that the solution of the discrete-domain FPT problem (6.25) is in the span of $\{\boldsymbol{D}^{-2}\boldsymbol{g}_i,\, i = 1, \dots, L\}$. This is reminiscent of Theorem 5.3, which tells us that the solution to the continuous-domain FPT problem (5.16) is in the span of the functions $\{\mathscr{D}^{-2}\varphi_i,\, i = 1, \dots, L\}$. Similarly, Theorem 6.8 states that solutions of the discrete-domain FPBP problem (6.25) are *discrete* $\boldsymbol{D}$*-splines*[18], just like (non-limit) solutions of the FPBP problem (5.23) are $\mathscr{D}$-splines when $\mathscr{D}$ is spline-admissible (see Theorem 5.4). There-

[18] As convex combinations of sparse *discrete* $\boldsymbol{D}$*-splines*.

fore, for the discrete problem (6.25) (respectively (6.26)) to approximate well (5.16) (respectively (5.23)), we would like that the vectors $\boldsymbol{D}^{-2}\boldsymbol{g}_i$ (respectively discrete $\boldsymbol{D}$-splines) converge in a certain sense towards the functions $\mathscr{D}^{-2}\varphi_i$ (respectively $\mathscr{D}$-splines) as the dimension N –i.e. the number of points in the discrete manifold Θ– grows to infinity. Although this seems to be the case in practice, such a convergence result has not been formally proven yet. Moreover, the fact that discrete Laplacians do not converge [30] in general towards the negative Laplace-Beltrami operator on $\mathbb{S}^{d-1}$ make us doubtful that such a convergence result could ever be established.

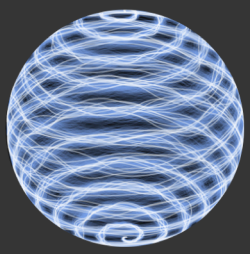
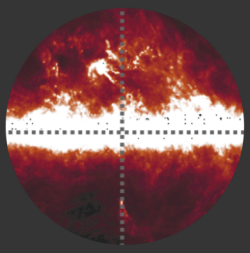
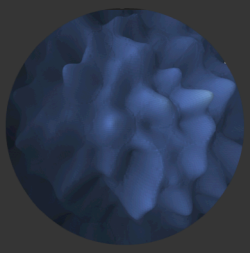
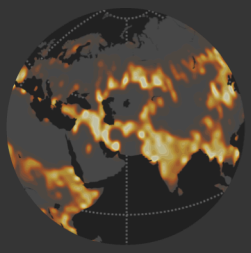
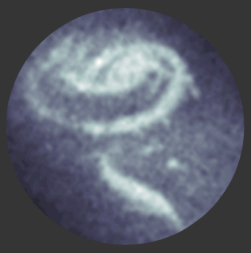
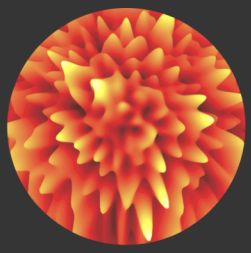
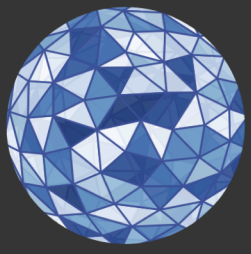
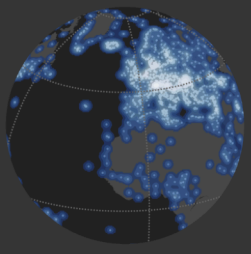
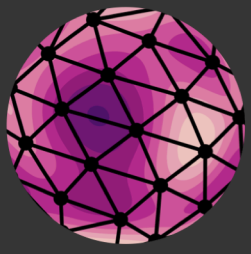
In contrast, the search-space discretisation schemes proposed in Section 1.1 and Section 1.2 yield solutions of the exact form predicted by Theorems 5.3 and 5.4 respectively. Unlike domain discretisation schemes (see Remark 6.9), they moreover do not require the sampling operator to be discretised, hence making it much easier to work with sampling operations inherent to the continuous domain, such as spatial sampling. For these reasons, we believe our search-space discretisation schemes to be much better indicated than domain discretisation schemes for solving FPT and FPBP problems in practice.

III

Algorithms & Applications

In this part, we discuss practical aspects of the spherical approximation framework from Part II. Some of the highlights of this part are the following:

- *In Chapter 7, we design efficient and provably convergent proximal algorithms for all discrete optimisation problems considered in Chapter 6. We moreover propose rules of thumb for setting their various hyperparameters and provide the proximal operators of most common cost functionals.*
- *In Chapter 8, we discuss the use of Wendland and Matérn pseudo-differential operators when designing gTikhonov and gTV penalties, and their convenient properties for practical purposes.*
- *In Chapter 9, we test the spherical approximation framework from Part II and novel algorithms from Chapter 7 on a variety of real and simulated datasets, coming from the fields of meteorology, forestry, radio astronomy and planetary sciences. The sampling functionals, cost functions and regularising strategies considered in each case are very diverse, showing the versatility of both our theoretical framework and algorithmic solutions. In the meteorology example, we moreover illustrate the superiority of continuous-domain vs. discrete-domain recovery, both in terms of accuracy and resolution.*

7 Optimisation Algorithms

In this chapter, we propose to solve the discrete problems (6.4), (6.14), (6.25) and (6.26) by means of provably convergent *fully-split proximal iterative methods* [134], which only involve simple matrix-vector multiplications and *proximal steps*. We treat the most general case where the cost function F is proximable but not necessarily differentiable with the *primal-dual splitting method (PDS)* introduced by Condat in his seminal work [43]. In the simpler (yet prevailing in practice) case where F is also *differentiable* and with β*-Lipschitz continuous* derivative, we leverage an optimal first-order method called *accelerated proximal gradient descent (APGD)* [16, 134], with faster convergence rate than the PDS method. For the sake of simplicity and without loss of generality, we consider the *real case* only, where $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{y} \in \mathbb{R}^L$ –i.e. the coefficients and data vector are assumed real. The complex case, less common in practice, can be handled similarly by identifying $\mathbb{C}^N$ and $\mathbb{C}^L$ with $\mathbb{R}^{2N}$ and $\mathbb{R}^{2L}$ respectively and reformulating optimisation problems (6.4), (6.14), (6.25) and (6.26) in terms of real operations only, using the techniques described in [148, Section 7.8]. For reference purposes, we provide in Table 7.2 page 133 a summary of the various algorithms investigated in this chapter.

1 The Primal-Dual Splitting Method

We review here the primal-dual splitting method (PDS) proposed in [43]. The latter is an iterative method aiming to solve the following *primal* optimisation problem:

$$\text{Find } \boldsymbol{x}^\star \in \underset{\boldsymbol{x}\in\mathcal{X}}{\arg\min} \; \{\mathcal{F}(\boldsymbol{x}) \; + \; \mathcal{G}(\boldsymbol{x}) + \mathcal{H}(\boldsymbol{K}\boldsymbol{x})\}\,, \tag{7.1}$$

under the following assumptions:

- J1 $\mathcal{X}$ and $\mathcal{Y}$ are two *finite dimensional*[1] *real Hilbert spaces* with respective *inner products* $\langle\cdot,\cdot\rangle_{\mathcal{X}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{Y}}$. We denote moreover the norms induced by both inner products on $\mathcal{X}$ and $\mathcal{Y}$ as $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$ respectively.
- J2 $\mathcal{F} : \mathcal{X} \to \mathbb{R}$ is *convex* and *differentiable*, with β*-Lipschitz continuous* gradient:

$$\|\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) - \boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}')\|_{\mathcal{X}} \leq \beta\|\boldsymbol{x} - \boldsymbol{x}'\|_{\mathcal{X}}, \qquad \forall(\boldsymbol{x},\boldsymbol{x}') \in \mathcal{X}^2, \tag{7.2}$$

[1] The algorithm presented in [43] is actually also valid for infinite dimensional vector spaces, in which case however the assumptions for the convergence of the method are slightly more stringent.

for some *Lipschitz constant* $\beta \in [0, +\infty[$.

J3 $\mathcal{G} : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and $\mathcal{H} : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ are two *proper, lower semi-continuous (lwsc)* and *convex functions* with *simple proximal operators*, in the sense that the latter admit a *closed-form* representation or can be computed *efficiently and with high precision*. As a reminder, the *proximal operator* of $\mathcal{G}$ (respectively $\mathcal{H}$ with the norm $\|\cdot\|_{\mathcal{Y}}$) is given by [134, Section 1.1]

$$\mathbf{prox}_{\mathcal{G}}(\boldsymbol{u}) := \arg\min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathcal{G}(\boldsymbol{x}) \quad + \quad \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_{\mathcal{X}}^2 \right\}, \qquad \forall \boldsymbol{u} \in \mathcal{X}. \tag{7.3}$$

J4 $\boldsymbol{K} : \mathcal{X} \to \mathcal{Y}$ is a *linear operator*, with *induced operator norm*:

$$\|\boldsymbol{K}\|_{\mathcal{X},\mathcal{Y}} = \sup_{\boldsymbol{x}\in\mathcal{X}, \|\boldsymbol{x}\|_{\mathcal{X}}=1} \|\boldsymbol{K}\boldsymbol{x}\|_{\mathcal{Y}}. \tag{7.4}$$

J5 The solution set in (7.1) is *non-empty*.

Vocabulary 7.1 — *Proximable Function.* *We say that a function is* proximable *if its proximity operator admits a closed-form representation.*

As explained in [43], PDS solves jointly the *primal* problem (7.1) and its associated *dual* [43, Equation 4] by combining them into the *saddle-point problem*

$$\text{Find } (\boldsymbol{x}^\star, \boldsymbol{z}^\star) \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{z}\in\Delta(\mathcal{H}^*)} \left\{ \mathcal{F}(\boldsymbol{x}) \; + \; \mathcal{G}(\boldsymbol{x}) \; - \; \mathcal{H}^*(\boldsymbol{z}) \; + \; \langle \boldsymbol{K}\boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{Y}} \right\}, \tag{7.5}$$

where $\mathcal{H}^* : \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ is the *convex conjugate* of $\mathcal{H}$, with *domain* $\Delta(\mathcal{H}^*) \subset \mathcal{Y}$ and defined as [127]

$$\mathcal{H}^*(\boldsymbol{z}) := \max_{\boldsymbol{y}\in\mathcal{Y}} \; \langle \boldsymbol{z}, \boldsymbol{y} \rangle_{\mathcal{Y}} - \mathcal{H}(\boldsymbol{y}), \quad \forall \boldsymbol{z} \in \mathcal{Y}. \tag{7.6}$$

Observe that the solutions of (7.5) w.r.t. the variable $\boldsymbol{x}$ are such that

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathcal{F}(\boldsymbol{x}) \; + \; \mathcal{G}(\boldsymbol{x}) \; + \; \mathcal{H}^{**}(\boldsymbol{K}\boldsymbol{x}) \right\},$$

where $\mathcal{H}^{**}(\boldsymbol{y}) := \max_{\boldsymbol{z}\in\Delta(\mathcal{H}^*)} \langle \boldsymbol{y}, \boldsymbol{z} \rangle_{\mathcal{Y}} - \mathcal{H}^*(\boldsymbol{z})$, $\boldsymbol{y} \in \mathcal{Y}$ is the conjugate of $\mathcal{H}^*$, called the *convex biconjugate*[2] of $\mathcal{H}$ [127]. When $\mathcal{H}$ is *convex* and regular enough –which is the case here– we have moreover, from the *Fenchel-Moreau theorem* [99], equality between $\mathcal{H}$ and its biconjugate: $\mathcal{H} = \mathcal{H}^{**}$. Therefore, solutions of the saddle-point problem (7.5) w.r.t. the variable $\boldsymbol{x}$ are indeed solutions of the primal problem (7.1). The saddle-point problem (7.5) presents however the advantage of splitting the complicated composite term $\mathcal{H}(\boldsymbol{K}\boldsymbol{x})$ in (7.1) into a sum of two simpler terms $\langle \boldsymbol{K}\boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{Y}} - \mathcal{H}^*(\boldsymbol{z})$ which are easier to optimise, since respectively differentiable and proximable. The proximal operator of the convex conjugate $\mathcal{H}^*$ is moreover given by *Moreau's identity* [134]

$$\mathbf{prox}_{\sigma\mathcal{H}^*}(\boldsymbol{z}) = \boldsymbol{z} - \sigma\, \mathbf{prox}_{\mathcal{H}/\sigma}(\boldsymbol{z}/\sigma), \qquad \forall \boldsymbol{z} \in \mathcal{Y}, \, \sigma > 0. \tag{7.7}$$

In [43], Condat proposes two PDS iterative methods for solving (7.5). In the

[2] The convex biconjugate function $\mathcal{H}^{**}$ of a function $\mathcal{H}$ can in general be interpreted geometrically as the *closed convex-hull* of the epigraph of $\mathcal{H}$, i.e. the largest lwsc convex function such that $\mathcal{H}^{**} \leq \mathcal{H}$ [200].

Algorithm 7.1: A primal-dual splitting method for solving (7.5).

1: **procedure** PDS($\tau, \sigma, \rho, \boldsymbol{x}_0, \boldsymbol{z}_0$)
2: **for all** $n \geq 1$ **do**
3: $\tilde{\boldsymbol{x}}_n = \mathbf{prox}_{\tau\mathcal{G}}\left(\boldsymbol{x}_{n-1} - \tau\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}_{n-1}) - \tau\boldsymbol{K}^*\boldsymbol{z}_{n-1}\right)$
4: $\tilde{\boldsymbol{z}}_n = \mathbf{prox}_{\sigma\mathcal{H}^*}\left(\boldsymbol{z}_{n-1} + \sigma\boldsymbol{K}\left[2\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_{n-1}\right]\right)$
5: $\boldsymbol{x}_n = \rho\tilde{\boldsymbol{x}}_n + (1-\rho)\boldsymbol{x}_{n-1}$
6: $\boldsymbol{z}_n = \rho\tilde{\boldsymbol{z}}_n + (1-\rho)\boldsymbol{z}_{n-1}$
7: **return** $\{(\boldsymbol{x}_n, \boldsymbol{z}_n)\}_{n\in\mathbb{N}}$

context of this thesis, we will work with the one described in Algorithm 7.1. The latter produces a sequence $\{(\boldsymbol{x}_n, \boldsymbol{z}_n)\}_{n\in\mathbb{N}} \subset \mathcal{X}\times\mathcal{Y}$ converging towards a solution pair $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$ of (7.5) by iterating rows 3 to 6 of Algorithm 7.1, starting from arbitrary initial guesses[3] $(\boldsymbol{x}_0, \boldsymbol{z}_0) \in \mathcal{X}\times\mathcal{Y}$. Notice that the update equations 3 and 4 are not too computationally intensive since they involve only the simple proximal operators of $\mathcal{G}$ and $\mathcal{H}^*$, as well as linear operations between the gradient $\boldsymbol{\nabla}\mathcal{F}$, the operator $\boldsymbol{K}$, its adjoint $\boldsymbol{K}^*$ and the primal/dual variables. The hyperparameters τ, σ in Algorithm 7.1 can be interpreted as *step sizes*, which control the amount of improvement in the primal and dual variable respectively, while the parameter ρ can be interpreted as a *momentum term*.[4] To ensure convergence of the method, these hyperparameters must verify the conditions listed in [43, Theorem 3.1 and Theorem 3.3], depending on whether the Lipschitz constant β is null[5] or not. Both results are provided hereafter for reference purposes.

[3] Typically chosen identically null or random.

[4] Condat also considers the more general case of a momentum varying across iterations.

[5] The Lipschitz constant β is null when $\mathcal{F}$ is an affine function or $\mathcal{F} = 0$.

Theorem 7.1 — *Convergence of the PDS Method ($\beta \neq 0$) [43].* Consider problem (7.5) under the assumptions J1 to J5 and let $\tau > 0$, $\sigma > 0$ and ρ be the *hyperparameters* of Algorithm 7.1. Suppose moreover that $\beta > 0$ and that the following holds:

K1 $\frac{1}{\tau} - \sigma\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}} \geq \frac{\beta}{2}$,

K2 $\rho \in]0, \delta[$, where $\delta := 2 - \frac{\beta}{2}\left(\frac{1}{\tau} - \sigma\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}}\right)^{-1} \in [1, 2[$.

Then, there exists a pair $(\boldsymbol{x}^\star, \boldsymbol{z}^\star) \in \mathcal{X}\times\mathcal{Y}$ solution to (7.5), s.t. the primal and dual sequences of estimates $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ *converge* towards $\boldsymbol{x}^\star$ and $\boldsymbol{z}^\star$ respectively, i.e.

$$\lim_{n\to+\infty} \|\boldsymbol{x}^\star - \boldsymbol{x}_n\|_{\mathcal{X}} = 0, \quad \text{and} \quad \lim_{n\to+\infty} \|\boldsymbol{z}^\star - \boldsymbol{z}_n\|_{\mathcal{Y}} = 0.$$

Proof. See [43, Section 4]. ■

Remark 7.1 *Note that Theorem 7.1 above is a specialisation of [43, Theorem 3.1] to the case where both $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional, the proximal and gradient steps are computed exactly, and the momentum term ρ is chosen constant across iterations as in Algorithm 7.1.*

Theorem 7.2 — *Convergence of the PDS Method ($\beta = 0$) [43].* Consider problem (7.5) under the assumptions J1 to J5 and let $\tau > 0$, $\sigma > 0$ and ρ be the hyperparameters of Algorithm 7.1. Suppose moreover that $\beta = 0$ and that the following holds:

L1 $\tau\sigma\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}} \leq 1$,

L2 $\rho \in [\epsilon, 2-\epsilon]$, for some $\epsilon > 0$.

Then, there exists a pair $(\boldsymbol{x}^\star, \boldsymbol{z}^\star) \in \mathcal{X} \times \mathcal{Y}$ solution to (7.5), s.t. the primal and dual sequences of estimates $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ *converge* towards $\boldsymbol{x}^\star$ and $\boldsymbol{z}^\star$ respectively, i.e.

$$\lim_{n\to+\infty} \|\boldsymbol{x}^\star - \boldsymbol{x}_n\|_{\mathcal{X}} = 0, \quad \text{and} \quad \lim_{n\to+\infty} \|\boldsymbol{z}^\star - \boldsymbol{z}_n\|_{\mathcal{Y}} = 0.$$

Proof. See [43, Section 4]. ■

Remark 7.2 *Again, Theorem 7.1 is a specialisation of [43, Theorem 3.3] to the case where the proximal and gradient steps are computed exactly, and the momentum term ρ is chosen constant across iterations as in Algorithm 7.1.*

In practice, the *convergence speed* of Algorithm 7.1 is improved by choosing σ and τ as large as possible and relatively *well-balanced* –so that both the primal and dual problems converge at the same pace. Consequently, we chose in our implementation of Algorithm 7.1 the perfectly balanced parameters $\sigma = \tau$ saturating the inequalities K1 and L1. For the scenario considered in Theorem 7.1 this yields:

$$\frac{1}{\tau} - \tau\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}} = \frac{\beta}{2} \quad \Longleftrightarrow \quad -2\tau^2\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}} - \beta\tau + 2 = 0,$$

which admits one positive root

$$\tau = \sigma = \frac{1}{\|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}}}\left(-\frac{\beta}{4} + \sqrt{\frac{\beta^2}{16} + \|\boldsymbol{K}\|^2_{\mathcal{X},\mathcal{Y}}}\right). \tag{7.8}$$

For the scenario considered in Theorem 7.2 finally, this yields

$$\tau = \sigma = \|\boldsymbol{K}\|^{-1}_{\mathcal{X},\mathcal{Y}}. \tag{7.9}$$

Note that the theoretical convergence rate of the PDS method has only been assessed in particular cases. For example, Chambolle and Pock have shown in [35, Theorem 1] that, in the case where $\mathcal{F} = 0$, the PDS method converges[6] at a suboptimal rate $\mathcal{O}(1/n)$. This rather slow convergence is also observed empirically in more general cases. In the subsequent section, we propose, for the common case $\mathcal{H} = 0$, a more efficient optimisation method called accelerated proximal gradient descent (APGD), with optimal convergence rate $o(1/n^2)$.

[6] More precisely they showed that the partial duality gap decreases at a rate inversely proportional to the number of iterations.

Remark 7.3 — *Stopping Criterion.* *In practice, we stop Algorithm 7.1 when the relative improvement in the primal variable $\boldsymbol{x}_n$ falls below a certain predetermined accuracy threshold $\varepsilon > 0$. This stopping criterion, which monitors*

improvement of the primal variable, is motivated by the fact that we are in this context only interested in solving the primal problem (7.1).

2 Accelerated Proximal Gradient Descent

Consider the specific case $\mathcal{H} = 0$ in optimisation problem (7.1). Then, the latter can be solved by means of the accelerated proximal gradient descent (APGD) method [8, 34], implemented in Algorithm 7.2. It has been shown in [8, Theorem 1] that, under the assumptions J1, J2, J3, J5, $0 < \tau \leq 1/\beta$ and $\mathfrak{d} > 2$, APGD achieves the following *optimal convergence rates*:

$$\lim_{n\to\infty} n^2 \left|\mathcal{J}(\boldsymbol{x}^\star) - \mathcal{J}(\boldsymbol{x}_n)\right| = 0 \qquad \& \qquad \lim_{n\to\infty} n^2 \|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\|_{\mathcal{X}}^2 = 0,$$

for some minimiser $\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \{\mathcal{J}(\boldsymbol{x}) := \mathcal{F}(\boldsymbol{x}) + \mathcal{G}(\boldsymbol{x})\} \neq \emptyset$. In other words, both the objective functional and the APGD iterates $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ converge at a rate $o(1/n^2)$. In our practical implementation of Algorithm 7.2, we chose the step size τ as large as possible $\tau = 1/\beta$ and set $\mathfrak{d}$ to the value $\mathfrak{d} = 75$. The latter choice was motivated by the results reported in [111, 113], which show significant practical acceleration for values of $\mathfrak{d}$ in the range $[50, 100]$.

Remark 7.4 — *Stopping Criterion.* *Similarly as for Algorithm 7.1, we stop in practice Algorithm 7.2 when the relative improvement $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\|_{\mathcal{X}}/\|\boldsymbol{x}_{n-1}\|_{\mathcal{X}}$ falls under a certain pre-determined accuracy threshold $\varepsilon > 0$.*

Algorithm 7.2: APGD method for solving (7.1) when $\mathcal{H} = 0$.

procedure APGD$(\tau, \mathfrak{d}, \boldsymbol{x}_0)$
  for all $n \geq 1$ **do**
    $\boldsymbol{z}_n = \mathbf{prox}_{\tau\mathcal{G}}(\boldsymbol{x}_{n-1} - \tau\nabla\mathcal{F}(\boldsymbol{x}_{n-1}))$
    $\boldsymbol{x}_n = \boldsymbol{z}_n + \frac{n-1}{n+\mathfrak{d}}(\boldsymbol{z}_n - \boldsymbol{z}_{n-1})$
  return $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$

3 Algorithms for Search Space Discretisation Schemes

In this section, we apply Algorithms 7.1 and 7.2 to the optimisation problems (6.4) and (6.14) obtained by the canonical search space discretisation schemes proposed in Sections 1.1 and 1.2 of Chapter 6 respectively. In all that follows, we assume that the Hilbert spaces $\mathbb{R}^N$ and $\mathbb{R}^L$ are equipped with their canonical inner products and induced norms. Moreover, all operator norms are defined w.r.t. these induced norms on $\mathbb{R}^N$ and $\mathbb{R}^L$. Since such canonical norms are unambiguous, we simplify their notations by dropping their subscripts.

3.1 gTikhonov Regularisation

Consider the (real) optimisation problem (6.4):

$$\text{Find } \boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{R}^L} \left\{F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \boldsymbol{x}^T \boldsymbol{G}\boldsymbol{x}\right\}, \tag{7.10}$$

where $\boldsymbol{y} \in \mathbb{R}^L$, $F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proper convex* and lwsc, $\boldsymbol{G} \in \mathbb{R}^{L \times L}$ and $\lambda > 0$. We propose in the subsequent sections various algorithms for solving (7.10) depending on the nature of the cost functional F.

3.1.1 Proximable Cost Functional

Assume that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proximable* but not necessarily smooth. Then, (7.10) can be seen as a specific instance of (7.1), with

$$\mathcal{X} = \mathcal{Y} = \mathbb{R}^L,\ \mathcal{F}(\boldsymbol{x}) = \lambda \boldsymbol{x}^T \boldsymbol{G} \boldsymbol{x},\ \mathcal{G}(\boldsymbol{x}) = 0\ , \mathcal{H}(\boldsymbol{z}) = E_{\boldsymbol{y}}(\boldsymbol{z}),\ \text{and}\ \boldsymbol{K} = \boldsymbol{G},$$

and

$$\boldsymbol{\nabla} \mathcal{F}(\boldsymbol{x}) = 2\lambda \boldsymbol{G} \boldsymbol{x}, \quad \forall \boldsymbol{x} \in \mathbb{R}^L \quad \text{and} \quad \beta = 2\lambda \|\boldsymbol{G}\|.$$

We can therefore solve (7.10) by specialising the generic Algorithm 7.1 to this particular setup. The resulting algorithm –with stopping criterion as in Remark 7.3– is implemented in Algorithm 7.3. The convergence condition K1 from Theorem 7.1 becomes in this case

$$\frac{1}{\tau} - \sigma \|\boldsymbol{G}\|^2 \geq \lambda \|\boldsymbol{G}\|,$$

and the rule of thumb (7.8) for setting the step sizes τ and σ yields

$$\tau = \sigma = \frac{1}{\|\boldsymbol{G}\|} \left(-\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + 1} \right).$$

Finally, condition K2 of Theorem 7.1 tells us that for such step sizes, the momentum parameter ρ should be chosen in the open interval $]0, 1[$.

Algorithm 7.3: PDS method for solving (7.10) when F is proximable but not necessarily smooth.

procedure PDS$(\boldsymbol{y}, \tau, \sigma, \rho, \boldsymbol{x}_0, \boldsymbol{z}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\tilde{\boldsymbol{x}}_n = \boldsymbol{x}_{n-1} - 2\tau\lambda \boldsymbol{G} \boldsymbol{x}_{n-1} - \tau \boldsymbol{G}^T \boldsymbol{z}_{n-1}$
    $\tilde{\boldsymbol{z}}_n = \mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*} \left(\boldsymbol{z}_{n-1} + \sigma \boldsymbol{G} \left[2\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_{n-1} \right] \right)$
    $\boldsymbol{x}_n = \rho \tilde{\boldsymbol{x}}_n + (1 - \rho) \boldsymbol{x}_{n-1}$
    $\boldsymbol{z}_n = \rho \tilde{\boldsymbol{z}}_n + (1 - \rho) \boldsymbol{z}_{n-1}$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon \|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

3.1.2 Smooth Cost Functional

Assume this time that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}$ is *differentiable* with γ-Lipschitz continuous derivative. Then, (7.10) can be seen as a

specific instance of (7.1), with

$$\mathcal{X} = \mathbb{R}^L,\ \mathcal{F}(\boldsymbol{x}) = E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}) + \lambda \boldsymbol{x}^T \boldsymbol{G}\boldsymbol{x},\ \text{and}\ \mathcal{G} = \mathcal{H} = 0,$$

and

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) = \boldsymbol{G}^T \boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}) + 2\lambda \boldsymbol{G}\boldsymbol{x}, \qquad \forall \boldsymbol{x} \in \mathbb{R}^L.$$

Note moreover that $\boldsymbol{\nabla}\mathcal{F}$ is Lipschitz continuous since, from the triangle inequality

$$\begin{aligned}\|\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) - \boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}')\| &\leq \|\boldsymbol{G}^T(\boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}) - \boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}'))\| + 2\lambda\|\boldsymbol{G}(\boldsymbol{x} - \boldsymbol{x}')\| \\ &\leq \left(\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{G}\|\right)\|\boldsymbol{x} - \boldsymbol{x}'\|, \qquad \forall (\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{R}^L \times \mathbb{R}^L.\end{aligned}$$

The Lipschitz constant β of $\boldsymbol{\nabla}\mathcal{F}$ is moreover such that $\beta \leq \gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{G}\|$. Since $\mathcal{H} = 0$, we can in this case solve (7.10), by means of the APGD method described in Algorithm 7.2. The resulting algorithm –with stopping criterion as in Remark 7.4– is implemented in Algorithm 7.4. The step size τ can optimally be chosen as $\tau = 1/\beta$ or underestimated as

$$\tau = (\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{G}\|)^{-1} \leq 1/\beta,$$

which may be easier to compute in practice.

Algorithm 7.4: APGD method for solving (7.10) when F is smooth.

procedure APGD$(\boldsymbol{y}, \tau, \mathfrak{d}, \boldsymbol{x}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\boldsymbol{z}_n = \boldsymbol{x}_{n-1} - \tau\boldsymbol{G}^T\boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}_{n-1}) - 2\tau\lambda\boldsymbol{G}\boldsymbol{x}_{n-1}$
    $\boldsymbol{x}_n = \boldsymbol{z}_n + \frac{n-1}{n+\mathfrak{d}}(\boldsymbol{z}_n - \boldsymbol{z}_{n-1})$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon\|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

3.2 gTV Regularisation

Consider the (real) optimisation problem (6.14):

$$\text{Find } \boldsymbol{x}^\star \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\arg\min}\left\{F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda\|\boldsymbol{x}\|_1\right\}, \tag{7.11}$$

where $\boldsymbol{y} \in \mathbb{R}^L$, $F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proper convex* and lwsc, $\boldsymbol{G} \in \mathbb{R}^{L\times N}$ and $\lambda > 0$. We propose in the subsequent sections various algorithms for solving (7.11) depending on the nature of the cost functional F.

3.2.1 Proximable Cost Functional

Assume that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proximable* but not necessarily smooth. Then, (7.11) can be seen as a specific instance

of (7.1), with

$$\mathcal{X} = \mathbb{R}^N,\ \mathcal{Y} = \mathbb{R}^L,\ \mathcal{F} = 0,\ \mathcal{G}(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_1\ , \mathcal{H}(\boldsymbol{z}) = E_{\boldsymbol{y}}(\boldsymbol{z}),\ \text{and}\ \boldsymbol{K} = \boldsymbol{G}.$$

Note that $\mathcal{G}$ is moreover simple, since its proximal operator is given by the so-called *soft-thresholding operator* [134, Chapter 6]:

$$\mathbf{prox}_{\tau\lambda\|\cdot\|_1}(\boldsymbol{x}) = \mathbf{soft}_{\tau\lambda}(\boldsymbol{x}) = \mathbf{max}(|\boldsymbol{x}| - \tau\lambda, 0)\,\mathbf{sgn}(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^N,\ \tau > 0, \tag{7.12}$$

where the functions $\mathbf{max} : \mathbb{R}^N \to \mathbb{R}^N$ and $\mathbf{sgn} : \mathbb{R}^N \to \mathbb{R}^N$ are the element-wise *maximum* and *signum* functions respectively. We can therefore solve (7.10), by specialising the generic Algorithm 7.1 to this particular setup. Since $\mathcal{F} = 0$, the convergence of Algorithm 7.5 is guaranteed this time by Theorem 7.2. In particular, the convergence condition L1 becomes in this case

$$\sigma\tau\|\boldsymbol{G}\|^2 \leq 1,$$

and the rule of thumb (7.9) for setting the step sizes τ and σ yields

$$\tau = \sigma = \|\boldsymbol{G}\|^{-1}.$$

Finally, condition L2 of Theorem 7.2 tells us that the momentum parameter ρ should be chosen in the closed interval $[\epsilon, 2-\epsilon]$, for some $\epsilon > 0$. For simplicity, we choose $\rho = 1$. The resulting algorithm –with stopping criterion as in Remark 7.3– is implemented in Algorithm 7.5.

Algorithm 7.5: PDS method for solving (7.11) when F is proximable but not necessarily smooth.

procedure PDS$(\boldsymbol{y}, \tau, \sigma, \boldsymbol{x}_0, \boldsymbol{z}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\boldsymbol{x}_n = \mathbf{soft}_{\tau\lambda}\left(\boldsymbol{x}_{n-1} - \tau\boldsymbol{G}^T\boldsymbol{z}_{n-1}\right)$
    $\boldsymbol{z}_n = \mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*}\left(\boldsymbol{z}_{n-1} + \sigma\boldsymbol{G}\left[2\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\right]\right)$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon\|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

3.2.2 Smooth Cost Functional

Assume this time that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}$ is *differentiable* with γ-Lipschitz continuous derivative. Then, (7.11) can be seen as a specific instance of (7.1), with

$$\mathcal{X} = \mathbb{R}^N,\ \mathcal{F}(\boldsymbol{x}) = E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}),\ \mathcal{G}(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_1,\ \text{and}\ \mathcal{H} = 0,$$

and

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) = \boldsymbol{G}^T\boldsymbol{\nabla}E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

The Lipschitz constant β of $\nabla\mathcal{F}$ is moreover such that $\beta \leq \gamma\|\boldsymbol{G}\|^2$. Again, $\mathcal{G}$ is simple, and its proximal operator is given by (7.12). Since $\mathcal{H} = 0$, we can solve (7.11) via the APGD Algorithm 7.2, which yields in this case the famous fast iterative soft-thresholding algorithm (FISTA) [16, 112, 113]. Again, the resulting algorithm –with stopping criterion as in Remark 7.4– is implemented in Algorithm 7.6. The step size τ can optimally be chosen as $\tau = 1/\beta$ or underestimated as

$$\tau = \frac{1}{\gamma\|\boldsymbol{G}\|^2} \leq \frac{1}{\beta},$$

which may be easier to compute in practice.

Algorithm 7.6: FISTA method for solving (7.11) when F is smooth.

procedure FISTA($\boldsymbol{y}, \tau, \mathfrak{d}, \boldsymbol{x}_0, \epsilon$)
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\boldsymbol{z}_n = \mathbf{soft}_{\tau\lambda}(\boldsymbol{x}_{n-1} - \tau\boldsymbol{G}^T\nabla E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}_{n-1}))$
    $\boldsymbol{x}_n = \boldsymbol{z}_n + \frac{n-1}{n+\mathfrak{d}}(\boldsymbol{z}_n - \boldsymbol{z}_{n-1})$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon\|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

4 Algorithms for Domain Discretisation Schemes

In this section, we apply Algorithms 7.1 and 7.2 to the optimisation problems (6.25) and (6.26) obtained by domain discretisation in Section 2.4 of Chapter 6. Again, we consider canonical norms on $\mathbb{R}^N$ and $\mathbb{R}^L$ as well as canonical operator norms, all noted without subscripts since unambiguous.

4.1 gTikhonov Regularisation

Consider the (real) optimisation problem (6.25):

$$\text{Find } \boldsymbol{x}^\star \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\arg\min} \left\{ F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda\|\boldsymbol{D}\boldsymbol{x}\|^2 \right\}, \tag{7.13}$$

where $\boldsymbol{y} \in \mathbb{R}^L$, $F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}\cup\{+\infty\}$ is *proper convex* and lwsc, $\boldsymbol{D} \in \mathbb{R}^{N\times N}$, and $\lambda > 0$. We propose in the subsequent sections various algorithms for solving (7.13) depending on the nature of the cost functional F.

4.1.1 Proximable Cost Functional

Assume that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proximable* but not necessarily smooth. Then, (7.13) can be seen as a specific instance of (7.1), with

$$\mathcal{X} = \mathbb{R}^N, \mathcal{Y} = \mathbb{R}^L, \ \mathcal{F}(\boldsymbol{x}) = \lambda\|\boldsymbol{D}\boldsymbol{x}\|^2, \ \mathcal{G}(\boldsymbol{x}) = 0\ , \mathcal{H}(\boldsymbol{z}) = E_{\boldsymbol{y}}(\boldsymbol{z}), \text{ and } \boldsymbol{K} = \boldsymbol{G},$$

and

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) = 2\lambda \boldsymbol{D}^T\boldsymbol{D}\boldsymbol{x}, \quad \forall \boldsymbol{x} \in \mathbb{R}^N \quad \text{and} \quad \beta = 2\lambda\|\boldsymbol{D}\|^2.$$

We can therefore solve (7.13), by specialising the generic Algorithm 7.1 to this particular setup. The resulting algorithm –with stopping criterion as in Remark 7.3– is implemented in Algorithm 7.7. The convergence condition K1 from Theorem 7.1 becomes in this case

$$\frac{1}{\tau} - \sigma\|\boldsymbol{G}\|^2 \geq \lambda\|\boldsymbol{D}\|^2,$$

and the rule of thumb (7.8) for setting the step sizes τ and σ yields

$$\tau = \sigma = \frac{\|\boldsymbol{D}\|^2}{\|\boldsymbol{G}\|^2}\left(-\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + \frac{\|\boldsymbol{G}\|^2}{\|\boldsymbol{D}\|^4}}\right).$$

Finally, condition K2 of Theorem 7.1 tells us that for such step sizes, the momentum parameter ρ should be chosen in the open interval $]0, 1[$.

Algorithm 7.7: PDS method for solving (7.13) when F is proximable but not necessarily smooth.

procedure PDS$(\boldsymbol{y}, \tau, \sigma, \rho, \boldsymbol{x}_0, \boldsymbol{z}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\tilde{\boldsymbol{x}}_n = \boldsymbol{x}_{n-1} - 2\tau\lambda\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{x}_{n-1} - \tau\boldsymbol{G}^T\boldsymbol{z}_{n-1}$
    $\tilde{\boldsymbol{z}}_n = \mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*}\left(\boldsymbol{z}_{n-1} + \sigma\boldsymbol{G}\left[2\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_{n-1}\right]\right)$
    $\boldsymbol{x}_n = \rho\tilde{\boldsymbol{x}}_n + (1-\rho)\boldsymbol{x}_{n-1}$
    $\boldsymbol{z}_n = \rho\tilde{\boldsymbol{z}}_n + (1-\rho)\boldsymbol{z}_{n-1}$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon\|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

4.1.2 Smooth Cost Functional

Assume this time that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}$ is *differentiable* with γ-Lipschitz continuous derivative. Then, (7.13) can be seen as a specific instance of (7.1), with

$$\mathcal{X} = \mathbb{R}^N, \ \mathcal{F}(\boldsymbol{x}) = E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}) + \lambda\|\boldsymbol{D}\boldsymbol{x}\|^2, \text{ and } \mathcal{G} = \mathcal{H} = 0.$$

and

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) = \boldsymbol{G}^T\boldsymbol{\nabla}E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}) + 2\lambda\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{x}, \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

We have moreover

$$\|\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) - \boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}')\| \leq \left(\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{D}\|^2\right)\|\boldsymbol{x} - \boldsymbol{x}'\|, \quad \forall (\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{R}^N \times \mathbb{R}^N,$$

and hence $\boldsymbol{\nabla}\mathcal{F}$ is β-Lipschitz continuous, with $\beta \leq \gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{D}\|^2$. Since $\mathcal{H} = 0$, we can in this case solve (7.13) by means of the APGD method described in Algorithm 7.2. The resulting algorithm –with stopping criterion as

in Remark 7.4– is implemented in Algorithm 7.8. The step size τ can optimally be chosen as $\tau = 1/\beta$ or underestimated as

$$\tau = (\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{D}\|^2)^{-1} \leq 1/\beta,$$

which may be easier to compute in practice.

Algorithm 7.8: APGD method for solving (7.13) when F is smooth.

procedure APGD$(\boldsymbol{y}, \tau, \mathfrak{d}, \boldsymbol{x}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\boldsymbol{z}_n = \boldsymbol{x}_{n-1} - \tau\boldsymbol{G}^T\nabla E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}_{n-1}) - 2\tau\lambda\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{x}_{n-1}$
    $\boldsymbol{x}_n = \boldsymbol{z}_n + \frac{n-1}{n+\mathfrak{d}}(\boldsymbol{z}_n - \boldsymbol{z}_{n-1})$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon\|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

4.2 gTV Regularisation

Consider the (real) optimisation problem (6.26):

$$\text{Find } \boldsymbol{x}^\star \in \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^N} \{F(\boldsymbol{y}, \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda\|\boldsymbol{D}\boldsymbol{x}\|_1\}, \tag{7.14}$$

where $\boldsymbol{y} \in \mathbb{R}^L$, $F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proper convex* and lwsc, $\boldsymbol{G} \in \mathbb{R}^{L\times N}$, $\boldsymbol{D} \in \mathbb{R}^{N\times N}$, and $\lambda > 0$. We propose in the subsequent sections various algorithms for solving (7.14) depending on the nature of the cost functional F.

4.2.1 *Proximable Cost Functional*

Assume that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$ is *proximable* but not necessarily smooth. Then, we proceed as in [43, Section 5] and see (7.14) as a specific instance of (7.1), with

$$\mathcal{Y} = \mathbb{R}^L \times \mathbb{R}^N, \ \mathcal{F} = 0, \ \mathcal{G} = 0,$$

$$\mathcal{H} : \begin{cases} \mathbb{R}^L \times \mathbb{R}^N \to \mathbb{R} \\ (\boldsymbol{z}, \boldsymbol{x}) \mapsto E_{\boldsymbol{y}}(\boldsymbol{z}) + \lambda\|\boldsymbol{x}\|_1 \end{cases} \quad \text{and} \quad \boldsymbol{K} : \begin{cases} \mathbb{R}^N \to \mathbb{R}^L \times \mathbb{R}^N \\ \boldsymbol{x} \mapsto (\boldsymbol{G}\boldsymbol{x}, \boldsymbol{D}\boldsymbol{x}). \end{cases} \tag{7.15}$$

We can therefore solve (7.14), by specialising the generic Algorithm 7.1 to this particular setup. To this end, we need an expression for the proximal operator of $\mathcal{H}^*$ and the adjoint of $\boldsymbol{K}$. First, note that the convex conjugate of $\mathcal{H}$ in (7.15) is given by

$$\begin{aligned} \mathcal{H}^*(\boldsymbol{z}, \boldsymbol{x}) &= \sup_{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{R}^L\times\mathbb{R}^N} \langle(\boldsymbol{z}, \boldsymbol{x}), (\boldsymbol{u}, \boldsymbol{v})\rangle_{\mathbb{R}^L\times\mathbb{R}^N} - \mathcal{H}(\boldsymbol{u}, \boldsymbol{v}) \\ &= \sup_{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{R}^L\times\mathbb{R}^N} \langle\boldsymbol{z}, \boldsymbol{u}\rangle_{\mathbb{R}^L} + \langle\boldsymbol{x}, \boldsymbol{v}\rangle_{\mathbb{R}^N} - E_{\boldsymbol{y}}(\boldsymbol{u}) - \underbrace{\lambda\|\boldsymbol{v}\|_1}_{:=J(\boldsymbol{v})} \end{aligned}$$

$$
\begin{aligned}
&= \left(\sup_{\boldsymbol{u} \in \mathbb{R}^L} \langle \boldsymbol{z}, \boldsymbol{u} \rangle_{\mathbb{R}^L} - E_{\boldsymbol{y}}(\boldsymbol{u}) \right) + \left(\sup_{\boldsymbol{v} \in \mathbb{R}^N} \langle \boldsymbol{x}, \boldsymbol{v} \rangle_{\mathbb{R}^N} - J(\boldsymbol{v}) \right) \\
&= E_{\boldsymbol{y}}^*(\boldsymbol{z}) + J^*(\boldsymbol{x}), \qquad \forall (\boldsymbol{z}, \boldsymbol{x}) \in \mathbb{R}^L \times \mathbb{R}^N.
\end{aligned}
$$

Since $\mathcal{H}^*$ is separable across the two variables $\boldsymbol{z}$ and $\boldsymbol{x}$, its proximal is easily obtained by (see [134, Section 2.1])

$$
\begin{aligned}
\mathbf{prox}_{\sigma \mathcal{H}^*}(\boldsymbol{z}, \boldsymbol{x}) &= (\mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*}(\boldsymbol{z}), \mathbf{prox}_{\sigma J^*}(\boldsymbol{x})) \\
&= (\mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*}(\boldsymbol{z}), \boldsymbol{x} - \sigma\, \mathbf{soft}_{\lambda/\sigma}(\boldsymbol{x}/\sigma)), \qquad \forall (\boldsymbol{z}, \boldsymbol{x}) \in \mathbb{R}^L \times \mathbb{R}^N,
\end{aligned}
$$

where the last equality results from Moreau's identity (7.7) and (7.12).

Next, we compute the adjoint of $\boldsymbol{K}$. We have

$$
\begin{aligned}
\langle \boldsymbol{K}\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v}) \rangle_{\mathbb{R}^L \times \mathbb{R}^N} &= \langle (\boldsymbol{G}\boldsymbol{x}, \boldsymbol{D}\boldsymbol{x}), (\boldsymbol{u}, \boldsymbol{v}) \rangle_{\mathbb{R}^L \times \mathbb{R}^N} \\
&= \langle \boldsymbol{G}\boldsymbol{x}, \boldsymbol{u} \rangle_{\mathbb{R}^L} + \langle \boldsymbol{D}\boldsymbol{x}, \boldsymbol{v} \rangle_{\mathbb{R}^N} \\
&= \left\langle \boldsymbol{x}, \boldsymbol{G}^T\boldsymbol{u} \right\rangle_{\mathbb{R}^N} + \left\langle \boldsymbol{x}, \boldsymbol{D}^T\boldsymbol{v} \right\rangle_{\mathbb{R}^N} \\
&= \left\langle \boldsymbol{x}, \boldsymbol{G}^T\boldsymbol{u} + \boldsymbol{D}^T\boldsymbol{v} \right\rangle_{\mathbb{R}^N}, \quad \forall (\boldsymbol{x}, (\boldsymbol{u}, \boldsymbol{v})) \in \mathbb{R}^N \times \left(\mathbb{R}^L \times \mathbb{R}^N \right),
\end{aligned}
$$

and hence from the definition of the adjoint we get

$$
\boldsymbol{K}^*(\boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{G}^T\boldsymbol{z} + \boldsymbol{D}^T\boldsymbol{x}, \qquad \forall (\boldsymbol{z}, \boldsymbol{x}) \in \mathbb{R}^L \times \mathbb{R}^N.
$$

Using these ingredients, we can finally implement the PDS algorithm for this specific setup, provided in Algorithm 7.9. Note that the update steps 7 and 8 of the dual variables in Algorithm 7.9 are independent, and can hence be executed in parallel. The convergence condition L1 becomes in this case

$$
\sigma\tau \|\boldsymbol{K}\|^2 \leq 1,
$$

where we have

$$
\|\boldsymbol{K}\|^2 = \|\boldsymbol{K}^*\boldsymbol{K}\| = \left\|\boldsymbol{G}^T\boldsymbol{G} + \boldsymbol{D}^T\boldsymbol{D}\right\| \leq \|\boldsymbol{G}\|^2 + \|\boldsymbol{D}\|^2.
$$

The step sizes τ and σ can hence be set according to the rule of thumb (7.9) as

$$
\tau = \sigma = \frac{1}{\sqrt{\|\boldsymbol{G}^T\boldsymbol{G} + \boldsymbol{D}^T\boldsymbol{D}\|}},
$$

or, for computational conveniency, underestimated as

$$
\tau = \sigma = \frac{1}{\sqrt{\|\boldsymbol{G}\|^2 + \|\boldsymbol{D}\|^2}}
$$

Finally, condition L2 of Theorem 7.2 tells us that the momentum parameter ρ should be chosen in the closed interval $[\epsilon, 2 - \epsilon]$, for some $\epsilon > 0$. For simplicity, we chose $\rho = 1$ in Algorithm 7.9.

Algorithm 7.9: PDS method for solving (7.14) when F is proximable but not necessarily smooth.

procedure $\mathrm{PDS}(\boldsymbol{y}, \tau, \sigma, \boldsymbol{x}_0, \boldsymbol{z}_0, \boldsymbol{v}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \tau \boldsymbol{G}^T \boldsymbol{z}_{n-1} - \tau \boldsymbol{D}^T \boldsymbol{v}_{n-1}$
    $\boldsymbol{u}_n = 2\boldsymbol{x}_n - \boldsymbol{x}_{n-1}$
    $\boldsymbol{z}_n = \mathbf{prox}_{\sigma E_{\boldsymbol{y}}^*}\left(\boldsymbol{z}_{n-1} + \sigma \boldsymbol{G}\boldsymbol{u}_n\right)$
    $\boldsymbol{v}_n = \boldsymbol{v}_{n-1} + \sigma \boldsymbol{D}\boldsymbol{u}_n - \sigma\, \mathbf{soft}_{\lambda/\sigma}\left(\frac{\boldsymbol{v}_{n-1}}{\sigma} + \boldsymbol{D}\boldsymbol{u}_n\right)$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon \|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

4.2.2 *Smooth Cost Functional*

Assume this time that the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}$ is *differentiable* with γ-Lipschitz continuous derivative. Then, (7.14) can be seen as a specific instance of (7.1), with

$$\mathcal{X} = \mathcal{Y} = \mathbb{R}^N,\ \mathcal{F}(\boldsymbol{x}) = E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}),\ \mathcal{G} = 0,\ \mathcal{H}(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_1,\ \boldsymbol{K} = \boldsymbol{D},$$

with

$$\boldsymbol{\nabla}\mathcal{F}(\boldsymbol{x}) = \boldsymbol{G}^T \boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^N, \quad \text{and} \quad \beta \leq \gamma \|\boldsymbol{G}\|^2.$$

We can therefore solve (7.14), by specialising the generic Algorithm 7.1 to this particular setup. The resulting algorithm –with stopping criterion as in Remark 7.3– is implemented in Algorithm 7.10. The convergence condition K1 from Theorem 7.1 becomes in this case

$$\frac{1}{\tau} - \sigma\|\boldsymbol{D}\|^2 \geq \frac{\beta}{2},$$

and the rule of thumb (7.8) for setting the step sizes τ and σ yields

$$\tau = \sigma = \frac{1}{\|\boldsymbol{D}\|^2}\left(-\frac{\beta}{4} + \sqrt{\frac{\beta^2}{16} + \|\boldsymbol{D}\|^2}\right).$$

For computational conveniency, we can underestimate the step sizes by replacing β by $\gamma\|\boldsymbol{G}\|^2$ in the above equation. Finally, condition K2 of Theorem 7.1 tells us that for such step sizes, the momentum parameter ρ should be chosen in the open interval $]0, 1[$.

4.3 Matrix-free Formulation

Observe that Algorithms 7.7 to 7.10 perform at each iteration multiple matrix-vector multiplications involving the discrete pseudo-differential operator $\boldsymbol{D}$. If done naively, this operation can be quite *computationally* and *memory* intensive since $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ and the resolution N of the point set can be quite large in practice. Hopefully, this operation can be performed efficiently when the

Algorithm 7.10: PDS method for solving (7.14) when F is smooth.

procedure PDS$(\boldsymbol{y}, \tau, \sigma, \rho, \boldsymbol{x}_0, \boldsymbol{z}_0, \epsilon)$
  $n = 0$
  repeat
    $n \leftarrow n + 1$
    $\tilde{\boldsymbol{x}}_n = \boldsymbol{x}_{n-1} - \tau \boldsymbol{G}^T \boldsymbol{\nabla} E_{\boldsymbol{y}}(\boldsymbol{G}\boldsymbol{x}_{n-1}) - \tau \boldsymbol{D}^T \boldsymbol{z}_{n-1}$
    $\boldsymbol{u}_n = 2\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_{n-1}$
    $\tilde{\boldsymbol{z}}_n = \boldsymbol{z}_{n-1} + \sigma \boldsymbol{D}\boldsymbol{u}_n - \sigma\, \mathbf{soft}_{\lambda/\sigma}\left(\frac{\boldsymbol{z}_{n-1}}{\sigma} + \boldsymbol{D}\boldsymbol{u}_n\right)$
    $\boldsymbol{x}_n = \rho\tilde{\boldsymbol{x}}_n + (1-\rho)\boldsymbol{x}_{n-1}$
    $\boldsymbol{z}_n = \rho\tilde{\boldsymbol{z}}_n + (1-\rho)\boldsymbol{z}_{n-1}$
  until $\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \leq \epsilon \|\boldsymbol{x}_{n-1}\|$
  return $\boldsymbol{x}_n$

pseudo-differential operator $\boldsymbol{D}$ takes the form of an *order-K polynomial* in terms of the *sparse Laplacian* $\boldsymbol{L}$ (see Section 2.3):

$$\boldsymbol{D} = \sum_{k=0}^{K} \theta_k \boldsymbol{L}^k.$$

In which case, the matrix-vector multiplication $\boldsymbol{D}\boldsymbol{x}$ (respectively $\boldsymbol{D}^T\boldsymbol{x}$) can be implemented as a cascade of multiplications between the *sparse* matrix $\boldsymbol{L}$ and the vector $\boldsymbol{x}$. In particular, if $\boldsymbol{z}_0 = \boldsymbol{x}$ and $\boldsymbol{x}_0 = \theta_0 \boldsymbol{x}$, then the output vector $\tilde{\boldsymbol{x}} = \boldsymbol{D}\boldsymbol{x}$ is given by the outcome $\boldsymbol{x}_K$ of the following recursion:

$$\begin{cases} \boldsymbol{z}_k &= \boldsymbol{L}\boldsymbol{z}_{k-1} \\ \boldsymbol{x}_k &= \boldsymbol{x}_{k-1} + \theta_k \boldsymbol{z}_k \end{cases}, \qquad k = 1, \ldots, K. \tag{7.16}$$

Such an implementation is said *matrix-free* [51] since it does not require forming nor storing the large matrix $\boldsymbol{D}$, but rather rely on sparse matrix-vector multiplications, which can be implemented in a memory and computationally efficient manner. As recommended in [139, 159], we consider for stability reasons an equivalent version of (7.16) provided in Algorithm 7.11. The weights $\{\tilde{\theta}_0, \ldots, \tilde{\theta}_K\} \subset \mathbb{R}$ in Algorithm 7.11 are such that

$$\sum_{k=0}^{K} \theta_k \boldsymbol{L}^k = \sum_{k=0}^{K} \tilde{\theta}_k T_k(\tilde{\boldsymbol{L}}),$$

where $T_k : [-1, 1] \to \mathbb{R}$ are *Chebyshev polynomials* and $\tilde{\boldsymbol{L}}$ is the Laplacian with rescaled and shifted spectrum in the interval $[-1, 1]$ [139]:

$$\tilde{\boldsymbol{L}} = \frac{2}{\lambda_{max}} \boldsymbol{L} - \boldsymbol{I}.$$

Note that steps 2, 3 and 6 of Algorithm 7.11 result from the recursion formula defining Chebyshev polynomials: $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, with $T_1(x) = x$

Algorithm 7.11: Sparse implementation of a matrix-vector product $\boldsymbol{D}\boldsymbol{x}$ involving a polynomial discrete pseudo-differential operator of the form $\boldsymbol{D} = \sum_{k=0}^{K} \tilde{\theta}_k T_k(\tilde{\boldsymbol{L}})$.

procedure FASTMULT($\tilde{\boldsymbol{L}}, K, \{\tilde{\theta}_1, \dots, \tilde{\theta}_K\}, \boldsymbol{x}$)
  $\boldsymbol{x}_0 = \boldsymbol{z}_0 = \boldsymbol{x}$
  $\boldsymbol{z}_1 = \tilde{\boldsymbol{L}}\boldsymbol{x}$
  $\boldsymbol{x}_1 = \tilde{\theta}_1 \boldsymbol{z}_1 + \tilde{\theta}_0 \boldsymbol{z}_0$
  for k=2,..., K **do**
    $\boldsymbol{z}_k = 2\tilde{\boldsymbol{L}}\boldsymbol{z}_{k-1} - \boldsymbol{z}_{k-2}$
    $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \tilde{\theta}_k \boldsymbol{z}_k$
  return $\boldsymbol{x}_K$ ▷ We have $\boldsymbol{x}_K = \boldsymbol{D}\boldsymbol{x}$.

and $T_0(x) = 1$.

5 Proximal Operators of Common Cost Functionals

In this section, we provide examples of common cost functionals which are *proximable* but non-smooth, and explain in which context they are used in practice. We moreover derive their proximal operators, allowing their use in Algorithms 7.3 to 7.10. Table 7.1 summarises the results of this section.

5.1 Exact Match

Consider the data-fidelity functional:

$$F(\boldsymbol{y}, \boldsymbol{z}) := \iota(\boldsymbol{z} - \boldsymbol{y}), \quad \forall \boldsymbol{z} \in \mathbb{R}^L,$$

where $\boldsymbol{y} \in \mathbb{R}^L$ and $\iota : \mathbb{R}^L \to \{0, +\infty\}$ defined in (5.33). This functional enforces an *exact match* between the predicted and observed samples, as required in the *interpolation problems* explored in Section 4 of Chapter 5. Such a functional is mainly useful in the context of noiseless data as it can lead to serious overfitting issues in the presence of noise. Its proximal operator is given, for all $\tau > 0$, by

$$\begin{aligned} \mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}) &= \arg\min_{\boldsymbol{x} \in \mathbb{C}^L} \iota(\boldsymbol{x} - \boldsymbol{y}) + \frac{1}{2\tau}\|\boldsymbol{z} - \boldsymbol{x}\|_{\mathbb{R}^L}^2 \\ &= \boldsymbol{y}, \qquad \forall \boldsymbol{z} \in \mathbb{R}^L, \end{aligned} \tag{7.17}$$

since $\iota(\boldsymbol{z} - \boldsymbol{y})$ is unbounded for every $\boldsymbol{z} \neq \boldsymbol{y}$.

5.2 ℓ_1-norm

Consider the data-fidelity functional:

$$F(\boldsymbol{y}, \boldsymbol{z}) := \|\boldsymbol{z} - \boldsymbol{y}\|_1, \quad \forall \boldsymbol{z} \in \mathbb{R}^L,$$

where $\boldsymbol{y} \in \mathbb{R}^L$ and $\|\cdot\|_1 : \mathbb{R}^L \to \mathbb{R}_+$ denotes the discrete ℓ_1-norm. This functional leads to *sparse* residuals, with *most* of the predicted samples match-

Name	$F(\boldsymbol{y},\boldsymbol{z}),\ \boldsymbol{y},\boldsymbol{z}\in\mathbb{R}^L$	$\mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}),\ \tau>0,\ \boldsymbol{z}\in\mathbb{R}^L$	Useful for
Exact Match	$\iota(\boldsymbol{z}-\boldsymbol{y})$	$\boldsymbol{y}$	Noiseless data, interpolation.
ℓ_1**-Norm**	$\|\boldsymbol{z}-\boldsymbol{y}\|_1$	$\mathbf{soft}_\tau(\boldsymbol{z}-\boldsymbol{y})+\boldsymbol{y}$	Strong outliers and heavy-tailed noise distributions.
ℓ_2**-Ball**	$\iota_{\mathcal{B}_{2,\epsilon}}(\boldsymbol{z}-\boldsymbol{y}),\ \epsilon>0$	$\epsilon\frac{\boldsymbol{z}-\boldsymbol{y}}{\|\boldsymbol{z}-\boldsymbol{y}\|}+\boldsymbol{y}$	Gaussian noise with known noise level.
ℓ_∞**-Norm**	$\|\boldsymbol{z}-\boldsymbol{y}\|_\infty$	See [134, Section 6.5.2]	Quantisation noise and compact noise distributions.
Generalised KL-Divergence	$\sum_{i=1}^L y_i\log\left(\frac{y_i}{z_i}\right)-y_i+z_i$ $\forall\boldsymbol{z},\boldsymbol{y}\in\mathbb{R}_+^L$	$\frac{1}{2}(\boldsymbol{z}-\tau+\sqrt{(\boldsymbol{z}-\tau)^2+4\boldsymbol{y}\tau})$	Count data with Poisson noise.

Table 7.1: Common data-fidelity functionals and their associated proximal operators.

ing exactly the observed samples, and a few –potentially large– misfits [128]. Such a functional is particularly useful in the context of *strong outliers* [4, 128], or more generally for noise distributions with *heavy tails*,[7] templated by the Laplace distribution [146]. Using the precomposition property of proximal operators and the know proximal operator of the ℓ_1-norm (7.12), it is easy to show that its proximal operator is given, for all $\tau>0$, by

[7] In imaging, we speak of *salt-and-pepper* noise.

$$\begin{aligned}\mathbf{prox}_{\tau F(\cdot,\boldsymbol{y})}(\boldsymbol{z}) &= \arg\min_{\boldsymbol{x}\in\mathbb{C}^L}\ \|\boldsymbol{x}-\boldsymbol{y}\|_1\ +\ \frac{1}{2\lambda}\|\boldsymbol{z}-\boldsymbol{x}\|_{\mathbb{R}^L}^2\\ &= \mathbf{soft}_\tau\left(\boldsymbol{z}-\boldsymbol{y}\right)\ +\ \boldsymbol{y},\qquad \forall\boldsymbol{z}\in\mathbb{R}^L.\end{aligned}\tag{7.18}$$

5.3 ℓ_2-ball

Consider the data-fidelity functional:

$$F(\boldsymbol{y},\boldsymbol{z}):=\iota_{\mathcal{B}_{2,\epsilon}}(\boldsymbol{z}-\boldsymbol{y}),\quad \forall\boldsymbol{z}\in\mathbb{R}^L,\tag{7.19}$$

where $\boldsymbol{y}\in\mathbb{R}^L$ and $\iota_{\mathcal{B}_{2,\epsilon}}:\mathbb{R}^L\to\mathbb{R}_+$ denotes the indicator function of the ℓ_2-ball on $\mathbb{R}^L$ with radius $\epsilon>0$:

$$\iota_{\mathcal{B}_{2,\epsilon}}(\boldsymbol{x})=\begin{cases}0 & \text{if } \|\boldsymbol{x}\|_{\mathbb{R}^L}\le\epsilon\\ +\infty & \text{otherwise,}\end{cases}\qquad\forall\boldsymbol{x}\in\mathbb{R}^L.\tag{7.20}$$

Such a functional is particularly useful in the context of *Gaussian white noise* with known standard deviation $\sigma>0$. Indeed, assume that $\boldsymbol{y}\sim\mathcal{N}(\tilde{\boldsymbol{y}},\sigma^2\boldsymbol{I}_L)$ with $\tilde{\boldsymbol{y}}\in\mathbb{R}^L$. Then, we have

$$\frac{\|\boldsymbol{y}-\tilde{\boldsymbol{y}}\|_{\mathbb{R}^L}^2}{\sigma^2}=\sum_{k=1}^L\frac{(y_i-\tilde{y}_i)^2}{\sigma^2}\sim\chi^2(L),$$

and hence

$$\mathbb{P}_{\boldsymbol{y}}(\{\boldsymbol{z}\in\mathbb{R}^L:\|\boldsymbol{y}-\boldsymbol{z}\|_{\mathbb{R}^L}^2\le\sigma^2Q_{\chi^2(L)}(1-\alpha)\}\ni\tilde{\boldsymbol{y}}\})$$

$$
\begin{aligned}
&= \mathbb{P}_{\boldsymbol{y}}(\{\boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{y} - \boldsymbol{z}\|_{\mathbb{R}^L} \leq \sigma\sqrt{Q_{\chi^2(L)}(1-\alpha)}\} \ni \tilde{\boldsymbol{y}}\}) \\
&= 1 - \alpha,
\end{aligned}
$$

and hence, if $\epsilon = \sigma\sqrt{Q_{\chi^2(L)}(1-\alpha)}$, the support of $\iota_{\mathcal{B}_{2,\epsilon}}(\cdot - \boldsymbol{y})$ has a probability $1-\alpha$ of containing the true mean $\tilde{\boldsymbol{y}} \in \mathbb{R}^L$. Using the precomposition property of proximal operators and the fact that the proximal operator of a convex set indicator function is the convex set orthogonal projection operator [134, Section 1.2], it is easy to show that the proximal operator of (7.19) is given, for all $\tau > 0$, by

$$
\mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}) = \epsilon\frac{\boldsymbol{z} - \boldsymbol{y}}{\|\boldsymbol{z} - \boldsymbol{y}\|_{\mathbb{R}^L}} + \boldsymbol{y}, \qquad \forall \boldsymbol{z} \in \mathbb{R}^L. \tag{7.21}
$$

5.4 ℓ_∞-norm

Consider the data-fidelity functional:

$$
F(\boldsymbol{y}, \boldsymbol{z}) := \|\boldsymbol{z} - \boldsymbol{y}\|_\infty = \max_{i=1,\dots,L} |z_i - y_i|, \quad \forall \boldsymbol{z} \in \mathbb{R}^L,
$$

where $\boldsymbol{y} \in \mathbb{R}^L$ and $\|\cdot\|_\infty : \mathbb{R}^L \to \mathbb{R}_+$ denotes the discrete ℓ_∞-norm. Such a functional is particularly useful in the context of *quantisation noise* [26], or more generally noise distributions with *compact support*, templated by the uniform distribution [146]. Using the precomposition property of proximal operators and the known proximal operator of the ℓ_∞-norm [134, Chapter 6], it is easy to show that its proximal operator is given, for all $\tau > 0$, by

$$
\begin{aligned}
\mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}) &= \arg\min_{\boldsymbol{x} \in \mathbb{R}^L} \|\boldsymbol{x} - \boldsymbol{y}\|_\infty + \frac{1}{2\tau}\|\boldsymbol{z} - \boldsymbol{x}\|^2_{\mathbb{R}^L} \\
&= \mathbf{prox}_{\tau\|\cdot\|_\infty}(\boldsymbol{z} - \boldsymbol{y}) + \boldsymbol{y}, \quad \forall \boldsymbol{z} \in \mathbb{R}^L.
\end{aligned} \tag{7.22}
$$

The proximal operator $\text{prox}_{\tau\|\cdot\|_\infty}$ does not admit a closed-form formula, but can however be evaluated very efficiently and with high accuracy (see [134, Section 6.5.2]).

5.5 Generalised Kullback-Leibler Divergence

Consider the data-fidelity functional:

$$
F(\boldsymbol{y}, \boldsymbol{z}) := D_{KL}(\boldsymbol{y}||\boldsymbol{z}) = \sum_{i=1}^{L} y_i \log\left(\frac{y_i}{z_i}\right) - y_i + z_i, \quad \forall \boldsymbol{z} \in \mathbb{R}^L_+, \tag{7.23}
$$

where $\boldsymbol{y} \in \mathbb{R}^L_+$ and $D_{KL}(\cdot||\cdot) : \mathbb{R}^L_+ \times \mathbb{R}^L_+ \to \mathbb{R}_+$ denotes the *generalised Kullback-Leibler (KL) divergence* [19, 20] for discrete positive vectors which do not necessarily sum to one.[8] In information theory, and in the case where $\mathbf{1}^T\boldsymbol{z} = \mathbf{1}^T\boldsymbol{y} = 1$ so that $\boldsymbol{z}$ and $\boldsymbol{y}$ can be interpreted as discrete probability distributions, the KL-divergence (7.23) can be interpreted as the *relative entropy* of $\boldsymbol{y}$ with respect to $\boldsymbol{z}$, i.e. the amount of information lost when using $\boldsymbol{z}$ to approximate $\boldsymbol{y}$. It is particularly useful in the context of *count data* with Poisson distribution

[8] Notice that if the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ represent discrete probability density functions, then the last two terms of (7.23) cancel out and we get back the traditional Kullback-Leibler divergence.

[19, 20], as encountered in positron emission tomography for example [161]. Indeed, (7.23) corresponds –up to an additive constant– to the likelihood of the data $\boldsymbol{y}$ where each component is independent with Poisson distribution and respective intensities given by the entries of $\boldsymbol{z}$. Its proximal operator is given, for all $\tau > 0$, by

$$\mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}\in\mathbb{R}_+^L} \sum_{i=1}^{L} y_i \log\left(\frac{y_i}{x_i}\right) - y_i + x_i \; + \; \frac{1}{2\tau}(z_i - x_i)^2 \tag{7.24}$$

Notice that (7.24) is the sum of L independent objective functionals, which can each be independently minimised by solving an optimisation problem of the form:

$$\hat{x} = \arg\min_{x>0} \quad y \log\left(\frac{y}{x}\right) - y + x \; + \; \frac{1}{2\tau}(z - x)^2, \tag{7.25}$$

for $y, z \geq 0$. Using the postcomposition, precomposition and affine addition properties of proximal operators and the known proximal operator of the *log-barrier function* in [134, Section 6.1.3] we find:

$$\hat{x} = \frac{z - \tau + \sqrt{(z-\tau)^2 + 4y\tau}}{2},$$

and finally

$$\mathbf{prox}_{\tau F(\boldsymbol{y},\cdot)}(\boldsymbol{z}) = \frac{\boldsymbol{z} - \tau + \sqrt{(\boldsymbol{z}-\tau)^2 + 4\boldsymbol{y}\tau}}{2}. \tag{7.26}$$

Name	Regularisation	Cost Functional	Discretisation Scheme	Type	Convergence Condition on τ, (σ)	Default Values of Hyperparameters τ, (σ), (ρ), $(\mathfrak{d})$
Algorithm 7.3	gTikhonov	Proximable	Search Space (7.10)	PDS	$\frac{1}{\tau} - \sigma\|\boldsymbol{G}\|^2 \geq \lambda\|\boldsymbol{G}\|$	$\tau = \sigma = \frac{1}{\|\boldsymbol{G}\|}\left(-\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4}+1}\right)$ $\rho = 0.4$
Algorithm 7.4	gTikhonov	Smooth	Search Space (7.10)	APGD	$\tau \leq 1/\beta$ $\left(\beta \leq \gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{G}\|\right)$	$\tau = \left(\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{G}\|\right)^{-1}$, $\mathfrak{d} = 75$
Algorithm 7.5	gTV	Proximable	Search Space (7.11)	PDS	$\sigma\tau\|\boldsymbol{G}\|^2 \leq 1$	$\tau = \sigma = \|\boldsymbol{G}\|^{-1}$
Algorithm 7.6	gTV	Smooth	Search Space (7.11)	APGD	$\tau \leq 1/\beta$ $\left(\beta \leq \gamma\|\boldsymbol{G}\|^2\right)$	$\tau = \left(\gamma\|\boldsymbol{G}\|^2\right)^{-1}$, $\mathfrak{d} = 75$
Algorithm 7.7	gTikhonov	Proximable	Domain (7.13)	PDS	$\frac{1}{\tau} - \sigma\|\boldsymbol{G}\|^2 \geq \lambda\|\boldsymbol{D}\|^2$	$\tau = \sigma = \frac{\|\boldsymbol{D}\|^2}{\|\boldsymbol{G}\|^2}\left(-\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + \frac{\|\boldsymbol{G}\|^2}{\|\boldsymbol{D}\|^4}}\right)$ $\rho = 0.4$
Algorithm 7.8	gTikhonov	Smooth	Domain (7.13)	APGD	$\tau \leq 1/\beta$ $\left(\beta \leq \gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{D}\|^2\right)$	$\tau = \left(\gamma\|\boldsymbol{G}\|^2 + 2\lambda\|\boldsymbol{D}\|^2\right)^{-1}$, $\mathfrak{d} = 75$
Algorithm 7.9	gTV	Proximable	Domain (7.14)	PDS	$\sigma\tau\|\boldsymbol{K}\|^2 \leq 1$, $\boldsymbol{K} : \boldsymbol{x} \mapsto (\boldsymbol{G}\boldsymbol{x}, \boldsymbol{D}\boldsymbol{x})$	$\tau = \sigma = \left(\sqrt{\|\boldsymbol{G}\|^2 + \|\boldsymbol{D}\|^2}\right)^{-1}$
Algorithm 7.10	gTV	Smooth	Domain (7.14)	PDS	$\frac{1}{\tau} - \sigma\|\boldsymbol{D}\|^2 \geq \frac{\beta}{2}$ $\left(\beta \leq \gamma\|\boldsymbol{G}\|^2\right)$	$\tau = \sigma = \frac{1}{\|\boldsymbol{D}\|^2}\left(-\frac{\gamma\|\boldsymbol{G}\|^2}{4} + \sqrt{\frac{\gamma^2\|\boldsymbol{G}\|^4}{16} + \|\boldsymbol{D}\|^2}\right)$ $\rho = 0.4$

Table 7.2: Summary of the various algorithms of Chapter 7. In the third column of this table, *smooth* means differentiable and with Lipschitz-continuous derivative, while *proximable* means that the proximal operator can be efficiently evaluated. The parameters between parentheses in the last two columns are specific to certain algorithms only. The Lipschitz constant of the smooth term of the optimisation problem solved by each algorithm is denoted by $\beta > 0$. Depending on the case, the smooth term can be composed of the cost functional, the regularisation term, or the sum of the two terms. In each case, we give convenient upper bounds on β for practical purposes. When the cost functional $E_{\boldsymbol{y}} = F(\boldsymbol{y}, \cdot) : \mathbb{R}^L \to \mathbb{R}$ is smooth, its Lipschitz constant is denoted by γ. Our practical implementations of Algorithms 7.3 to 7.10 set the hyperparameters to their default values, obtained for PDS from the rule of thumbs (7.8) and (7.9), and provided in the last column of this table.

133

8

Practical Spherical Splines

One key insight of Theorem 5.4 is that the solutions of the FPBP problem (5.23) are $\mathscr{D}$-splines. As such, they inherit all their analytical properties from the zonal Green kernel associated to the pseudo-differential operator $\mathscr{D}$ used in the gTV regularisation term. To a lesser extent, this is also true for the FPT problem (5.16) whose unique solution can be expressed as a linear combination of the sampling functionals convolved twice with the zonal Green kernel of $\mathscr{D}$ (see Theorem 5.3). For practical purposes, it is hence important to choose the pseudo-differential operator in agreement with the desired analytical properties of the solution(s). In Example 4.1, we have for example introduced Sobolev operators $\mathscr{D}_\beta := [\mathrm{Id} - \Delta_{\mathbb{S}^{d-1}}]^\beta$, whose associated zonal Green kernels reproduce, for $\beta > (d-1)/2$ (see Lemma 5.5), the *Sobolev spaces*

$$\mathscr{H}^\beta(\mathbb{S}^{d-1}) = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sum_{n\in\mathbb{N}} (1+n(n+d-2))^\beta \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2 < +\infty \right\}.$$

The latter are nested RKHSs,

$$\mathscr{H}^\gamma(\mathbb{S}^{d-1}) \subset \mathscr{H}^\beta(\mathbb{S}^{d-1}) \subset \mathscr{L}^2(\mathbb{S}^{d-1}), \quad \forall \gamma \geq \beta > \frac{d-1}{2},$$

containing functions with β-increasing degrees of smoothness. For example, a function $f \in \mathscr{H}^\beta(\mathbb{S}^{d-1})$ for $\beta \in \mathbb{N}$ is differentiable up to order β, with all its derivatives up to that order square-integrable. Sobolev operators seem hence particularly well-suited to enforce a certain degree of smoothness in the solutions of FPT or FPBP problems. Unfortunately, the Sobolev zonal Green kernel, given from (4.11) by

$$\psi_\beta(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \sum_{n=0}^{+\infty} \frac{N_d(n)}{\mathfrak{a}_d \left(1+n(n+d-2)\right)^\beta} P_{n,d}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle), \qquad \forall \boldsymbol{r}, \boldsymbol{s} \in \mathbb{S}^{d-1}, \quad (8.1)$$

does not admit a convenient closed-form expression, making Sobolev operators –and consequently Sobolev splines– very cumbersome to work with in practice. In this chapter, we hence discusss two kernels, named *Matérn* and *Wendland*

kernels, with similar smoothness properties as the Sobolev kernel (8.1), but much better suited for practical purposes since expressible in terms of simple functions and well-localised in space. Both Matérn and Wendland kernels are obtained from *restrictions* of scaled *radial kernels* to the hypersphere [64, 104]:

$$\begin{aligned}\psi_\beta^\epsilon(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) &= \frac{1}{\epsilon^n}\Psi_\beta\left(\epsilon^{-1}\|\boldsymbol{r}-\boldsymbol{s}\|_{\mathbb{R}^d}\right) \\ &= \frac{1}{\epsilon^n}\Psi_\beta\left(\epsilon^{-1}\sqrt{2-2\langle \boldsymbol{r}, \boldsymbol{s}\rangle}\right), \quad \forall (\boldsymbol{r},\boldsymbol{s}) \in \mathbb{S}^{d-1}\times\mathbb{S}^{d-1}, \end{aligned} \tag{8.2}$$

where $0 < \epsilon \leq 1$ is a *scale parameter* and $\Psi_\beta : \mathbb{R}_+ \to \mathbb{R}$, $\beta > (d-1)/2$, is such that the kernel $\Psi_\beta(\|\boldsymbol{x}-\boldsymbol{y}\|_{\mathbb{R}^d})$, $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d$, reproduces the Euclidean Sobolev space $\mathscr{H}^{\beta+1/2}(\mathbb{R}^d)$ (see [104, Section 2] for more details, and Sections 1 and 2 for examples of such radial basis functions).

The resulting kernels (8.2) are zonal by construction, with Fourier-Legendre coefficients $\{\hat{\psi}_\beta^\epsilon[n]\}_{n\in\mathbb{N}}$ verifying [104, Section 2]:

$$c_1(1+\epsilon n)^{-2\beta} \leq \hat{\psi}_\beta^\epsilon[n] \leq c_2(1+\epsilon n)^{-2\beta}, \quad \forall n \geq 0. \tag{8.3}$$

From (8.3), we deduce that $\hat{\psi}_\beta^\epsilon[n] > 0$, $\forall n \in \mathbb{N}$, and $\hat{\psi}_\beta^\epsilon[n] = \Theta(n^{-p})$ with $p = 2\beta > d-1$. Hence, the kernels (8.2) can moreover be interpreted (see Definition 4.1) as the zonal Green kernels of a family of spline-admissible[1] pseudo-differential operators given by:

$$\mathscr{D}_\beta^\epsilon : \begin{cases} \mathscr{S}(\mathbb{S}^{d-1}) \to \mathscr{S}(\mathbb{S}^{d-1}) \\ h \mapsto \mathscr{D}_\beta^\epsilon h := \displaystyle\sum_{n=0}^{+\infty} \frac{1}{\hat{\psi}_\beta^\epsilon[n]} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m Y_n^m\right]. \end{cases} \tag{8.4}$$

[1] Since $p = 2\beta > d-1$, we have indeed from Proposition 4.5 that $\mathscr{D}_\beta^\epsilon$ is spline-admissible.

Still thanks to (8.3), it is moreover possible to show [64, 104] that, for a given $\beta > (d-1)/2$, the norms $\|f\|_{\mathscr{D}_\beta^\epsilon} = \sum_{n\in\mathbb{N}} \hat{\psi}_\beta^\epsilon[n]^{-1} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2$, are all *equivalent* to the canonical *Sobolev norm*

$$\|f\|_{\mathscr{D}_\beta} = \sum_{n\in\mathbb{N}} (1+n(n+d-2))^\beta \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2.$$

The native RKHS

$$\mathscr{N}_{\mathscr{D}_\beta^\epsilon} = \left\{ f \in \mathscr{S}'(\mathbb{S}^{d-1}) : \sum_{n\in\mathbb{N}} \frac{1}{\hat{\psi}_\beta^\epsilon[n]} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2 < +\infty \right\},$$

contains therefore exactly the same elements as the Sobolev space $\mathscr{H}^\beta(\mathbb{S}^{d-1})$:

$$\mathscr{N}_{\mathscr{D}_\beta^\epsilon} = \mathscr{H}^\beta(\mathbb{S}^{d-1}), \qquad \forall \beta > (d-1)/2,\ \epsilon \in]0,1].$$

In conclusion, the zonal Green kernels ψ_β^ϵ in (8.2) reproduce the Sobolev space

$\mathscr{H}^\beta(\mathbb{S}^{d-1})$ when the latter is equipped with the inner product :

$$\langle h, g \rangle_{\mathscr{D}^\epsilon_\beta} = \sum_{n=0}^{+\infty} \frac{1}{\hat{\psi}^\epsilon_\beta[n]} \left[\sum_{m=1}^{N_d(n)} \hat{h}_n^m \overline{\hat{g}_n^m} \right],$$

and can hence be used as a replacement for the Sobolev zonal Green kernel to build practical spherical splines. In the subsequent sections, we give examples of kernels ψ^ϵ_β as in (8.2), called the *Matérn* and *Wendland* kernels, and plotted in Fig. 8.3.

Remark 8.1 — *About the Scale Parameter* ϵ. *For a fixed $\beta > (d-1)/2$, we have seen that the kernels ψ^ϵ_β for $0 < \epsilon \leq 1$ all reproduce the Sobolev space $\mathscr{H}^\beta(\mathbb{S}^{d-1})$. As such, one could question the relevancy of this parameter in the construction (8.2) proposed in [64, 104]. Such doubts are however dispelled when considering approximation errors made by projecting functions from $\mathscr{H}^\beta(\mathbb{S}^{d-1})$ into specific spline spaces $\mathfrak{S}_{\mathscr{D}^\epsilon_\beta}(\mathbb{S}^{d-1}, \Xi_M)$ with fixed knot sets $\Xi_M \subset \mathbb{S}^{d-1}$. Indeed, it was shown in [64, 104] that the approximation error is proportional to the quantity $(\Theta_{\Xi_M}/\epsilon)^\beta$, where $\Theta_{\Xi_M} > 0$ is the nodal width of Ξ_M defined in (6.7) page 95. As such, choosing ϵ at least as large as the nodal width Θ_M helps in reducing the approximation error.*

Remark 8.2 *The previous developments illustrate well the two dual ways in which splines can be built. The first approach, adopted in Chapter 4, consists of starting from a known pseudo-differential operator and computing its zonal Green kernel. The latter may however not admit a convenient closed-form expression, as is the case for the Sobolev zonal Green kernel (8.1) which can only be expressed as an infinite series. The second approach, adopted above, starts from a kernel with known analytical expression and shows that it corresponds indeed to the Green kernel of some pseudo-differential operator. The latter may however not be expressible in terms of standard pseudo-differential operators anymore as is the case in (8.4).*

1 Matérn Kernel and Matérn Spherical Splines

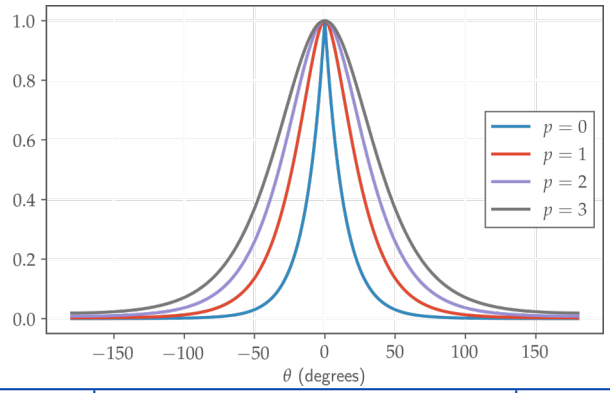
The *Matérn functions* $S^\epsilon_\nu : \mathbb{R}_+ \to \mathbb{R}$ are defined as [143, Chapter 4, p. 84]

$$S^\epsilon_\nu(r) := \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\epsilon}\right)^\nu K_\nu\left(\frac{r}{\epsilon}\right), \quad \forall r > 0,$$

where ν and ϵ are *nonnegative parameters*, Γ is the *Gamma function* and K_ν is the *modified Bessel function of the second kind* [1, Section 9.6]. For half integers $\nu = p + 1/2$ with $p \in \mathbb{N}$, it is possible to write the Matérn function as the product of an exponential and a polynomial of order p. We have notably:

- $\nu = 1/2$, $p = 0$:

$$S^\epsilon_{1/2}(r) = \exp\left(-\frac{r}{\epsilon}\right), \quad \forall r > 0,$$



$p \in \mathbb{N}$	$\nu = p + 1/2$	**Reproducing Zonal Kernel** $\psi^\epsilon_\beta(r(t))$ $r(t) = \sqrt{2-2t}$, $t \in [-1,1]$, $0 < \epsilon \leq 1$	β**-Sobolev Space** $\beta = \nu + (d-1)/2$	
			$d = 2$	$d = 3$
0	$1/2$	$\exp(-r/\epsilon)$	$\mathscr{H}^1(\mathbb{S}^1)$	$\mathscr{H}^{1.5}(\mathbb{S}^2)$
1	$3/2$	$[1 + r/\epsilon]\exp(-r/\epsilon)$	$\mathscr{H}^2(\mathbb{S}^1)$	$\mathscr{H}^{2.5}(\mathbb{S}^2)$
2	$5/2$	$[3 + 3r/\epsilon + r^2/\epsilon^2]\exp(-r/\epsilon)$	$\mathscr{H}^3(\mathbb{S}^1)$	$\mathscr{H}^{3.5}(\mathbb{S}^2)$
3	$7/2$	$[15 + 15r/\epsilon + 6r^2/\epsilon^2 + r^3/\epsilon^3]\exp(-r/\epsilon)$	$\mathscr{H}^4(\mathbb{S}^1)$	$\mathscr{H}^{4.5}(\mathbb{S}^2)$

Figure 8.1: Matérn reproducing kernels for various spherical Sobolev spaces, for $d = 2, 3$. The plots are for $\epsilon = 0.25$.

- $\nu = 3/2$, $p = 1$:

$$S^\epsilon_{3/2}(r) = \left[1 + \frac{r}{\epsilon}\right] \exp\left(-\frac{r}{\epsilon}\right), \quad \forall r > 0,$$

- $\nu = 5/2$, $p = 2$:

$$S^\epsilon_{5/2}(r) = \left[3 + \frac{3\,r}{\epsilon} + \frac{r^2}{\epsilon^2}\right] \exp\left(-\frac{r}{\epsilon}\right), \quad \forall r > 0,$$

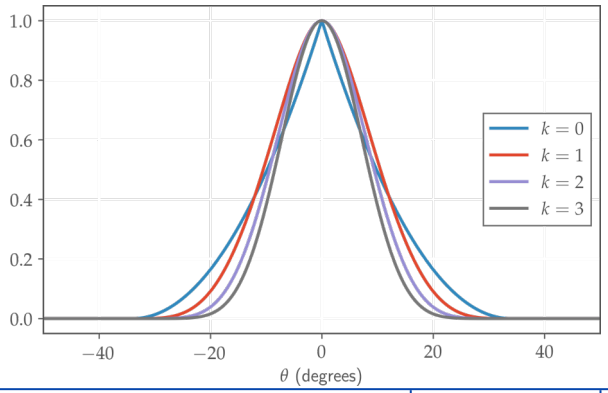
- $\nu = 7/2$, $p = 3$:

$$S^\epsilon_{7/2}(r) = \left[15 + \frac{15\,r}{\epsilon} + \frac{6\,r^2}{\epsilon^2} + \frac{r^3}{\epsilon^3}\right] \exp\left(-\frac{r}{\epsilon}\right), \quad \forall r > 0.$$

In the limit $\nu \to \infty$, the Matérn function converges towards the *Gaussian function* [143, Chapter 4, p. 84]:

$$S^\epsilon_\infty(r) = \exp\left(-\frac{r^2}{2\epsilon^2}\right), \quad \forall r > 0.$$

For practical purposes, $\nu \geq 7/2$ yield Matérn functions almost indistinguishable from the Gaussian function [143, Chapter 4, p. 84]. It was shown in [104] that the radial kernels $S^\epsilon_{\nu(\beta)}(\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^d})$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ reproduce the Sobolev spaces $\mathscr{H}^{\beta+1/2}(\mathbb{R}^d)$ for $\nu(\beta) = \beta - (d-1)/2$, $\beta > (d-1)/2$ and $0 < \epsilon \leq 1$. From [104, Lemma 2.1] and the above developments, the restriction of these radial



$k \in \mathbb{N}$	**Reproducing Zonal Kernel** $\psi^{\epsilon}_{3,k}(r(t))$ $r(t) = \sqrt{2-2t}$, $t \in [-1,1]$, $0 < \epsilon \leq 1$	β**-Sobolev Space** $\beta = 3/2 + k$	**Smoothness**
0	$(1-r/\epsilon)_+^2$	$\mathscr{H}^{1.5}(\mathbb{S}^2)$	$\mathscr{C}^0$
1	$(1-r/\epsilon)_+^4(1+4r/\epsilon)$	$\mathscr{H}^{2.5}(\mathbb{S}^2)$	$\mathscr{C}^2$
2	$(1-r/\epsilon)_+^6(3+18r/\epsilon+35r^2/\epsilon^2)$	$\mathscr{H}^{3.5}(\mathbb{S}^2)$	$\mathscr{C}^4$
3	$(1-r/\epsilon)_+^8(15+120r/\epsilon+375r^2/\epsilon^2+480r^3/\epsilon^3)$	$\mathscr{H}^{4.5}(\mathbb{S}^2)$	$\mathscr{C}^6$

Figure 8.2: Wendland reproducing kernels for various spherical Sobolev spaces and $d = 3$. The support of $\psi^{\epsilon}_{3,k}(r(t))$ is $r(t) \in [0, \epsilon]$. The plots are for $\epsilon = 0.6$.

kernels to $\mathbb{S}^{d-1}$ hence yields zonal kernels

$$\psi^{\epsilon}_{\beta}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = S^{\epsilon}_{\nu(\beta)}(\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}), \quad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \tag{8.5}$$

which reproduce the spherical Sobolev spaces $\mathscr{H}^{\beta}(\mathbb{S}^{d-1})$ for $\nu(\beta) = \beta - (d-1)/2$, $\beta > (d-1)/2$ and $0 < \epsilon \leq 1$. Fig. 8.1 lists the Matérn zonal Green kernels for various Sobolev spaces in the specific cases where $d = 2, 3$. Examples of *Matérn splines* are moreover plotted in Fig. 8.3.

2 Wendland Kernel and Wendland Spherical Splines

Wendland's functions $\phi_{d,k} : \mathbb{R}_+ \to \mathbb{R}$, $k \in \mathbb{N}$ are constructed by repeatedly applying and integral operator $\mathcal{I}$ to *Askey's truncated power functions* ϕ_l:

$$\phi_{d,k}(r) := (\mathcal{I}^k \phi_l)(r), \ k \in \mathbb{N}, \ l := \lfloor d/2 \rfloor + k + 1, \ \phi_l(r) := (1-r)_+^l, \ a_+ := \max(a, 0),$$

where $\mathcal{I}$ is given by :

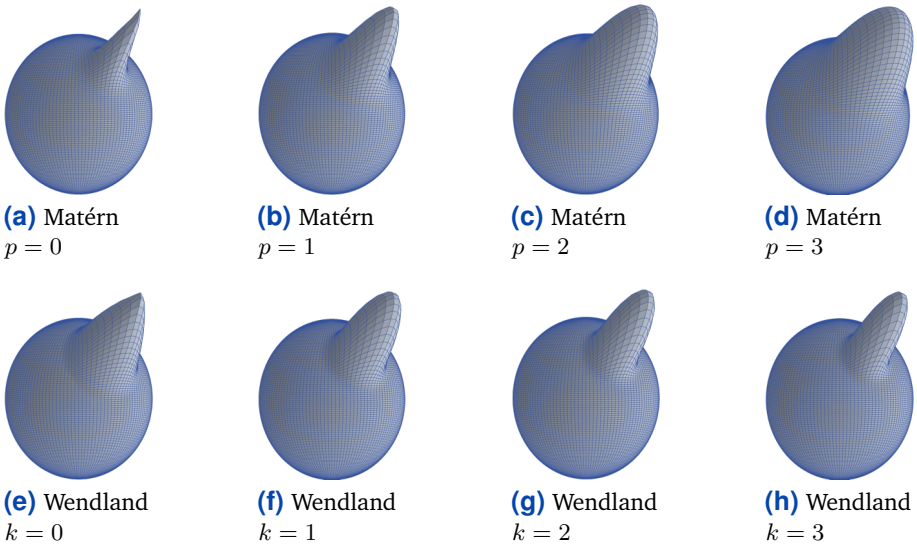
$$(\mathcal{I}\phi)(r) = \int_r^{+\infty} t\phi(t)\, dt, \quad r \geq 0.$$

It can be shown [201] that Wendland's functions can be represented as:

$$\phi_{d,k}(r) = (1-r)_+^{l+k} p_{k,l}(r), \quad r \geq 0,$$

where $p_{k,l}$ is a polynomial of degree k whose coefficients depend on l.

These functions are *compactly supported* piecewise polynomials with support



(a) Matérn $p = 0$ **(b)** Matérn $p = 1$ **(c)** Matérn $p = 2$ **(d)** Matérn $p = 3$

(e) Wendland $k = 0$ **(f)** Wendland $k = 1$ **(g)** Wendland $k = 2$ **(h)** Wendland $k = 3$

Figure 8.3: Figs. 8.3a to 8.3d: Matérn spline (8.5) for $\beta = p + 3/2$, $\epsilon = 0.1$, $p = 0, 1, 2, 3$ and focus direction $\boldsymbol{r} = (1, -1, 1)/\sqrt{3}$. Figs. 8.3e to 8.3h: Wendland spline (8.6) for $\beta = k + 3/2$, $\epsilon = 0.6$, $k = 0, 1, 2, 3$ and focus direction $\boldsymbol{r} = (1, -1, 1)/\sqrt{3}$.

$[0, 1]$ which yield positive definite radial kernels in $\mathbb{R}^d$ with minimal degree and prescribed smoothness [39, 201]. They have been introduced by Wendland [190] in the context of high-dimensional approximation/interpolation. For $d \geq 3$, Wenland's radial kernels $\phi_{d,k}(\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^d})$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ were moreover proven [39, 201] to reproduce Sobolev spaces of the form $\mathscr{H}^{k+(d+1)/2}(\mathbb{R}^d)$. A similar result was established in the spherical setup [64, 126], stating that restrictions of scaled Wendland's radial kernels to the sphere

$$\psi^{\epsilon}_{d,k}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \phi_{d,k}\left(\frac{\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}}{\epsilon}\right), \quad (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \tag{8.6}$$

yield zonal kernels reproducing spherical Sobolev spaces $\mathscr{H}^{k+d/2}(\mathbb{S}^{d-1})$ for $d \geq 3$. In the case $d = 2$, similar results can be obtained via a generalisation of Wendland's functions called the *missing Wendland's functions*[201]. These are however significantly more complicated to work with and will hence not be investigated in this work. Examples of Wendland zonal Green kernels and their associated RKHSs are provided in Figs. 8.2 and 8.3 for $d = 3$.

3 Computational Advantages of Matérn and Wendland Splines

Having rapide decay, the Matérn and Wendland kernels present multiple computational advantages in practice, listed in the subsequent sections.

3.1 Sparse Spline Synthesis

Notice that the pseudo-differential operators associated to the Matérn and Wendland zonal Green kernels are positive definite by (8.3). From Propo-

sition 4.6 we can hence write a Matérn or Wendland spline with knot set $\Xi_N = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\}$ and coefficients $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N] \in \mathbb{C}^N$ as

$$\mathfrak{s}(\boldsymbol{r}) = \sum_{n=1}^{N} \alpha_n \psi_\beta^\epsilon(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle) = \Psi(\boldsymbol{\alpha})(\boldsymbol{r}), \quad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{8.7}$$

where Ψ is the synthesis operator (see Definition 6.1) associated to the family of functions $\{\psi_\beta^\epsilon(\langle \cdot, \boldsymbol{r}_n \rangle),\ \boldsymbol{r}_n \in \Xi_N\}$, and ψ_β^ϵ denotes the Matérn or Wendland kernel respectively. From the rapid decays of the Matérn and Wendland kernels, we have moreover:

$$\psi_\beta^\epsilon(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle) \simeq 0, \quad \text{if} \quad \sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle} \geq R(\epsilon),$$

for some chord distance $R(\epsilon) \leq \sqrt{2}$, proportional to the scale parameter $\epsilon \in]0, 1]$. For the Wendland kernel (8.6) for example, we have $R(\epsilon) = \epsilon$:

$$\psi_\beta^\epsilon(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle) = 0, \quad \text{if} \quad \sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle} \geq \epsilon. \tag{8.8}$$

Therefore, if the scale parameter ϵ is chosen small enough, then the nth term of the summation in (8.7) will be zero outside of a local neighbourhood[2] of the knot $\boldsymbol{r}_n$. This fact can be leveraged in practice for designing efficient *sparse synthesis schemes* for Matérn and Wendland splines. As explained in Remark 6.6, this is particularly relevant for the gTV search space discretisation scheme from Theorem 6.5.

[2] This neighbourhood is defined as the points in $\mathbb{S}^{d-1}$ with a chord distance to $\boldsymbol{r}_n$ smaller than $R(\epsilon) \ll \sqrt{2}$.

3.2 Sparse Gram Matrices

In many experimental setups, the rapid decays of the Matérn and Wendland kernels cause the Gram matrices $\boldsymbol{G}$ in Theorems 6.2 and 6.5 to be *sparse*, allowing them to be conveniently implemented as such in the various algorithms of Chapter 7. For example, consider Theorem 6.5 in the context of the pseudo-differential operator associated to the Wendland kernel ψ_β^ϵ and spatial sampling functionals $\{\varphi_l = \delta_{\boldsymbol{\rho}_l}, l = 1, \dots, L\}$ with sampling directions $\{\boldsymbol{\rho}_l,\ l = 1, \dots, L\} \subset \mathbb{S}^{d-1}$. Then, the entries of the Gram matrix $\boldsymbol{G} \in \mathbb{R}^{L \times N}$ are given by

$$G_{ln} = \psi_\beta^\epsilon(\langle \boldsymbol{\rho}_l, \boldsymbol{r}_n \rangle), \qquad l = 1, \dots, L, \quad n = 1, \dots, N.$$

From (8.8) it is easy to see that, for ϵ small enough and the point sets $\{\boldsymbol{\rho}_l,\ l = 1, \dots, L\}$ and $\{\boldsymbol{r}_n,\ n = 1, \dots, N\}$ reasonably well distributed over $\mathbb{S}^{d-1}$, most of the entries of $\boldsymbol{G}$ are null. This behaviour extends to many spatially-localised measurement processes such as local averages or filtrations (see Chapter 9 for real-life examples).

3.3 Fast Spherical Convolution

In certain cases, the search space discretisation schemes considered in Theorems 6.2 and 6.5 require computing the spherical convolution between a zonal Green kernel $\psi_{\mathscr{D}}$ and some functions $\{\varphi_i,\ i = 1 \dots, L\} \subset \mathscr{L}^2(\mathbb{S}^{d-1})$.

When the zonal Green kernel has relatively small support –as is the case for Matérn and Wendland kernels– such a convolution can be implemented very efficiently using tools from Section 2 of Chapter 6 and Section 4.3 of Chapter 7. Indeed, consider an equidistributed spherical point set $\Theta = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\}$ as in Example 6.2. Then, provided there are a sufficient number of points N, we can approximate with high accuracy the Lebesgue measure μ on $\mathbb{S}^{d-1}$ by the measure $\nu_N = (1/N)\sum_{n=1}^N \delta_{\boldsymbol{r}_n}$ (since by definition of an equidistributed point set, the measure ν_N converges in the weak* sense towards μ). This yields $\forall f \in \mathscr{L}^2(\mathbb{S}^{d-1})$,

$$\begin{aligned} g(\boldsymbol{r}) = (\psi_{\mathscr{D}} * f)(\boldsymbol{r}) &= \int_{\mathbb{S}^{d-1}} \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) f(\boldsymbol{s}) \mu(d\boldsymbol{s}) \\ &\simeq \int_{\mathbb{S}^{d-1}} \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) f(\boldsymbol{s}) \nu_N(d\boldsymbol{s}) \\ &= \frac{1}{N}\sum_{n=1}^N \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_n\rangle) f(\boldsymbol{r}_n), \qquad \forall \boldsymbol{r} \in \mathbb{S}^{d-1}. \end{aligned}$$

Evaluating the convolution for all directions in the point set Θ gives us

$$g(\boldsymbol{r}_m) = \frac{1}{N}\sum_{n=1}^N \psi_{\mathscr{D}}(\langle \boldsymbol{r}_m, \boldsymbol{r}_n\rangle) f(\boldsymbol{r}_n), \qquad m = 1, \dots, N,$$

or in matrix notations:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}, \tag{8.9}$$

where $\boldsymbol{g} = [g(\boldsymbol{r}_1), \dots, g(\boldsymbol{r}_N)] \in \mathbb{C}^N$, $\boldsymbol{f} = [f(\boldsymbol{r}_1), \dots, f(\boldsymbol{r}_N)] \in \mathbb{C}^N$ and $H_{mn} = \psi_{\mathscr{D}}(\langle \boldsymbol{r}_m, \boldsymbol{r}_n\rangle)$, $m, n = 1, \dots, N$. Using the formalism from Section 2.3 of Chapter 6, it is moreover possible to interpret (8.9) as a linear transformation of some signal $\boldsymbol{f}$ defined over the spherical tessellation graph associated to the point set Θ. Since $\psi_{\mathscr{D}}$ is zonal and has small support, the linear operator $\boldsymbol{H}$ is moreover very well approximated by a graph filter with finite taps $K \ll N$ [159]. Such filters take necessarily the form of a K-order polynomial of the graph Laplacian $\boldsymbol{L}$ [47, 48]:

$$\boldsymbol{H} \simeq \sum_{k=0}^K \theta_k \boldsymbol{L}^k, \tag{8.10}$$

where $\boldsymbol{\theta} = [\theta_0, \dots, \theta_K] \in \mathbb{C}^{K+1}$ are some coefficients obtained by solving:

$$\boldsymbol{\theta} = \underset{\boldsymbol{\eta} \in \mathbb{C}^{K+1}}{\arg\min} \left\| \boldsymbol{H} - \sum_{k=0}^K \eta_k \boldsymbol{L}^k \right\|_F.$$

From the representation (8.10) of $\boldsymbol{H}$, we can then use Algorithm 7.11 from Section 4.3 to implement (8.9) efficiently as a cascade of multiplications between the *sparse* matrix $\boldsymbol{L}$ and the vector $\boldsymbol{f}$. This provides us with an efficient

scheme for approximating spherical convolutions between a kernel with small support and an arbitrary function in $\mathscr{L}^2(\mathbb{S}^{d-1})$.

9 Test Cases

In this chapter, we test the spherical approximation framework proposed in Chapters 5 and 6 on a mix of real and simulated datasets originating from a variety of real-life spherical approximation problems encountered in *environmental sciences*, *radio astronomy* and *planetary sciences*. In all these applications, various sampling and cost functionals are investigated, demonstrating the versatility and genericity of the approximation framework. A summary of all experiments investigated in this chapter is available in Table 9.1. Interactive versions of the spherical maps provided in this chapter are moreover available at the following link: matthieumeo.github.io.

1 Sea Surface Temperature Anomalies

In this example, we propose to establish global maps of sea surface temperature anomalies for the month of January 2011. Such maps are used in environmental sciences to monitor global climate change as well as manage the population of marine species and ecosystems particularly sensitive to fluctuations in water temperature. The data consists of 6745 simulated anomalies sampled at various points across the globe by drifting floats of the *Argo fleet* [7, 98], and corrupted by Gaussian white noise. The various maps produced are obtained by means of canonically discretised FPT and FPBP problems (6.4) and (6.14), as well as their discrete domain counterparts (6.25) and (6.26). Since the noise distribution is Gaussian, we consider the indicator function of an ℓ_2-ball as cost functional. The latter being nonsmooth but proximable (see Section 5.3 of Chapter 7), we make use of the PDS Algorithms 7.3, 7.5, 7.7 and 7.9 to solve optimisation problems (6.4), (6.14), (6.25) and (6.26) respectively. Motivated by the discussion in Chapter 8, we consider a Matérn pseudo-differential operator for the gTikhonov and gTV regularisation terms in the continuous FPT and FPBP problems. For their discrete domain counterparts, we consider a discrete Sobolev operator (see Section 2.3).

1.1 Background

Sea surface temperature is usually defined as the temperature of the one millimetre upper layer of the oceans, reflecting the thermal energy stored in the

Experiment Name	**Sea Surface Temperature Anomaly**	**Wildfires and Deforestation**	**Planck and the CMB**	**Lunar Elemental Abundance Maps**
Discussed in Section	1, 2 (Appendix A)	2	3	4
Results in Figures	9.2, 9.3, A.1, A.2	9.5, 9.7	9.9c, 9.9g	9.12, 9.13
Field of Application	Meteorology	Forestry	Radio astronomy	Planetary science
Data Source	NASA's Aqua [3], Argo [97, 98]	LAI [169], Fire [170]	Simulated, Planck [54]	PDS [102], Pixon [195]
Data in Fig(s).	9.1	9.4, 9.6	9.9b, 9.9f	9.10
Sampling Functional	Dirac (spatial sampling)	Rectangular step function	Squared Jinc (9.8), Gaussian	Kappa (9.11)
Number of Samples (L)	6745	24000	768, 9248	14986
Cost Function	ℓ_2-ball	KL-divergence, Quadratic	KL-divergence	ℓ_2-ball
Regularisation	gTikhonov, gTV	gTV	gTV	gTikhonov
Discretisation	Domain, Search space ($N = 7386$)	Search space ($N = 210216$)	Search space ($N = 118181$) ($N = 652997$)	Search space
Green Kernel	Matern ($p = 1$), Wendland ($k = 1$)	Wendland ($k = 1$)	Wendland ($k = 1$)	Matern ($p = 0$)
Noise Model	Gaussian	Poisson	Gaussian	Poisson
PSNR (dB)	10	–	30	–
Algorithms	7.3, 7.5, 7.7, 7.9, 7.11	7.6, 7.5	7.5, Dirty Imaging [181]	7.3, Pixon [141]

Table 9.1: Summary of the various experiments presented in Chapter 9.

latter. Sea surface temperatures departing from long-term averages (typically 12 years) are called *temperature anomalies*. While some anomalies are transient and simply due to ocean circulation patterns (such as *El Niño* and *La Niña*), others persist over many years and can hence be potential indicators of global climate changes [145]. Sea surface temperature anomalies are also very important in the monitoring and management of marine ecosystems particularly sensitive to water temperature fluctuations. For example, above-average sea water temperatures can result in *coral bleaching*, a phenomenon suspected to be responsible of the disappearance of between 29 and 50% of the Great Barrier Reef in 2016 [80]. Similarly, high water temperatures are contributing factors to *harmful algal blooms*, which lead to oxygen depletion in natural waters, with disastrous consequences on marine life [74].

1.2 Data Description

For this experiment, we simulated sea surface temperature anomalies by sampling at 6745 locations the global map of sea surface temperature anomalies produced by *NASA's Aqua* satellite [136] in January 2011 [3]. These anomalies, which serve here as a ground truth, were derived by comparing the sea surface temperatures recorded in January 2011 by NASA's Aqua satellite to the 12-year-averaged historical data for the same month collected by the *Pathfinder* satellite [188] between 1985 and 1997. The resulting map is depicted in Fig. 9.1a. The 6745 sampling locations were chosen as the positions of all floats from the *Argo fleet* [7] during the month of January 2011, obtained from [97] and curated by the authors of [98]. Argo is an international program, initiated in the early 2000's that uses 4000 drifting floats to monitor temperature, salinity and currents in the Earth's oceans. The samples were further polluted by Gaussian white noise with *peak signal to noise ratio (PSNR)* 10 dB. The resulting samples are plotted in Fig. 9.1b.

1.3 Data Model

Let $f : \mathbb{S}^2 \to \mathbb{R}$ denote the sea surface temperature anomaly function defined at every location on the globe (modelled as the unit sphere $\mathbb{S}^2$). Since temperatures typically have *smooth variations* at the surface of the Earth, we assume f to be an element of some Sobolev space $\mathscr{H}^\beta(\mathbb{S}^2)$, with $\beta > 1$ (see Chapter 8).

The $L = 6745$ measurements $\{y_1, \dots, y_L\} \subset \mathbb{R}$ correspond here to noisy anomaly records collected by the Argo floats across the globe at locations

$$\{\boldsymbol{p}_1, \dots, \boldsymbol{p}_L\} \subset \mathbb{S}^2.$$

Assuming a Gaussian white noise model, the float records can moreover be modelled as realisations of independent Gaussian random variables $\{Y_1, \dots, Y_L\}$, centred around the true temperature anomalies obtained by *ideal spatial sampling* of f:

$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}(f(\boldsymbol{p}_i), \sigma^2), \tag{9.1}$$

where $\mathcal{N}$ denotes the Gaussian distribution and $\sigma^2 > 0$ is the (unknown) noise variance, assumed uniform. Note that we have

$$\mathbb{E}[Y_i] = f(\boldsymbol{p}_i) = \langle \delta_{\boldsymbol{p}_i} | f \rangle, \qquad i = 1, \dots, L,$$

which fits well in our generic data model (5.1) page 67, if we choose the sampling functionals as Dirac measures $\delta_{\boldsymbol{p}_i}$. Note moreover that spatial sampling is indeed well-defined for f since for $\beta > 1$ the Sobolev space $\mathscr{H}^\beta(\mathbb{S}^2)$ is an RKHS.

1.4 Methods

1.4.1 Continuous Domain Methods

(a) gTikhonov Regularisation

We consider first recovering f by means of the following FPT problem:

$$f^\star = \arg \min_{f \in \mathscr{H}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2)} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{\Phi}(f)) \quad + \quad \|\mathscr{D}^\epsilon_{2.5} f\|_2^2 \right\}, \tag{9.2}$$

where:

- $\iota_{\mathcal{B}_{2,\rho}} : \mathbb{R}^L \to \{0\} \cup \{+\infty\}$ is the indicator function (7.20) of the ℓ_2-ball with radius $\rho = 0.5\% \times \|\boldsymbol{y}\|_2$. As explained in Section 5.3 of Chapter 7, the indicator function in (9.2) defines, under the Gaussian noise model (9.1), a confidence region containing the true samples $\mathbb{E}[\boldsymbol{y}]$ with probability $1 - \alpha$, for some $0 < \alpha < 1$ dependent on ρ.
- $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2) \to \mathbb{R}^L$ is the *sampling operator* given by

  $$\boldsymbol{\Phi}(f) = [\langle \delta_{\boldsymbol{p}_1} | f \rangle, \dots, \langle \delta_{\boldsymbol{p}_L} | f \rangle], \quad \forall f \in \mathscr{H}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2).$$

  Note that $\boldsymbol{\Phi}$ is well defined since $\mathscr{H}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2)$ is an RKHS (see Chapter 8).
- $\mathscr{D}^\epsilon_{2.5} : \mathscr{S}'(\mathbb{S}^2) \to \mathscr{S}'(\mathbb{S}^2)$ is the pseudo-differential operator associated to the Matérn zonal Green kernel with fixed scale $\epsilon \simeq 0.017$ –corresponding to an angular resolution[1] of approximately 4°:

  $$\psi^\epsilon_{2.5}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = S^\epsilon_{3/2}(\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}), \qquad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2. \tag{9.3}$$

[1] The angular resolution is measured here as the full width at half maximum (FWHM) of the Matérn kernel.

From Theorems 6.2 and 5.3, the solution to optimisation problem (9.2) is unique and given by:

$$f^\star(\boldsymbol{r}) = \sum_{l=1}^{L} x_l^\star \, \psi^\epsilon_{2.5} * \psi^\epsilon_{2.5}(\langle \boldsymbol{r}, \boldsymbol{p}_l \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^2,$$

where $*$ denotes the spherical convolution[2] operator (see Definition 3.3) and $\boldsymbol{x}^\star = [x_1^\star, \dots, x_L^\star] \in \mathbb{R}^L$ is the unique solution to the discrete optimisation problem:

$$\boldsymbol{x}^\star = \arg \min_{\boldsymbol{x} \in \mathbb{R}^L} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) \quad + \quad \boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} \right\}. \tag{9.4}$$

[2] As discussed in Section 3.3, spherical convolution with the Wendland kernel can be implemented efficiently.

Entries of the matrix $\boldsymbol{H} \in \mathbb{R}^{L \times L}$ are moreover given by

$$H_{lk} = \psi^\epsilon_{2.5} * \psi^\epsilon_{2.5}(\langle \boldsymbol{p}_l, \boldsymbol{p}_k \rangle), \qquad \forall l, k \in [\![1, L]\!].$$

We solve (9.4) using Algorithm 7.3. Since the Matérn kernel is spatially localised (see Fig. 8.1), the matrix $\boldsymbol{H}$ is in practice *sparse* (as discussed in Section 3.2 of Chapter 8) and is implemented as such in the iterations of the numerical solver for computational and storage efficiency.

(b) gTV Regularisation

Next, we consider recovering f by means of the following FPBP problem:

$$f^\star \in \arg\min_{f \in \mathcal{M}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2)} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{\Phi}(f)) \quad + \quad \|\mathscr{D}^\epsilon_{2.5} f\|_{TV} \right\}, \tag{9.5}$$

where the sampling operator $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2) \to \mathbb{R}^L$ is this time given by:

$$\boldsymbol{\Phi}(f) = [\langle f | \delta_{\boldsymbol{p}_1} \rangle, \dots, \langle f | \delta_{\boldsymbol{p}_L} \rangle], \quad \forall f \in \mathcal{M}_{\mathscr{D}^\epsilon_{2.5}}(\mathbb{S}^2).$$

Again, $\boldsymbol{\Phi}$ is well-defined over $\mathcal{M}^\epsilon_{\mathscr{D}_{2.5}}(\mathbb{S}^2)$ since the Matérn kernel $\psi^\epsilon_{2.5}$ has continuous traces, and hence from Proposition 5.2, the predual $\mathscr{C}^\epsilon_{\mathscr{D}_{2.5}}(\mathbb{S}^2)$ contains all Dirac measures.

From the discussion in Section 1.2 of Chapter 6, solutions of the optimisation problem (9.5) can be approximated as quasi-uniform Matérn splines:

$$f^\star(\boldsymbol{r}) = \sum_{n=1}^{N} x^\star_n \, \psi^\epsilon_{2.5}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^2,$$

where $N = 7386$, $\Xi_N = \{\boldsymbol{r}_n, \, n = 1, \dots, N\} \subset \mathbb{S}^2$ is a Fibonacci lattice (see Example 6.1) and $\boldsymbol{x}^\star = [x^\star_1, \dots, x^\star_N] \in \mathbb{R}^N$ is some solution to the discrete optimisation problem:

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{G}\boldsymbol{x}) \quad + \quad \|\boldsymbol{x}\|_1 \right\}. \tag{9.6}$$

The matrix $\boldsymbol{G} \in \mathbb{R}^{L \times N}$ is moreover given by

$$G_{ln} = \psi^\epsilon_{2.5}(\langle \boldsymbol{p}_l, \boldsymbol{r}_n \rangle), \; \forall (l, n) \in [\![1, L]\!] \times [\![1, N]\!].$$

We solve (9.6) using the PDS Algorithm 7.5. Again, since the Matérn kernel is spatially localised, the matrix $\boldsymbol{G}$ is in practice sparse and is implemented as such in Algorithm 7.5 for computational and storage efficiency.

1.4.2 Discrete Domain Methods

For comparison purposes, we also consider recovering f by means of the domain discretisation schemes described in Section 2 of Chapter 6. To this end, we consider the restriction $\boldsymbol{f} \in \mathbb{R}^N$ of f to the discrete set of directions $\Xi_N = \{\boldsymbol{r}_n, \, n = 1, \dots, N\} \subset \mathbb{S}^2$, where Ξ_N is the same Fibonacci lattice as in Paragraph (b). We recover $\boldsymbol{f}$ via the discrete domain counterparts of (9.2) and (9.5), given in this case by:

$$\boldsymbol{f}^\star = \arg\min_{\boldsymbol{f} \in \mathbb{R}^N} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{J}\boldsymbol{f}) \quad + \quad \|\boldsymbol{D}\boldsymbol{f}\|_2^2 \right\}, \tag{9.7}$$

and

$$\boldsymbol{f}^\star \in \arg\min_{\boldsymbol{f} \in \mathbb{R}^N} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{J}\boldsymbol{f}) \quad + \quad \|\boldsymbol{D}\boldsymbol{f}\|_1 \right\}, \tag{9.8}$$

respectively. The entries of the *sensing matrix* $\boldsymbol{J} \in \mathbb{R}^{L\times N}$ are defined as

$$J_{ij} = \delta_{n_i j}, \qquad \forall i = 1,\dots,L,\ j = 1,\dots,N,$$

where δ_{ij} is the *Kronecker delta* and

$$n_i = \arg\min_{n=1,\dots,N} \|\boldsymbol{p}_i - \boldsymbol{r}_n\|_2, \quad i = 1,\dots,L.$$

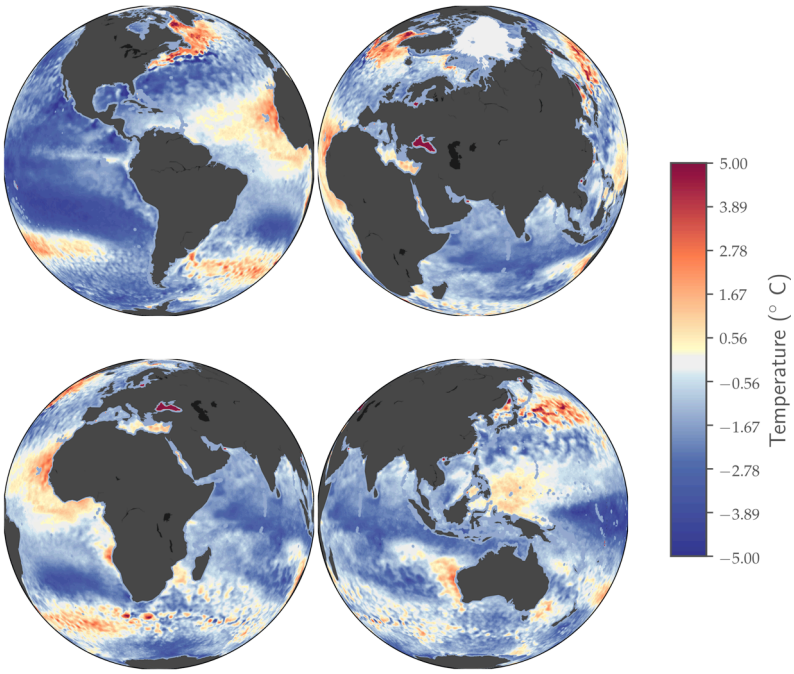
Roughly speaking, $\boldsymbol{J}$ corresponds to a discrete sampling matrix on the lattice Ξ_N, where the off-lattice sampling locations $\{\boldsymbol{p}_i, i = 1,\dots,L\}$ have been mapped to their closest neighbour in Ξ_N. The discrete pseudo-differential operator $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ finally is chosen as the discrete Sobolev operator $\boldsymbol{D} = (\boldsymbol{I}_N + \boldsymbol{L})^2$, where $\boldsymbol{L}$ is the Laplacian of the spherical tessellation graph associated to the point set Ξ_N. Note that $\boldsymbol{D} = (\boldsymbol{I}_N + \boldsymbol{L})^{2.5}$ would have been a more canonical choice since $\mathscr{D}^{\epsilon}_{2.5}$ in Section 1.4.1 is equivalent to the Sobolev operator $(\mathrm{Id} - \Delta_{\mathbb{S}^2})^{2.5}$ (see discussion in Chapter 8 and Fig. 8.2). However, computing $\boldsymbol{D} = (\boldsymbol{I}_N + \boldsymbol{L})^{2.5}$ requires computing the eigenvalue decomposition of the matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$, which is often impossible in practice due to the size of the latter. Moreover, such a choice of discrete pseudo-differential operator would make Algorithms 7.7 and 7.9 used to solve (9.7) and (9.8) much more computationally and memory intensive since the latter could no longer perform matrix-vector products involving $\boldsymbol{D}$ with the matrix free Algorithm 7.11. Indeed, this algorithm was designed in Section 4.3 for discrete pseudo-differential operators taking the form of polynomials of $\boldsymbol{L}$, and $\boldsymbol{D} = (\boldsymbol{I}_N + \boldsymbol{L})^{2.5}$ is *not* a polynomial in $\boldsymbol{L}$.

1.5 Results

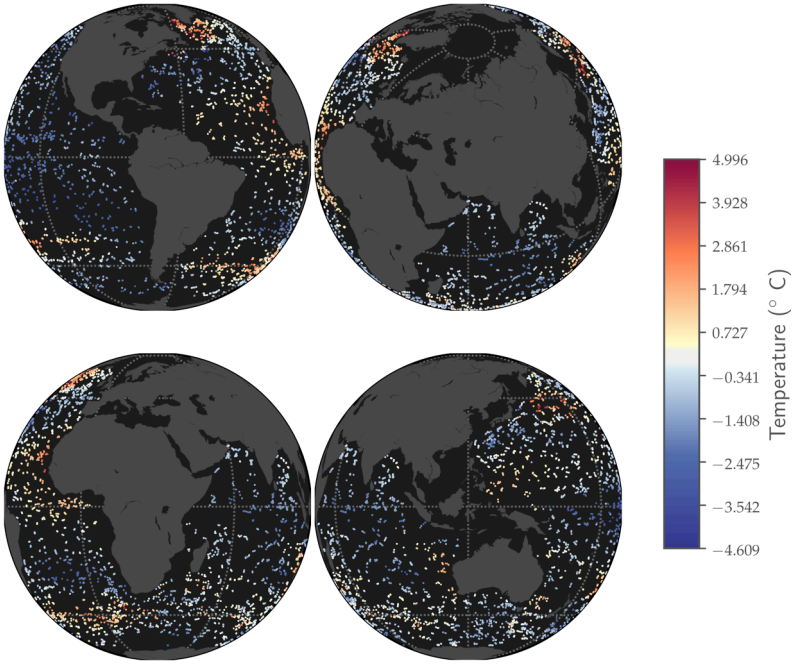
The various estimates of the sea surface temperature anomaly function obtained by solving (9.2), (9.5), (9.7) and (9.8) are provided in Figs. 9.2a, 9.2b, 9.3a and 9.3b respectively. The smoothing induced by the gTikhonov regularisation is clearly visible for both the continuous and discrete domains estimates in Figs. 9.2a and 9.3a respectively. The nature of this smoothing seem however different for the two estimates: in Fig. 9.2a, the large scale structures of the actual anomaly map are clearly visible, while in Fig. 9.3a, only the strongest features remain.

In contrast, the continuous and discrete gTV estimates in Figs. 9.2b and 9.3b capture far more of the fine fluctuations in the true anomaly map: see for example the eastern coast and southern tip of Africa, as well as the regions surrounding Greenland, Japan or the Indian ocean. This time, both estimates exhibit much more similar features. The discrete gTV estimate in Fig. 9.3b however appears rougher than the continuous gTV estimate in Fig. 9.2b due to the clearly visible pixelisation artefacts.

Remark 9.1 *In Appendix A, we consider replacing the Matérn pseudo-differential operator $\mathscr{D}^{\epsilon}_{2.5}$ in (9.2) and (9.5) with an equivalent Wendland pseudo-differential operator. This yields two new estimates of the surface temperature anomaly function provided in Figs. A.1 and A.2. Not surprisingly, they appear very similar*

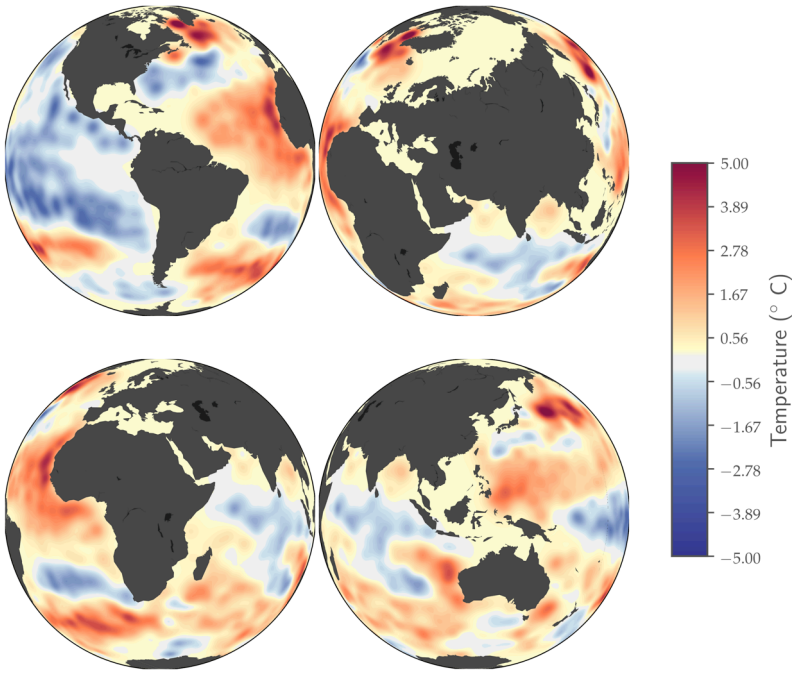


(a) Global map of sea surface temperature anomalies in January 2011 produced from *NASA's Aqua* satellite data.

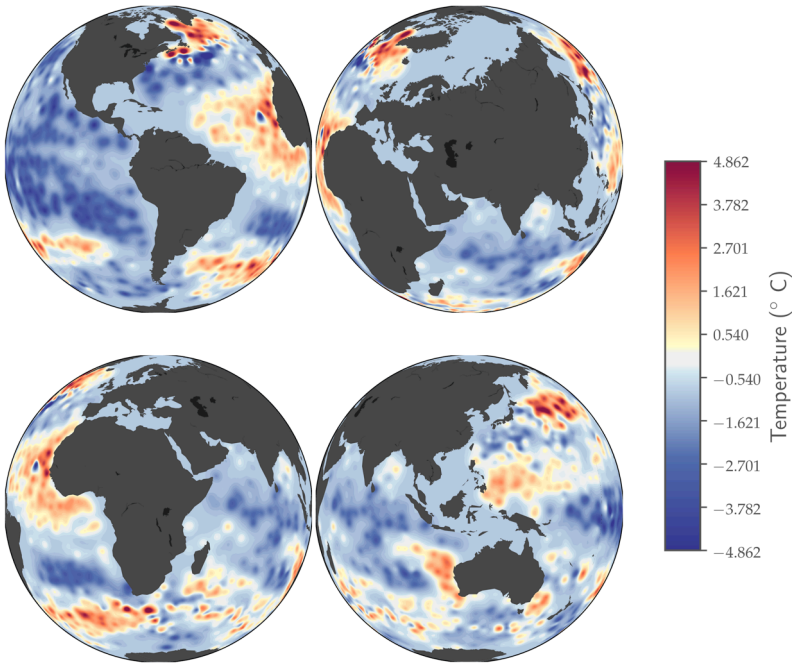


(b) Simulated anomalies recorded by Argo floats in January 2011. Float locations are marked by dots coloured according to the recorded anomaly (red=warmer temperatures, blue=colder temperatures).

Figure 9.1: The data for the experiments in Section 1 consists of 6745 anomalies sampled from a global sea surface temperature map produced from *NASA's Aqua* satellite data in January 2011. The sample locations correspond to the locations of the Argo drifting floats during that month.

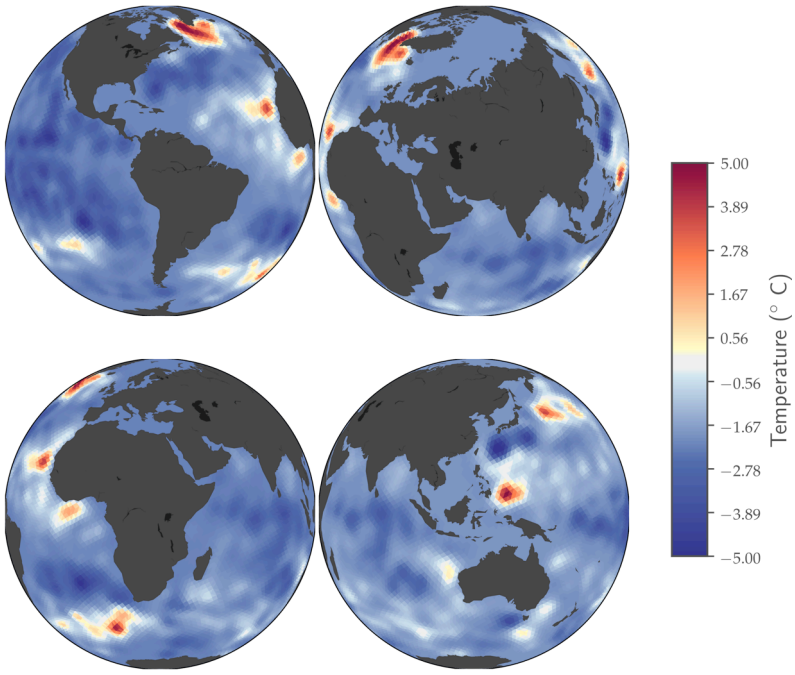


(a) Sea surface temperature anomaly function obtained by solving the FPT problem (9.2) with gTikhonov regularisation.

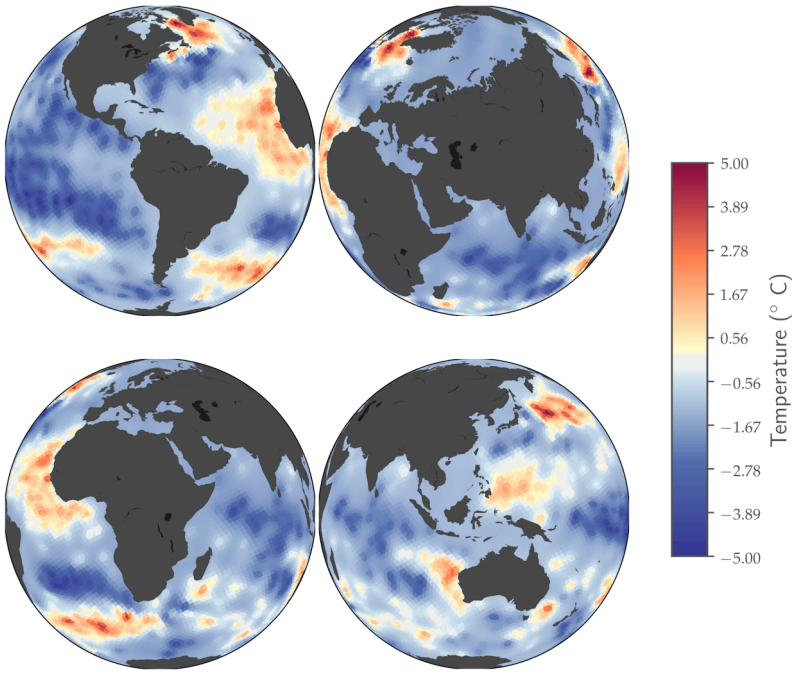


(b) Sea surface temperature anomaly function obtained by solving the FPBP problem (9.5) with gTV regularisation.

Figure 9.2: Estimates of the sea surface temperature anomaly function obtained with the continuous domain methods from Section 1.4.1.



(a) Sea surface temperature anomaly function obtained by solving the discrete problem (9.7), with discrete gTikhonov regularisation.



(b) Sea surface temperature anomaly function obtained by solving the discrete problem (9.8), with discrete gTV regularisation.

Figure 9.3: Estimates of the sea surface temperature anomaly function obtained with the discrete domain methods from Section 1.4.2.

to those of Figs. 9.2a and 9.2b: the two kernels can be used interchangeably in practice.

2 Wildfires and Deforestation

In this example, we propose to establish global density maps of trees and wildfires across the globe for the year 2016. Tree density maps are used in environmental sciences to monitor deforestation and illegal logging, as well as assess the amount of vegetal photosynthesis. Similarly, wildfire maps allow scientists to better understand atmospheric chemistry and its impact on climate. In both experiments, the data used consists of tree and fire counts recorded by NASA's Aqua and Terra satellites. The resolution of the raw data is moreover deliberately reduced by a factor of 3 by binning the counts in patches of angular size $\simeq 1.5° \times 1.5°$. The goal of this resolution reduction is to show that the lost resolution can be successfully recovered by spline-based approximation. In both cases, two density maps are obtained by solving with Algorithms 7.5 and 7.6 respectively two FPBP problems: one with a least-squares cost functional, and one with a KL-divergence cost functional –ideally suited for count data with Poisson-like distribution (see Section 5.5 of Chapter 7).

2.1 Background

Home to 80% of terrestrial species, forests host most of Earth's biodiversity and contribute largely to its preservation [63]. Indeed, tree canopies play a crucial role in the regulation of the water cycle, creation of litter and exchange of energy between the ground and the atmosphere, which are all contributing factors to the overall good health of an ecosystem [45]. Changes to forest habitats can lead to the extinction of endangered species and disrupt the entire food chain equilibrium. But forests are more than animal shelters: they also protect humans from natural hazards such as floods or droughts[63]. In addition, forests represent natural and cost-efficient solutions for mitigating climate-change, and can provide 30% of the solution for keeping global warming below 2 °C [63]. Indeed, the *photosynthesis*, used by trees and plants to convert light into energy, cleans the atmosphere by absorbing carbon emissions and releasing oxygen.
In order to enlighten policy-makers and hopefully stem the current environmental crisis, it is hence crucial to monitor deforestation and understand its causes, such as agricultural conversion, commodity production, urbanisation, illegal logging or fires. Fires deserve perhaps a special attention as they contribute largely to the overall greenhouse gas emissions, with an estimated contribution of 30% to the net annual increase in the concentration of atmospheric carbon dioxide [62, 170].

2.2 Data Description

For this experiment, we worked with the *Leaf Area Index (LAI)* [94] and *Fire* [89] data products provided by NASA for the year 2016. The datasets, available at [169, 170], provide monthly counts of trees and active fires respectively at a

resolution of 0.1 degrees square. The counts are estimated from multispectral images captured by the *MODIS* aboard NASA's *Terra* and *Aqua* satellites [197]. For a better visual appreciation of the results, we aggregated the data from every month of 2016 (see Figs. 9.4a and 9.6a) and binned it to a lower resolution of approximately 1.5 degrees square (this corresponds to a reduction of resolution by a factor 3). The goal of this resolution reduction is to show that the lost resolution can be successfully recovered by spline-based approximation. We further corrupted the binned data with Poisson noise, a common noise model for count data. The processed data for both experiments is displayed in Figs. 9.4b and 9.6b.

2.3 Data Model

In both cases, one wishes to estimate the spatial density of trees (respectively fires) f at the surface of the Earth, using counts $\{y_1, \dots, y_L\} \subset \mathbb{N}$ from non-overlapping equal-angle patches $\{B_1, \dots, B_L\} \subset \mathbb{S}^2$ tiling the sphere. Since the data at hand consists of counts, a Poisson noise model is a suitable choice. This can be achieved by modelling tree locations as random occurrences of some *spatial Poisson point process* [38, 172], often used in spatial statistics to model random spatial scattering of objects [144, 161]. The distribution of a Poisson point process is entirely determined by its *intensity measure* $\Lambda \in \mathcal{M}(\mathbb{S}^2)$, which counts the expected number of objects (in this case trees or fires) observed in a given region of the sphere. The sought spatial density of the Poisson process is then given –assuming it exists– by the *Radon-Nikodym derivative* [157] $f : \mathbb{S}^2 \to \mathbb{R}$ of Λ, also called *density* or *intensity* function of the point process. Similarly as in Section 1, we assume f to be an element of the RKHS $\mathscr{H}^\beta(\mathbb{S}^2)$ for some $\beta > 1$. With such a formalism, the reported counts can then be seen as realisations of $L = 24000$ independent Poisson random variables $\{Y_1, \dots, Y_L\}$:

$$Y_i \overset{\text{ind}}{\sim} \text{Poisson}\left(\lambda_i\right), \qquad i = 1, \dots, L, \tag{9.9}$$

with rates $\lambda_i > 0$ given by:

$$\lambda_i = \int_{\mathbb{S}^2} f(\boldsymbol{r}) \chi_{B_i}(\boldsymbol{r}) d\boldsymbol{r} = \int_{B_i} f(\boldsymbol{r}) d\boldsymbol{r}, \qquad i = 1, \dots, L, \tag{9.10}$$

and where $\chi_{B_i} \in \mathscr{L}^2(\mathbb{S}^2)$ are the *characteristic functions* of the surveyed patches $\{B_i,\, i = 1, \dots, L\} \subset \mathbb{S}^2$. We can reinterpret the rates in (9.10) as generalised samples of f:

$$Y_i \overset{\text{ind}}{\sim} \text{Poisson}\left(\langle f, \chi_{B_i} \rangle_{\mathbb{S}^2}\right), \qquad i = 1, \dots, L,$$

where $\mathbb{E}[Y_i] = \langle f, \chi_{B_i} \rangle_{\mathbb{S}^2}$, $i = 1, \dots, L$, hence yielding a data model falling into the scope of the generalised sampling framework (5.1).

2.4 Methods

2.4.1 KL-Divergence Cost Function

Since the data consists of counts, we consider recovering f by means of the following FPBP problem:

$$f^\star \in \arg\min_{f \in \mathcal{M}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2)} \left\{ D_{KL}(\boldsymbol{y}||\boldsymbol{\Phi}(f)) \quad + \quad \lambda \|\mathscr{D}^\epsilon_{3,1} f\|_{TV} \right\}, \tag{9.11}$$

where:

- D_{KL} denotes the *generalised Kullback-Leibler divergence* defined in Section 5.5 of Chapter 7. As explained there, the KL-divergence cost function can be shown to be proportional to the negative log-likelihood of the data $\boldsymbol{y} = [y_1, \dots, y_L] \in \mathbb{R}^L$ under the Poisson data model (9.9).
- $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2) \to \mathbb{R}^L$ is the sampling operator given by:

$$\boldsymbol{\Phi}(f) = [\langle f|\chi_{B_1}\rangle, \dots, \langle f|\chi_{B_L}\rangle], \quad \forall f \in \mathcal{M}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2).$$

- λ is a strictly positive constant, tuned manually.
- $\mathscr{D}^\epsilon_{3,1} : \mathscr{S}'(\mathbb{S}^2) \to \mathscr{S}'(\mathbb{S}^2)$ is the pseudo-differential operator associated to the Wendland zonal Green kernel with a scale $\epsilon \simeq 0.026$ –corresponding again to an angular resolution[3] of approximately $1°$:

$$\psi^\epsilon_{3,1}(\langle \boldsymbol{r}, \boldsymbol{s}\rangle) = \phi_{3,1}\left(\frac{\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s}\rangle}}{\epsilon}\right), \qquad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2.$$

[3] The angular resolution is measured here as the full width at half maximum (FWHM) of the Wendland kernel.

Note that since the sampling functions are square-integrable and $\mathscr{D}^\epsilon_{3,1}$ is invertible and with spectral growth order $p = 2.5 > (d-1)/2 = 1$ (see Chapter 8), we can use Proposition A.1 to show that $\{\chi_{B_i},\, i = 1, \dots, L\} \subset \mathscr{C}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2)$ and hence the sampling operator $\boldsymbol{\Phi}$ is indeed well defined.

From the discussion in Section 1.2 of Chapter 6, the solutions of (9.11) can be approximated as quasi-uniform Wendland splines:

$$f^\star(\boldsymbol{r}) = \sum_{n=1}^{N} x^\star_n \, \psi^\epsilon_{3,1}(\langle \boldsymbol{r}, \boldsymbol{r}_n\rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^2,$$

where $N = 210216$, $\Xi_N = \{\boldsymbol{r}_n,\, n = 1, \dots, N\} \subset \mathbb{S}^2$ is a Fibonacci lattice (see Example 6.1) and $\boldsymbol{x}^\star = [x^\star_1, \dots, x^\star_N] \in \mathbb{R}^N$ is some solution to the discrete optimisation problem:

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ D_{KL}(\boldsymbol{y}||\boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \|\boldsymbol{x}\|_1 \right\}. \tag{9.12}$$

From Theorem 6.5, the matrix $\boldsymbol{G} \in \mathbb{R}^{L \times N}$ is moreover given by

$$G_{ln} = (\psi^\epsilon_{3,1} * \chi_{B_l})(\boldsymbol{r}_n), \qquad \forall (l, n) \in [\![1, L]\!] \times [\![1, N]\!].$$

We solve (9.12) using Algorithm 7.5. Since the Wendland kernel and the patches have compact support, the matrix $\boldsymbol{G}$ is in practice sparse and is imple-

mented as such in Algorithm 7.5 for computational and storage efficiency.

2.4.2 Quadratic Cost Function

For sufficiently large rates, the Poisson distribution can be well approximated by a Gaussian distribution. This motivates the use of a least-squares data functional in (9.11), yielding:

$$f^\star \in \arg\min_{f \in \mathcal{M}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2)} \left\{ \|\boldsymbol{y} - \boldsymbol{\Phi}(f)\|_2^2 \quad + \quad \lambda \|\mathscr{D}^\epsilon_{3,1} f\|_{TV} \right\}. \tag{9.13}$$

The discrete optimisation problem (9.12) then becomes:

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ \|\boldsymbol{y} - \boldsymbol{G}\boldsymbol{x}\|_2^2 \quad + \quad \lambda \|\boldsymbol{x}\|_1 \right\}, \tag{9.14}$$

which can be solved efficiently using Algorithm 7.6.

2.5 Results

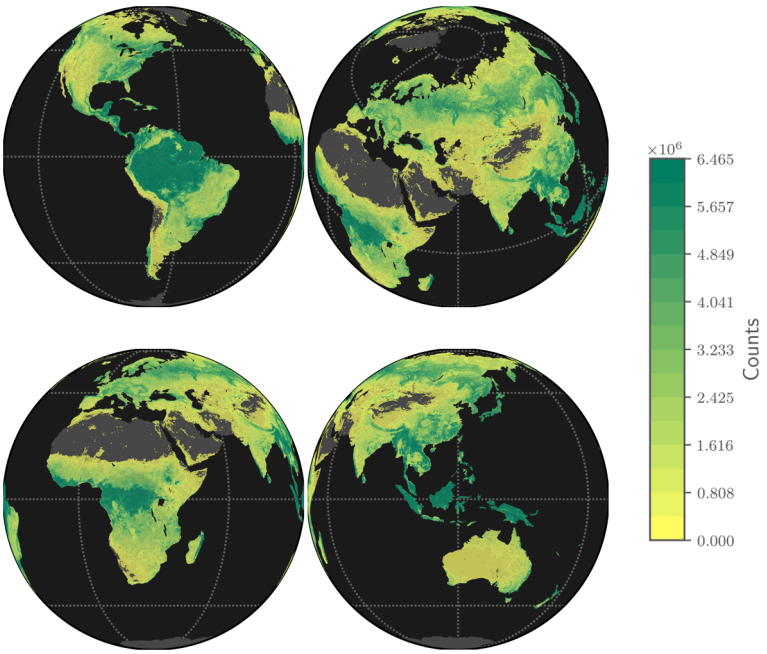
The tree and fire density functions estimated with both recovery strategies (9.11) and (9.13) are provided in Figs. 9.5a and 9.5b and Figs. 9.7a and 9.7b respectively. For both experiments, we observe that the recovered estimates (with KL-divergence and least-squares cost functions respectively) have a much higher resolution than the original corrupted binned counts in Figs. 9.4b and 9.6b, recovering almost the natural resolution of the unprocessed data in Figs. 9.4a and 9.6a. We observe however that the KL-divergence cost function seems to better recover regions with low intensity count, such as the Arabian Peninsula or Australia. In contrast, the least-squares data-fidelity functional has a tendency of yielding sparser density estimates, where all low intensity count regions are set to zero. This behaviour was already observed in image restoration under Poisson noise [187].

3 Planck and the Cosmic Microwave Background

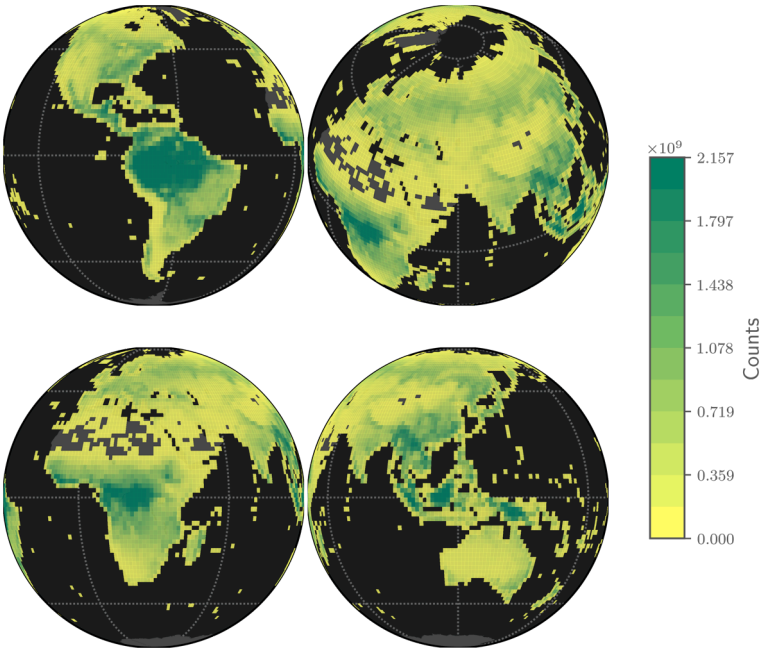
In this example, we propose to recover *full-sky intensity maps* from the raw measurements of radio telescopes such as Planck [2]. Such maps display the intensity (or equivalently the temperature) of every astronomical radio source across the celestial sphere. In this example, we use a realistic physical model to simulate radio data from first a simplistic point source sky model and then a more realistic sky model built from high-resolution images from Planck. We recover the sky intensity maps by solving with Algorithm 7.5 a FPBP problem with a KL-divergence data-fidelity term. We moreover compare the accuracy and resolution of the recovered map to the *dirty map*, a common radio astronomy imaging product obtained by naive *smoothing* of the data [181].

3.1 Background

The brightest celestial objects in the universe are *stars* such as the Sun. The light they shine is the result of the thermonuclear fusion of hydrogen and helium in their core, with peak emission wavelength proportional to their temperature.

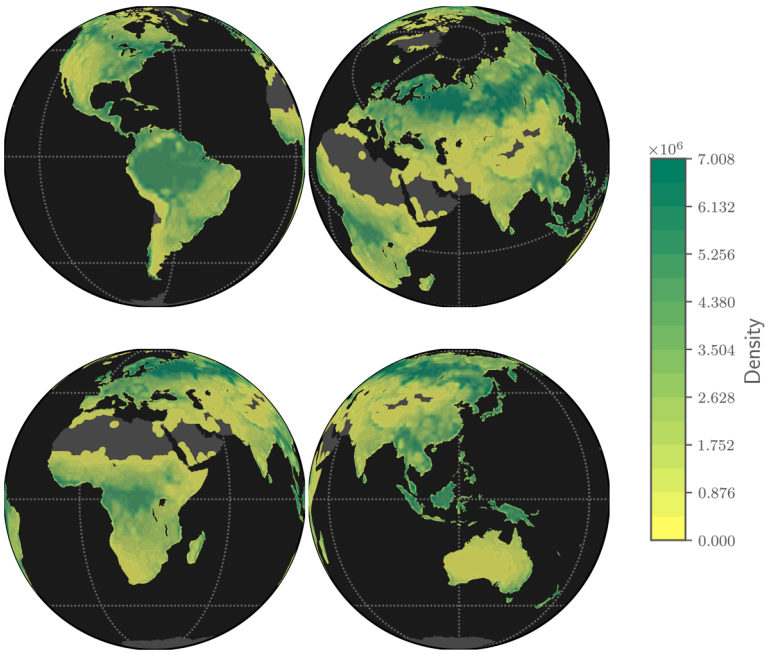


(a) Aggregated tree counts at full resolution (0.1 degree).

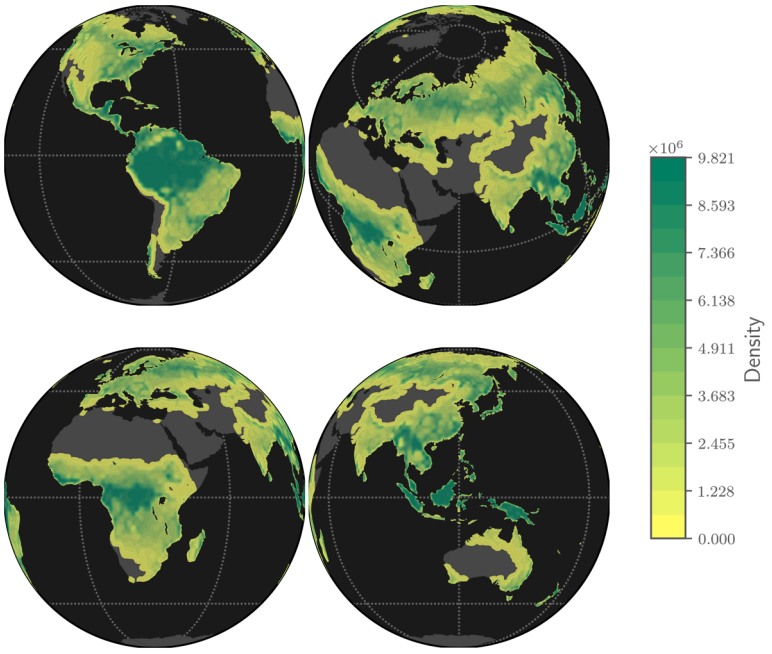


(b) Aggregated tree counts with reduced resolution (1.5 degree) and additional Poisson corruption.

Figure 9.4: Aggregated tree counts for the year 2016 produced from MODIS data, a sensor aboard *NASA's Terra/Aqua* satellites.

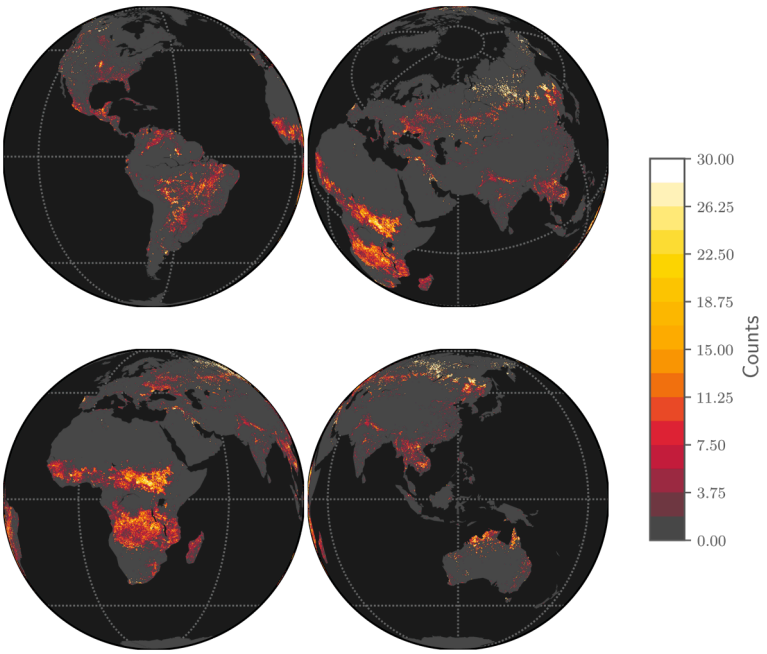


(a) KL-divergence cost function.

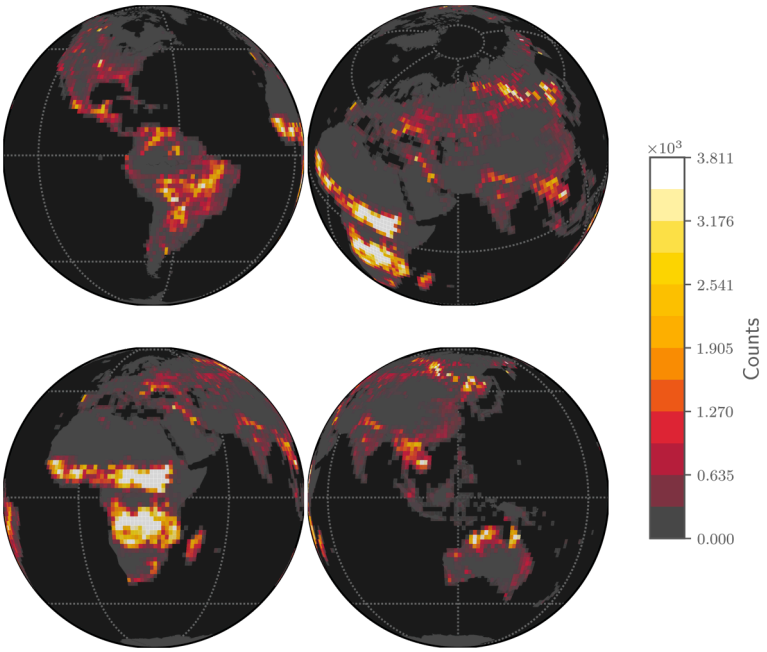


(b) Quadratic cost function.

Figure 9.5: Estimates of the tree density function obtained by solving the FPBP optimisation problems (9.12) and (9.14), with KL-divergence **(a)** and quadractic **(b)** cost functions respectively.

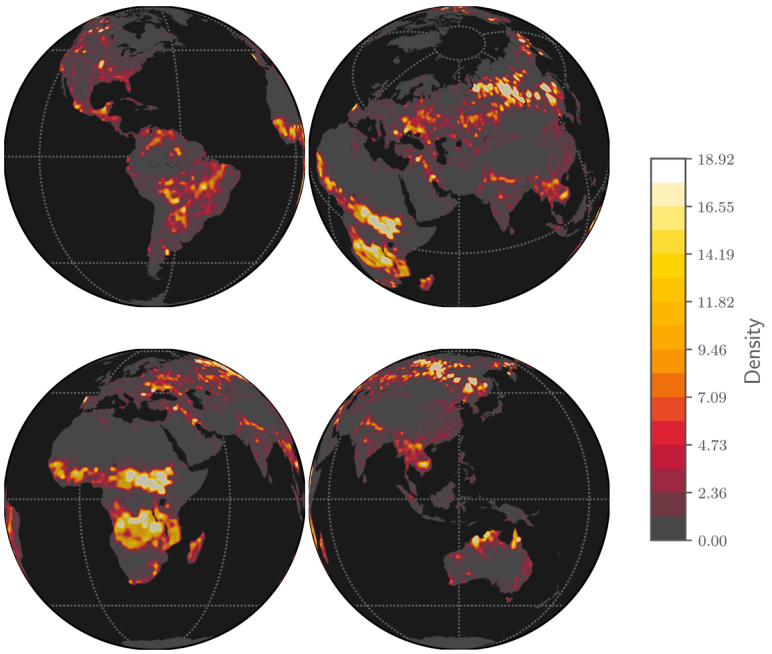


(a) Aggregated fire counts at full resolution (0.1 degree).

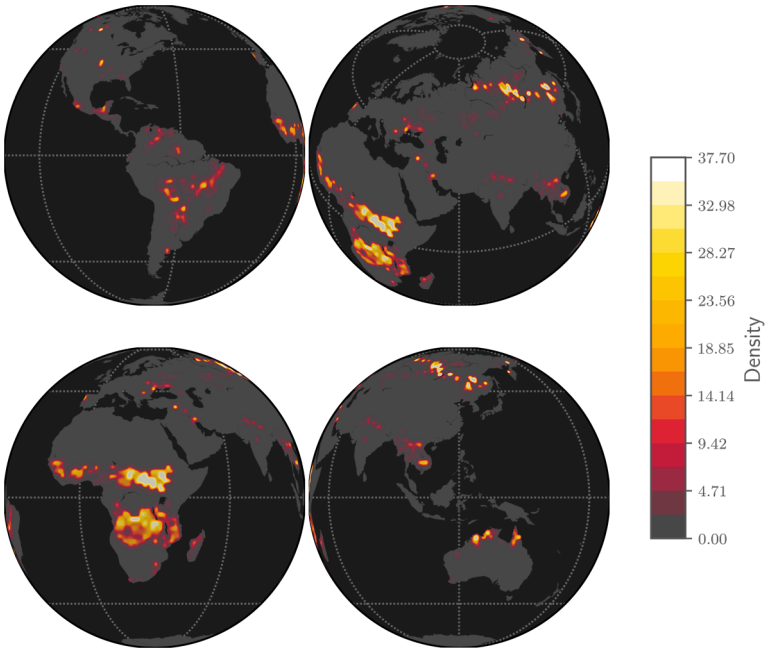


(b) Aggregated fire counts with reduced resolution (1.5 degree) and additional Poisson corruption.

Figure 9.6: Aggregated fire counts for the year 2016 produced from MODIS data, a sensor aboard *NASA's Terra/Aqua* satellites.



(a) KL-divergence cost function.



(b) Quadratic cost function.

Figure 9.7: Estimates of the fire density function obtained by solving the FPBP optimisation problems (9.12) and (9.14), with KL-divergence **(a)** and quadractic **(b)** cost functions respectively.

Most stars emit in the visible range: *red dwarfs*, which are relatively cold ($\leq$ 3'500 K), appear *red* to our eyes, while *supergiant stars*, much hotter ($\geq$ 10'000 K), appear *blue*. The light coming from distant stars however can appear shifted towards larger wavelengths for an observer on Earth. This phenomenon, called *redshift*, is a simple manifestation of the *Doppler effect* due to the expansion of the universe. From *Hubble's law*, the further a celestial object is from us, the faster it is moving away from us, and hence the more its emission spectrum is redshifted. At some point, the light reaching us from distant stars falls into the infrareds and below (microwaves and radiowaves), making it impossible to observe them with traditional optical telescopes. To probe the universe deeper, astronomers hence need radio telescopes, capable of observing radio emissions. A rudimentary radio telescope was first built by Karl Jansky in 1930, who observed radiation coming from the *Milky Way*. This major discovery gave birth to the field of radio astronomy and opened a myriad of new horizons for astronomers and cosmologists. Nowadays, modern radio telescopes often take the form of *dish antennae* with a central feed, such as Planck [2], which was sent into space to observe the cosmic microwave radiation background (CMB) [90] and test cosmological models.

3.2 Data Description

We simulate realistic radio data from two sky models. The first one is a point source sky model, taking the form of a stream of 120 Diracs on the sphere, with uniformly distributed directions and log-normal distributed intensities (see Fig. 9.9a). The second is a high resolution intensity map of the sky at 857 GHz produced by the Planck Collaboration using the data from the entire Planck mission (see Fig. 9.9e). It is available on the *Planck Legacy Archive* platform [54]. In both cases, the data (see Figs. 9.9b and 9.9f respectively) consists of noisy directional samples[4] of the sky intensity map, convolved with a model of the point spread function of the radio telescope, in this case the *beamshape* [81] of the dish antenna.

[4] Distributed according to a Fibonacci tessellation. For the point source case, we chose $L = 768$, while for the Planck model, we chose $L = 9248$.

For the point source case, we modelled the point spread function $\varphi : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}_+$ by a *squared jinc function* (see Fig. 9.8):

$$\varphi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \frac{\lambda_0^2 J_1^2\left(2\pi R\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}/\lambda_0\right)}{R^2(2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle)}, \qquad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2, \tag{9.15}$$

where J_1 is the *Bessel function of the first kind* with order 1, $\lambda_0 = 3$ and $R = 9$. Equation (9.15) corresponds [82, 85] to the far-field beamshape of an ideal circular aperture with radius $R = 9$ m, steered towards a direction $\boldsymbol{s} \in \mathbb{S}^2$ and operating at a wavelength $\lambda_0 = 3$ m.

For the Planck sky model, we considered a wavelength $\lambda_0 = 10$ cm and a dish with radius $R = 1$ m. Moreover, we approximated (9.15) by a Gaussian beam with properly chosen scale $\sigma > 0$:

$$\varphi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \exp\left(\frac{\langle \boldsymbol{r}, \boldsymbol{s} \rangle - 1}{\sigma^2}\right), \qquad (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2,$$

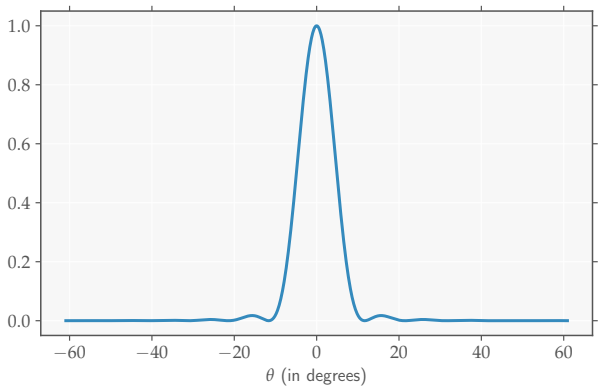


Figure 9.8: Beamshape at 100 MHz of a dish antenna with ideal circular aperture of radius 9m.

allowing us to perform efficiently the spherical convolution of the high resolution Planck map with the instrument beam. To this end, we used the routine named `healpy.sphtfunc.smoothing` from the `healpy` Python3 package [202].

In both cases finally, we corrupted the directional samples of the blurred intensity maps with Gaussian white noise of PSNR of 30 dB.

3.3 Data Model

In both cases, one wishes to estimate the intensity field f of stars on the celestial sphere using samples $\{y_1, \ldots, y_L\} \subset \mathbb{R}$ obtained by steering the dish antenna towards various directions $\{\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_L\} \subset \mathbb{S}^2$. Similarly as before, we assume f to be an element of the Sobolev space $\mathscr{H}^\beta(\mathbb{S}^2)$ for some $\beta > 1$. With a Gaussian white noise assumption, the samples recorded by the dish antenna can be modelled as Gaussian random variables:

$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}\left((\varphi * f)(\boldsymbol{\rho}_i), \sigma^2\right), \qquad i = 1, \ldots, L, \tag{9.16}$$

for some $\sigma > 0$. We can reinterpret the means in (9.16) as generalised samples of f:

$$\mathbb{E}[Y_i] = \int_{\mathbb{S}^2} \varphi(\langle \boldsymbol{r}, \boldsymbol{\rho}_i \rangle) f(\boldsymbol{r}) d\boldsymbol{r}, \qquad i = 1, \ldots, L,$$

hence yielding a data model falling into the scope of the generalised sampling framework (5.1).

3.4 Methods

3.4.1 gTV Regularisation

We consider recovering f by means of the FPBP problem:

$$f^\star \in \arg \min_{f \in \mathcal{M}_{\mathscr{D}^\epsilon_{3,1}}(\mathbb{S}^2)} \left\{ D_{KL}(\boldsymbol{y} || \boldsymbol{\Phi}(f)) \quad + \quad \lambda \| \mathscr{D}^\epsilon_{3,1} f \|_{TV} \right\}, \tag{9.17}$$

where:

- $\lambda > 0$ is fixed manually,
- $\boldsymbol{\Phi} : \mathcal{M}_{\mathscr{D}^{\epsilon}_{3,1}}(\mathbb{S}^2) \to \mathbb{R}^L$ is the sampling operator given by

$$\boldsymbol{\Phi}(f) = [\langle f | \varphi(\langle \cdot, \boldsymbol{\rho}_1 \rangle) \rangle, \dots, \langle f | \varphi(\langle \cdot, \boldsymbol{\rho}_L \rangle) \rangle], \quad \forall f \in \mathcal{M}_{\mathscr{D}^{\epsilon}_{3,1}}(\mathbb{S}^2).$$

  Note that, with the same arguments as in Section 2.4.1, it is possible to show that $\boldsymbol{\Phi}$ is well defined, i.e. $\{\varphi(\langle \cdot, \boldsymbol{\rho}_i \rangle),\, i = 1, \dots, L\} \subset \mathscr{C}_{\mathscr{D}^{\epsilon}_{3,1}}(\mathbb{S}^2)$.
- $\mathscr{D}^{\epsilon}_{3,1} : \mathscr{S}'(\mathbb{S}^2) \to \mathscr{S}'(\mathbb{S}^2)$ is the pseudo-differential operator associated to the Wendland zonal Green kernel with scales ϵ corresponding to an angular resolution of approximately 2° and 1° for the point source and Planck sky models respectively:

$$\psi^{\epsilon}_{3,1}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \phi_{3,1}\left(\frac{\sqrt{2 - 2\,\langle \boldsymbol{r}, \boldsymbol{s} \rangle}}{\epsilon} \right), \qquad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2.$$

The KL-divergence cost function in (9.17) helps to better recover low-intensity sources in the sky. From the discussion in Section 1.2 of Chapter 6, solutions to the optimisation problem (9.17) can be approximated by quasi-uniform Wendland splines:

$$f^\star(\boldsymbol{r}) = \sum_{n=1}^{N} x^\star_n \, \psi^{\epsilon}_{3,1}(\langle \boldsymbol{r}, \boldsymbol{r}_n \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^2,$$

where $\Xi_N = \{\boldsymbol{r}_n,\, n = 1, \dots, N\} \subset \mathbb{S}^2$ is a Fibonacci lattice[5] (see Example 6.1) and $\boldsymbol{x}^\star = [x^\star_1, \dots, x^\star_N] \in \mathbb{R}^N$ is some solution to the discrete optimisation problem:

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \left\{ D_{KL}(\boldsymbol{y} || \boldsymbol{G}\boldsymbol{x}) \quad + \quad \lambda \|\boldsymbol{x}\|_1 \right\}. \tag{9.18}$$

[5] For the point source case, we chose $N = 118181$, while for the Planck model, we chose $N = 652997$.

The matrix $\boldsymbol{G} \in \mathbb{R}^{L \times N}$ is moreover given by

$$G_{ln} = \psi^{\epsilon}_{3,1} * \varphi(\langle \boldsymbol{\rho}_l, \boldsymbol{r}_n \rangle),\ \forall (l, n) \in [\![1, L]\!] \times [\![1, N]\!].$$

We solve (9.18) using Algorithm 7.5. Since Wendland splines and the instrument beamshape φ are well localised in space, the matrix $\boldsymbol{G}$ is in practice sparse and is implemented as such in Algorithm 7.5 for computational and storage efficiency.

3.4.2 Dirty Image

For comparison purposes, we also produce the *dirty image*, commonly used in radio astronomy [181]. In our context, the latter is obtained by interpolating the samples with the telescope beamshape:

$$f_D(\boldsymbol{r}) = \sum_{i=1}^{L} y_i \, \varphi(\langle \boldsymbol{r}, \boldsymbol{\rho}_i \rangle), \quad \boldsymbol{r} \in \mathbb{S}^2. \tag{9.19}$$

Notice that if the functions $\varphi(\langle \boldsymbol{r}, \boldsymbol{\rho}_i \rangle)$ were all *orthogonal* (which is certainly not the case in general) then (9.19) would be the orthogonal projection of f

onto span $\{\varphi(\langle\cdot, \boldsymbol{\rho}_1\rangle), \ldots, \varphi(\langle\cdot, \boldsymbol{\rho}_L\rangle)\}$. In the absence of orthogonality, (9.19) can be interpreted as a *smoothing* of the true sky intensity.

3.5 Results

The sky intensity estimates obtained by spline-based imaging (9.18) and dirty imaging (9.19) are available in Figs. 9.9c and 9.9d for the point source sky model and Figs. 9.9g and 9.9h for the Planck sky model. In both cases, we observe that the sky intensity estimates obtained by spline approximation have far greater resolution than dirty sky estimates. This is particularly obvious for the point source sky model where neighbouring celestial sources, fused together in the dirty estimate Fig. 9.9d, are successfully resolved in the spline-based estimate Fig. 9.9c. Similarly, the extended structures in Fig. 9.9g appear much sharper than in Fig. 9.9h.

4 Lunar Elemental Abundance Maps

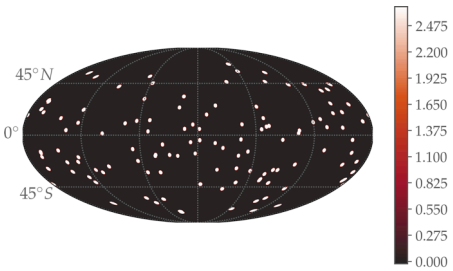
In this example, we build global distribution maps of radioactive elements on the surface of the Moon using real data collected by NASA's *Lunar Prospector (LP)* probe [101, 103]. Such maps, called *elemental abundance maps*, are used by scientists to retrace the Moon's geologic history [101]. For example, abundance of Thorium (Th) reveals past magmatic activity and differentiation [195]. In this experiment, the data consists in 377367 geolocalised gamma ray counts, obtained by orbital gamma-ray spectroscopy (GRS) and recorded over a period of approximately six months by the Lunar Prospector probe on its lowest orbit.[6] To reduce the size of the data, the latter was moreover binned on a Fibonacci Delaunay tessellation consisting of 14986 triangular cells with average angular diameter of 1°. The abundance map is obtained by solving with Algorithm 7.3 an FPT problem with ℓ_2-ball cost function. For comparison purposes, we also provide an abundance map obtained with the state-of-the-art *Pixon method* [141] and reproduced from the data provided in the supplementary material of [195].

[6] Which was at an average altitude of 30 km.

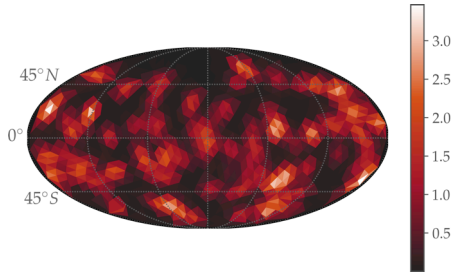
4.1 Background

The *giant-impact* hypothesis[7] suggests that the Moon was formed by the accretion of debris originating from a collision between the proto-Earth and *Theia*, an hypothesised planetoid of the size of Mars (approximately 6500 km in diameter) in the early Solar system [24, 49]. This hypothesis, currently favoured by the scientific community, is notably supported by Moon rock samples collected by the *Apollo* missions [24]. These indicate that the primordial Moon's crust was largely liquified, forming the so-called *lunar magma ocean*. This past magmatic activity is well explained by the *giant-impact* hypothesis given the high energy that such a collision would generate. In order to understand better the geologic history of the Moon, NASA launched the Lunar Prospector (LP) mission [24, 101, 103] in January 1998. The latter orbited around the Moon for two years, probing its surface by means of gamma-ray spectroscopy (GRS) in search of a material called KREEP (potassium (K), rare earth elements (REE),

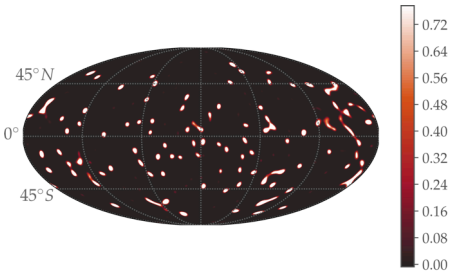
[7] The giant-impact hypothesis is sometimes also called *Big Splash* or *Theia Impact*.



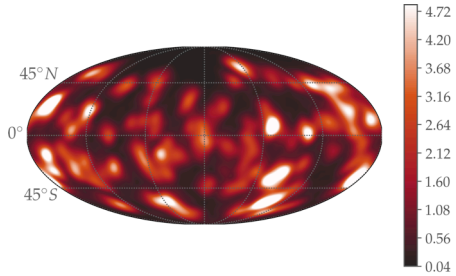
(a) Point sources convolved with a 2° wide Gaussian beam.



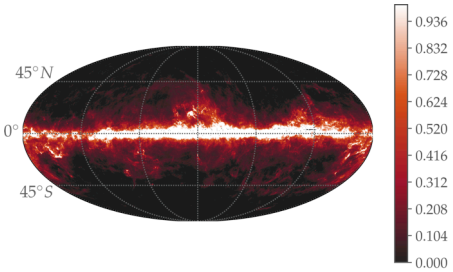
(b) Raw samples from the radio telescope ($L = 768$).



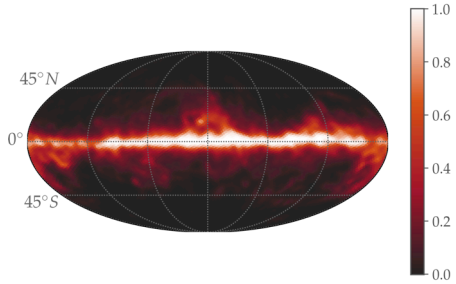
(c) Sparse spline approximation, solution of (9.18).



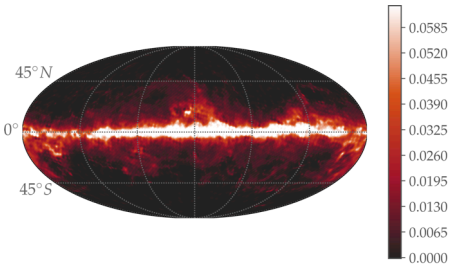
(d) Dirty field obtained by the smoothing (9.19).



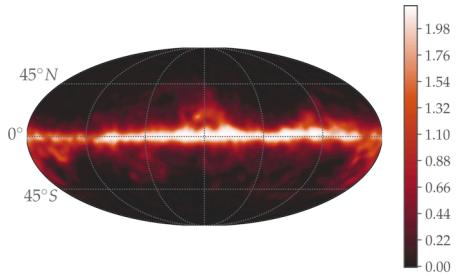
(e) Planck sky model.



(f) Raw samples from the radio telescope ($L = 9248$).



(g) Sparse spline approximation, solution of (9.18).



(h) Dirty field obtained by the smoothing (9.19).

Figure 9.9: Spline-based and dirty imaging of radio sources. In Figs. 9.9a to 9.9d we investigate a point source sky model with 120 sources. In Figs. 9.9e to 9.9h we use high-resolution data from the entire Planck mission as sky model. In both cases, the sparse spline approximation of the sky intensity field appear much sharper than the one obtained by dirty imaging. All images were enhanced by contrast stretching for better visualisation.

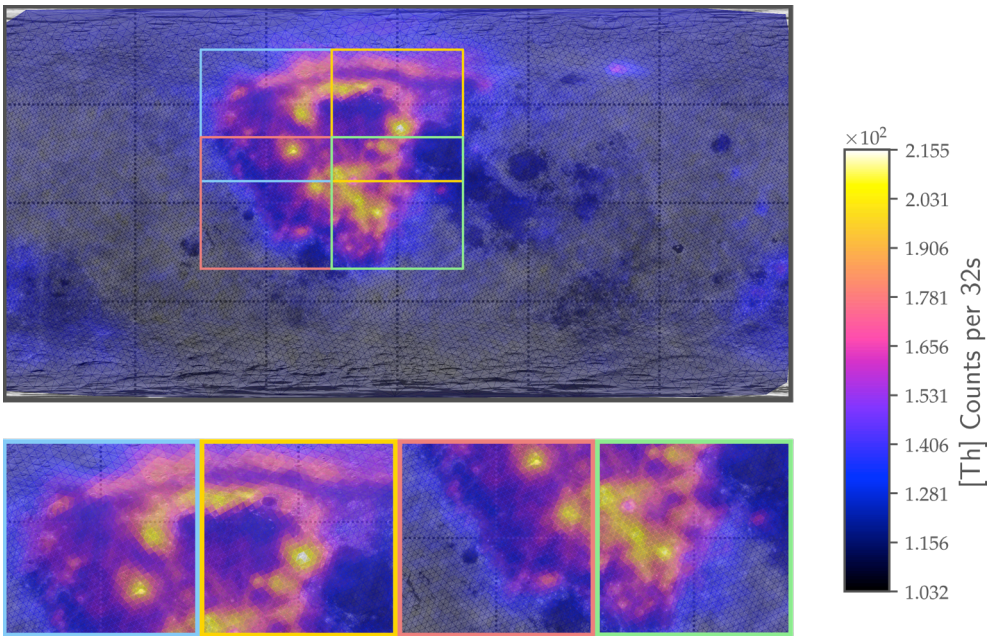


Figure 9.10: Per-bin average of gamma-ray counts within the Th line (2.62 ± 0.2 MeV), during the LOW2 phase of the Lunar Prospector mission.

phosphorus (P)), of which lunar magma was largely composed [101]. Among the various chemical elements involved in the composition of KREEP, the Thorium (Th) is the most easily observed. Indeed, its 2.61 MeV peak in the Moon's gamma ray spectrum is both strong and well separated from other peaks [194]. In this experiment, we will hence build an abundance map for Thorium.

4.2 Data Description

For this experiment, we worked with reduced spectrometer data collected during the LOW2 phase of the Lunar Prospector mission, which lasted for a total of 180.023 days [103]. During this period, the spacecraft was on its lowest orbit at an average altitude of 30 km. The data is available in the NASA Planetary Data System (Goesciences Node) [102] and consists of 377367 counts of gamma rays originating from the natural decay of radioactive elements in the top 10 centimetres of the lunar crust [194]. Since we are interested here in building an abundance map for Thorium only, we removed all counts with energy outside the Th line: 2.62 ± 0.2 MeV [101]. The counts are obtained over integration times of 32 seconds, and geolocalised on the surface of the Moon using the spacecraft's position at the time of the observation [103]. In order to reduce the data size to something more manageable, we furthermore bin the data on a Fibonacci Delaunay tessellation, consisting in 14986 triangular cells with average angular diameter of 1°. The average number of counts in each bin is displayed in Fig. 9.10.

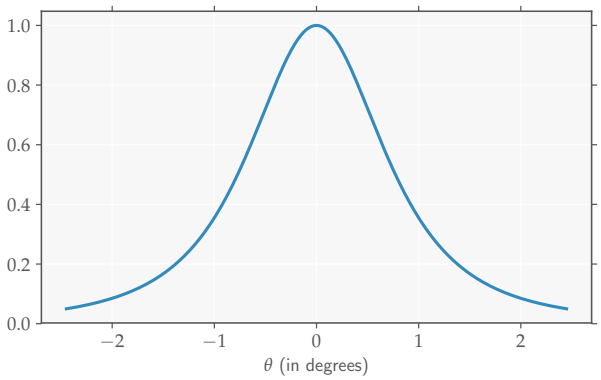


Figure 9.11: Point spread function of the orbital gamma-ray spectrometer onboard the Lunar Prospector probe at an altitude of $h = 30$ km.

4.3 Data Model

Using the formalism introduced in Section 2.3, it is possible to model the random emissions of gamma rays at the surface of the Moon as occurences of some *spatial Poisson point process* with intensity function $f : \mathbb{S}^2 \to \mathbb{R}_+$, assumed here to belong to the RKHS $\mathscr{H}^\beta(\mathbb{S}^2)$, for some $\beta > 1$. Denoting by N_i the number of observations in each of the Fibonacci bins, the binned counts $\{y_1, \dots, y_L\} \subset \mathbb{R}_+$ can be seen as realisations of independent Poisson random variables $\{Y_1, \dots, Y_L\}$:

$$Y_i \overset{\text{ind}}{\sim} \text{Poisson}(N_i \lambda_i), \qquad i = 1, \dots, L,$$

The rates $\lambda_i > 0$ are moreover given by:

$$\lambda_i = (\varphi * f)(\boldsymbol{r}_i) = \int_{\mathbb{S}^2} \varphi(\langle \boldsymbol{r}, \boldsymbol{r}_i \rangle) f(\boldsymbol{r}) d\boldsymbol{r}, \qquad i = 1, \dots, L,$$

where $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\} \subset \mathbb{S}^2$ are the centres of each bin. The function $\varphi : [-1, 1] \to \mathbb{R}_+$ is the point spread function (PSF) of the orbital gamma-ray spectrometer which, as demonstrated in [101], is well fitted by a *kappa function* (see Fig. 9.11):

$$\varphi(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \left[1 + \frac{R^2 \arccos^2(\langle \boldsymbol{r}, \boldsymbol{s} \rangle)}{2\sigma(h)^2} \right]^{-\kappa(h)-1},$$

where $R = 1737.1$ km is the radius of the Moon, h is the altitude (in km) of the spacecraft, and [101]

$$\sigma(h) = 0.704h + 1.39, \qquad \kappa(h) = -4.87 \times 10^{-4} h + 0.631.$$

From properties of the Poisson distribution we have finally:

$$\frac{1}{N_i} \mathbb{E}[Y_i] = \varphi * f(\boldsymbol{r}_i), \qquad i = 1, \dots, L,$$

which falls indeed into the scope of the generalised sampling framework (5.1).

4.4 Methods

4.4.1 *gTikhonov Regularisation*

We consider recovering f by means of the following FPT problem:

$$f^\star = \arg \min_{f \in \mathscr{H}_{\mathscr{D}^\epsilon_{1.5}}(\mathbb{S}^2)} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{\Phi}(f)) \quad + \quad \|\mathscr{D}^\epsilon_{1.5} f\|_2^2 \right\}, \tag{9.20}$$

where:

- $\iota_{\mathcal{B}_{2,\rho}} : \mathbb{R}^L \to \{0\} \cup \{+\infty\}$ is the indicator function (7.20) of the ℓ_2-ball with radius $\rho = 3\% \times \|\boldsymbol{y}\|_2$.
- $\boldsymbol{\Phi} : \mathscr{H}_{\mathscr{D}^\epsilon_{1.5}}(\mathbb{S}^2) \to \mathbb{R}^L$ is the *sampling operator* given by

  $$\boldsymbol{\Phi}(f) = [\langle \varphi_1 | f \rangle, \ldots, \langle \varphi_L | f \rangle], \quad \forall f \in \mathscr{H}_{\mathscr{D}^\epsilon_{1.5}}(\mathbb{S}^2),$$

  where $\langle \varphi_i | f \rangle = \langle f, \varphi(\langle \cdot, \boldsymbol{r}_i \rangle) \rangle = \varphi * f(\boldsymbol{r}_i)$, $i = 1, \ldots, L$. Note that $\boldsymbol{\Phi}$ is well defined since, from Proposition 5.1, we have $\mathscr{L}^2(\mathbb{S}^2) \subset \mathscr{H}'_{\mathscr{D}^\epsilon_{1.5}}(\mathbb{S}^2)$.
- $\mathscr{D}^\epsilon_{1.5} : \mathscr{S}'(\mathbb{S}^2) \to \mathscr{S}'(\mathbb{S}^2)$ is the pseudo-differential operator associated to the Matérn zonal Green kernel with fixed scale $\epsilon \simeq 0.016$:

  $$\psi^\epsilon_{1.5}(\langle \boldsymbol{r}, \boldsymbol{s} \rangle) = \exp\left(- \frac{\sqrt{2 - 2\langle \boldsymbol{r}, \boldsymbol{s} \rangle}}{\epsilon} \right), \qquad \forall (\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{S}^2 \times \mathbb{S}^2.$$

From Theorems 6.2 and 5.3, the solution to optimisation problem (9.20) is unique and, using properties of zonal kernels, can be written as:

$$f^\star(\boldsymbol{r}) = \sum_{l=1}^{L} x_l^\star \, (\psi^\epsilon_{1.5} * \psi^\epsilon_{1.5} * \varphi)(\langle \boldsymbol{r}, \boldsymbol{r}_l \rangle), \quad \forall \boldsymbol{r} \in \mathbb{S}^2,$$

where $*$ denotes the spherical convolution[8] operator (see Definition 3.3) and $\boldsymbol{x}^\star = [x_1^\star, \ldots, x_L^\star] \in \mathbb{R}^L$ is the unique solution to the discrete optimisation problem:

$$\boldsymbol{x}^\star = \arg \min_{\boldsymbol{x} \in \mathbb{R}^L} \left\{ \iota_{\mathcal{B}_{2,\rho}}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) \quad + \quad \boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} \right\}. \tag{9.21}$$

Entries of the matrix $\boldsymbol{H} \in \mathbb{R}^{L \times L}$ are moreover given by

$$H_{lk} = (\varphi * \psi^\epsilon_{1.5} * \psi^\epsilon_{1.5} * \varphi)(\langle \boldsymbol{r}_l, \boldsymbol{r}_k \rangle), \qquad \forall l, k \in [\![1, L]\!].$$

We solve (9.21) using Algorithm 7.3. Since the Matérn kernel is spatially localised (see Fig. 8.1), the matrix $\boldsymbol{H}$ is in practice *sparse* (as discussed in Section 3.2 of Chapter 8) and is implemented as such in the iterations of the numerical solver for computational and storage efficiency.

[8] As discussed in Section 3.3, spherical convolution with the Wendland kernel can be implemented efficiently.

4.4.2 *Pixon Method*

For comparison purposes, we also provide an abundance map obtained with the state-of-the-art *Pixon method* [141] and reproduced from the data provided in the supplementary material of [195]. The Pixon method is a discrete Bayesian method, reported to achieve from 1.5 to 2 times better spatial resolution than other deconvolution methods in planetary sciences [195]. It is locally adaptive:

during the reconstruction pixels are grouped together into *pixons*, whose size is modified so as to minimise a local misfit statistic [195]. Note that the image produced in [195] was obtained from the aggregated data from the phases `LOW1` and `LOW2` of the Lunar Prospector mission [103], which spanned on a period of 220.506 days (against 180.023 days for the data we used). Moreover, the data was binned on a finer grid with cells of angular size 0.5°.

4.5 Results

The Thorium abundance maps obtained by solving (9.21) and the Pixon method are provided in Figs. 9.12 and 9.13 respectively. Both maps successfully sharpen the empirical distribution of the data in Fig. 9.10 and suppress the statistical noise polluting it. The Pixon map Fig. 9.13 however appears slightly sharper than the gTikhonov map Fig. 9.12. It moreover exhibits fine details not distinguishable in Fig. 9.12 (especially in the regions surrounded by the green and yellow boxes respectively). This could be due to the local adaptivity of the Pixon method, and the fact that the data used to produce this estimate was more abundant and binned on a finer grid. Nevertheless, the sharpness of the gTikhonov map Fig. 9.12 is quite remarkable, especially given the known limitations of gTikhonov regularisation (see Section 3.3 of Chapter 5) and the relative simplicity of the optimisation problem (9.21).

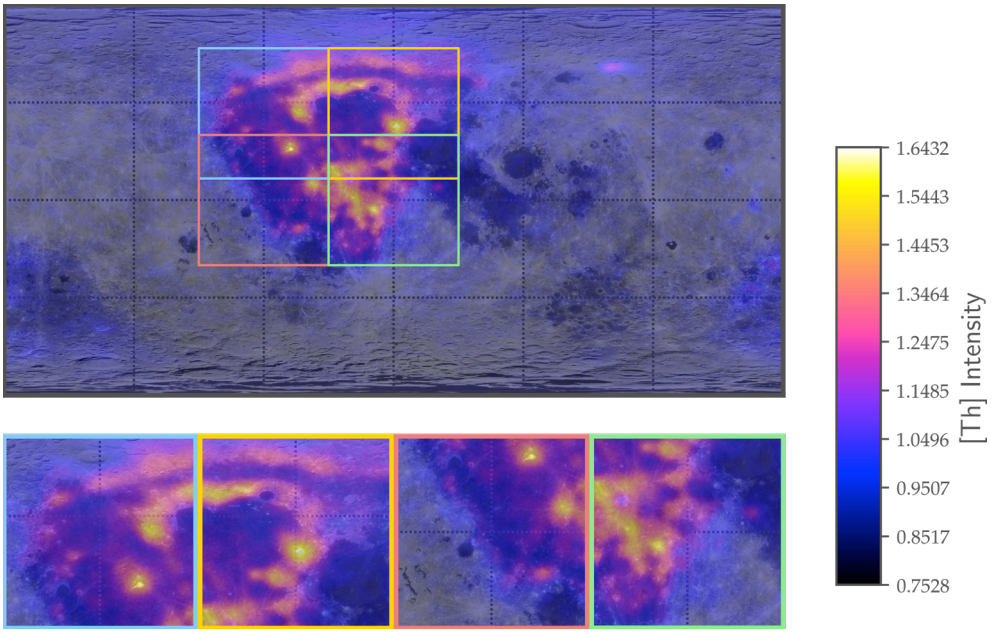


Figure 9.12: Estimate of the Thorium density function obtained by solving the FPT problem (9.21).

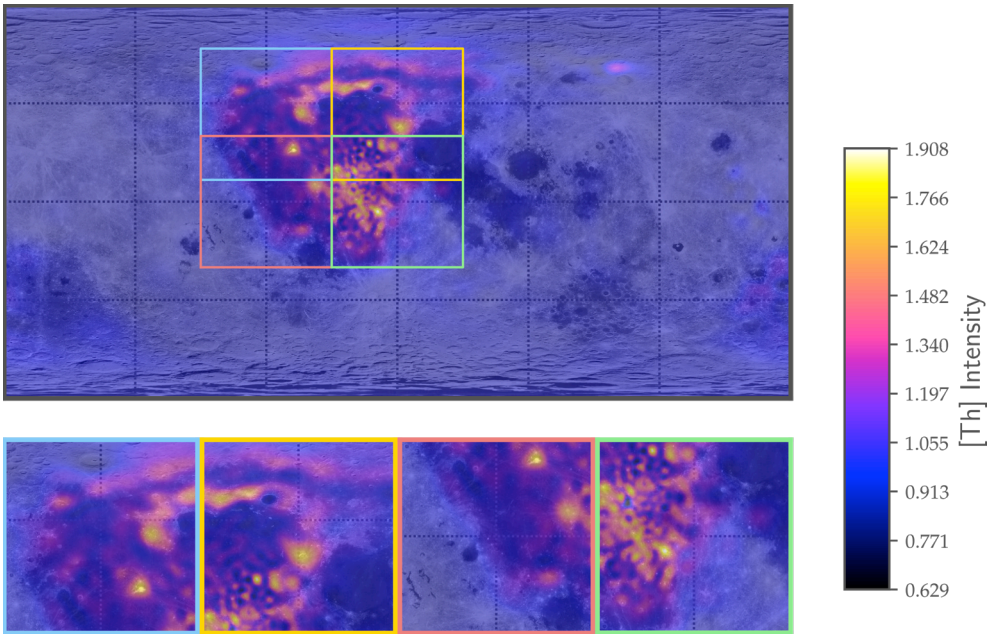
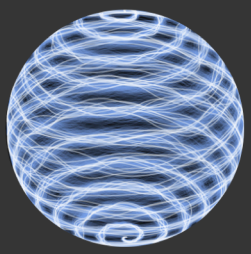
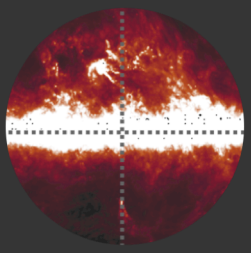


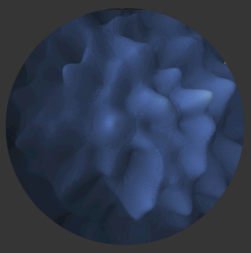
Figure 9.13: Estimate of the Thorium density function obtained with the Pixon method [141]. This figure was reproduced from the data provided in the supplementary material of [195].

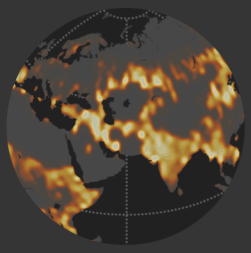
IV

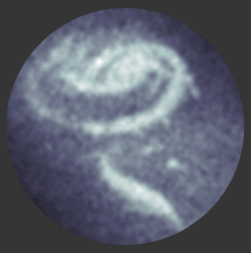
Further Topics & Conclusion

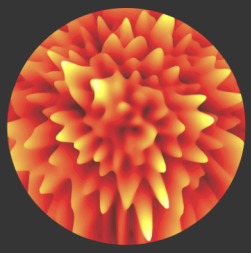


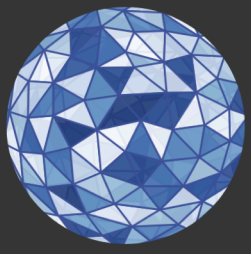


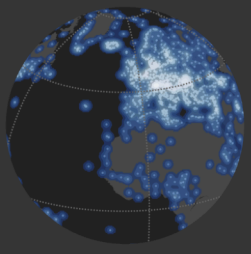


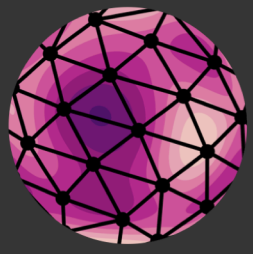












In this part, we discuss further topics and conclude this thesis. Some of the highlights of this part are the following:

- *In Chapter 10, we design an efficient and locally convergent algorithm for recovering the spatial innovations of periodic Dirac streams with finite rates of innovation. This algorithm is envisioned as an alternative to the quasi-uniform spline discretisation scheme proposed in Chapter 6 for gTV regularised functional inverse problems.*
- *In Chapter 11, we show how the convergence speed of proximal algorithms can be "boosted" by means of recurrent neural networks, for purposes of real-time acoustic imaging.*
- *In Chapter 12 finally, we reflect back on the trajectory of this thesis and outline a few prospective research avenues building on top of the material of this thesis.*

10

Generalised Sampling of FRI Signals*

In this chapter, we introduce a non-convex optimisation algorithm, baptised Cadzow plug-and-play gradient descent (CPGD), allowing the estimation of the spatial innovations of a *periodic Dirac stream* with *finite rate of innovation* [25] from generalised measurements of the latter. The algorithm is *extremely simple* and *very efficient*, outperforming the state-of-the-art algorithm proposed in [130]. Unlike the latter, CPGD is moreover provably *locally convergent*. As discussed in Section 1, this algorithm could be used for the purpose of estimating extreme point solutions to FPBP problems formulated over the circle $\mathbb{S}^1$. However, it has much wider applicability, and is hence presented here in a more general context (in particular, we consider Dirac streams with arbitrary period $T > 0$). Note that the notations of this chapter have been adapted to the conventions generally adopted in the finite rate of innovation (FRI) framework [25, 130, 184], and may hence differ slightly from those of Parts I to III of this thesis.

*The material presented in this chapter is the result of joint work with A. Besson, P. Hurley and M. Vetterli, and is the topic of [162], currently under submission.

1 Motivation in the Context of this Thesis

A common belief about generalised total variation regularisation is that it enforces *sparsity* in the variations of the functions recovered by functional penalised basis pursuit. As shown by Theorem 5.4, this is actually only true for very specific FPBP solutions, namely extreme point solutions, which –assuming that the gTV pseudo-differential operator $\mathscr{D}$ is spline-admissible– take the form of *sparse* $\mathscr{D}$-splines, i.e. with less innovations K than the total number L of measurements:

$$f^\star(\boldsymbol{r}) = \sum_{k=1}^{K} \alpha_k \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_k \rangle), \qquad \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{10.1}$$

with $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K] \in \mathbb{C}^K$ and $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_K\} \subset \mathbb{S}^{d-1}$. Extreme point solutions are hence, in virtue of *Occam's razor principle* (see Chapter 5), particularly interesting since relatively *simple*: they use as little as possible degrees of freedom to fit the data. It is hence desirable to come up with algorithmic solutions capable of approximating specifically the extreme point solutions

(10.1) of an FPBP problem. Of course, one could always approximate (10.1) by a quasi-uniform $\mathscr{D}$-spline, and solve the discrete PBP problem (6.14), as proposed in Theorem 6.5. Unfortunately, there is a priori no reason that the spline obtained in this way would approximate specifically an extreme point solutions of the continuous FPBP problem. Indeed, the quasi-uniform $\mathscr{D}$-spline approximation scheme leveraged in (6.14) is meant to approximate well *any* non-limit FPBP solution, irrespective of it being an extreme point or not. Moreover, the quasi-uniform $\mathscr{D}$-spline approximation of (10.1) for a fixed knot set size N may not necessarily have sparse innovations –i.e. bounded by the number of measurements– hence making it very hard to distinguish extreme point solutions from regular interior point solutions once discretised by means of quasi-uniform $\mathscr{D}$-splines.

In this chapter, we therefore propose an alternative discretisation strategy yielding solutions with guaranteed sparse form (10.1). The idea is to enforce the sparse parametric form (10.1) by replacing the unknown f in the FPBP problem (5.23) by a *non-uniform* $\mathscr{D}$-spline of the form $\sum_{k=1}^{L} \alpha_k \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_k \rangle)$, whose amplitudes and knot directions are both assumed unknown and *learnt* from the data. Similarly as in Theorem 5.4, it can be shown that the resulting discrete optimisation problem then takes the form:

$$\boldsymbol{\alpha}^\star, \{\boldsymbol{r}_1^\star, \dots, \boldsymbol{r}_L^\star\} = \underset{\substack{\{\boldsymbol{r}_1,\dots,\boldsymbol{r}_L\} \subset \mathbb{S}^{d-1} \\ \boldsymbol{\alpha} \in \mathbb{C}^L}}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}\left(\sum_{k=1}^{L} \alpha_k \psi_{\mathscr{D}}(\langle \boldsymbol{r}, \boldsymbol{r}_k \rangle)\right)\right) + \lambda \|\boldsymbol{\alpha}\|_1 \right\}, \tag{10.2}$$

where $\boldsymbol{y} \in \mathbb{C}^L$ and F, $\boldsymbol{\Phi}$, $\mathscr{D}$ and λ are as in Theorem 5.4. Note that from the definition of a spline and the fact that $\mathscr{D}$ is invertible, we can moreover rewrite (10.2) as

$$\boldsymbol{\alpha}^\star, \{\boldsymbol{r}_1^\star, \dots, \boldsymbol{r}_L^\star\} = \underset{\substack{\{\boldsymbol{r}_1,\dots,\boldsymbol{r}_L\} \subset \mathbb{S}^{d-1} \\ \boldsymbol{\alpha} \in \mathbb{C}^L}}{\arg\min} \left\{ F\left(\boldsymbol{y}, \boldsymbol{\Phi}\mathscr{D}^{-1}\left(\sum_{k=1}^{L} \alpha_k \delta_{\boldsymbol{r}_k}\right)\right) + \lambda \|\boldsymbol{\alpha}\|_1 \right\}. \tag{10.3}$$

The problem is hence to find a *Dirac stream* $\sum_{k=1}^{L} \alpha_k \delta_{\boldsymbol{r}_k}$ whose unknown innovations minimise (10.3). Since the directions $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\}$ have a nonlinear dependency on the data and are optimised over the continuous domain $\mathbb{S}^{d-1}$, (10.3) appears like a very hard optimisation problem. Fortunately, the generalised finite rate of innovation (FRI) sampling framework [25, 130, 184], shows that, at least for the circle when $d = 2$ (see [52] for extensions to the sphere when $d = 3$), the optimisation problem can actually be *decoupled*. Indeed, the directions $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_L\}$ can be learnt by finding the roots of the so-called *annihilating filter*, which is independent of the Dirac amplitudes. Once the directions are found, the latter can simply be plugged into (10.3) and the amplitudes are easily obtained by means of Algorithms 7.5 and 7.6 for example, depending on the nature of the cost function F. Unfortunately, finding the annihilating filter of a Dirac stream from generalised measurements of

the latter is a *non-convex* problem [130]. As a result, the only optimisation procedure [130] available in the literature for carrying out this task in full generality is very computationally intensive, and lacks convergence guarantees. In this chapter, we propose a novel optimisation procedure, baptised *Cadzow plug-and-play gradient descent (CPGD)*, which is more efficient than the state of the art and has local convergence guarantees.

2 Introduction to FRI

Sampling theorems are at the foundation of modern digital signal processing [148, 183] as they permit to navigate conveniently between the *analog* and *digital* worlds. The most famous of these theorems is undoubtedly the *Shannon sampling theorem* [158] which states that *bandlimited* signals can be recovered exactly from their discrete samples, provided a sufficient sampling rate. This major result has had tremendous impact on the field of signal processing and by extension on many fields of natural sciences. But this unanimous celebration has lead many scientists to start thinking about sampling theory exclusively in terms of bandlimitedness, which was nothing but a *sufficient* condition for a signal to admit a discrete representation. In reality, sampling theorems can also be devised for non-bandlimited signals, as long as they possess *finitely many* degrees of freedom (df).

This remarkable fact was brought to the attention of the signal processing community by Vetterli et al. in their seminal work [184], where they introduced the finite rate of innovation (FRI) framework, concerned with the sampling of *sparse* non-bandlimited signals such as the prototypical sparse signal, namely the T-periodic[1] *Dirac stream*:

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - Tk'), \quad \forall t \in \mathbb{R}, \tag{10.4}$$

[1] In the context of Section 1, we have of course $T = 2\pi$.

with $x_k \in \mathbb{C}$ and $t_k \in [0, T[$. In the FRI framework, the sparsity of a signal is measured in terms of its *rate of innovation*, defined as the number of degrees of freedom per unit of time. For instance, the Dirac stream (10.4) has $2K$ degrees of freedom $\{x_k, t_k\}_{k=1,\dots,K}$ per period T, yielding a finite rate of innovation of $\rho = 2K/T$. Intuitively, any lossless sampling scheme for (10.4) must therefore have a sampling rate at least as large as the rate of innovation ρ, or it will be impossible to fix all the degrees of freedom. In [25], Blu et al. described a sampling scheme achieving the second best sampling rate after the critical innovation rate, permitting to perfectly recover the signal innovations from the knowledge of $2K + 1$ consecutive Fourier coefficients of x.

Unfortunately, this scheme is very sensitive to noise perturbations in the collected samples. This is because the recovery of the innovations t_k relies on the resolution of a so-called *annihilating equation*, whose solvability requires the Toeplitz matrix built from the Fourier coefficients to be rank deficient. While this structural constraint is guaranteed to hold in the case of noiseless recovery of Dirac streams, it can break in the presence of noise, inevitable in practical applications. As a remedy to this stability issue, Blu et al. proposed

to *denoise* the collected samples prior to solving the annihilating equation. To this end, they leveraged the well-known *Cadzow algorithm* [33] which aims to retrieve the closest rank-deficient Toeplitz matrix to a high-dimensional embedding of the data via an *alternating projection method*. When upgraded with this extra denoising step, simulations results from Blu et al. in [25] revealed that the overall accuracy of the recovery procedure remains very good for a signal to noise ratio (SNR) as low as 5 dB. While the Cadzow algorithm empirically provides accurate results after a few iterations, its theoretical convergence has however not been demonstrated to date, due to the non-convex nature of the space of rank-deficient matrices. Condat and Hirabayashi [44] revisited Cadzow denoising as a structured low-rank approximation (SLRA) problem and proposed a *Douglas-Rachford splitting* algorithm to solve it [42], which has higher accuracy. Unfortunately, the gain comes at the price of significantly higher computational cost, the Douglas-Rachford splitting method requiring many more iterations to converge than Cadzow denoising. In addition to their somewhat heuristic nature, neither Cadzow denoising nor its upgrade can handle more general types of input measurements as considered in the generalised finite rate of innovation (genFRI) framework introduced by Pan et al. in [130]. The latter extends FRI to very generic cases, where the measurements are related to the unknown Fourier coefficients of signals satisfying the annihilating property by a linear map. In such configurations, it is therefore necessary to estimate both the Fourier coefficients and their corresponding annihilating filter. Pan et al. proposed to perform this joint estimation task by solving a non-convex constrained optimisation problem which recovers the Fourier coefficients, required to minimise a quadratic data-fidelity term, and their corresponding annihilating filter coefficients. The annihilating equation linking the two unknowns is moreover explicitly enforced as a constraint. They suggested to solve this optimisation problem via an iterative alternating minimisation algorithm with multiple random initialisations [130]. However, the proposed algorithm comes without convergence guarantees, requires fine tuning of many hyper-parameters, and is computationally intensive.

In this chapter, we propose to solve the implicit genFRI problem via proximal gradient descent (PGD) [16, 134]. We first consider PGD with *exact* proximal steps which is shown to converge towards critical points of the implicit genFRI problem. The latter is however impractical since the proximal step involved at each iteration does not have a closed-form expression. We therefore consider an *inexact* PGD [71], with proximal steps approximated by means of alternating projections, which amount to Cadzow denoising in the case of injective forward matrices $\boldsymbol{G}$. Such an approach is reminiscent of the plug-and-play (PnP) framework in which proximal operators involved in first-order iterative methods are replaced by generic denoisers [73, 154, 182]. For this reason, we baptise our reconstruction algorithm Cadzow plug-and-play gradient descent (CPGD). We demonstrate that CPGD converges locally towards fixed points of the update equation for injective matrices $\boldsymbol{G}$. Through simulations of irregular and noisy time sampling of periodic stream of Diracs we show that CPGD is more acurate and efficient by several orders of magnitude than the procedure proposed by Pan et al. in [130].

3 Preliminaries

In this section we introduce a linear operator, baptised *Toeplitzification operator*,[2] which transforms a vector into a Toeplitz matrix. This operator will be used in the regularisation term of our implicit genFRI optimisation problem. We then briefly review the method of alternating projections [55] as well as the FRI [184] framework and Cadzow denoising [44].

[2] The alternative appellation *Toeplitzication* is used in [44].

3.1 Toeplitzification Operator

Assume that we are given an arbitrary vector $\boldsymbol{x} \in \mathbb{C}^N$, $N = 2M+1$, with entries indexed as follows:

$$\boldsymbol{x} = [x_{-M}, x_{-M+1}, \dots, x_{M-1}, x_M]^\mathsf{T}.$$

Then, for any $P \leq M$, we can embed $\boldsymbol{x}$ into the space $\mathbb{T}_P$ of Toeplitz matrices of $\mathbb{C}^{(N-P)\times(P+1)}$ by means of the following *Toeplitzification operator*:

$$T_P : \begin{cases} \mathbb{C}^N & \to \mathbb{T}_P \subset \mathbb{C}^{(N-P)\times(P+1)} \\ \boldsymbol{x} & \mapsto [T_P(\boldsymbol{x})]_{i,j} := x_{-M+P+i-j}, \end{cases} \tag{10.5}$$

where $i = 1, \dots, N-P$, $j = 1, \dots, P+1$. Note from (10.5) that the value of an entry $[T_P(\boldsymbol{x})]_{i,j}$ of the matrix $T_P(\boldsymbol{x})$ depends only on the distance $i-j$ between the row and column indexes: $T_P(\boldsymbol{x})$ is therefore a *Toeplitz* matrix and the vector $\boldsymbol{x}$ is called its *generator*. The Toeplitzification operator (10.5) can be used to implement linear convolutions. Indeed, it can be shown (see Section 1 of Appendix B) that the multiplication of $T_P(\boldsymbol{x})$ with a vector $\boldsymbol{u} = [u_1, \cdots, u_{P+1}]^\mathsf{T} \in \mathbb{C}^{P+1}$ returns the valid part of the convolution between the two zero-padded sequences $\tilde{x} = \left[\dots, 0, x_{-M}, \dots, \boxed{x_0}, \dots, x_M, 0, \dots\right] \in \mathbb{C}^{\mathbb{Z}}$ and $\tilde{u} = \left[\dots, \boxed{0}, u_1, \dots, u_{P+1}, 0, \dots\right] \in \mathbb{C}^{\mathbb{Z}}$.

3.2 Inverse Toeplitzification Operator

The *inverse Toeplitzification* operator is the pseudo-inverse of the Toeplitzification operator, mapping a Toeplitz matrix $\boldsymbol{H} \in \mathbb{C}^{(N-P)\times(P+1)}$ to its generator $\boldsymbol{h} \in \mathbb{C}^N$. As we shall prove in Proposition 10.2, inverse Toeplitzification is achieved by averaging across each diagonal of $T_P(\boldsymbol{x})$. It is interesting to note that this operation is also leveraged in Cadzow denoising as described in [25], in order to map back the data from its high dimensional Toeplitz embedding. The formal interpretation of this inverse map as the pseudo-inverse of the Toeplitzification operator proposed hereafter is nevertheless not proven in [25], nor anywhere else we may be aware of.
To compute the pseudo-inverse of T_P, we first need an expression for its adjoint map, detailed in the proposition hereafter.

Proposition 10.1 — *Adjoint operator of* T_P. *The adjoint operator* T_P^* *of* T_P *defined in* (10.5) *is given by*

$$T_P^* : \begin{cases} \mathbb{C}^{(N-P)\times(P+1)} \to \mathbb{C}^N \\ \boldsymbol{H} \mapsto h_j = \displaystyle\sum_{i=k+j-1-P} H_{ik}, \quad j = 1, \ldots, N. \end{cases} \tag{10.6}$$

Proof. Consider a matrix $\boldsymbol{H} \in \mathbb{C}^{(N-P)\times(P+1)}$ and define the following Frobenius inner product

$$\begin{aligned} \langle T_P(\boldsymbol{x}), \boldsymbol{H} \rangle_F = \operatorname{tr}\left(T_P^H(\boldsymbol{x})\boldsymbol{H}\right) &= \sum_{i=1}^{N-P}\sum_{k=1}^{P+1} \overline{T_P(\boldsymbol{x})}_{ik} H_{ik} \\ &= \sum_{i=1}^{N-P}\sum_{k=1}^{P+1} \bar{x}_{-M+P+i-k} H_{ik} \overset{s=i-k+P}{=} \sum_{s=0}^{N-1} x_{-M+s} \left(\sum_{i=k+s-P} H_{ik} \right) \end{aligned} \tag{10.7}$$

The term $\displaystyle\sum_{i=k+(s-P)} H_{ik}$ sums the elements of $\boldsymbol{H}$ along lines with equation $i = k + (s - P)$. These lines have slope 1 and intercept $b = s - P$. Notice that these lines have non-null intersection with the lattice $(k, i) \in [1, P+1] \times [1, N-P]$ for $b \in [-P, N-P-1]$. Indeed, the two extreme cases occur when the lines hit the points $(1, N-P)$ and $(P+1, 1)$. This happens respectively when $1 + b = N - P \Rightarrow b = N - P - 1$ and $P + 1 + b = 1 \Rightarrow b = -P$. Since $s \in [0, N-1]$ the intercept b varies indeed in the range $[-P, N-P-1]$ and each term in the summation is non-null. The summation $\displaystyle\sum_{i=k+(s-P)} H_{ik}$ corresponds then to summing across each diagonal of $\boldsymbol{H}$. We finally get:

$$\langle T_P(\boldsymbol{x}), \boldsymbol{H} \rangle_F = \langle \boldsymbol{x}, T_P^*(\boldsymbol{H}) \rangle,$$

with

$$T_P^* : \begin{cases} \mathbb{C}^{(N-P)\times(P+1)} \to \mathbb{C}^N \\ \boldsymbol{H} \mapsto h_j = \displaystyle\sum_{i=k+j-1-P} H_{ik}, \quad j = 1, \ldots, N. \end{cases}$$

■

Note that the adjoint map T_P^* proceeds by summing across each diagonal of the input matrix $\boldsymbol{H}$. We are now ready to derive an expression for the (left) pseudo-inverse of T_P, described in the proposition hereafter.

Proposition 10.2 — *Pseudo-Inverse of* T_P. *The* pseudo-inverse

$$T_P^\dagger : \mathbb{C}^{(N-P)\times(P+1)} \to \mathbb{C}^N$$

of T_P defined in (10.5) is given by

$$T_P^\dagger = \mathbf{\Gamma}^{-1} T_P^*, \tag{10.8}$$

where $\mathbf{\Gamma} \in \mathbb{C}^{N\times N}$ is a diagonal matrix with diagonal entries given by:

$$\Gamma_{i,i} = \begin{cases} i & \text{for } 1 \leq i \leq P, \\ P+1 & \text{for } P < i \leq N-P, \\ N+1-i & \text{for } N-P < i \leq N. \end{cases} \tag{10.9}$$

Proof. From (10.6) and the definition of T_P, it is straightforward to observe that the operator $\mathbf{\Gamma} = T_P^* T_P : \mathbb{C}^N \to \mathbb{C}^N$ is a diagonal matrix, with diagonal entries given by:

$$\Gamma_{i,i} = \begin{cases} i & \text{for } i \leq P \\ P+1 & \text{for } P < i \leq N-P \\ N+1-i & \text{for } N-P < i \leq N. \end{cases} \tag{10.10}$$

The operator $T_P^\dagger = \mathbf{\Gamma}^{-1} T_P^*$ is hence a left inverse for T_P:

$$T_P^\dagger T_P = \mathbf{\Gamma}^{-1} T_P^* T_P = (T_P^* T_P)^{-1} T_P^* T_P = \boldsymbol{I}_N. \tag{10.11}$$

Moreover, the latter is actually the pseudo-inverse of T_P. Indeed, we have trivially:

$$T_P T_P^\dagger T_P = T_P, \qquad T_P^\dagger T_P T_P^\dagger = T_P^\dagger, \qquad (T_P^\dagger T_P)^* = T_P^\dagger T_P.$$

Finally, we have

$$(T_P T_P^\dagger)^* = T_P \mathbf{\Gamma}^{-*} T_P^* = T_P T_P^\dagger, \tag{10.12}$$

since $\mathbf{\Gamma}$ is diagonal and hence symmetric. $T_P^\dagger$ verifies thus the definition of the pseudo-inverse of T_P. ■

Observe that the composition of T_P^* and $\mathbf{\Gamma}^{-1}$ in the expression of the pseudo-inverse (10.8) corresponds indeed to a diagonal averaging: T_P^* first sums across each diagonal of the matrix $\boldsymbol{H} \in \mathbb{C}^{(N-P)\times(P+1)}$ and $\mathbf{\Gamma}^{-1}$ then divides the sums by the number of elements on each diagonal.

3.3 The Method of Alternating Projections

In this section we briefly discuss the method of alternating projections (MAP) [55], central to Cadzow denoising. It is used in computational mathematics to approximate projections onto intersecting sets. In its simplest form proposed by von Neumann in 1933 [186], the MAP performs a cascade of n projection steps onto subsets $\{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$ of some Hilbert space $\mathcal{H}$, starting from a point $z \in \mathcal{H}$:

$$\check{z} = \left[\Pi_{\mathcal{M}_K} \cdots \Pi_{\mathcal{M}_1}\right]^n (z). \tag{10.13}$$

In (10.13), $\Pi_{\mathcal{M}_k}$ denotes the orthogonal projection map onto $\mathcal{M}_k$, defined for $k = 1, \dots, K$ as

$$\Pi_{\mathcal{M}_k} : \begin{cases} \mathcal{H} \to \mathcal{M}_k, \\ z \mapsto \arg \min\limits_{x \in \mathcal{M}_k} \|z - x\|, \end{cases}$$

for some norm $\|\cdot\|$ on $\mathcal{H}$. In the case of closed linear subspaces $\{\mathcal{M}_1, \dots, \mathcal{M}_K\}$, von Neumann and Halperin showed that [12, 75, 186]

$$\lim_{n \to \infty} \left\| \left[\Pi_{\mathcal{M}_K} \cdots \Pi_{\mathcal{M}_1}\right]^n (z) - \Pi_{\bigcap_{k=1}^K \mathcal{M}_k}(z) \right\| = 0, \quad \forall z \in \mathcal{H}. \tag{10.14}$$

The MAP equation (10.13) can hence be used to approximate the complex projection map $\Pi_{\bigcap_{k=1}^K \mathcal{M}_k}$. For closed convex sets $\{\mathcal{M}_1, \dots, \mathcal{M}_K\}$, Bregman [55] showed moreover the weak convergence of the MAP towards a point in the intersection $\bigcap_{k=1}^K \mathcal{M}_k$. Strong convergence towards the actual projection was achieved by Dysktra's MAP [13], one of the most popular variant to von Neumann's original algorithm. In the case of non-convex intersecting sets finally, the convergence of the MAP has only been established *locally* [5, 107, 108, 175]. For example, Andersson et al. considered in [5] the case of two (potentially non-convex) finite-dimensional manifolds $\mathcal{M}_1, \mathcal{M}_2 \subset \mathcal{H}$ and showed the following local convergence result [5, Theorem 1.6]:

Theorem 10.3 — *Local Convergence of MAP for Non-Convex Sets [5].* Let $x \in \mathcal{M}_1 \cap \mathcal{M}_2$ be *non-tangential*, i.e. the angle between $\mathcal{M}_1$ and $\mathcal{M}_2$ at x is positive.[a] Then, for $z \in \mathcal{H}$ and $\epsilon > 0$, there exists $\delta \geq 0$ such that, if $\|x - z\| \leq \delta$,

$$\left[\Pi_{\mathcal{M}_2}\Pi_{\mathcal{M}_1}\right]^n (z) \overset{n \to \infty}{\to} z_\infty \in \mathcal{M}_1 \cap \mathcal{M}_2,$$

and

$$\|z_\infty - \Pi_{\mathcal{M}_1 \cap \mathcal{M}_2}(z)\| < \epsilon \|x - \Pi_{\mathcal{M}_1 \cap \mathcal{M}_2}(z)\|.$$

[a] See [5, Definition 4.2] and [5, Definition 4.3] for a precise definition of the angle between two manifolds and the concept of non-tangentiality.

Roughly speaking, Theorem 10.3 states that if the starting point z is close enough to a non-tangential point of $\mathcal{M}_1 \cap \mathcal{M}_2$ (which as explained in [5] are all but very exceptional points of $\mathcal{M}_1 \cap \mathcal{M}_2$), then the MAP converges to a point in $\mathcal{M}_1 \cap \mathcal{M}_2$. Moreover, the error $\|z_\infty - \Pi_{\mathcal{M}_1 \cap \mathcal{M}_2}(z)\|$ can be made arbitrarily small with respect to $\|x - \Pi_{\mathcal{M}_1 \cap \mathcal{M}_2}(z)\|$. Theorem 10.3 is however difficult to apply in practice since the value of δ guaranteeing a relative error below a given threshold ϵ is unknown. The MAP is hence often used as a heuristic in non-convex settings with no convergence guarantees. This is notably the case of Cadzow denoising, discussed further in Section 3.5.

3.4 FRI in a Nutshell

The classical FRI framework, introduced in [184], aims at estimating the innovations $\{(x_k, t_k),\ k = 1, \dots, K\} \subset \mathbb{C} \times [0, T[$, of a T-periodic[3] stream of

[3] In the context of Section 1, we have of course $T = 2\pi$.

Diracs:

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - Tk'), \quad \forall t \in \mathbb{R}.$$

In standard FRI, the estimation procedure is divided into two stages. The locations t_k are first estimated by a nonlinear method, and then arranged into a Vandermonde system whose solution yields the Dirac amplitudes [25]. The recovery of the locations t_k relies on the so-called *annihilating equation*, dating from Prony's work [147], which cancels out the Fourier series coefficients of x by convolving them with a particular filter, called the *annihilating filter*. The latter is defined as the finite-tap sequence $h = [\cdots, 0, h_0, h_1, \ldots, h_K, 0, \cdots] \in \mathbb{C}^{\mathbb{Z}}$, with z-transform vanishing at roots $\{u_k := e^{-j2\pi t_k/T},\ k = 1, \ldots, K\}$:

$$H(z) = \sum_{k=0}^{K} h_k z^{-k} = \prod_{k=1}^{K} (1 - u_k z^{-1}). \tag{10.15}$$

For such a filter, we have indeed

$$(\hat{x} * h)_m = \sum_{k=0}^{K} h_k \hat{x}_{m-k} = \sum_{k'=1}^{K} x_{k'} \left(\sum_{k=0}^{K} h_k u_{k'}^{-k} \right) u_{k'}^{m} = 0,\ m \in \mathbb{Z}, \tag{10.16}$$

where $\hat{x}_m = \sum_{k=1}^{K} x_k u_k^m, m \in \mathbb{Z}$, are the Fourier coefficients of x in (10.4). Notice that the roots u_k of the z-transform $H(z)$ in (10.15) of h are in one-to-one correspondence with the locations t_k. Recovering them amounts to estimating the coefficients $\boldsymbol{h} = [h_0, \ldots, h_K] \in \mathbb{C}^{K+1}$ of h from the annihilating equation (10.16). Assuming, for instance, that we dispose of $N = 2M + 1$ consecutive Fourier coefficients of x, e.g. $\boldsymbol{x} = [\hat{x}_{-M}, \ldots, \hat{x}_M] \in \mathbb{C}^{2M+1}$, we can extract the $N - K$ equations from (10.16) corresponding to the convolution indices $m = -M + K, \ldots, M$, and use the Toeplitzification operator[4] defined in (10.5) to form the following matrix equation:

$$T_K(\boldsymbol{x})\boldsymbol{h} = \boldsymbol{0}_{N-K}, \quad \|\boldsymbol{h}\| \neq 0. \tag{10.17}$$

[4] Remember the link between the Toeplitzification operator and convolution discussed in Section 3.1.

Observe that any nontrivial element of the nullspace of $T_K(\boldsymbol{x})$ is a solution to (10.17). For $M \geq K$, it can be shown [25] that $T_K(\boldsymbol{x})$ has rank K and therefore a nontrivial nullspace with dimension 1. Up to a multiplicative constant, the annihilating equation (10.17) admits hence a unique solution. The latter is obtained numerically by means of *total least-squares* [25], which computes the eigenvector associated to the smallest[5] eigenvalue of $T_K(\boldsymbol{x})$. In the critical case $M = K$, the matrix $T_K(\boldsymbol{x})$ is *square*, while in the *oversampling* case $M > K$ it is *rectangular and tall*. As explained in [25], oversampling makes the estimation procedure more resilient to potential noise perturbations in the Fourier coefficients. In such cases, Blu et al. recommend moreover to perform Cadzow denoising on the Fourier coefficients $\boldsymbol{x}$ (see Section 3.5) as well as replace (10.17) by a more general annihilating equation:

[5] An eigenvalue exactly equal to zero may in practice be impossible to obtain due to numerical inaccuracies.

$$T_P(\boldsymbol{x})\tilde{\boldsymbol{h}} = \boldsymbol{0}_{N-P}, \quad \|\tilde{\boldsymbol{h}}\| \neq 0. \tag{10.18}$$

with $K \leq P \leq M$,[6] and $\tilde{\boldsymbol{h}} \in \mathbb{C}^{P+1}$. The generalised annihilating equation (10.18) presents the advantage of involving more coefficients from $\boldsymbol{x}$, hence offering resilience to noise. Again, it is possible to show that $T_P(\boldsymbol{x})$ has rank K, and hence a nontrivial nullspace with dimension $P+1-K$. Solutions to (10.18) are hence nonunique in this case, but all equally valid for practical purposes.

[6] The case $P = M$ is reported to yield the best empirical results in [25].

3.5 Cadzow Denoising

For strong noise perturbations, the generalised annihilating equation (10.18) may fail to admit a nontrivial solution. Indeed, noisy generators $\boldsymbol{x}$ can yield full column rank matrices $T_P(\boldsymbol{x})$ with trivial nullspace. As a potential cure, Blu et al. proposed to denoise the Fourier coefficients $\boldsymbol{x}$ prior to solving the annihilating equation. This denoising step attempts to transform $T_P(\boldsymbol{x})$ into a Toeplitz matrix with rank at most K, thus guaranteeing the existence of non-trivial solutions to (10.18). This operation is carried out by means of *Cadzow denoising* [44], an alternating projection method (see Section 3.3) applied heuristically to the subspace $\mathbb{T}_P$ of Toeplitz matrices and the subset $\mathcal{H}_K$ of matrices with rank at most K:

$$\mathcal{H}_K := \left\{ \boldsymbol{M} \in \mathbb{C}^{(N-P)\times(P+1)} \mid \operatorname{rank} \boldsymbol{M} \leq K \right\}. \tag{10.19}$$

Using the notations introduced in Sections 3.1, 3.2 and 3.3, Cadzow denoising can be seen as processing the noisy coefficients $\boldsymbol{x}$ as follows:

$$\check{\boldsymbol{x}} = T_P^{\dagger} \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \right]^n T_P(\boldsymbol{x}), \tag{10.20}$$

for some suitable $n \in \mathbb{N}$. Note that the inverse Toeplitzification operator $T_P^{\dagger}$ applied after the alternating projection method is used to recover the denoised Fourier coefficients $\check{\boldsymbol{x}} \in \mathbb{C}^N$. Since $\mathcal{H}_K$ is a *non-convex* set the convergence of the MAP in (10.20) is not guaranteed. Nevertheless, experimental results [25, 44] suggest that Cadzow denoising almost always converges after a few iterations (typically $n \leq 20$), which could theoretically[7] be explained by the *local* convergence result in Theorem 10.3. We conclude this section by providing closed-form expressions for the projection operators $\Pi_{\mathbb{T}_P}$ and $\Pi_{\mathcal{H}_K}$, needed in (10.20).

[7] As explained in Section 3.3, the assumptions of Theorem 10.3 are in practice very difficult to verify in practice.

3.5.1 *Projection onto* $\mathbb{T}_P$

The orthogonal projection operator onto the subspace $\mathbb{T}_P \subset \mathbb{C}^{(N-P)\times(P+1)}$ of rectangular Toeplitz matrices can be written in terms of the Toeplitzification operator and its pseudo-inverse as:

$$\Pi_{\mathbb{T}_P} = T_P T_P^{\dagger} = T_P \boldsymbol{\Gamma}^{-1} T_P^{*}.$$

Proof. The operator T_P is actually a surjection onto the subspace $\mathbb{T}_P$ of rectangular Toeplitz matrices with size $(N-P)\times(P+1)$. Indeed, it is easy to see that every such matrix can be written as in (10.5) for some generator $\boldsymbol{x} \in \mathbb{C}^N$. Moreover, we have from (10.11) that $T_P^{\dagger} T_P = \boldsymbol{I}_N$ and hence from [183, Theo-

rem 2.29], $T_P T_P^\dagger$ is a projection operator onto the range $\mathbb{T}_P$ of T_P. Since $T_P T_P^\dagger$ is moreover self-adjoint from (10.12), it is actually an *orthogonal* projection operator, which achieves the proof. ■

3.5.2 Projection onto $\mathcal{H}_K$

The orthogonal projection operator onto the space $\mathcal{H}_K$ of matrices with rank at most K is given by the *Eckart-Young-Minsky theorem* [53]. The latter indeed states that the projection map

$$\Pi_{\mathcal{H}_K}(\boldsymbol{X}) = \arg\min_{\boldsymbol{H}\in\mathcal{H}_K} \|\boldsymbol{X}-\boldsymbol{H}\|_F, \quad \boldsymbol{X}\in\mathbb{C}^{(N-P)\times(P+1)},$$

can be computed in closed-form as:

$$\Pi_{\mathcal{H}_K}(\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{\Lambda}_K\boldsymbol{V}^*, \quad \boldsymbol{X}\in\mathbb{C}^{(N-P)\times(P+1)}, \tag{10.21}$$

where $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*$ is the singular value decomposition (SVD) of $\boldsymbol{X}$, and $\boldsymbol{\Lambda}_K$ is the diagonal matrix of sorted singular values truncated to the K strongest ones. Note that the output of the projection map is unique as long as the $K-$th and $(K+1)-$th largest singular values are different. Fortunately, this is almost always the case in practice, due to numerical inaccuracy.

4 Generalised FRI as an Inverse Problem

4.1 Generalised FRI

In Section 3.4, we described a procedure for recovering the locations t_k from consecutive Fourier coefficients of x. Remains now the issue of computing these Fourier coefficients from a collection of arbitrary *measurements* $\boldsymbol{y}\in\mathbb{C}^L$ of x, $L\geq N$. Blu et al. treated the simple scenario of measurements resulting from regular time sampling with ideal low-pass prefiltering [25]. In such a case, they showed that, for a well chosen prefilter bandwidth, the Fourier coefficients could simply be obtained by applying a discrete Fourier transform to the measurements $\boldsymbol{y}$. For more general measurement types, the situation is more complex, and the Fourier coefficients $\boldsymbol{x}\in\mathbb{C}^N$ must in general be estimated by solving a linear inverse problem:

$$\boldsymbol{y}=\boldsymbol{G}\boldsymbol{x}+\boldsymbol{n}, \tag{10.22}$$

where the forward matrix $\boldsymbol{G}\in\mathbb{C}^{L\times N}, L\geq N$, is application dependent, and $\boldsymbol{n}$ accounts for additive noise, usually assumed to be a white Gaussian random vector. In [130], Pan et al. have proposed the generalised finite rate of innovation (genFRI) optimisation problem for inverting (10.22). The latter is a non-convex constrained optimisation problem which jointly recovers the Fourier coefficients $\boldsymbol{x}\in\mathbb{C}^N$ –required to minimise a quadratic data-fidelity term– and their corresponding annihilating filter coefficients $\boldsymbol{h}\in\mathbb{C}^{P+1}$. The annihilating equation linking the two unknowns is moreover explicitly enforced

as a constraint, yielding an optimisation problem of the form:

$$\min_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \boldsymbol{h}\in\mathbb{C}^{P+1}}} \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \quad \text{subject to} \quad \begin{cases} T_P(\boldsymbol{x})\,\boldsymbol{h} = \boldsymbol{0}_{N-P}, \\ \|\boldsymbol{h}\| = 1. \end{cases} \tag{10.23}$$

[8] In [130], the authors also propose less natural normalisation strategies with improved empirical performances.

The normalisation[8] $\|\boldsymbol{h}\| = 1$ is used to exclude trivial solutions to the annihilating equation in (10.23). In the case of known noise level, the authors propose a relaxed version of (10.23), and suggest to solve it via an iterative alternating minimisation algorithm with multiple random initialisations. The latter comes however without convergence guarantees and is computationally very intensive.

4.2 Implicit Generalised FRI

The annihilating equation constraint in (10.23) can be thought of as regularising the genFRI problem. Indeed, minimising the quadratic term $\|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2$ alone in the presence of noise would not necessarily yield Fourier coefficients $\boldsymbol{x}$ with non-trivial annihilating filter, which the annihilating constraint enforces explicitly. Unfortunately, this regularisation also complicates significantly the optimisation procedure. Indeed, it requires the introduction of an extra unknown variable with non-linear dependency on the data, namely the annihilating filter $\boldsymbol{h}$. Moreover, the non-linear constraint $T_P(\boldsymbol{x})\,\boldsymbol{h} = \boldsymbol{0}_{N-P}$ is highly non-convex, and state-of-the-art algorithms, such as alternating minimisation or gradient descent [134], may suffer from getting trapped in local minima [36]. To circumvent these issues, we propose the following *implicit* formulation of the genFRI problem, in which only the Fourier coefficients are recovered:

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \quad \text{subject to} \quad \begin{cases} \operatorname{rank} T_P(\boldsymbol{x}) \leq K, \\ \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} \leq \rho, \end{cases} \tag{10.24}$$

where $K \leq P \leq M$, $\rho \in]0,+\infty]$, and $\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}$ is the norm induced by the diagonal and positive definite matrix $\boldsymbol{\Gamma} \in \mathbb{C}^{N\times N}$ in (10.10):

$$\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} := \sqrt{\boldsymbol{x}^H\boldsymbol{\Gamma}\boldsymbol{x}}, \quad \forall \boldsymbol{x} \in \mathbb{C}^N. \tag{10.25}$$

Similarly to (10.23), the quadratic term $\|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2$ in (10.24) is used to guarantee high fidelity of the recovered coefficients with the observed data. Unlike (10.23) however, (10.24) leverages a regularising rank constraint on $T_P(\boldsymbol{x})$ which does not explicitly involve the unknown annihilating filter. As already discussed in Section 3.5 in the context of Cadzow denoising, requiring $T_P(\boldsymbol{x})$ to be of rank at most K is indeed a sufficient condition for the generalised annihilating equation (10.18) to admit nontrivial solutions. This implicit regularisation greatly simplifies the genFRI problem, since it decouples the problem of estimating the Fourier coefficients from the problem of estimating the annihilating filter. The normalisation constraint $\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} \leq \rho$ finally, requires the recovered Fourier coefficients to have finite weighted energy (10.25). As shall be seen in Section 5, it can be relaxed when the forward matrix $\boldsymbol{G}$ is injective by

setting $\rho = +\infty$. Indeed, it is only used to ensure that the objective functional in (10.24) is *coercive* in underdetermined cases, where the forward matrix $\boldsymbol{G}$ has a nontrivial null space. Coercivity is indeed a key assumption [109] for the convergence of the proximal gradient descent method envisioned in Section 5.1 to solve (10.24). We conclude this section by noting that the choice of $\boldsymbol{\Gamma}$ as weighting matrix in the energy normalisation constraint is arbitrary and purely motivated by computational considerations. Indeed, any choice of positive definite weighting matrix in (10.25) would have been suitable for the sole purpose of making the objective functional coercive. As explained in Section 5.2 however, defining the weighting matrix as $\boldsymbol{\Gamma}$ greatly simplifies the computations involved at each iteration of the numerical solver proposed in Section 5.1.

5 Optimisation Algorithm

5.1 Non-Convex Proximal Gradient Descent

The optimisation problem (10.24) can be rewritten in unconstrained form as:

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \;+\; \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right) \;+\; \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}\left(\boldsymbol{x}\right), \tag{10.26}$$

where $\mathcal{H}_K$ is the non-convex set of matrices with rank lower than K defined in (10.19), $\mathbb{B}_\rho^{\boldsymbol{\Gamma}} := \{\boldsymbol{x}\in\mathbb{C}^N : \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} \leq \rho\}$ is the $\boldsymbol{\Gamma}$-ball with radius $\rho > 0$, and $\iota_{\mathcal{H}_K} : \mathbb{C}^{(N-P)\times(P+1)} \to \{0,+\infty\}$, $\iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}} : \mathbb{C}^N \to \{0,+\infty\}$ are indicator functions with domains $\mathcal{H}_K$ and $\mathbb{B}_\rho^{\boldsymbol{\Gamma}}$, respectively. Observe that the unconstrained optimisation problem (B.8) can be written as a sum between a *convex* and *differentiable* quadratic term

$$F(\boldsymbol{x}) := \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2, \qquad \boldsymbol{x}\in\mathbb{C}^N,$$

and a *non-convex* and *non-differentiable* term

$$H(\boldsymbol{x}) := \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right) \;+\; \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}(\boldsymbol{x}), \qquad \boldsymbol{x}\in\mathbb{C}^N.$$

It is moreover easy to see that the gradient of F

$$\nabla F(\boldsymbol{x}) = 2\boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x}\in\mathbb{C}^N, \tag{10.27}$$

is β-*Lipschitz continuous* with respect to the $\boldsymbol{\Gamma}$-norm (10.25), with Lipschitz constant given by

$$\begin{aligned}
\beta &= 2\|\boldsymbol{G}^H\boldsymbol{G}\|_{\boldsymbol{\Gamma}} \\
&= \sup\left\{2\|\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{x}\|_{\boldsymbol{\Gamma}} : \boldsymbol{x}\in\mathbb{C}^N,\, \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}=1\right\} \\
&= \sup\left\{2\left\|\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\tilde{\boldsymbol{x}}\right\|_2 : \tilde{\boldsymbol{x}}\in\mathbb{C}^N,\, \|\tilde{\boldsymbol{x}}\|_2=1\right\} \\
&= 2\left\|\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right\|_2.
\end{aligned} \tag{10.28}$$

It is hence possible to optimise (B.8) by means of proximal gradient descent (PGD) [134], an iterative method alternating between gradient and proximal

steps according to the following update equation:

$$\boldsymbol{x}_{k+1} \in \operatorname{prox}_{\tau H}^{\boldsymbol{\Gamma}}\left(\boldsymbol{x}_k - \tau \nabla F\left(\boldsymbol{x}_k\right)\right), \tag{10.29}$$

for $k \geq 0$, $\boldsymbol{x}_0 \in \mathbb{C}^N$, $\tau > 0$ and $\operatorname{prox}_{\tau H}^{\boldsymbol{\Gamma}}$ defined in (10.30). Given a current estimate $\boldsymbol{x}_k \in \mathbb{C}^N$, the update equation (B.2) decreases the value of the objective function (B.8) by selecting a *proximal point* [134] –with respect to H– of a target located at a distance τ from $\boldsymbol{x}_k$ along the direction of steepest descent $-\nabla F(\boldsymbol{x}_k)$. The operator mapping a point $\boldsymbol{x} \in \mathbb{C}^N$ to its proximal points with respect to H is called *proximal operator*, and is defined as [134]

$$\operatorname{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) : \begin{cases} \mathbb{C}^N \to \mathscr{P}\left(\mathbb{C}^N\right), \\ \boldsymbol{x} \mapsto \arg\min\limits_{\boldsymbol{z} \in \mathbb{C}^N} \dfrac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}}^2 + H(\boldsymbol{z}), \end{cases} \tag{10.30}$$

where $\mathscr{P}(\mathbb{C}^N)$ is the *power set* of $\mathbb{C}^N$, and $\tau > 0$ controls the relative importance of H with respect to the squared distance to $\boldsymbol{x}$ measured in terms of the $\boldsymbol{\Gamma}$-norm (10.25). The function H being non-convex, the proximal operator (10.30) will in general return *multiple* proximal points, which can all be used interchangeably in (B.2). The convergence of the sequence $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$ of PGD iterates (B.2) towards critical points of (B.8) is established in the following theorem.

Theorem 10.4 — *Convergence of PGD for Arbitrary $\boldsymbol{G}$.* Assume that $\rho \in]0, +\infty[$ in (B.8), and $\tau < 1/\beta$ with β defined in (B.2). Then, any *limit point* $\boldsymbol{x}_\star$ of the sequence $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$ generated by (B.2) is a *local minimum* of (B.8).

Proof. The proof of this theorem is given in Section 2 of Appendix B. ■

As stated by Theorem 10.5 hereafter, the convergence of PGD furthermore extends to the case $\rho = +\infty$, at least for injective forward matrices $\boldsymbol{G}$. Setting $\rho = +\infty$ in (10.24) can be interpreted as dropping the energy normalisation constraint, since $\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} \leq +\infty$ is trivially verified and hence the associated indicator function $\iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}$ in (B.8) is always null.

Theorem 10.5 — *Convergence of PGD for Injective $\boldsymbol{G}$.* Assume that $\rho = +\infty$ in (B.8), $\tau < 1/\beta$ with β defined in (B.2), and $\boldsymbol{G} \in \mathbb{C}^{L \times N}$ in (B.8) is *injective*, i.e. $\ker(\boldsymbol{G}) = \{\boldsymbol{0}_N\}$. Then, any *limit point* $\boldsymbol{x}_\star$ of the sequence $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$ generated by (B.2) is a *local minimum* of (B.8).

Proof. The proof of this theorem is given in Section 2 of Appendix B. ■

A practical implication of Theorem 10.5 is that, for injective forward matrices $\boldsymbol{G}$, PGD applied to the following relaxed implicit genFRI problem is convergent:

$$\min_{\boldsymbol{x} \in \mathbb{C}^N} \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \; + \; \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right), \tag{10.31}$$

where $F(\boldsymbol{x}) := \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2$, and $H(\boldsymbol{x}) := \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right)$. As discussed in Sec-

tion 5.2, (10.31) should always be favoured over (B.8) for injective forward matrices $\boldsymbol{G}$, since solving it via PGD requires less computations at each proximal step.

5.2 Cadzow PnP Gradient Descent

As seen in the previous section, PGD requires the computation of the proximal operator (10.30) at each iteration, which amounts to finding a minimiser to the following non-convex optimisation problem:

$$\check{\boldsymbol{x}} \in \arg\min_{\boldsymbol{z}\in\mathbb{C}^N} \left\{ \frac{1}{2\tau}\|\boldsymbol{x}-\boldsymbol{z}\|_{\boldsymbol{\Gamma}}^2 + \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{z}\right)\right) + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}\left(\boldsymbol{z}\right) \right\}, \tag{10.32}$$

for some input $\boldsymbol{x} \in \mathbb{C}^N$. Observe that the proximal step (10.32) can be seen as a generalised *projection* step, aiming to find a point $\check{\boldsymbol{x}}$ as close as possible –in terms of the $\boldsymbol{\Gamma}$-norm[9]– from $\boldsymbol{x}$ while verifying some convex and non-convex constraints specified by the indicator functions. This is formalised by Proposition 10.6, which shows that solutions to (10.32) can be identified with those of an orthogonal projection problem:

[9] Observe that the weighting matrix $\boldsymbol{\Gamma}$ puts more emphasis on the coefficients that appear more often in the Toeplitz matrix $T_P(\boldsymbol{x})$.

Proposition 10.6 — *Proximal Operator as MAP.* *The proximal operator (10.30) of $H(\boldsymbol{x}) := \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right) + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}(\boldsymbol{x})$, for $\rho \in]0,+\infty]$ and $K \leq P \leq M$ is given by*

$$\mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) = T_P^{\dagger}\Pi_{\mathbb{T}_P\cap\mathcal{H}_K\cap\mathbb{B}_\rho}T_P(\boldsymbol{x}), \quad \forall \boldsymbol{x}\in\mathbb{C}^N, \tag{10.33}$$

where $\mathbb{B}_\rho := \{\boldsymbol{X} \in \mathbb{C}^{(N-P)\times(P+1)} : \|\boldsymbol{X}\|_F \leq \rho\}$ and $\Pi_{\mathbb{T}_P\cap\mathcal{H}_K\cap\mathbb{B}_\rho}$ is the orthogonal projection operator *onto $\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho$ with respect to the* Frobenius norm:

$$\Pi_{\mathbb{T}_P\cap\mathcal{H}_K\cap\mathbb{B}_\rho}(\boldsymbol{X}) : \begin{cases} \mathbb{C}^{(N-P)\times(P+1)} \to \mathcal{P}\left(\mathbb{C}^{(N-P)\times(P+1)}\right), \\ \boldsymbol{X} \mapsto \arg\min_{\boldsymbol{H}\in\mathbb{T}_P\cap\mathcal{H}_K\cap\mathbb{B}_\rho} \|\boldsymbol{X}-\boldsymbol{H}\|_F . \end{cases}$$

Proof. Recall the definition of the proximal set associated to a point $\boldsymbol{x} \in \mathbb{C}^N$:

$$\mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}\in\mathbb{C}^N} \left\{ \frac{1}{2\tau}\|\boldsymbol{x}-\boldsymbol{z}\|_{\boldsymbol{\Gamma}}^2 + \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{z}\right)\right) + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}\left(\boldsymbol{z}\right) \right\}. \tag{10.34}$$

When mapped via the Toeplitzification operator T_P, the proximal set (10.34) becomes

$$\begin{aligned} &T_P\left(\mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x})\right) = \\ &= \left\{T_P(\check{\boldsymbol{x}}),\ \check{\boldsymbol{x}} \in \mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x})\right\} \\ &= \left\{\check{\boldsymbol{X}} \in \mathbb{T}_P,\ T_P^{\dagger}(\check{\boldsymbol{X}}) \in \mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x})\right\} \\ &= \arg\min_{\boldsymbol{Z}\in\mathbb{T}_P} \left\{ \frac{1}{2\tau}\|T_P^{\dagger}(\boldsymbol{Z})-\boldsymbol{x}\|_{\boldsymbol{\Gamma}}^2 + \iota_{\mathcal{H}_K}\left(\boldsymbol{Z}\right) + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}}\left(T_P^{\dagger}(\boldsymbol{Z})\right) \right\} \end{aligned}$$

$$= \arg \min_{\boldsymbol{Z} \in \mathbb{T}_P \cap \mathcal{H}_K} \left\{ \frac{1}{2\tau} \|T_P^\dagger(\boldsymbol{Z}) - \boldsymbol{x}\|_{\boldsymbol{\Gamma}}^2 + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}} \left(T_P^\dagger(\boldsymbol{Z}) \right) \right\}, \tag{10.35}$$

where we have used the fact that $T_P^\dagger T_P(\boldsymbol{z}) = \boldsymbol{z}$ for all $\boldsymbol{z} \in \mathbb{C}^N$. We have moreover:

$$\begin{aligned}
\|T_P^\dagger(\boldsymbol{Z}) - \boldsymbol{x}\|_{\boldsymbol{\Gamma}}^2 =& \langle \boldsymbol{\Gamma} T_P^\dagger(\boldsymbol{Z}) - \boldsymbol{x}, T_P^\dagger(\boldsymbol{Z}) - \boldsymbol{x} \rangle_2 \\
=& \langle \boldsymbol{\Gamma} T_P^\dagger(\boldsymbol{Z}), T_P^\dagger(\boldsymbol{Z}) \rangle_2 + \langle \boldsymbol{\Gamma}\boldsymbol{x}, \boldsymbol{x} \rangle_2 - \langle \boldsymbol{\Gamma} T_P^\dagger(\boldsymbol{Z}), \boldsymbol{x} \rangle_2 - \langle \boldsymbol{\Gamma}\boldsymbol{x}, T_P^\dagger(\boldsymbol{Z}) \rangle_2 \\
=& \langle \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{Z}), \boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{Z}) \rangle_2 + \langle T_P^* T_P(\boldsymbol{x}), \boldsymbol{x} \rangle_2 \\
& - \langle \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{Z}), \boldsymbol{x} \rangle_2 - \langle \boldsymbol{\Gamma}\boldsymbol{x}, \boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{Z}) \rangle_2 \\
=& \langle \boldsymbol{Z}, T_P \boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{Z}) \rangle_F + \langle T_P(\boldsymbol{x}), T_P(\boldsymbol{x}) \rangle_F - \langle \boldsymbol{Z}, T_P(\boldsymbol{x}) \rangle_F \\
& - \langle T_P(\boldsymbol{x}), \boldsymbol{Z} \rangle_F \\
=& \langle \boldsymbol{Z}, \Pi_{\mathbb{T}_P}(\boldsymbol{Z}) \rangle_F + \|T_P(\boldsymbol{x})\|_F^2 - \langle \boldsymbol{Z}, T_P(\boldsymbol{x}) \rangle_F - \langle T_P(\boldsymbol{x}), \boldsymbol{Z} \rangle_F \\
=& \langle \boldsymbol{Z}, \boldsymbol{Z} \rangle_F + \|T_P(\boldsymbol{x})\|_F^2 - \langle \boldsymbol{Z}, T_P(\boldsymbol{x}) \rangle_F - \langle T_P(\boldsymbol{x}), \boldsymbol{Z} \rangle_F \\
=& \|\boldsymbol{Z}\|_F^2 + \|T_P(\boldsymbol{x})\|_F^2 - \Re \left(\langle \boldsymbol{Z}, T_P(\boldsymbol{x}) \rangle_F \right) \\
=& \|\boldsymbol{Z} - T_P(\boldsymbol{x})\|_F^2, \qquad \forall \boldsymbol{Z} \in \mathbb{T}_P,
\end{aligned} \tag{10.36}$$

where we have used the fact that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^H = T_P^* T_P$ and $T_P \boldsymbol{\Gamma}^{-1} T_P^* = \Pi_{\mathbb{T}_P}$ (see Propositions 10.1 and 10.2). With similar arguments, we have $\forall \boldsymbol{Z} \in \mathbb{T}_P$:

$$\begin{aligned}
\left\| T_P^\dagger(\boldsymbol{Z}) \right\|_{\boldsymbol{\Gamma}} \leq \rho &\Leftrightarrow \sqrt{\langle \boldsymbol{\Gamma} T_P^\dagger(\boldsymbol{Z}), T_P^\dagger(\boldsymbol{Z}) \rangle_2} \leq \rho \\
&\Leftrightarrow \sqrt{\langle \boldsymbol{Z}, \boldsymbol{Z} \rangle_F} \leq \rho \\
&\Leftrightarrow \|\boldsymbol{Z}\|_F \leq \rho.
\end{aligned}$$

so that

$$\iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}} \left(T_P^\dagger(\boldsymbol{Z}) \right) = \iota_{\mathbb{B}_\rho} (\boldsymbol{Z}), \quad \forall \boldsymbol{Z} \in \mathbb{T}_P, \tag{10.37}$$

where $\mathbb{B}_\rho := \left\{ \boldsymbol{Z} \in \mathbb{C}^{(N-P)\times(P+1)} : \|\boldsymbol{Z}\|_F \leq \rho \right\}$. Plugging (10.36) and (10.37) into (10.35) hence yields

$$\begin{aligned}
& T_P \left(\mathrm{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) \right) = \\
&= \arg \min_{\boldsymbol{Z} \in \mathbb{T}_P \cap \mathcal{H}_K} \left\{ \frac{1}{2\tau} \|T_P^\dagger(\boldsymbol{Z}) - \boldsymbol{x}\|_{\boldsymbol{\Gamma}}^2 + \iota_{\mathbb{B}_\rho^{\boldsymbol{\Gamma}}} \left(T_P^\dagger(\boldsymbol{Z}) \right) \right\} \\
&= \arg \min_{\boldsymbol{Z} \in \mathbb{T}_P \cap \mathcal{H}_K} \left\{ \frac{1}{2\tau} \|\boldsymbol{Z} - T_P(\boldsymbol{x})\|_F^2 + \iota_{\mathbb{B}_\rho} (\boldsymbol{Z}) \right\} \\
&= \arg \min_{\boldsymbol{Z} \in \mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho} \left\{ \frac{1}{2\tau} \|\boldsymbol{Z} - T_P(\boldsymbol{x})\|_F^2 \right\} \\
&= \arg \min_{\boldsymbol{Z} \in \mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho} \|\boldsymbol{Z} - T_P(\boldsymbol{x})\|_F \\
&= \Pi_{\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho} T_P(\boldsymbol{x}).
\end{aligned} \tag{10.38}$$

Using the fact that $T_P^\dagger T_P = \boldsymbol{I}_N$ we can finally rewrite (10.38) as

$$\operatorname{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) = T_P^\dagger \Pi_{\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho} T_P(\boldsymbol{x}),$$

which completes the proof. ■

Equation (10.38) provides us with a practical way of computing the proximal set (10.30) associated to a point $\boldsymbol{x} \in \mathbb{C}^N$. Unfortunately, the orthogonal projection operator $\Pi_{\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho}$ admits no simple closed-form expression. We therefore propose to approximate it by the method of alternating projections (MAP) (see Section 3.3):

$$\Pi_{\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho} \simeq \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}\right]^n, \tag{10.39}$$

for some $n \in \mathbb{N}$. Observe that when $\rho = +\infty$ (which is possible for injective matrices $\boldsymbol{G}$, see Theorem 10.5) we have $\Pi_{\mathbb{B}_\rho} = \mathrm{Id}$ and hence the right-hand side of (10.39) simplifies to $[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K}]^n$. Note that since $\mathcal{H}_K$ is non-convex, the convergence as n grows to infinity of the product $\left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}\right]^n$ towards the actual projection map $\Pi_{\mathbb{T}_P \cap \mathcal{H}_K \cap \mathbb{B}_\rho}$ is not guaranteed in general (see discussion in Section 3.3). For the specific case $\rho = +\infty$ however, it is possible to apply Theorem 10.3 to show the local convergence of the MAP (10.39):

Corollary 10.7 — *Convergence of Approximate Proximal Operator.* *Let $\boldsymbol{Z} \in \mathcal{H}_K \cap \mathbb{T}_P$ be a non-tangential point [5, Definition 4.3].*
Then, for $\boldsymbol{X} \in \mathbb{C}^{(N-P)\times(P+1)}$ and $\epsilon > 0$, there exists $\delta \geq 0$ such that, if $\|\boldsymbol{X} - \boldsymbol{Z}\|_F \leq \delta$,

1. $\left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K}\right]^n (\boldsymbol{X}) \overset{n\to\infty}{\to} \boldsymbol{X}_\infty \in \mathcal{H}_K \cap \mathbb{T}_P$,
2. $\|\boldsymbol{X}_\infty - \Pi_{\mathcal{H}_K \cap \mathbb{T}_P}(\boldsymbol{X})\|_F < \epsilon \|\boldsymbol{X} - \Pi_{\mathcal{H}_K \cap \mathbb{T}_P}(\boldsymbol{X})\|_F$.

Proof. Similarly to the proof of [6, Theorem 7], Corollary 10.7 is obtained by applying Theorem 10.3 to the manifolds $\mathcal{M}_1 = \mathcal{R}_K$ of matrices with rank exactly K –which is dense in $\mathcal{H}_K$ [5, Proposition 2.1]– and $\mathcal{M}_2 = \mathbb{T}_P$. For more details, see the proof of [6, Theorem 7], which discusses the local convergence of the MAP for $\mathcal{H}_K \cap \mathbb{H}_P$ where $\mathbb{H}_P$ denotes the space of rectangular Hankel matrices. Since Hankel matrices are just reflected Toeplitz matrices, their analysis extends easily to the case of Toeplitz matrices. ■

Roughly speaking, Corollary 10.3 states that, if applied to a matrix $\boldsymbol{X}$ close enough to a non-tangential point of $\mathbb{T}_P \cap \mathcal{H}_K$ (which as discussed in [5] for the case of Hankel matrices are all but very exceptional matrices of $\mathbb{T}_P \cap \mathcal{H}_K$), the MAP (10.39) converges to a point in $\mathbb{T}_P \cap \mathcal{H}_K$. Moreover, the error

$$\|\boldsymbol{X}_\infty - \Pi_{\mathcal{H}_K \cap \mathbb{T}_P}(\boldsymbol{X})\|_F$$

can be made arbitrarily small with respect to $\|\boldsymbol{X} - \Pi_{\mathcal{H}_K \cap \mathbb{T}_P}(\boldsymbol{X})\|_F$. While difficult to verify in practice, the local convergence result Corollary 10.7 reassures us however on the well-foundedness of approximation (10.39).

Algorithm 10.12: Cadzow plug-and-play gradient descent (CPGD)

procedure CPGD($\boldsymbol{y}, \boldsymbol{x}_0, \tau, n, \rho, \epsilon$)
  $k = 0$
  repeat
    $k \leftarrow k + 1$
    $\boldsymbol{z}_k = \boldsymbol{x}_{k-1} - 2\tau \boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{x}_{k-1} - \boldsymbol{y})$
    if $\rho = +\infty$ **then**
      $\boldsymbol{x}_k = T_P^\dagger \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K}\right]^n T_P(\boldsymbol{z}_k)$
    else
      $\boldsymbol{x}_k = T_P^\dagger \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}\right]^n T_P(\boldsymbol{z}_k)$
  until $\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\| \leq \epsilon \|\boldsymbol{x}_{k-1}\|$
  return $\boldsymbol{x}_n$

Plugging (10.39) into (10.38) finally yields the following approximate proximal step:

$$\operatorname{prox}_{\tau H}^{\boldsymbol{\Gamma}}(\boldsymbol{x}) \simeq T_P^\dagger \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}\right]^n T_P(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{C}^N, \tag{10.40}$$

for some $n \geq 0$. The PGD algorithm with approximate proximal step (10.40) is provided in Algorithm 10.12. Observe that when $\rho = +\infty$, (10.40) reduces to Cadzow denoising (10.20). The effect of heuristic (10.39) is hence to replace the proximal step in the PGD iterations by a generic *denoising* step. Such an approach is reminiscent of the plug-and-play (PnP) framework [154, 182] in image processing, which leverages deep learning-based denoisers to approximate complex projection or proximal operators [73]. For this reason, we baptise our algorithm Cadzow plug-and-play gradient descent (CPGD). In the next section, we study the convergence of Algorithm 10.12.

5.3 Local Fixed-Point Convergence of CPGD

In Section 5.1, we established Theorems 10.4 and 10.5 which show the convergence of PGD towards critical points of (B.8). However, such results required the computation of exact proximal steps (10.30) in the PGD iterations, and do not apply to CPGD which leverages the inexact proximal step (10.40). Convergence of PGD in non-convex setups with inexact proximal steps was studied in [71, 199]. The results established in both papers require the proximal step approximation errors incurred at each iteration to be decreasing and summable, which may not necessarily be the case for the MAP approximation (10.39). It is nevertheless possible to demonstrate that the iterations of CPGD are locally *contractive*, and therefore locally convergent towards a fixed point by the Banach contraction principle.

Theorem 10.8 — *CPGD is a Local Contraction.* Let $\mathcal{R}_K \subset \mathbb{C}^{(N-P)\times(P+1)}$ be the set of matrices of rank exactly $K \leq P \leq \lfloor N/2 \rfloor$, and $U_{\tau,n} : \mathbb{C}^N \to \mathbb{C}^N$

the update CPGD map

$$U_{\tau,n}(\boldsymbol{x}) := H_n\left(\boldsymbol{x} - \tau\nabla F(\boldsymbol{x})\right), \qquad \boldsymbol{x} \in \mathbb{C}^N, \tag{10.41}$$

with $H_n(\boldsymbol{x}) := T_P^{\dagger}[\Pi_{\mathbb{T}_P}\Pi_{\mathcal{H}_K}\Pi_{\mathbb{B}_\rho}]^n T_P(\boldsymbol{x})$. Let $\boldsymbol{G} \in \mathbb{C}^{L\times N}$ be injective, and $\boldsymbol{\Gamma}$ be the diagonal and positive definite matrix defined in (10.10). Define

$$\alpha := 2\lambda_{min}\left(\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right),$$
$$\beta := 2\lambda_{max}\left(\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right),$$

where $\lambda_{min}(\boldsymbol{M})$ and $\lambda_{max}(\boldsymbol{M})$ denote the minimum and maximum eigenvalues of a matrix $\boldsymbol{M}$ respectively.

Then, $U_{\tau,n}$ is locally well-defined (single-valued) and *Lipschitz continuous* with respect to the $\boldsymbol{\Gamma}$-norm

$$\|U_{\tau,n}(\boldsymbol{x}) - U_{\tau,n}(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} \leq L_\tau\|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}},$$

for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$ such that $T_P(\boldsymbol{x}), T_P(\boldsymbol{z})$ are in some neighbourhood of some matrix $\boldsymbol{R} \in \mathcal{R}_K$. The *Lipschitz constant* L_τ is given by

$$L_\tau = \max\left\{|1 - \tau\alpha|, |1 - \tau\beta|\right\}.$$

Moreover, $U_{\tau,n}$ is *contractive*, i.e. $0 < L_\tau < 1$, for $0 < \tau < 2/\beta$, and minimised for $\tau = 2/(\alpha + \beta)$.

Proof. The proof of this theorem is given in Section 3 of Appendix B. ■

The following corollary shows the local convergence of CPGD towards a fixed-point of the update map (10.41):

Corollary 10.9 — *CPGD Converges Locally.* *With the same notations as in Theorem 10.8, assume that all CPGD iterates $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ are such that*

$$\{T_P(\boldsymbol{x}_{k+1}), T_P(\boldsymbol{x}_k)\} \subset \mathcal{U}_k, \quad \forall k \in \mathbb{N}, \tag{10.42}$$

for some neighbourhood $\mathcal{U}_k$ of some point $\boldsymbol{R}_k \in \mathcal{R}_K$. Assume further that $0 < \tau < 2/\beta$. Then, $\boldsymbol{x}_k \overset{k\to\infty}{\to} \boldsymbol{x}_\star$ where $\boldsymbol{x}_\star \in \mathbb{C}^N$ is a fixed-point *of $U_{\tau,n}$, i.e. $U_{\tau,n}(\boldsymbol{x}_\star) = \boldsymbol{x}_\star$. Moreover, we have*

$$\|\boldsymbol{x}_\star - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} \leq \frac{L_\tau^k}{1 - L_\tau}\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}}, \quad \forall k \geq 1. \tag{10.43}$$

Proof. First, we note that from Theorem 10.8, we have under the assumptions of the corollary that

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} \leq \|U_{\tau,n}(\boldsymbol{x}_k) - U_{\tau,n}(\boldsymbol{x}_{k-1})\|_{\boldsymbol{\Gamma}} \leq L_\tau\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\|_{\boldsymbol{\Gamma}},$$

for all $k \geq 1$ and hence by induction

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} \leq L_\tau^k \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}}, \quad \forall k \geq 1. \tag{10.44}$$

By assumption $0 < \tau < 2/\beta$ and therefore $0 < L_\tau < 1$. We deduce hence from (10.44) that $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Let $j, k \in \mathbb{N}$ with $j > k$:

$$\begin{aligned}
\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} &\leq \sum_{m=k}^{j-1} \|\boldsymbol{x}_{m+1} - \boldsymbol{x}_m\|_{\boldsymbol{\Gamma}} \\
&\leq \sum_{m=k}^{j-1} L_\tau^m \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}} \\
&= \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}} L_\tau^k \sum_{m=0}^{j-1-k} L_\tau^m \qquad (10.45) \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}} L_\tau^k \sum_{m=0}^{\infty} L_\tau^m \\
&= \frac{L_\tau^k}{1 - L_\tau} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}}.
\end{aligned}$$

For every $\epsilon > 0$, we can choose a $J \in \mathbb{N}$ such that

$$L_\tau^J < \frac{\epsilon(1 - L_\tau)}{\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}}},$$

and hence for all $j > k > J$

$$\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} < \epsilon.$$

The sequence $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ is hence a Cauchy sequence, and since $\mathbb{C}^N$ is complete, it converges towards a limit point $\boldsymbol{x}_\star \in \mathbb{C}^N$. We have moreover, since $U_{\tau,n}$ is continuous

$$\boldsymbol{x}_\star = \lim_{n\to\infty} \boldsymbol{x}_k = \lim_{n\to\infty} U_{\tau,n}(\boldsymbol{x}_{k-1}) = U_{\tau,n}\left(\lim_{n\to\infty} \boldsymbol{x}_{k-1}\right) = U_{\tau,n}(\boldsymbol{x}_\star),$$

and hence $\boldsymbol{x}_\star$ is a fixed-point of $U_{\tau,n}$. Note moreover that, from (10.45) we get

$$\begin{aligned}
\|\boldsymbol{x}_\star - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} &= \lim_{j\to+\infty} \|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\boldsymbol{\Gamma}} \\
&\leq \lim_{j\to+\infty} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}} L_\tau^k \sum_{m=0}^{j-1-k} L_\tau^m \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}} L_\tau^k \sum_{m=0}^{+\infty} L_\tau^m = \frac{L_\tau^k}{1 - L_\tau} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\boldsymbol{\Gamma}},
\end{aligned}$$

which proves (10.43) of Corollary 10.9. ■

Remark 10.1 — *Speed of Convergence.* *From Theorem 10.8 and (10.9), we*

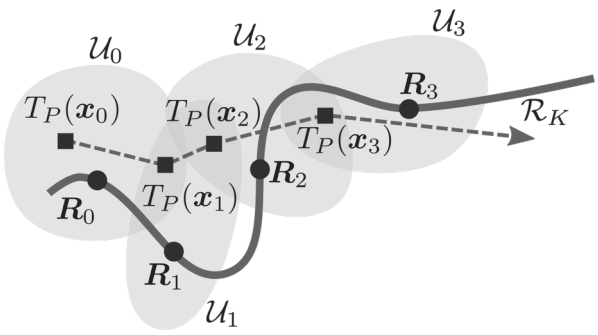


Figure 10.1: Geometric interpretation of condition (10.42) in Corollary 10.9.

see that the sequence $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ *converges the fastest when* L_τ *is minimised, i.e.* $\tau = 2/(\alpha+\beta)$.

Remark 10.2 — *Fixed Points vs. Critical Points.* *Note that Corollary 10.9 is a much weaker result than Theorems 10.4 and 10.5. Indeed, Corollary 10.9 only shows the local convergence of CPGD towards fixed points of* $U_{\tau,n}$*, which may not necessarily be critical points of the optimisation problem (B.8). Theorems 10.4 and 10.5 on the other hand, show the global convergence of PGD with exact proximal step towards critical points of (B.8). This is however the price to pay for computing the proximal step (10.32) efficiently in practice.*

Remark 10.3 — *Geometric Interpretation of Condition (10.42).* *Roughly speaking, Corollary 10.9 guarantees the convergence of CPGD towards a fixed point of the update map (10.41), provided that the forward matrix* $\boldsymbol{G}$ *is injective, and that any two consecutive lifted estimates* $T_P(\boldsymbol{x}_k)$, $T_P(\boldsymbol{x}_{k+1})$, *are in a common neighbourhood* $\mathcal{U}_k$ *of some matrix* $\boldsymbol{R}_k \in \mathcal{R}_K$*. Note that this is much less stringent than requiring the entire lifted path* $\{T_P(\boldsymbol{x}_k)\}_{k\in\mathbb{N}}$ *to belong to some neighbourhood* $\mathcal{U}$ *of some fixed matrix* $\boldsymbol{R} \in \mathcal{R}_K$*. Indeed, condition (10.42) allows for the lifted estimates to travel from one neighbourhood of the manifold* $\mathcal{R}_K$ *to another, provided that every visited neighbourhood contains at least two consecutive lifted estimates (see Fig. 10.1 for an illustration). This condition, although difficult to verify in practice, seems however likely to hold for* $\rho = +\infty$*, small enough step sizes, large enough* n *and* $\boldsymbol{x}_0 = \boldsymbol{0}_N$*. Indeed, in such a case, we have:*

- $T_P(\boldsymbol{x}_0) \in \mathcal{H}_K$ *is in some neighbourhood of* $\mathcal{R}_K$ *since* $\mathcal{R}_K$ *is dense in* $\mathcal{H}_K$*.*
- *For* n *large enough,* $T_P(\boldsymbol{x}_k)$ *is very likely to be in some neighbourhood of* $\mathcal{R}_K$*, since the denoising step in the update map (10.41) makes* $T_P(\boldsymbol{x}_k)$ *close to be in the intersection* $\mathcal{H}_K \cap \mathbb{T}_P$ *(see Corollary 10.7).*
- *For a small enough step size* τ*,* $T_P(\boldsymbol{x}_k)$ *and* $T_P(\boldsymbol{x}_{k+1})$ *are likely to belong to the same neighbourhood of* $\mathcal{R}_K$*.*

6 Experimental Results

In this section we validate the CPGD method numerically, considering as a testbed the scenario of irregular time sampling from [130, Section IV.A]. More

precisely, we define a 1-periodic stream of $K = 9$ Diracs (see Fig. 10.2):

$$x(t) = \sum_{m\in\mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - m), \quad \forall t \in \mathbb{R}, \tag{10.46}$$

where the amplitudes $x_k \in \mathbb{R}_+$ and locations $t_k \in [0, 1[$ are random, with log-normal and uniform[10] distributions respectively. We then generate $N = 2M+1$ noisy samples as

$$y_n = \sum_{m=-M}^{M} \hat{x}_m \exp(j2\pi m\theta_n) \quad + \quad \epsilon_n, \qquad n = 1, \dots, N, \tag{10.47}$$

where $\hat{x}_m = \sum_{k=1}^{K} x_k \exp(-j2\pi m t_k)$ are the Fourier coefficients of the Dirac stream x, $\{\theta_n\}_{n=1,\dots,N} \subset [0, 1[$ are chosen uniformly[11] at random, and $\{\epsilon_n\}_{n=1,\dots,N}$ are independent realisations of a Gaussian random variable $\mathcal{N}(0, \sigma^2)$. As explained in [130, Section IV.A], the samples y_n thus obtained correspond to noisy samples of the low-pass filtered[12] Dirac stream x at irregular times $\{\theta_n\}_{n=1,\dots,N} \subset [0, 1[$ (see Fig. 10.2). Using the formalism of Section 4, we can rewrite (10.47) as

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{10.48}$$

where $\boldsymbol{x} = [\hat{x}_{-M}, \dots, \hat{x}_M] \in \mathbb{C}^N$, $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_N] \in \mathbb{R}^N$, and $\boldsymbol{G} \in \mathbb{C}^{N\times N}$ is given by

$$\boldsymbol{G} = \begin{bmatrix} e^{-j2\pi M\theta_1} & \dots & 1 & \dots & e^{j2\pi M\theta_1} \\ e^{-j2\pi M\theta_2} & \dots & 1 & \dots & e^{j2\pi M\theta_2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ e^{-j2\pi M\theta_{N-1}} & \dots & 1 & \dots & e^{j2\pi M\theta_{N-1}} \\ e^{-j2\pi M\theta_N} & \dots & 1 & \dots & e^{j2\pi M\theta_N} \end{bmatrix}.$$

Note that from the periodicity of complex exponentials, it is possible to flip the columns of $\boldsymbol{G}$ so as to rewrite it as a Vandermonde matrix [130]. This shows that $\boldsymbol{G}$ is *injective* –since the irregular time samples are all distinct. From the samples $\boldsymbol{y}$ and the data model (10.48), we consider recovering the Fourier coefficients $\boldsymbol{x} \in \mathbb{C}^N$ by means of three algorithms:

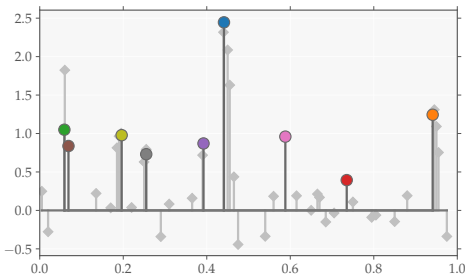
- The CPGD algorithm 10.12 with $\rho = +\infty$ (since $\boldsymbol{G}$ is injective) and step size $\tau = 1.5/\beta$ (where β is as in Theorem 10.8).
- The state-of-the-art algorithm of Pan et al. [130], referred to hereafter as GenFRI. We use the Python 3 implementation of GenFRI provided by Pan et al. in their official Github repository [76]. We moreover set the number of random initialisations to its default value 50.
- The baseline method, referred to hereafter as LS-Cadzow, which consists in naively applying Cadzow denoising to the least-squares estimate of the Fourier coefficients

$$\begin{cases} \boldsymbol{x}_{\text{LS}} = \underset{\boldsymbol{x}\in\mathbb{C}^N}{\arg\min} \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2, \\ \boldsymbol{x}_{\text{LS-Cadzow}} = T_P^{\dagger} \left[\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K}\right]^n T_P\left(\boldsymbol{x}_{\text{LS}}\right). \end{cases} \tag{10.49}$$

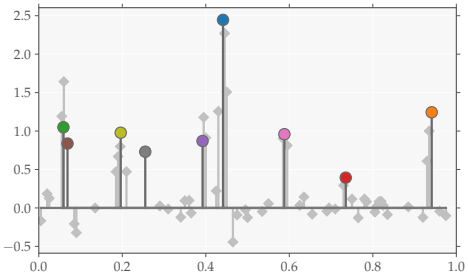
[10] To avoid degenerate cases, the Diracs are furthermore required to have a minimum separation distance of 1% of the total period.

[11] To avoid degenerate cases, the sampling locations are furthermore required to have a minimal separation distance of 0.5% of the total period.

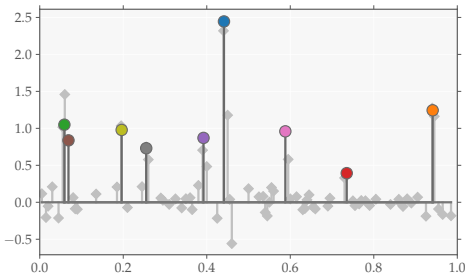
[12] Where the low-pass filter is chosen as an *ideal* low-pass filter with bandwidth $2M + 1$.



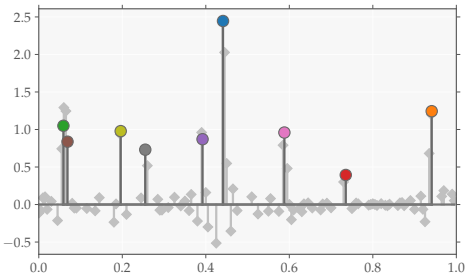
(a) $\beta = 2$



(b) $\beta = 3$



(c) $\beta = 4$



(d) $\beta = 5$

Figure 10.2: Dirac stream with $K = 9$ sources (dark grey, round coloured heads) and noiseless irregular time samples (light grey, diamond heads), for various oversampling parameters β.

We solve the least-squares optimisation problem in (10.49) by means of the `lstsq` function in the Python 3 package `numpy` [129], with cut-off ratio `rcond` $= 5 \times 10^{-5}$.

For CPGD and GenFRI, we fix the maximum number of iterations to 500 and consider that convergence is reached if the iterate norm is changed by less than 0.08% from one iteration to the other. For Cadzow denoising, we fix the number of iterations to 10 both for LS-Cadzow and CPGD. The reconstruction accuracy is assessed by matching the true Dirac locations $\{t_k\}_{k=1,\dots,K}$ to the recovered ones, denoted by $\{\omega_k\}_{k=1,\dots,K}$. To do so, we proceed as explained in Section 3.4 and infer the Dirac locations $\{\omega_k\}_{k=1,\dots,K}$ from the z-transform roots of the annihilating filter associated to the Fourier coefficients estimated by each method.[13] Then, we solve by means of the *Hungarian algorithm*[14] [96] the following matching problem

$$\min \left\{ \frac{1}{K} \sum_{k=1}^{K} d(t_k, \omega_{j_k}),\ j_1, \dots, j_K \in \{1, \dots, K\} \right\}, \tag{10.50}$$

where $d(t, \omega) = \min\{|t - \omega|, 1 - |t - \omega|\}\ \forall t, \omega \in [0, 1[$ is the canonical distance on the periodised interval $[0, 1[$. Finally, we report the average positioning error, corresponding to the value of the cost function $\sum_{k=1}^{K} c(t_k, \omega_{i_k})/K$ for the indices $\{i_1, \dots, i_K\}$ solutions to the matching problem (10.50). This metric is computed for 192 noise realisations, different $M = \beta K$ with the oversampling factor $\beta \in \{2, 3, 4, 5\}$ (see Figs. 10.2a, 10.2b, 10.2c and 10.2d respectively) and various noise levels

$$\sigma = \max_{k=1,\dots,K} |x_k| \times \exp\left(-\frac{\text{PSNR}}{10}\right),$$

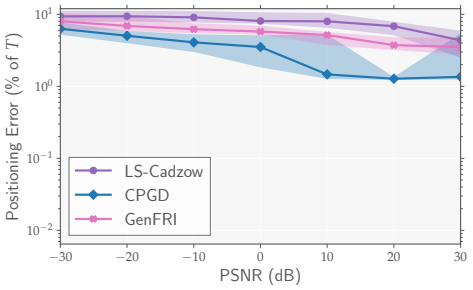
where the peak signal to noise ratio PSNR ranges from -30 to 30 decibels. The results of the experiments are displayed on Figs. 10.3, 10.4, 10.5 and 10.6. In Fig. 10.3, 10.4 and 10.6 we plot –for different oversampling factors and PSNR, the median and inter-quartile region of the empirical distributions of the average positioning error, reconstruction time and number of iterations of the three methods respectively. The reported reconstruction times are for a dual-socket Intel E5-2680v3 (2x 12C/24T) @ 2.5GHz with 256GB RAM. In Fig. 10.5, we plot –for each source, different oversampling factors and PSNR, the median and inter-quartile region of the empirical distribution of the source location as estimated by the three methods against the true source location. All empirical distributions are obtained over 192 independent noise realisations.

In terms of reconstruction accuracy, Figs. 10.3 and 10.5 reveal that CPGD is superior to GenFRI which is itself superior to the baseline method LS-Cadzow in nearly all cases, with the exception of very low PSNRs ($\sim$ -30 dB), where the three methods have comparable reconstruction accuracy. For oversampling parameters $\beta \geq 4$ and a PSNR larger than 0, CPGD is moreover more accurate than GenFRI and LS-Cadzow by a few orders of magnitude (from 1 to 3 orders of magnitude depending on the PSNR). Fig. 10.5 also reveals that the CPGD locations are much less variable than the GenFRI or LS-Cadzow locations.

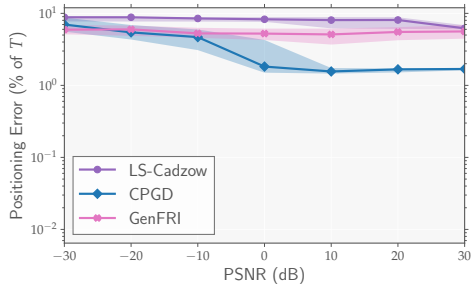
[13] See [44] for additional details on the procedure for recovering the Dirac locations from the annihilating filter coefficients.

[14] The Hungarian algorithm is available in the Python 3 package `scipy` [87].

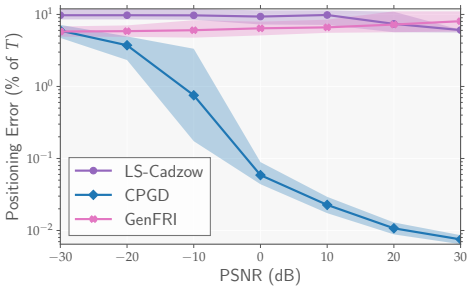
In terms of speed and number of iterations, Fig. 10.4 reveals that LS-Cadzow is superior to GenFRI by approximately three orders of magnitude, and superior to CPGD by one to two orders of magnitude. CPGD is itself superior to GenFRI by approximately one order of magnitude (two for high PSNRs and $\beta = 5$). The reconstruction time and number of iterations for CPGD tend to decrease as the PSNR increases, and are more variable for small PSNRs. In comparison, both GenFRI and LS-Cadzow have relatively constant reconstruction times and number of iterations, with low to none variability for various PSNRs and oversampling factors.



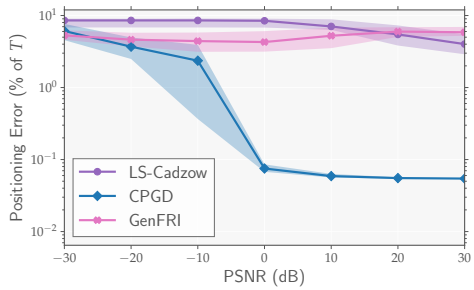
(a) $\beta = 2$



(b) $\beta = 3$

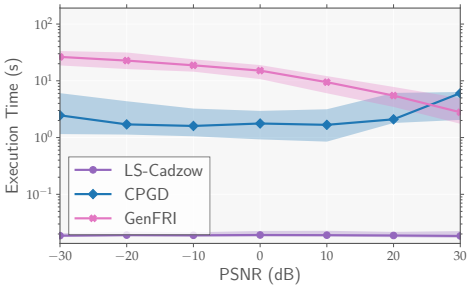


(c) $\beta = 4$

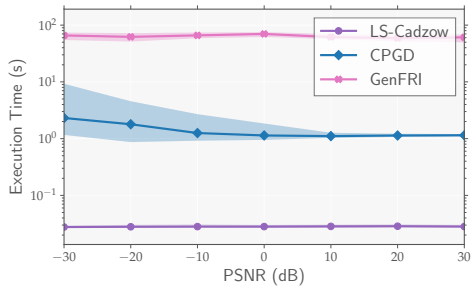


(d) $\beta = 5$

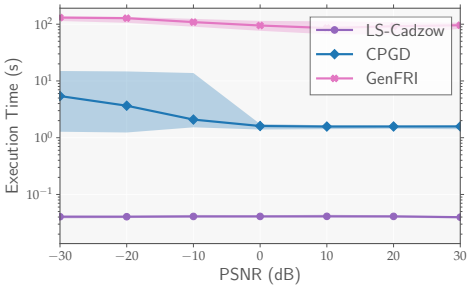
Figure 10.3: Positioning error (10.50) (in percent of period) for LS-Cadzow, CPGD and GenFRI, various oversampling parameters $\beta \in \{2, 3, 4, 5\}$ and a PSNR in $\{-30, -20, -10, 0, 10, 20, 30\}$ dB. For each case, plain lines and shaded areas represent respectively the median and inter-quartile region of the positioning error's empirical distribution obtained from 192 independent noise realisations.



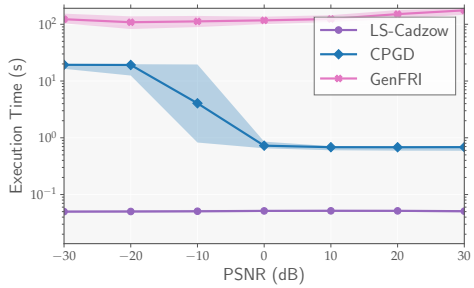
(a) $\beta = 2$



(b) $\beta = 3$

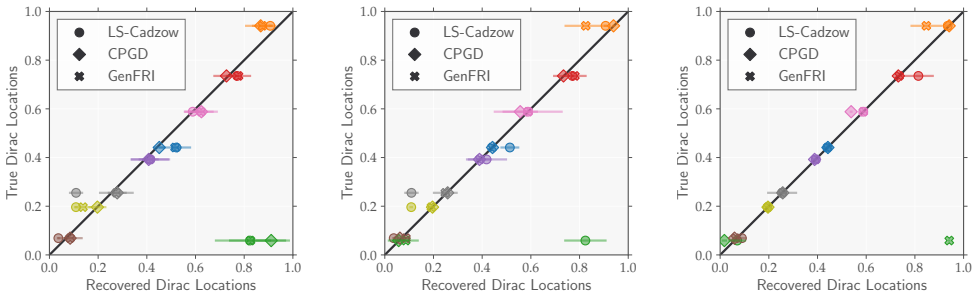


(c) $\beta = 4$

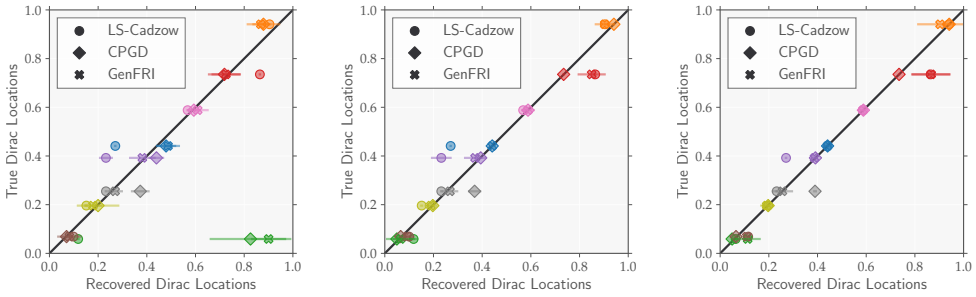


(d) $\beta = 5$

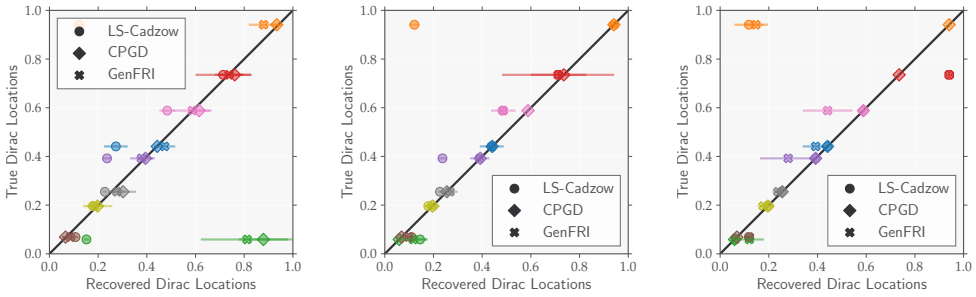
Figure 10.4: Reconstruction time for LS-Cadzow, CPGD and GenFRI, various oversampling parameters $\beta \in \{2, 3, 4, 5\}$ and a PSNR in $\{-30, -20, -10, 0, 10, 20, 30\}$ dB. For each case, plain lines and shaded areas represent respectively the median and inter-quartile region of the reconstruction time's empirical distribution obtained from 192 independent noise realisations. The reported reconstruction times are for a dual-socket Intel E5-2680v3 (2x 12C/24T) @ 2.5GHz with 256GB RAM.



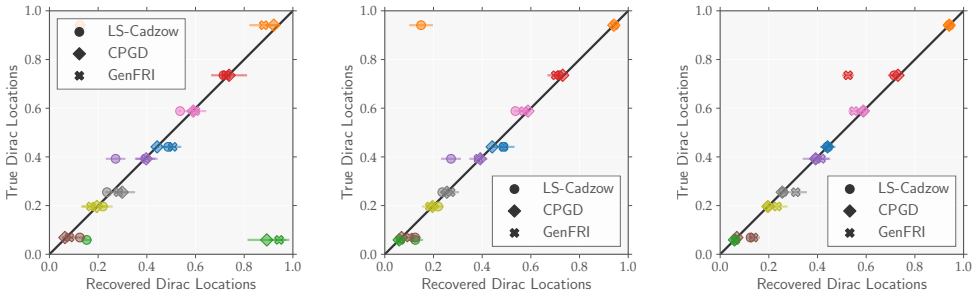
(a) $\beta = 2$, PSNR=-30 dB. **(b)** $\beta = 2$, PSNR=0 dB. **(c)** $\beta = 2$, PSNR=30 dB.



(d) $\beta = 3$, PSNR=-30 dB. **(e)** $\beta = 3$, PSNR=0 dB. **(f)** $\beta = 3$, PSNR=30 dB.

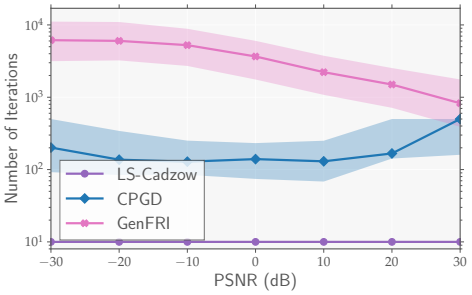


(g) $\beta = 4$, PSNR=-30 dB. **(h)** $\beta = 4$, PSNR=0 dB. **(i)** $\beta = 4$, PSNR=30 dB.

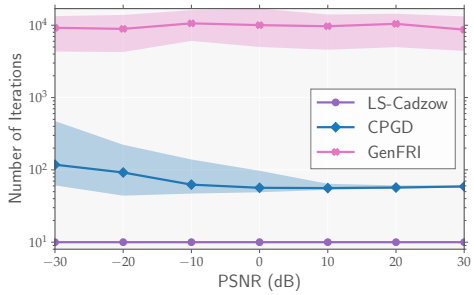


(j) $\beta = 5$, PSNR=-30 dB. **(k)** $\beta = 5$, PSNR=0 dB. **(l)** $\beta = 5$, PSNR=30 dB.

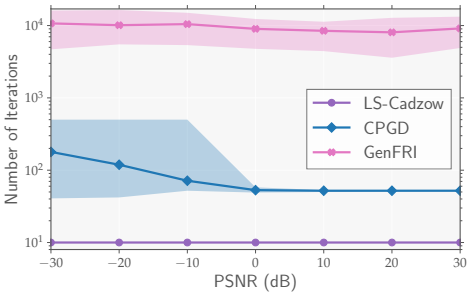
Figure 10.5: Actual vs. recovered Dirac locations for LS-Cadzow, CPGD and GenFRI, various oversampling parameters $\beta \in \{2, 3, 4, 5\}$, a PSNR in $\{-30, 0, 30\}$ dB. For each case and each source (denoted by different colours), the markers and horizontal lines represent respectively the median and inter-quartile region of the estimated locations' empirical distribution obtained from 192 noise realisations. The closer a marker is from the line $y = x$ (in dark grey), the better the recovery is.



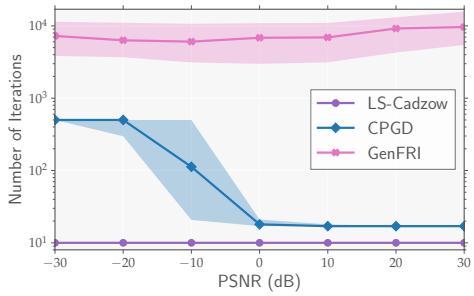
(a) $\beta = 2$



(b) $\beta = 3$



(c) $\beta = 4$



(d) $\beta = 5$

Figure 10.6: Number of iterations for LS-Cadzow, CPGD and GenFRI, various oversampling parameters $\beta \in \{2, 3, 4, 5\}$ and a PSNR in $\{-30, -20, -10, 0, 10, 20, 30\}$ dB. For each case, plain lines and shaded areas represent respectively the median and inter-quartile region of the empirical distribution of the number of iterations obtained from 192 independent noise realisations.

11

RNN-Powered Spherical Approximation*

*The material presented in this chapter is the result of joint work with S. Kashani, P. Hurley and M. Vetterli, and is the topic of [166].

In Chapter 7 we have introduced various iterative proximal algorithms for solving spherical approximation problems in practice. While computationally efficient, such algorithms can in certain applications be too slow to reach convergence. This is notably the case in the context of *real-time imaging*, where dozens of images per second are typically produced. In this chapter, we propose to boost proximal methods by approximating them by *finite-depth neural-networks* with *recurrent architectures*. The idea is to replace the various gradient and proximal steps from proximal methods by a cascade of recurrent layers with trainable parameters. Roughly speaking, the trained recurrent neural-network (RNN) improves convergence speed by navigating more efficiently in the search space via learnt "shortcuts". For concreteness, we illustrate this approach for proximal gradient descent (PGD) applied to the problem of live *acoustic imaging*. However, the methodology could easily adapted to different applications and any of the algorithms from Chapter 7.

1 Introduction

1.1 Motivation

An acoustic camera (AC) [32, 77, 86, 92] is a multi-modal imaging device that allows one to visualise in real-time sound emissions from every *direction* in space. This is typically achieved by overlaying on the live video from an optical camera a heatmap representing the intensity of the ambient directional sound field, recovered from the simultaneous recordings of a *microphone array* [17, 142]. Most commercial acoustic cameras recover the sound intensity field by combining linearly the correlated microphone recordings with a delay and sum (DAS) *beamformer* [142, Chapter 5]. The beamformer acts as an *angular filter* [81, 82], steering sequentially the array sensitivity pattern –or *beamshape*– towards various directions where the sound intensity field is probed. Acoustic images obtained this way are cheap to compute, but are *blurred* by the beamshape of the microphone array, and hence exhibit poor angular resolution [31, 160, 165]. The severity of this blur can be shown [181] to be proportional to the ratio λ/D, where D is the diameter of the microphone array and λ the sound wavelength. Because of the relatively large wavelengths of acoustic

waves in the audible range, this blur can be significant in practice: a 30 cm diameter microphone array has an angular resolution at 5 kHz (an $E\flat$) of approximately 10 degrees, against $7{\cdot}10^{-4}$ degrees for a standard optical camera at 790 THz (violet). Moreover, acoustic cameras are often deployed in confined environments [124], requiring them to be as *compact* and *portable* as possible, which limits[1] further the achievable angular resolution.

[1] Remember that the blur spread is inversely proportional to the microphone array diameter.

The advent of *compressed sensing* techniques [58, 151] –and their wide adoption in imaging sciences [21, 119, 192]– have inspired algorithmic solutions [31, 40, 41, 160] to the acoustic imaging problem, promising vastly improved angular resolutions. Unfortunately, these methods proved ill-suited for real-time purposes. Indeed, they often rely on iterative solvers, such as proximal gradient descent (PGD) [134] or its accelerated variants [16, 112]. While exhibiting a fast convergence rate [16], such methods still require on the order of a few dozen iterations to converge in practice, making them unable to cope with the high refresh-rate[2] of acoustic cameras. For this reason, and despite their clear superiority in terms of resolving power, nonlinear imaging methods have not yet replaced the suboptimal DAS imager in the software stack of commercial acoustic cameras.

[2] An acoustic camera typically updates the acoustic image a dozen times per second.

The recent eruption of *deep learning* [37, 122, 198] in the field of imaging sciences may however seal the fate of DAS for good. Indeed, this new imaging paradigm leverages *neural-networks* [105] to reduce dramatically the image formation time. Unlike compressed-sensing methods which proceed iteratively, neural-networks *encode* the image reconstruction process in a cascade of linear and nonlinear transformations *trained* on a very large number of input/output example pairs. Once properly trained, a neural-network can be efficiently evaluated for some input data to produce images of high quality, with similar accuracy and resolution as state-of-the-art compressed sensing methods [122]. Network architectures used for inverse imaging [37, 73, 83, 138, 198] are most often convolutional neural-network (CNNs), directly adapted from generic architectures developed for *image classification* and *segmentation* [152]. While suitable for image processing tasks such as *denoising*, *super-resolution* or *deblurring* [23, 135], such architectures are ill-suited [122] for more complex image reconstruction problems where the input data may not consist of an image, as is the case in *biomedical imagery* [21, 119], *interferometry* [192] or acoustic imaging. Moreover, and particularly limiting for our current purposes, standard convolutional architectures cannot handle images with non-Euclidean domains [47] such as *spherical* maps [139] produced by omnidirectional acoustic or optical cameras.

To overcome these limitations, *recurrent* architectures [70, 110, 122, 173] have been proposed, by *unrolling* iterative convex optimisation algorithms. Such networks are not only able to handle non-image inputs, but also have greater interpretability than generic CNNs. For example, Gregor and LeCun proposed in their pioneering work [70] a recurrent neural-network (RNN) dubbed LISTA[3], inspired from the popular iterative soft-thresholding algorithm (ISTA) [16].[4] Their network can be seen as generalising ISTA, allowing for the normally fixed gradient and proximal steps occurring at each iteration of the algorithm to be learnt from the data: update steps of ISTA are replaced

[3] LISTA stands for learned iterative soft-thresholding algorithm.

[4] ISTA is an instance of proximal gradient descent for *penalised basis pursuit* problems [178].

by a cascade of recurrent layers with trainable parameters. The depth of the resulting RNN is typically much smaller than the number of iterations required for ISTA to converge. Roughly speaking, the network is learning *shortcuts* in the reconstruction space, allowing it to achieve a prescribed reconstruction accuracy faster than gradient-based iterative methods.[5]

While the effectiveness of ISTA was verified on small images from the MNIST dataset (784 pixels) [70], its application to large-scale imaging problems remains challenging. This is mainly due to the huge number of weights parametrising the network which, in the fully-connected case, grows as the number of pixels to the square. Storing[6] –let alone learning– all those weights quickly becomes intractable for increasing resolutions. As a potential fix, Gregor and LeCun recommended sparsifying the network by *pruning* layer connections. While they showed that such a pruning could reduce the number of parameters in the network by 80% without affecting too much the performance of the latter, this is still insufficient for large-scale problems, and additional structure must be considered on network layers. Such structure is however often very dependent on the problem at hand.

[5] Of course, such shortcuts will most likely only be valid for the distribution of inputs and outputs implicitly defined by the training set, which should hence be carefully crafted for the network to generalise well in practice.

[6] For a 1 megapixel image, the weights parametrising the network would be approximately 8 Gb in size.

1.2 Contributions

In this work, we propose the first realistic architecture of a LISTA neural-network adapted to acoustic imaging. Our custom architecture, dubbed *DeepWave*, is capable of rendering high-resolution spherical maps of real-life sound intensity fields in milliseconds. DeepWave is tailored to the acoustic imaging problem, leveraging fully its underlying structure so as to minimise the number of network parameters. The latter is easy to train, with a typical training time of less than an hour on a general-purpose CPU. Unlike most state-of-the-art neural-network architectures, it moreover readily supports *complex*-valued input vectors, making it capable of directly processing the raw correlated microphone recordings. Assuming a microphone array with M microphones, the instantaneous covariance matrix $\hat{\boldsymbol{\Sigma}} \in \mathbb{C}^{M\times M}$ of the microphone recordings is processed by the network as follows (see also Fig. 11.1):

$$\boldsymbol{x}^l = \sigma\left(P_{\boldsymbol{\theta}}\left(\boldsymbol{L}\right)\boldsymbol{x}^{l-1} + \left[\overline{\boldsymbol{B}} \circ \boldsymbol{B}\right]^H \mathrm{vec}(\hat{\boldsymbol{\Sigma}}) - \boldsymbol{\tau}\right), \qquad l = 1, \dots, L, \tag{11.1}$$

where $\mathrm{vec} : \mathbb{C}^{M\times M} \to \mathbb{C}^{M^2}$ is the *vectorisation operator* and $\circ$ denotes the *Khatri-Rao product* (see Appendix 1 for definitions). The neurons $\{\boldsymbol{x}^1, \dots, \boldsymbol{x}^L\} \subset \mathbb{R}_+^N$ at the output of each layer l of the depth L neural-network correspond to the acoustic image as it is processed by the network, with N the number of pixels. The neuron $\boldsymbol{x}^0 \in \mathbb{R}_+^N$ defines the initial state of the network. The non-linear *activation function*[7] $\sigma : \mathbb{R} \to \mathbb{R}$ induces *sparsity* in the acoustic image, and is inspired by the proximal operator of an elastic-net penalty [134]. The remaining quantities, namely $P_{\boldsymbol{\theta}}(\boldsymbol{L})$, $\boldsymbol{B}$ and $\boldsymbol{\tau}$ are trainable parameters of the network, with various roles:

[7] Typified by a rectilinear unit.

- *Deblurring:* the matrix $P_{\boldsymbol{\theta}}(\boldsymbol{L}) := \sum_{k=0}^{K} \theta_k \boldsymbol{L}^k \in \mathbb{R}^{N\times N}$ can be interpreted as a *deblurring matrix*, cleaning potential artefacts from the array beamshape. Following the approach of [139], it is defined as a poly-

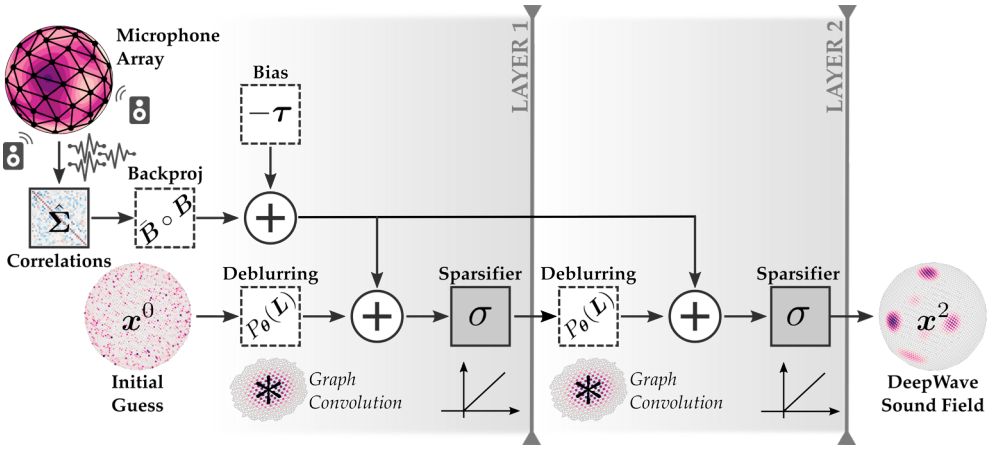


Figure 11.1: DeepWave's recurrent architecture (11.1) for $L = 2$ layers and random initialisation. Learnable parameters of the network are denoted by dashed boxes. Affine operations are denoted by white boxes and nonlinear activations by grey boxes.

nomial of the *graph Laplacian* $\boldsymbol{L} \in \mathbb{R}^{N\times N}$ based on the *connectivity graph* of the spherical tessellation in use, with learnable coefficients $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_K] \in \mathbb{R}^{K+1}$ (see Section 2.3 of Chapter 6). Such parametrisation permits notably the interpretation of $P_{\boldsymbol{\theta}}(\boldsymbol{L})$ as a finite-support filter defined on the tessellation graph. Moreover, fast graph convolution algorithms are available for such filters [47] (see Algorithm 7.11).

- *Back-projection:* the operation $\left[\overline{\boldsymbol{B}} \circ \boldsymbol{B}\right]^H \mathrm{vec}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}\left(\boldsymbol{B}^H \hat{\boldsymbol{\Sigma}} \boldsymbol{B}\right)$ (C.8) is a *back-projection*, mapping the raw microphone correlations to the image domain. Thanks to the convenient Khatri-Rao structure, this linear operation depends only on the matrix $\boldsymbol{B} \in \mathbb{C}^{M\times N}$.
- *Bias:* the vector $\boldsymbol{\tau} \in \mathbb{R}^N$ is a non-uniform *bias*, boosting or shrinking the neurons of the network. Since only positive neurons are activated by the nonlinearity σ, this biasing operation helps sparsify the final acoustic image.

The total number of learnable coefficients in DeepWave is *linear* in the number of pixels. The rationale behind DeepWave's architecture is detailed in Section 2, with theoretical justifications for the structures of the deblurring and back-projection linear operators. In Section 3, we discuss network training, including initialisation and regularisation. We moreover derive the *forward* and *backward-propagation* recursions[8] for our custom architecture, required for forming gradient steps. Finally, we test the architecture on synthetic as well as real data acquired with the *Pyramic array* [22, 155]. DeepWave is shown to have similar resolving power as state-of-the-art compressed-sensing methods, with a computational overhead similar to the DAS imager. To our knowledge, this is the first time a nonlinear imager of the kind achieves real-time performance on a standard computing platform. While developed primarily for acoustic cameras, DeepWave can easily be applied in neighbouring array signal processing fields [95], including radio astronomy, radar and sonar technologies.

[8] DeepWave implementation can be found on https://github.com/imagingofthings/DeepWave.

2 Network Architecture

In this section, we proceed similarly to [70, 110, 173] and construct DeepWave by studying the update equations of an iterative solver, namely proximal gradient descent applied to acoustic imaging.

2.1 Proximal Gradient Descent for Acoustic Imaging

In all that follows, we model the sound intensity field as a discrete *spherical map* with resolution N, specified by an *intensity vector* $\boldsymbol{x} \in \mathbb{R}_+^N$ defined over an *equidistributed point set* $\Theta = \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_N\} \subset \mathbb{S}^2$ (see Section 2 of Chapter 6). Using the classical far-field array signal processing data model [95, 164, 181], we propose to recover the sound intensity map by solving the following convex optimisation problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{R}_+^N} \frac{1}{2} \left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{A}\, \mathrm{diag}(\boldsymbol{x}) \boldsymbol{A}^H \right\|_F^2 \quad + \quad \lambda \left[\gamma \|\boldsymbol{x}\|_1 + (1-\gamma)\|\boldsymbol{x}\|_2^2 \right], \tag{11.2}$$

where $\|\cdot\|_F$ denotes the *Frobenius norm*, $\gamma \in]0,1[$ and $\lambda > 0$ are hyperparameters, and $\hat{\boldsymbol{\Sigma}} \in \mathbb{C}^{M \times M}$ is the empirical covariance matrix of the microphone recordings. In a far-field context, the *forward map* $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ –linking the intensity vector to the microphone recordings– is commonly modelled by the so-called *steering matrix* [95]: $A_{mn} := \exp\left(-2\pi j \langle \boldsymbol{p}_m, \boldsymbol{r}_n \rangle / \lambda_0\right)$, where $\{\boldsymbol{p}_1, \dots, \boldsymbol{p}_M\} \subset \mathbb{R}^3$ are the microphone locations and $\lambda_0 > 0$ the sound wavelength. Using properties (C.5) and (C.6) of the *vectorisation operator* and the Frobenius norm [84, 181], problem (11.2) can be re-written in vectorised form as:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{R}_+^N} \frac{1}{2} \left\| \mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}\right) - \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right) \boldsymbol{x} \right\|_2^2 \quad + \quad \lambda \left[\gamma \|\boldsymbol{x}\|_1 + (1-\gamma)\|\boldsymbol{x}\|_2^2 \right], \tag{11.3}$$

where $\circ$ denotes the *Khatri-Rao product* (see Definition C.3). Problem (11.3) is an *elastic-net penalised least-squares problem* [203], which seeks an optimal[9] trade-off between *data-fidelity* and *group-sparsity*. Group-sparsity is in this context better suited than traditional sparsity since acoustic sources are often diffuse. It is worth noting that, since the elastic-net functional is strictly convex for $\gamma \in [0,1[$, problem (11.3) admits a unique solution. The latter can moreover be approximated by means of proximal gradient descent (PGD) [16], whose update equations are given here by (see Section 2 of Appendix C):

[9] The notion of optimality is defined here by the penalty parameter λ.

$$\boldsymbol{x}^k = \mathbf{ReLu}\left(\frac{\boldsymbol{x}^{k-1} - \alpha \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \left[\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right) \boldsymbol{x}^{k-1} - \mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}\right) \right] - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma) + 1} \right), \quad k \geq 1, \tag{11.4}$$

where $\boldsymbol{x}^0 \in \mathbb{R}^N$ is arbitrary, $\alpha \leq 1 / \left\| \overline{\boldsymbol{A}} \circ \boldsymbol{A} \right\|_2^2$ is the *step size* and $\mathbf{ReLu}(x) := \max(x, 0)$ is the *rectified linear unit* [106], applied element-wise to a real vec-

[10] Note that with $\boldsymbol{x}^0 \in \mathbb{R}^N$, every gradient step produces a real vector.

tor.[10] The sequence of iterates $\{\boldsymbol{x}^k\}_{k\in\mathbb{N}}$ defined in (11.4) reduces the objective function in (11.3) at a rate $O(1/k)$ [16]. Accelerated variants of proximal gradient descent have been proposed [16], which modify (11.4) with an extra *momentum term*:

$$\begin{cases} \boldsymbol{y}^k &= \mathbf{ReLu}\left(\dfrac{\boldsymbol{x}^{k-1} - \alpha\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \left[\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)\boldsymbol{x}^{k-1} - \mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}\right)\right] - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1}\right), \\ \boldsymbol{x}^k &= \boldsymbol{y}^k \;+\; \omega^k\left(\boldsymbol{y}^k - \boldsymbol{y}^{k-1}\right) \end{cases} \tag{11.5}$$

where $k \geq 1$ and the *momentum sequence* $\{\omega^k\}_{k\in\mathbb{N}}$ can be designed in various ways [34, 112]. In our experiments, we will use (11.5) as a baseline for speed comparisons, where ω^k is updated according to Chambolle and Dossal's strategy [34]: $\omega^k = (k-1)/(k+d), \quad k \geq 0$, with $d = 50$ [112]. The accelerated proximal gradient descent (APGD) method thus obtained is the fastest reported in the literature, with convergence rate $o(1/k^2)$ [112] (see Section 2 of Chapter 7). Finally, we leverage the formulae $\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)\boldsymbol{x} = \mathrm{vec}(\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{x})\boldsymbol{A}^H)$ (C.5), and $\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \mathrm{vec}(\boldsymbol{R}) = \mathrm{diag}(\boldsymbol{A}^H\boldsymbol{R}\boldsymbol{A})$ (C.8), to compute gradient steps efficiently in (11.5).

2.2 DeepWave : a PGD-inspired RNN for Fast Acoustic Imaging

In practice PGD is terminated according to some stopping criterion. The intensity map $\boldsymbol{x}^L$ obtained after L iterations of (11.4) can then be seen as the output of an RNN with depth L and intermediate neurons linked by the recursion formula:

$$\boldsymbol{x}^l = \mathbf{ReLu}\left(\boldsymbol{\mathcal{D}}\boldsymbol{x}^{l-1} + \boldsymbol{\mathcal{B}}\,\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}\right) - \boldsymbol{\tau}\right), \qquad l = 1, \ldots, L. \tag{11.6}$$

We call this RNN the *oracle RNN*, since its weights $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{N\times M^2}$ and $\boldsymbol{\tau} \in \mathbb{R}^N$ are not learnt but simply *given* to us by identifying (11.6) with (11.4):

$$\boldsymbol{\mathcal{D}} = \frac{1}{\beta}\left[I - \alpha\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)\right], \quad \boldsymbol{\mathcal{B}} = \frac{\alpha}{\beta}\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H, \quad \boldsymbol{\tau} = \frac{\lambda\alpha\gamma}{\beta}\mathbf{1}_N, \tag{11.7}$$

where $\beta = 2\lambda\alpha(1-\gamma)+1$. An analysis of (11.7) allows us moreover to interpret physically the affine operations performed by the oracle RNN. The matrix $\boldsymbol{\mathcal{B}}$ first is a *back-projection* operator, mapping the vectorised correlation matrix into a spherical map by applying the adjoint of the forward operator used in (11.3). The resulting spherical map is called a *dirty map*, and is equivalent to the DAS image [181, Section 5.2][193]. The matrix $\boldsymbol{\mathcal{D}}$ then is a *deblurring operator*, which subtracts at each iteration a fraction of the array beamshape from the spherical map, hence *cleaning* the latter of blur artefacts. The vector $\boldsymbol{\tau}$ finally is an affine *shrinkage operator*, which biases uniformly the spherical

map. The latter permits –in conjunction with the rectified linear unit– the *sparsification* of the spherical map and hence improve its angular resolution.
Since the oracle RNN is merely a reinterpretation of PGD, it inherits all its properties. In particular, it is capable of solving (11.3) with high accuracy for arbitrary input correlation matrices. Unfortunately, this great generalisability is typically obtained at the price of a very large number[11] of layers L, resulting in impractical reconstruction times. If one is however willing to sacrifice some of this generalisability, it is possible to reduce drastically the network depth by unfreezing the weights $\boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\tau}$ in (11.6), and allowing them to be *learnt* for some specific input distribution. This idea was first explored in the context of sparse coding by Gregor and LeCun [70], resulting in the LISTA network. A fully-connected architecture, corresponding to unconstrained $\boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{B}}$ and $\boldsymbol{\tau}$, would however result in $\mathcal{O}(N^2)$ weights to be learnt, which is unfeasible in large-scale acoustic imaging problems. To overcome this issue, we propose in the next paragraphs a parsimonious parametrisation of $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{B}}$. The resulting RNN architecture, dubbed DeepWave, is given in (11.1) and depicted in Fig. 11.1.

[11] Even with momentum acceleration, PGD typically requires more than 50 iterations to converge. The oracle RNN obtained by unrolling PGD will consequently be very deep.

2.2.1 Parametrisation of the Deblurring Operator

Our parametrisation of $\boldsymbol{\mathcal{D}}$ is motivated by the following result, characterising the oracle deblurring kernel for *spherical microphone arrays* [142].

Proposition 11.1 — *Oracle Deblurring Operator for Spherical Microphone Arrays.* *Consider a* spherical microphone array*, with diameter D and microphone directions $\{\tilde{\boldsymbol{p}}_1, \dots, \tilde{\boldsymbol{p}}_M\} \subset \mathbb{S}^2$, forming a* equidistributed spherical point set*. Then, we have $\forall i, j \in \{1, \dots, N\}$*

$$\left[I - \alpha \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)\right]_{ij} \simeq \left[\delta_{ij} - \alpha M^2 \operatorname{sinc}^2\left(\frac{D}{\lambda_0}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|\right)\right], \quad (11.8)$$

where λ_0 is the wavelength*, δ_{ij} denotes the* Kronecker delta *and*

$$\operatorname{sinc}(x) := \frac{\sin(\pi x)}{\pi x}$$

is the cardinal sine*. Moreover, the approximation (11.8) is* extremely good *for $M \geq 3\lfloor \frac{2\pi D}{\lambda_0} \rfloor^2$.*

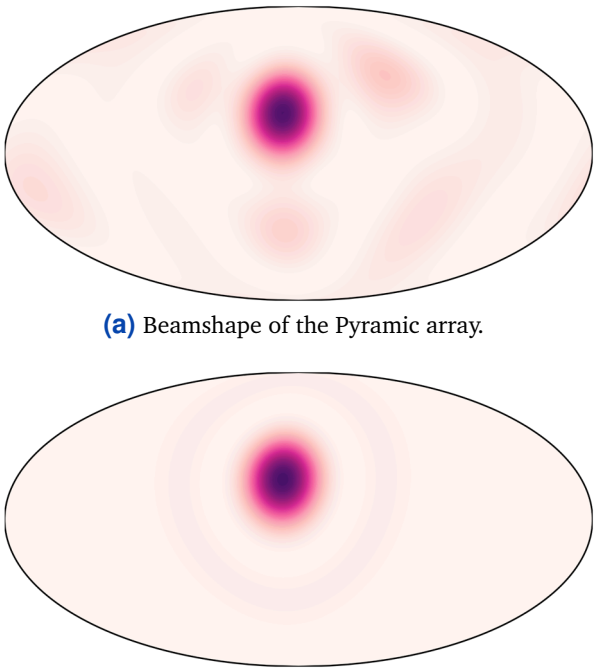
Proof. To prove (11.8), it is sufficient to show that

$$\left[\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)\right]_{ij} \simeq M^2 \operatorname{sinc}^2\left(\frac{D}{\lambda_0}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|\right). \quad (11.9)$$

To this end, we first use (C.11) and obtain:

$$\left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right)^H \left(\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right) = \left|\boldsymbol{A}^H \boldsymbol{A}\right|^{\odot 2}. \quad (11.10)$$

For a spherical array with diameter D and microphone directions $\{\tilde{\boldsymbol{p}}_1, \dots, \tilde{\boldsymbol{p}}_M\} \subset$



(a) Beamshape of the Pyramic array.

(b) Approximate beamshape obtained with (11.9).

Figure 11.2: Accuracy of approximation (11.9) for the Pyramic array [155] ($D = 30$[cm], $M = 48$) at 1 kHz.

$\mathbb{S}^2$, we get moreover from the definition of the steering matrix that:

$$[\boldsymbol{A}^H \boldsymbol{A}]_{ij} = \sum_{m=1}^{M} \exp\left(j\frac{\pi D}{\lambda_0}\langle \boldsymbol{r}_i - \boldsymbol{r}_j, \tilde{\boldsymbol{p}}_m\rangle\right), \qquad i,j = 1,\dots,N. \tag{11.11}$$

Since the microphone directions are assumed to form an equidistributed spherical point set (such as the Fibonacci or HEALPix tessellations discussed in Section 2 of Chapter 6), we can interpret (11.11) as a quadrature rule on the sphere [125, Chapter 3], yielding:

$$\frac{4\pi}{M}\sum_{m=1}^{M} \exp\left(j\frac{\pi D}{\lambda_0}\langle \boldsymbol{r}_i - \boldsymbol{r}_j, \tilde{\boldsymbol{p}}_m\rangle\right) \simeq \int_{\mathbb{S}^2} \exp\left(j\frac{\pi D}{\lambda_0}\langle \boldsymbol{r}_i - \boldsymbol{r}_j, \tilde{\boldsymbol{p}}\rangle\right) d\tilde{\boldsymbol{p}} \tag{11.12}$$

$$= 4\pi\,\mathrm{sinc}\left(\frac{D}{\lambda_0}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|\right), \tag{11.13}$$

where the second equality (11.13) follows from the result on [168, p. 154]. From (11.13), (11.11) and (11.10) we obtain (11.9) from which (11.8) trivially follows.
Regarding the quality of the approximation (11.9) finally, we use the approximate bandlimitedness of complex plane-waves in the spherical domain [142, Chapter 2]. Indeed, quadrature rules such as (11.12) are almost exact for bandlimited functions [142, Chapter 3], provided a high-enough number of quadrature points M. For example, a function with spherical harmonic bandwidth $L \in \mathbb{N}$ is extremely well approximated by the HEALPix quadrature rule

for $M \geq 3L^2$ [67]. In our case, the *plane-wave expansion* [142, Chapter 2] gives us

$$\exp\left(j\frac{\pi D}{\lambda_0}\langle \boldsymbol{r}_i - \boldsymbol{r}_j, \tilde{\boldsymbol{p}}\rangle\right) = 4\pi \sum_{l=0}^{+\infty}\sum_{k=-l}^{l} j^l(2l+1)j_l\left(\frac{\pi D}{\lambda_0}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|_2\right)\overline{Y_l^k}(\tilde{\boldsymbol{r}}_{ij})Y_l^k(\tilde{\boldsymbol{p}}),$$

where j_l are *spherical Bessel functions*, Y_l^k *spherical harmonics*, and $\tilde{\boldsymbol{r}}_{ij} = (\boldsymbol{r}_i - \boldsymbol{r}_j)/\|\boldsymbol{r}_i - \boldsymbol{r}_j\|_2^2$ [142]. Since $j_l(x) \simeq 0$ for $l \geq x$ [142, Chapter 2] we have hence that complex plane-waves are approximately bandlimited with bandwidth $L = \lfloor \frac{\pi D}{\lambda_0}\|\boldsymbol{r}_i - \boldsymbol{r}_j\|_2 \rfloor \leq \lfloor \frac{2\pi D}{\lambda_0} \rfloor$. As a result, choosing $M \geq 3\lfloor \frac{2\pi D}{\lambda_0} \rfloor^2$ makes the approximation (11.9) very accurate. ■

Remark 11.1 *While proven for spherical arrays only, approximation (11.9) (and hence (11.8)) remains quite accurate in practice, even for non-spherical microphone arrays such as the Pyramic array used in our real-world experiments [155]. In Fig. 11.2, we investigated visually the quality of the approximation (11.9) for the Pyramic array at 1 kHZ. To this end, we plotted a row of* $\left|\boldsymbol{A}^H\boldsymbol{A}\right|^{\odot 2}$ *(which corresponds to the beamshape of the instrument for a particular direction [181]) with and without approximation. We observe that the approximation is already very good, even if the Pyramic array possesses only* $M = 48$ *microphones against the 90 required by Proposition 11.1 for an optimal approximation accuracy at this frequency.*

Proposition 11.1 tells us that, for spherical arrays with sufficient number of microphones[12], the oracle deblurring operator $\boldsymbol{\mathcal{D}}$ in (11.7) corresponds actually to a sampled *zonal kernel*: $[\boldsymbol{\mathcal{D}}]_{ij} = \kappa(\|\boldsymbol{r}_i - \boldsymbol{r}_j\|)$ for some $\kappa : \mathbb{R}_+ \to \mathbb{R}$. Since zonal kernels are used to define spherical convolutions (see Chapter 3), $\boldsymbol{\mathcal{D}}$ can hence be seen as a *discrete convolution operator* over the discrete domain $\Theta = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N\}$. Its bandwidth is moreover essentially finite, since coefficients $[\boldsymbol{\mathcal{D}}]_{ij}$ decay as $1/\|\boldsymbol{r}_i - \boldsymbol{r}_j\|^2$. As discussed in [47, 139], discrete spherical convolution operators with finite scope can be efficiently represented and implemented by means of the *graph signal processing* [159] techniques discussed in Section 2.3 of Chapter 6. This leads us to consider the following parametrisation:

$$\boldsymbol{\mathcal{D}} = P_{\boldsymbol{\theta}}(\boldsymbol{L}) := \sum_{k=0}^{K} \theta_k \boldsymbol{L}^k,$$

where $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_K] \in \mathbb{R}^{K+1}$, K controls the scope of the discrete convolution and $\boldsymbol{L} \in \mathbb{R}^{N\times N}$ is the *Laplacian* associated to the spherical tessellation graph of Θ (see Section 2.3 of Chapter 6). Note that with this parametrisation, the number of parameters characterising $\boldsymbol{\mathcal{D}}$ drops from N^2 to $K+1$, with $K \ll N$.

[12] For a spherical array with diameter $D = 30$ cm operating at 1 kHz, $M \geq 90$ is sufficient.

2.2.2 Parametrisation of the Back-projection Operator

The oracle back-projection operator (11.7) admits a factorisation in terms of the Khatri-Rao product. We decide hence to equip $\boldsymbol{\mathcal{B}}$ with a similar structure: $\boldsymbol{\mathcal{B}} = (\overline{\boldsymbol{B}} \circ \boldsymbol{B})^H$ for some learnable matrix $\boldsymbol{B} \in \mathbb{C}^{M\times N}$. With such a parametrisation, the number of parameters characterising $\boldsymbol{\mathcal{B}}$ drops from NM^2 to NM. The

Algorithm 11.13: DeepWave forward propagation.

Input: $\hat{\boldsymbol{\Sigma}}_t,\ \boldsymbol{x}_t^0,\ \hat{\boldsymbol{x}}_t,\ \boldsymbol{\theta},\ \boldsymbol{B},\ \boldsymbol{\tau},\ \sigma$
Output: $\mathcal{L}_t \in \mathbb{R}_+,\ \{\boldsymbol{s}_t^l\} \subset \mathbb{R}^N$

$\boldsymbol{y}_t \leftarrow \mathrm{diag}(\boldsymbol{B}^H \hat{\boldsymbol{\Sigma}}_t \boldsymbol{B}) - \boldsymbol{\tau}$
for l **in** $[1, \dots, L]$ **do**
    $\boldsymbol{s}_t^l \leftarrow P_{\boldsymbol{\theta}}(\boldsymbol{L})\boldsymbol{x}_t^{l-1} + \boldsymbol{y}_t$
    $\boldsymbol{x}_t^l \leftarrow \sigma(\boldsymbol{s}_t^l)$
$\mathcal{L}_t \leftarrow \frac{1}{2} \left\| \hat{\boldsymbol{x}}_t - \boldsymbol{x}_t^L \right\|_2^2 / \left\| \hat{\boldsymbol{x}}_t \right\|_2^2$

Algorithm 11.14: DeepWave backward propagation.

Input: $\hat{\boldsymbol{\Sigma}}_t,\ \boldsymbol{x}_t^0,\ \hat{\boldsymbol{x}}_t,\ \boldsymbol{\theta},\ \boldsymbol{B},\ \sigma,\ \{\boldsymbol{s}_t^l\}_{l=1,\dots,L}$
Output: $\partial\boldsymbol{\theta} \in \mathbb{R}^{K+1},\ \partial\boldsymbol{B} \in \mathbb{C}^{M\times N},\ \partial\boldsymbol{\tau} \in \mathbb{R}^N$

$(\partial\boldsymbol{x}, \partial\boldsymbol{\theta}, \partial\boldsymbol{\tau}) \leftarrow \left(\left(\sigma(\boldsymbol{s}_t^L) - \hat{\boldsymbol{x}}_t\right) / \left\|\hat{\boldsymbol{x}}_t\right\|_2^2, \boldsymbol{0}, \boldsymbol{0} \right)$
for l **in** $[L, \dots, 1]$ **do**
    $\partial\boldsymbol{s} \leftarrow \mathrm{diag}\left(\sigma'(\boldsymbol{s}_t^l)\right)\ \partial\boldsymbol{x}$
    $\partial\boldsymbol{x} \leftarrow P_{\boldsymbol{\theta}}(\boldsymbol{L})\partial\boldsymbol{s}$
    $\partial\boldsymbol{\tau} \leftarrow \partial\boldsymbol{\tau} - \partial\boldsymbol{s}$
    $[\partial\boldsymbol{\theta}]_k \leftarrow [\partial\boldsymbol{\theta}]_k + \partial\boldsymbol{s}^T\, T_k(\boldsymbol{L})\, \sigma(\boldsymbol{s}_t^{l-1})$
$\partial\boldsymbol{B} \leftarrow -2\hat{\boldsymbol{\Sigma}}_t \boldsymbol{B}\, \mathrm{diag}\,(\partial\boldsymbol{\tau})$

Figure 11.3: Forward and backward algorithms to compute gradients of $\mathcal{L}_t$ with respect to $\boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{\tau}$. For notational simplicity we use the shorthand $\partial\boldsymbol{\alpha} = \partial\mathcal{L}_t/\partial\boldsymbol{\alpha}$, and assume $\sigma(\boldsymbol{s}_t^0) = \boldsymbol{x}_t^0$.

Khatri-Rao structure guarantees moreover real-valued –and hence physically-interpretable– dirty maps.

3 Network Training

To facilitate the description of the training procedure, we adopt the following shorthand notations.

- $\mathrm{DeepWave}(\boldsymbol{\Omega}, L)$ denotes a specific instance of the DeepWave network (11.1) with parameters $\boldsymbol{\Omega} := \{\boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{\tau}\}$ and depth L.
- $\mathrm{APGD}(\alpha, \lambda, \gamma)$ denotes an instance of APGD (11.5), with tuning parameters $(\alpha, \lambda, \gamma) \in \mathbb{R}_+^3$.

The network parameters are chosen as minimisers of the following optimisation problem:

$$\hat{\boldsymbol{\Omega}} \in \underset{\substack{\boldsymbol{\theta}\in\mathbb{R}^{K+1} \\ \boldsymbol{B}\in\mathbb{C}^{M\times N} \\ \boldsymbol{\tau}\in\mathbb{R}^N}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T} \underbrace{\frac{\left\|\hat{\boldsymbol{x}}_t - \boldsymbol{x}_t^L(\boldsymbol{\Omega})\right\|_2^2}{2\left\|\hat{\boldsymbol{x}}_t\right\|_2^2}}_{:=\mathcal{L}_t} + \underbrace{\frac{\lambda_{\boldsymbol{\theta}}}{2(K+1)}\|\boldsymbol{\theta}\|_2^2}_{:=\mathcal{L}_{\boldsymbol{\theta}}} + \underbrace{\frac{\lambda_{\boldsymbol{B}}}{2MN}\|\boldsymbol{B}\|_F^2}_{:=\mathcal{L}_{\boldsymbol{B}}} + \cdots$$
$$\cdots + \underbrace{\frac{\lambda_{\boldsymbol{\tau}}}{2N}\left\|\boldsymbol{L}^{1/2}\boldsymbol{\tau}\right\|_2^2}_{:=\mathcal{L}_{\boldsymbol{\tau}}}. \tag{11.14}$$

The quantities $\{\boldsymbol{x}_t^L(\boldsymbol{\Omega})\}_t$ and $\{\hat{\boldsymbol{x}}_t\}_t$ in (11.14) correspond respectively to the outputs of $\mathrm{DeepWave}(\boldsymbol{\Omega}, L)$ and $\mathrm{APGD}(\alpha, \lambda, \gamma)$ with identical example input data $\{(\hat{\boldsymbol{\Sigma}}_t, \boldsymbol{x}_t^0)\}_t$. The first term $\frac{1}{T}\sum_{t=1}^T \mathcal{L}_t$ is a data-fidelity term, which attempts to bring $\hat{\boldsymbol{x}}_t$ and $\boldsymbol{x}_t^L(\boldsymbol{\Omega})$ as close as possible from one another.[13] The additional terms $\mathcal{L}_{\boldsymbol{\theta}}, \mathcal{L}_{\boldsymbol{B}}, \mathcal{L}_{\boldsymbol{\tau}}$ are smoothing *regularisers*, fighting against *overfitting*, a common issue in deep learning. Since the shrinkage operator $\boldsymbol{\tau}$ can be interpreted as a signal on the spherical tessellation graph associated to Θ, the smoothing term $\mathcal{L}_{\boldsymbol{\tau}}$ is defined via the graph Laplacian $\boldsymbol{L} \in \mathbb{R}^{N\times N}$, as is

[13] in a mean relative squared-error sense.

customary in graph signal processing (see Section 2.3 of Chapter 6).
Optimisation of (11.14) is carried out by stochastic gradient descent (SGD) with momentum acceleration [174]. Gradients of $\mathcal{L}_t$ with respect to $\boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{\tau}$ are efficiently evaluated using *reverse-mode algorithmic differentiation* [15] and are given in Algorithms 11.13 and 11.14 (see 3 of Appendix C for a derivation). While random initialisation of neural-networks is a common practice in deep learning [174], this strategy failed for our specific architecture, leading to poor validation loss and considerably increased training times. Instead, we hence use the oracle parameters (11.7) to initialise SGD:

$$\boldsymbol{\theta}^0 := \underset{\boldsymbol{\theta}\in\mathbb{R}^{K+1}}{\arg\min} \|P_{\boldsymbol{\theta}}(\boldsymbol{L}) - \boldsymbol{\mathcal{D}}\|_F^2, \quad \boldsymbol{B}^0 := \sqrt{\frac{\alpha}{\beta}}\boldsymbol{A}, \quad \boldsymbol{\tau}^0 := \frac{\lambda\alpha\gamma}{\beta}\mathbf{1}_N. \tag{11.15}$$

For greater numerical stability during training, we proceed as in Section 4.3 of Chapter 7 and reparameterise the deblurring filter as $P_{\boldsymbol{\theta}}(\tilde{\boldsymbol{L}}) = \sum_{k=0}^{K} \theta_k T_k(\tilde{\boldsymbol{L}})$, where $T_k(\cdot)$ is the Chebychev polynomial of order k and $\tilde{\boldsymbol{L}}$ is the normalised Laplacian with spectrum in $[-1, 1]$. Finally, we substitute the **ReLu** activation function by a scaled rectified $\tanh$ to avoid the exploding gradient problem [137].[14]

[14] An alternative is to use a *truncated* **ReLu**. Given initialisation strategy eq. (11.15), network training will still converge with similar step sizes as those used with $\tanh$ non-linearities.

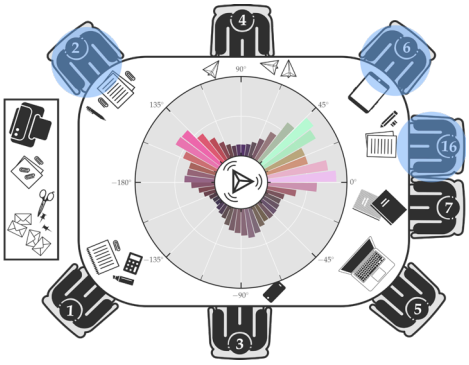
4 Experimental Results

In this section, we compare the accuracy, resolution and runtime performance of DeepWave to DAS and APGD on real-world (RW) and simulated (SIM) datasets. More comprehensive dataset descriptions and additional results, including an ablation study, are provided in Appendices 4 to 6.

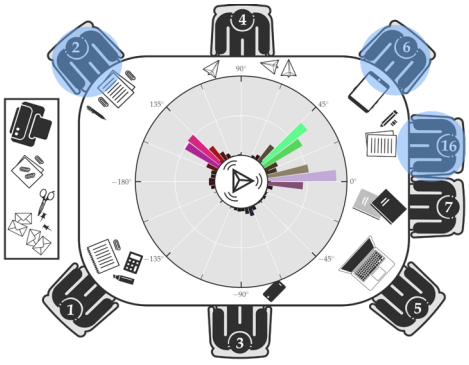
4.1 Real-data Experiments

Dataset 1 [131] (RW) reproduces a conference room setup depicted in Figs. 11.4a and 11.4b, where 8 people[15] are gathered around a table and speak either in turns or simultaneously (with at most 3 concurrent speakers). Recordings of the conversation are collected by the 48-element Pyramic array [155] (Fig. 11.4f) positioned at the centre of the table. Since human speech is wide-band, the audible range [1500, 4500] Hz in the latter are pre-processed every 100 ms and split into 9 uniform bins to form a suitable training set $\{(\hat{\boldsymbol{\Sigma}}_t, \hat{\boldsymbol{x}}_t, \boldsymbol{x}_t^0)\}_t$ of 2760 data points per frequency band for DeepWave (with $N = 2234$). Frequency channels are processed independently by each algorithm. DeepWave is trained by splitting the data points into a training and validation set (respectively 80% and 20% in size). For each frequency band, we chose an architecture with 5 layers.
In Fig. 11.4, Fig. C.5 and Table C.1 respectively, we compare the accuracy and runtime of DeepWave, DAS and APGD. A video showing the evolution in time of DeepWave and DAS azimuthal sound fields (as in Figs. 11.4a and 11.4b) is also available online at https://www.youtube.com/watch?v=PwB3CS2rHdI.
In terms of resolution, DeepWave and APGD perform similarly, outperforming DAS by approximately 27%. The mean contrast scores for DeepWave and DAS over the test set of Dataset 1 are 0.99 ($\pm$0.0081) and 0.89 ($\pm$0.07), respec-

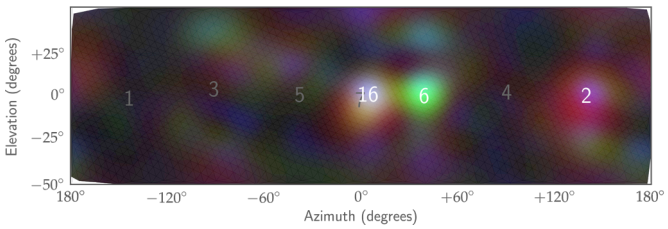
[15] The 8 people are represented in the experiment by loadspeakers playing male and female speech samples.



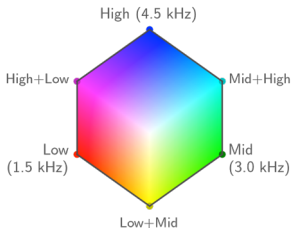
(a) DAS azimuthal sound field.



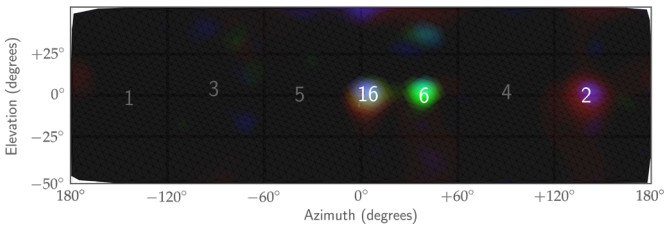
(b) DeepWave azimuthal sound field.



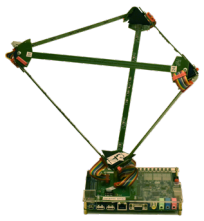
(c) DAS spherical sound field (resolution: 25.3°, contrast: 0.78).



(d) Frequency-colour mapping.

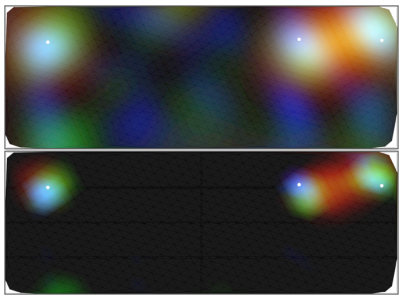


(e) DeepWave spherical sound field (resolution: 18.5°, contrast: 0.97).

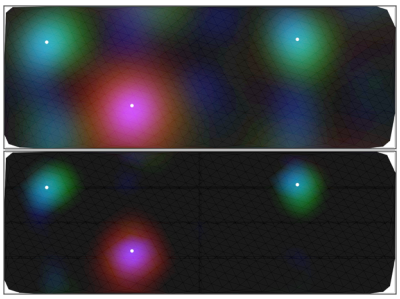


(f) Pyramic array.

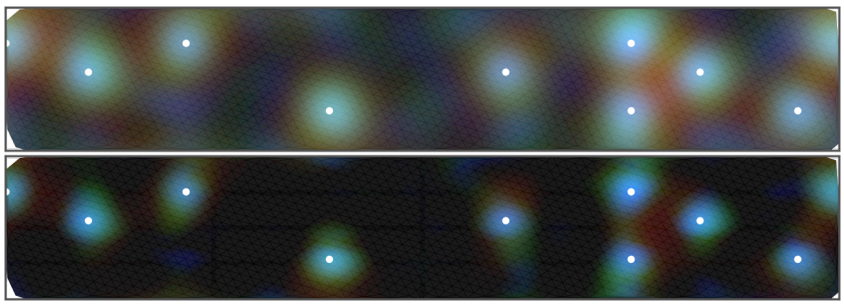
Figure 11.4: Snapshots at time $t = 1.7$ s of the sound intensity fields produced by DeepWave and DAS for the Pyramic recordings with speakers 2, 6 and 16 active. Sound frequencies range from 1.5 to 4.5 kHz and were mapped to true colours (see Fig. 11.4d, colour shades correspond to lower intensities). The spherical maps of DAS and DeepWave are plotted in Figs. 11.4c and 11.4e, respectively. In Figs. 11.4a and 11.4b we plot the azimuthal projections of Figs. 11.4c and 11.4e, respectively.



(a) DAS vs. DeepWave sound fields for Dataset 2.



(b) DAS vs. DeepWave sound fields for Dataset 2.



(c) DAS vs. DeepWave sound fields for synthetic data trained on Dataset 2.

Figure 11.5: Snapshots of the sound intensity fields produced by DeepWave and DAS when trained on Dataset 2 (with 10 held-out source directions). Each subplot contains a DAS image (top) and a DeepWave image (bottom). The frequency color mapping is identical to Fig. 11.4d. Figs. 11.5a and 11.5b show azimuthal sound field slices on $[-20°, 150°]$ using real-world covariance matrices with sources from unseen directions during training. Fig. 11.5c shows a full 360° sound field on a synthetic covariance matrix from unseen directions during training. Elevations span $[-15°, +15°]$.

tively. Note that since the metrics used for assessing resolution and contrast[16] are not perfectly reflective of human-eye perception, the reported image quality improvements appear even more striking through visual inspection of the sound intensity fields (see for example Fig. 11.4).

[16] As is customary, resolution is measured as the *width at half-maximum* of the impulse response of the algorithms. Contrast is measured as the difference between the maximum and mean of the greyscale image.

Dataset 2 [150] (RW) consists of 2700 template recordings from the Pyramic array taken in an anechoic chambre at an angular resolution of 2 degrees in azimuth and three different elevations (-15, 0, 15 degrees). Recordings contain both male and female speech samples to cover a wide audible range. The audio samples can be combined to simulate complex multi-source sound fields, hence we leverage this property to augment the dataset to 5700 distinct recordings with one, two, or three active speakers simultaneously. The raw time-series are then pre-processed as for **Dataset 1** to obtain a training set of 151,980 data points per frequency band (with $N = 1568$). Network training is identical to that of **Dataset 1**, except that 10 azimuth directions are also witheld from the training set to assess how well the network generalises to emissions from unseen directions.
Figs. 11.5a and 11.5b show sample DAS and DeepWave reconstructions with

real sources from directions withheld from the training set. Similarly, Fig. 11.5c shows sample reconstructions when the network is trained on real data but tested on synthetic narrow-band covariance matrices induced by sources from directions absent from the training set. In both cases we see that DeepWave outperforms DAS in resolution and contrast (i.e. sharper blobs and darker background).

4.2 Further experiments

Dataset 3 (SIM) finally is a simulated dataset with recordings from a spherical microphone array using a narrow-band point-source data-model at 2 kHz [181]. The sources are randomly positioned over a 120° field-of-view, with up to 10 concurrent sources per recording. Experiment results available in Fig. C.6 corroborate the real-data results, hence showing that DeepWave generalises well to a large number of sources with unconstrained positions. We further investigated in Fig. C.7 the influence of network depth, and concluded that 5 or 6 layers are generally sufficient for the investigated dataset. In terms of runtimes finally, DeepWave and DAS both reach real-time requirements (6.5 ms and 2.0 ms respectively), largely outperforming APGD (211 ms). (See Table C.1 for more details.)

12 Conclusion

1 The Trajectory of this Thesis

As this thesis is coming to an end, let us reflect back on its positioning and trajectory. We started off Chapter 1 with an environmental manifesto, outlining the role that spherical approximation techniques could play in mitigating the current environmental crisis. A subsequent goal was then to empower scientists and practitioners by providing them with more performant algorithms for sensing, processing and recovering signals defined on the surface of the Earth, or more generally any spherical surface.

This pragmatic motivation was however quickly challenged by the first few chapters of this dissertation. In the latter, we took indeed a 180 degree turn, asking for the reader to take a small leap of faith while embarking on a journey into theory. Chapter 2 reviewed some key concepts from functional analysis and established very general representer theorems, pertaining to optimisation in abstract Hilbert and Banach spaces. Chapters 3 and 4 used the theory of spherical harmonic analysis to construct spherical pseudo-differential operators and spherical splines.

In Chapter 5 the reasons for this mathematical prelude became clearer. We showed how the various concepts of Chapters 2 to 4 came naturally into play when revisiting spherical approximation problems as specific instances of generic functional inverse problems on the sphere. Unlike ad-hoc discrete methods traditionally favoured by practitioners, functional inverse problems present the advantage of being directly formulated in the continuous spherical domain, the natural domain for analogue spherical signals encountered in nature. In Theorems 5.3 and 5.4 we showed that, if regularised by means of gTikhonov and gTV regularisation, functional inverse problems admitted *finite dimensional* solutions, which could hence be estimated in practice despite being defined over a *continuous* domain. For gTikhonov regularisation, we showed in Theorem 5.3 that the solution was *unique* and could be expressed as a linear combination of the sampling linear functionals –modelling the instrument– primitived twice with respect to the regularising pseudo-differential operator $\mathscr{D}$. For gTV regularisation, we showed in Theorem 5.4 that the solutions were convex combinations of spherical $\mathscr{D}$-splines with *sparse* innovations, i.e. less than available measurements. These two representer theorems not only

allowed us to compare the effects of both regularisation strategies, but also inspired in Chapter 6 two canonical search space discretisation schemes, *exact* for gTikhonov regularisation and with controlled approximation error for gTV regularisation (see Theorems 6.2 and 6.5 respectively).

In Chapter 7, we proposed algorithmic solutions adapted from the primal-dual splitting method and APGD to solve the discrete optimisation problems resulting from both discretisation schemes. The proposed algorithms were shown to be *computationally efficient*, provably *convergent* and compatible with *most common cost functionals* –including non-differentiable ones, such as the KL-divergence often used in the context of Poisson noise.

In Chapter 8, we introduced the last ingredient to our spherical approximation framework, namely the Wendland and Matérn splines, particularly convenient for practical purposes.

In Chapter 9, we were finally in a position to deliver on our promises from Chapter 1. We tested our continuous-domain spherical approximation framework and novel algorithms on a variety of real and simulated datasets, coming from the fields of meteorology, forestry, radio astronomy and planetary sciences. The sampling functionals, cost functions and regularisation strategies considered in each case were diverse, showing the versatility of both the theoretical framework and algorithmic solutions. In the meteorology example, we moreover illustrated the superiority of continuous-domain vs. discrete-domain recovery, both in terms of accuracy and resolution. This superiority was partially explained by the fact that continuous-domain methods could, unlike their discrete-domain counterparts, directly process the irregular spatial samples without resorting to *ad-hoc* gridding steps.

The last part of the thesis discussed related topics and paved the road towards promising new research avenues. In Chapter 10 we designed an efficient and locally convergent algorithm for recovering the spatial innovations of periodic Dirac streams with finite rates of innovation. If generalised to the sphere –and more generally the hypersphere, this algorithm could be envisioned as an alternative to the quasi-uniform spline discretisation scheme proposed in Chapter 6 for gTV regularised functional inverse problems.

In Chapter 11 finally, we showed how the convergence speed of proximal algorithms could be "boosted" by means of recurrent neural networks for purposes of real-time imaging. For simplicity, we illustrated the scheme with the specific case of PGD applied to acoustic imaging. The described methodology could easily be generalised to different spherical approximation tasks as well as more general algorithms such as the primal-dual splitting methods of Chapter 7.

In conclusion, we hope that the contributions of this thesis will spark interest among the community of practitioners, and give rise to the development of new reconstruction algorithms for spherical approximation problems. For readers looking for inspiration on how to continue this work further, we provide in the subsequent section a few additional research avenues.

2 Prospective Research Avenues

2.1 Robust Non-Convex Cost Functionals

A common remedy to misspecified noise models and/or strong outliers in data consists in using *non-convex* cost functionals [117, 123]. For example, *Tukey's bisquare function* is commonly used in robust *M-estimation* theory [121] to eliminate the effect of strong outliers. Being non-convex –as a matter of fact it is *semistrictly quasi-convex*,[1] this cost functional is unfortunately unusable in the context of the spherical approximation framework described in this thesis. Indeed, the abstract results Corollary 2.10 and Theorem 2.12 from Chapter 2 were both established under the assumption of a convex cost functional. It is hence necessary to extend these results to the case of non-convex cost functionals, or at least to the case of semistrictly/strictly quasi-convex cost functionals. We have good reasons to believe that this can be achieved:

- For Corollary 2.10, the convexity of the cost functional is exclusively used to show the existence and unicity of the solution. The form of the solutions should hence be unaffected by the use of a non-convex cost functional.
- For Theorem 2.12, the convexity of the cost functional is primarily used to show that the solution set is non-empty, convex and weak* compact –and hence from Theorem 2.11, the weak* convex hull of its extreme points. As explained in [28, Remark 3.5], it is still possible to invoke [28, Theorem 3.1] to characterise the form of the extreme points under the assumption of semistrict quasi-convexity of the cost functional. Moreover, the convexity of the solution set is still guaranteed for quasi-convex cost functionals (see [28, Remark 3.10]). It should therefore be possible to show the weak* compacity of the solution set in this case too, with similar arguments as in [72, Proposition 8].

Note that the use of non-convex cost functionals would also require adapting the primal-dual splitting and APGD methods from Chapter 7, whose convergence was only shown for convex cost functionals. In the case where the cost functional is differentiable and with Lipschitz continuous derivative,[2] one possibility could be to use the non-convex extension of APGD introduced in [109].

[1] See Remark 3.5 of [28] for a definition of strict, semistrict and standard quasi-convexity.

[2] Which is for example the case for Tukey's bisquare function.

2.2 Spherical Gaussian White Noise

Some spherical fields encountered in nature are well modelled by *Gaussian random fields* [114, 115, 180]. This is notably the case in radio astronomy where the source amplitudes typically fluctuate randomly and independently from one another [164, 181], making the source field resemble a (complex-valued) spherical *Gaussian white noise* [114, 115] entirely determined by an unknown *control measure* $\nu \in \mathcal{M}(\mathbb{S}^{d-1})$. The typical estimation task consists then in recovering this control measure using K independent realisations $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_K\} \subset \mathbb{C}^L$ of a *random measurement vector* of *integral white noises*:

$$\boldsymbol{Y} = \left[\int_{\mathbb{S}^{d-1}} \varphi_1(\boldsymbol{r}) f(d\boldsymbol{r}), \cdots, \int_{\mathbb{S}^{d-1}} \varphi_L(\boldsymbol{r}) f(d\boldsymbol{r}) \right],$$

where $\{\varphi_1, \dots, \varphi_L\} \subset \mathscr{L}^2(\mathbb{S}^{d-1}, \nu)$. Considering the empirical covariance matrix $\hat{\mathbf{\Sigma}} \in \mathbb{C}^{L \times L}$ of the data, we can show that

$$\mathbb{E}\left[\hat{\Sigma}_{ij}\right] = \int_{\mathbb{S}^{d-1}} \varphi_i(\boldsymbol{r})\overline{\varphi_j(\boldsymbol{r})}\nu(d\boldsymbol{r}) = \langle \nu | \varphi_i \overline{\varphi_j} \rangle, \qquad i, j = 1, \dots, L. \tag{12.1}$$

Since (12.1) falls into the scope of the generalised sampling framework from Chapter 5, we could consider recovering ν by means of an FPT or FPBP problem with sampling operator constructed from data model (12.1).

2.3 Spherical Fields Varying in Time

In certain applications, it can be interesting to monitor the evolution of a spherical field through time. This is notably the case in environmental monitoring tasks such as the meteorology and forestry examples from Chapter 9. To this end, one possibility consists in extending the approximation framework of this thesis to generalised functions defined over the product domain $\mathbb{S}^{d-1} \times [0, T[$, where the interval $[0, T[\subset \mathbb{R}$ represents the time-window of interest. Such an extension would notably require the definition of *spatio-temporal pseudo-differential operators*. Following Definition 4.1, the latter could be defined as "roughening" operators diagonalised by the Fourier basis on $\mathbb{S}^{d-1} \times [0, T[$, given by

$$Y_n^m(\boldsymbol{r}) \exp\left(j\frac{2\pi}{T}kt\right), \quad n \in \mathbb{N},\ m = 1, \dots, N_d(n),\ k \in \mathbb{Z}.$$

Note that for efficiency reasons, it could be beneficial to consider only spatio-temporal pseudo-differential operators with *separable* symbols and hence separable Green kernels. Finally, Theorems 5.3 and 5.4 should be relatively easy to extend to this particular setup, since they are based on the abstract results Corollary 2.10 and Theorem 2.12 which are in no way specific to the spherical domain.

2.4 Vector-Valued Spherical Fields

The spherical approximation framework from Chapter 5 is limited to *scalar* spherical fields $f : \mathbb{S}^{d-1} \to \mathbb{C}$. In some applications however, one may wish to recover *vector-valued* spherical fields $\boldsymbol{f} : \mathbb{S}^{d-1} \to \mathbb{C}^n$, for some $n \in \mathbb{N}$. This is typically the case in meteorology or oceanography, where wind and oceanic currents are studied via vector velocity maps. Since a vector-valued spherical field $\boldsymbol{f} : \mathbb{S}^{d-1} \to \mathbb{C}^n$ can be seen as tuple $(f_1, \dots, f_n)$ of scalar spherical fields, one could envision recovering such vector-valued spherical fields by finding solutions to the following FPT problem:

$$\min_{(f_1,\dots,f_n) \in \Pi_{k=1}^n \mathscr{H}_{\mathscr{D}_k}(\mathbb{S}^{d-1})} \left\{ F(\boldsymbol{y}, \boldsymbol{\Phi}(f_1, \dots, f_n)) + \lambda \sum_{k=1}^n \|\mathscr{D}_k f_k\|_2^2 \right\}, \tag{12.2}$$

where:

- $\Pi_{k=1}^n \mathscr{H}_{\mathscr{D}_k}(\mathbb{S}^{d-1})$ denotes the *direct product* of the Hilbert spaces $\mathscr{H}_{\mathscr{D}_k}(\mathbb{S}^{d-1})$, equipped with its canonical inner product norm,

- $\{\mathscr{D}_k : \mathscr{S}(\mathbb{S}^{d-1}) \to \mathscr{S}(\mathbb{S}^{d-1})\}_{k=1,\dots,n}$ are suitable *spherical pseudo-differential operators*,
- $F : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_+ \cup \{+\infty\}$ is some *proper* and *convex* cost functional,
- $\boldsymbol{y} \in \mathbb{C}^L$ is some *measurement vector*,
- λ is a *real positive constant*,
- $\boldsymbol{\Phi} : \Pi_{k=1}^n \mathscr{H}_{\mathscr{D}_k}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ is a *sampling operator* given by:

$$\boldsymbol{\Phi}(f_1, \dots, f_n) = \sum_{k=1}^n (\langle \varphi_{1,k} | f_k \rangle, \dots, \langle \varphi_{L,k} | f_k \rangle),$$

where $\{(\varphi_{i,1}, \dots, \varphi_{i,n})\}_{i=1,\dots,L} \subset \Pi_{k=1}^n \mathscr{H}'_{\mathscr{D}_k}(\mathbb{S}^{d-1})$ are some linearly independent sampling functionals.

It should then be doable to derive a representer theorem for (12.2) using Corollary 2.10. Similarly, one could consider recovering vector-valued spherical fields by means of the following FPBP problem:

$$\min_{(f_1,\dots,f_n) \in \Pi_{k=1}^n \mathcal{M}_{\mathscr{D}_k}(\mathbb{S}^{d-1})} \left\{ F(\boldsymbol{y}, \boldsymbol{\Phi}(f_1, \dots, f_n)) + \lambda \sum_{k=1}^n \|\mathscr{D}_k f_k\|_{TV} \right\}, \quad (12.3)$$

with this time, a sampling operator $\boldsymbol{\Phi} : \Pi_{k=1}^n \mathcal{M}_{\mathscr{D}_k}(\mathbb{S}^{d-1}) \to \mathbb{C}^L$ given by:

$$\boldsymbol{\Phi}(f_1, \dots, f_n) = \sum_{k=1}^n (\langle f_k | \varphi_{1,k} \rangle, \dots, \langle f_k | \varphi_{L,k} \rangle),$$

for some linearly independent sampling functionals $\{(\varphi_{i,1}, \dots, \varphi_{i,n})\}_{i=1,\dots,L}$ in $\Pi_{k=1}^n \mathscr{C}_{\mathscr{D}_k}(\mathbb{S}^{d-1})$. Again, provided that one can characterise the extreme points of the unit ball of the composite gTV norm $\sum_{k=1}^n \|\mathscr{D}_k f_k\|_{TV}$ on $\Pi_{k=1}^n \mathcal{M}_{\mathscr{D}_k}(\mathbb{S}^{d-1})$, it should be possible to use Theorem 2.12 so as to derive a representer theorem for (12.3). Note that an optimisation problem very similar to (12.3) was considered in [9] in the case of vector fields defined over $\mathbb{R}^d$.

2.5 Biased Measurements

The generalised sampling framework introduced in Chapter 5 assumes an *unbiased* measurement process, i.e. $\mathbb{E}[\boldsymbol{y}] = \boldsymbol{\Phi}(f)$. In practice however, it can happen that the measurements provided by each sensor are systematically biased yielding a new data model of the form

$$\mathbb{E}[\boldsymbol{y}] = \boldsymbol{\Phi}(f) + \boldsymbol{\mu},$$

where $\boldsymbol{\mu} \in \mathbb{C}^L$ is an unknown vector describing the bias introduced at each sensor. In such a case, it is possible to update the FPT and FPBP problems from Chapter 5 so as to jointly estimate the spherical field and the unknown sensor biases:

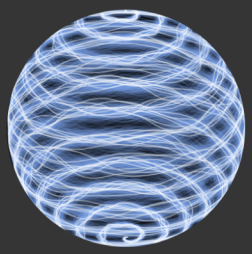
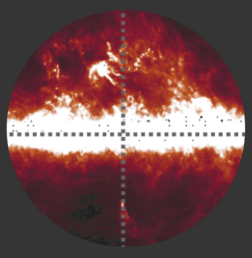
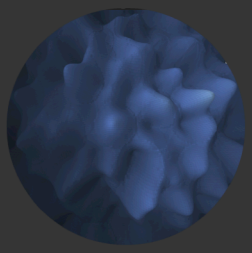
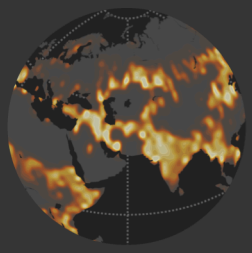
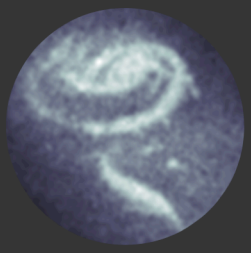
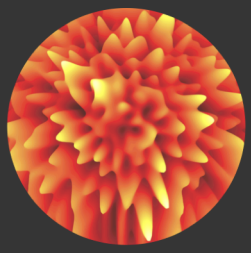
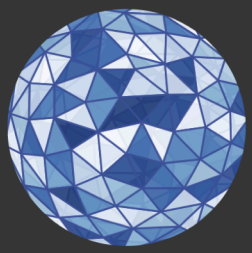
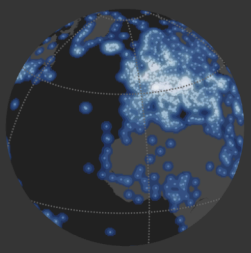
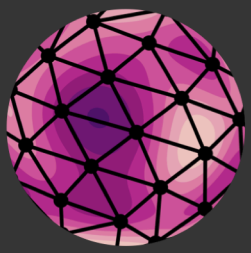
$$\min_{(f,\boldsymbol{\mu}) \in \mathscr{H}_{\mathscr{D}}(\mathbb{S}^{d-1}) \times \mathbb{C}^L} \left\{ F(\boldsymbol{y}, \boldsymbol{\Phi}(f) + \boldsymbol{\mu}) + \lambda \left(\|\mathscr{D} f\|_2^2 + \|\boldsymbol{D}\boldsymbol{\mu}\|_2^2 \right) \right\}, \quad (12.4)$$

$$\min_{(f,\boldsymbol{\mu})\in\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})\times\mathbb{C}^L} \{F(\boldsymbol{y}, \boldsymbol{\Phi}(f)+\boldsymbol{\mu}) + \lambda\left(\|\mathscr{D}f\|_{TV} + \|\boldsymbol{D}\boldsymbol{\mu}\|_1\right)\}, \tag{12.5}$$

where $\boldsymbol{D} \in \mathbb{C}^{L\times L}$. Again, it should be relatively easy to reinterpret (12.4) and (12.5) as specific instances of the generic optimisation problems (2.14) and (2.17) from Corollary 2.10 and Theorem 2.12 respectively. This would allow us to derive representer theorems for these two cases as well, provided that the extreme points of the regularisation ball in (12.5) can indeed be characterised.

V

Appendices

A Supplementary Material to Chapter 9

1 Sufficient Condition for $\mathscr{L}^2(\mathbb{S}^{d-1}) \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$

Proposition A.1 — *Sufficient Condition for $\mathscr{L}^2(\mathbb{S}^{d-1}) \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$.* *Let $\mathscr{D}$ be a* pseudo-differential operator *as in Definition 4.1 with trivial nullspace and* spectral growth order $p > (d-1)/2$, *and $(\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1}), \|\mathscr{D}\cdot\|_{TV})$ the space defined in eq. (5.5) equipped with the gTV norm. Then, all square-integrable functions are included in the predual $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ of $\mathcal{M}_{\mathscr{D}}(\mathbb{S}^{d-1})$, i.e.*

$$\mathscr{L}^2(\mathbb{S}^{d-1}) \subset \mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1}).$$

Proof. From Proposition 5.2, a function $f \in \mathscr{L}^2(\mathbb{S}^{d-1})$ is in $\mathscr{C}_{\mathscr{D}}(\mathbb{S}^{d-1})$ is there exists $\eta \in \mathscr{C}(\mathbb{S}^{d-1})$ s.t. $f = \mathscr{D}\eta$. Since $\mathscr{D}$ is assumed invertible, this is equivalent to requiring that $\mathscr{D}^{-1}f \in \mathscr{C}(\mathbb{S}^{d-1})$, which is guaranteed if the series of functions

$$(\mathscr{D}^{-1}f)(\boldsymbol{r}) = \sum_{n\in\mathbb{N}} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} \hat{f}_n^m Y_n^m(\boldsymbol{r}), \quad \boldsymbol{r} \in \mathbb{S}^{d-1}, \tag{A.1}$$

converges *uniformly* (see [125, Theorem 2.14]). To show that (A.1) is uniformly convergent, we consider its remainder for some $N \in \mathbb{N}$. Then, from the addition theorem 3.2 and the Cauchy-Schwarz inequality we get, for each $\boldsymbol{r} \in \mathbb{S}^{d-1}$:

$$\begin{aligned}\left|\sum_{n=N}^{+\infty} \frac{1}{\hat{D}_n} \sum_{m=1}^{N_d(n)} \hat{f}_n^m Y_n^m(\boldsymbol{r})\right| &\leq \left|\sum_{n=N}^{+\infty} \frac{\sum_{m=1}^{N_d(n)} |Y_n^m(\boldsymbol{r})|^2}{|\hat{D}_n|^2}\right| \left|\sum_{n=N}^{+\infty} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2\right| \\ &= \left|\sum_{n=N}^{+\infty} \frac{N_d(n)}{\mathfrak{a}_d |\hat{D}_n|^2}\right| \left|\sum_{n=N}^{+\infty} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2\right|.\end{aligned}$$

Since $f \in \mathscr{L}^2(\mathbb{S}^{d-1})$ we have trivially $\lim_{N\to+\infty} \left|\sum_{n=N}^{+\infty} \sum_{m=1}^{N_d(n)} |\hat{f}_n^m|^2\right| = 0$. Moreover, since $|\hat{D}_n| = \Theta(n^p)$ we have from (3.3) $N_d(n)|\hat{D}_n|^{-2} = \mathcal{O}(n^{d-2-2p})$. Since $p > (d-1)/2 \Rightarrow d-2-2p < -1$ we have hence that the series

$\sum_{n\in\mathbb{N}} \frac{N_d(n)}{\mathfrak{a}_d|\hat{D}_n|^2}$ is convergent and hence is remainder tends to zero. We have hence

$$\left|(\mathscr{D}^{-1}f)(\boldsymbol{r}) - \sum_{n=0}^{N-1}\frac{1}{\hat{D}_n}\sum_{m=1}^{N_d(n)}\hat{f}_n^m Y_n^m(\boldsymbol{r})\right| = \left|\sum_{n=N}^{+\infty}\frac{1}{\hat{D}_n}\sum_{m=1}^{N_d(n)}\hat{f}_n^m Y_n^m(\boldsymbol{r})\right|$$
$$\leq \left|\sum_{n=N}^{+\infty}\frac{N_d(n)}{\mathfrak{a}_d|\hat{D}_n|^2}\right|\left|\sum_{n=N}^{+\infty}\sum_{m=1}^{N_d(n)}|\hat{f}_n^m|^2\right| \overset{N\to+\infty}{\longrightarrow} 0.$$

Moreover, since the upper bound is independent on $\boldsymbol{r}$ the convergence is uniform, which achieves the proof. ■

2 Sea Surface Temperature Anomalies

In this section, we consider the same setup as in Section 1 of Chapter 9, but replace the pseudo-differential operator $\mathscr{D}^{\epsilon}_{2.5}$ in (9.2) and (9.5) with the pseudo-differential operator $\mathscr{D}^{\eta}_{3,1}$ associated to the Wendland zonal Green kernel with scale $\eta \simeq 0.09$ –corresponding again to an angular resolution[1] of approximately 4°:

[1] The angular resolution is measured here as the full width at half maximum (FWHM) of the Wednland kernel.

$$\psi^{\eta}_{3,1}(\langle \boldsymbol{r},\boldsymbol{s}\rangle) = \phi_{3,1}\left(\frac{\sqrt{2-2\langle \boldsymbol{r},\boldsymbol{s}\rangle}}{\eta}\right), \qquad \forall(\boldsymbol{r},\boldsymbol{s})\in\mathbb{S}^2\times\mathbb{S}^2.$$

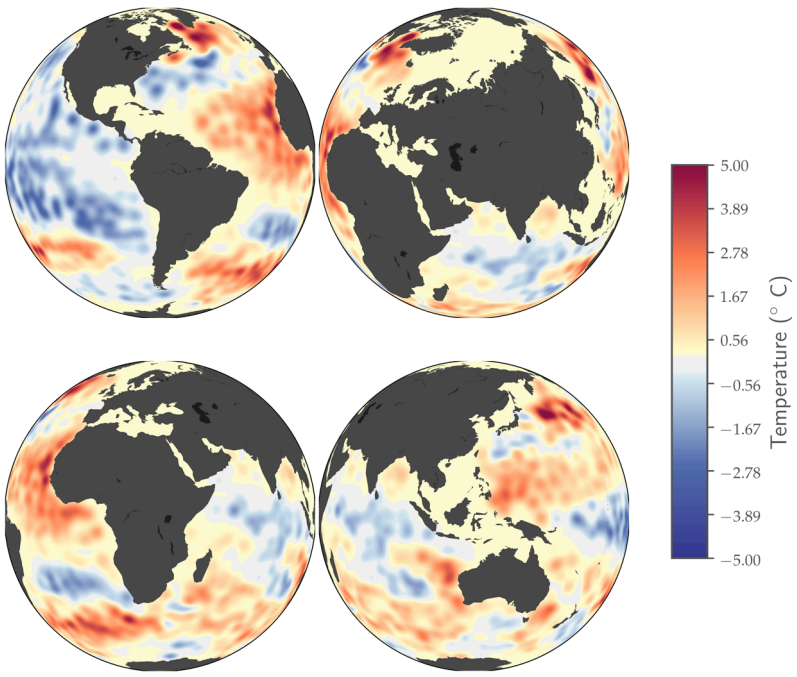


Figure A.1: Sea surface temperature anomaly function obtained by solving the FPT problem (9.2), with Wendland gTikhonov regularisation.

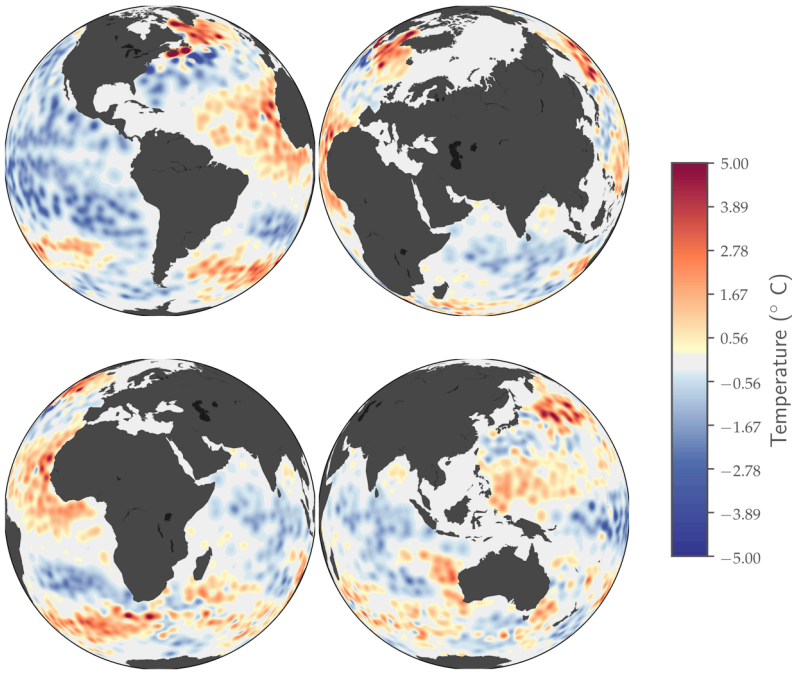


Figure A.2: Sea surface temperature anomaly function obtained by solving the FPBP problem (9.5), with Wendland gTV regularisation.

Note that $\mathscr{D}^{\varepsilon}_{2.5}$ and $\mathscr{D}^{\eta}_{3,1}$ are both equivalent (in a sense described in Chapter 8) to the Sobolev operator $(\mathrm{Id} - \Delta_{\mathbb{S}^2})^{2.5}$. The estimates obtained by solving (9.2) and (9.5) with the redefined Wendland gTikhonov and gTV regularisation terms are provided in Figs. A.1 and A.2 respectively. Not surprisingly, they look very similar to those displayed in Figs. 9.2a and 9.2b, with the only difference that the spherical map in Fig. A.2 appears slightly sparser than the one in Fig. 9.2b.

B

Main Proofs of Chapter 10

1 The Toeplitzification Operator and Convolutions

Consider the Toeplitzification operator defined in (10.5). When multiplied with a vector $\boldsymbol{u} = [u_1, \cdots u_{P+1}] \in \mathbb{C}^{P+1}$, the matrix $T_P(\boldsymbol{x})$ returns the valid part of the convolution between the two zero-padded sequences:

$$\tilde{x} = \left[\cdots, 0, x_{-M}, \cdots, \boxed{x_0}, \cdots, x_M, 0, \cdots\right] \in \mathbb{C}^{\mathbb{Z}}$$

and

$$\tilde{u} = \left[\cdots, \boxed{0}, u_1, \cdots, u_{P+1}, 0, \cdots\right] \in \mathbb{C}^{\mathbb{Z}}.$$

Indeed,

$$(\tilde{x} * \tilde{u})[k] = \sum_{j\in\mathbb{Z}} \tilde{x}_{k-j}\tilde{u}_j = \sum_{j=1}^{P+1} \tilde{x}_{k-j}u_j.$$

The valid part corresponds to the indices i for which all the terms in the summation are non-zero. This is the case when

$$k \in [-M + P + 1, \ldots, M + 1].$$

Consider $i = k + M - P$ we get that the valid part of the convolution is given by

$$\begin{aligned}(\tilde{x} * \tilde{u})[i - M + P] &= \sum_{j=1}^{P+1} x_{-M+P+i-j}u_j, \\ &= \sum_{j=1}^{P+1} [T_P(\boldsymbol{x})]_{i,j}\, u_j, \quad i = 1, \ldots, N - P,\end{aligned}$$

which corresponds precisely to $T_P(\boldsymbol{x})\boldsymbol{u}$.

2 Proofs of Theorems 10.4 and 10.5

The proofs of Theorems 10.4 and 10.5 rely on the following lemma, adapted from [109, Theorem 1], which establishes the convergence of PGD in a general setup:

Lemma B.1 — *Convergence of PGD.* *Consider the norm* $\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, $\boldsymbol{x} \in \mathbb{R}^n$, *induced by some inner product* $\langle \cdot, \cdot \rangle$ *on* $\mathbb{R}^n$. *Consider moreover the general problem:*

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \Phi(\boldsymbol{x}) = F(\boldsymbol{x}) + H(\boldsymbol{x}), \tag{B.1}$$

where $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *and* $H : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *are potentially* non-convex *functions such that:*

1. F *is a* proper *function, i.e. its domain is non-empty,* differentiable *and with* Lipschitz continuous *gradient for some Lipschitz constant* $0 \leq \beta < +\infty$,
$$\|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\| \leq \beta \|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$
2. H *is a* proper *and lower semi-continuous (lwsc) function, potentially* non-smooth.
3. $\Phi = F + H$ *is* coercive, *i.e.* Φ *is bounded from below and*
$$\lim_{\|\boldsymbol{x}\| \to +\infty} \Phi(\boldsymbol{x}) = +\infty.$$

Then, the iterates $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$ *generated by the proximal gradient descent (PGD) applied to (B.1):*

$$\boldsymbol{x}_{k+1} \in prox_{\tau H}\left(\boldsymbol{x}_k - \tau \nabla F\left(\boldsymbol{x}_k\right)\right), \quad k \geq 0, \tag{B.2}$$

with $\tau < 1/\beta$ *and* $\boldsymbol{x}_0 \in \mathbb{R}^n$, *are* bounded. *Moreover, any limit point* $\boldsymbol{x}_\star$ *of* $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$ *is a* local minimum *of* Φ.

Proof. The lemma is easily shown by specifying the proof of [109, Theorem 1] to the non-accelerated case. For the sake of completeness, it is provided hereafter. From the definition of the proximal operator,

$$\text{prox}_{\tau H}(\boldsymbol{x}) : \begin{cases} \mathbb{R}^n \to \mathcal{P}\left(\mathbb{R}^n\right), \\ \boldsymbol{x} \mapsto \arg\min_{\boldsymbol{z} \in \mathbb{R}^n} \dfrac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{z}\|^2 + H(\boldsymbol{z}), \end{cases}$$

we can reinterpret (B.2) as

$$\boldsymbol{x}_{k+1} \in \arg\min_{\boldsymbol{z} \in \mathbb{R}^n} \frac{1}{2\tau} \|\boldsymbol{z} - \boldsymbol{x}_k\|^2 + \langle \nabla F(\boldsymbol{x}_k), \boldsymbol{z} - \boldsymbol{x}_k \rangle + H(\boldsymbol{z}). \tag{B.3}$$

We have hence

$$\frac{1}{2\tau} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 + \langle \nabla F(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \rangle + H(\boldsymbol{x}_{k+1}) \leq H(\boldsymbol{x}_k).$$

From the Lipschitz continuity of ∇F we have moreover

$$\begin{aligned}\Phi(\boldsymbol{x}_{k+1}) \leq & H(\boldsymbol{x}_{k+1}) + F(\boldsymbol{x}_k) + \langle \nabla F(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \rangle \\ & + \frac{\beta}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \\ \leq & H(\boldsymbol{x}_k) - \frac{1}{2\tau}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 - \langle \nabla F(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \rangle \\ & + F(\boldsymbol{x}_k) + \langle \nabla F(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \rangle + \frac{\beta}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \\ = & \Phi(\boldsymbol{x}_k) - \left(\frac{1}{2\tau} - \frac{\beta}{2}\right)\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2. \end{aligned} \tag{B.4}$$

Since $\tau < 1/\beta$ we have hence $(1/2\tau - \beta/2) \geq 0$ and

$$\Phi(\boldsymbol{x}_{k+1}) \leq \Phi(\boldsymbol{x}_k) \leq \Phi(\boldsymbol{x}_0), \quad \forall k \geq 1.$$

The sequence $\{\Phi(\boldsymbol{x}_k)\}_{k\in\mathbb{N}}$ is hence bounded and since Φ is coercive so is $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$. The sequence $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ admits hence limit points. Moreover, since $\Phi(\boldsymbol{x}_k)$ is decreasing and bounded from below, it takes the same value $\Phi_\star \in \mathbb{R}$ at all of these limit points. Summing (B.4), we obtain hence:

$$\left(\frac{1}{2\tau} - \frac{\beta}{2}\right)\sum_{k=0}^{+\infty}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \Phi(\boldsymbol{x}_0) - \Phi_\star < +\infty.$$

Since $\tau < 1/\beta$, we have necessarily $\sum_{k=0}^{+\infty}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 < +\infty$, which yields

$$\lim_{k\to+\infty}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| = 0. \tag{B.5}$$

From the optimality condition (B.3) and Items 1 and 3 of Proposition 1 of the supplementary material of [109], we have moreover

$$\begin{aligned}\mathbf{0}_n \in & \nabla F(\boldsymbol{x}_k) + \frac{1}{\tau}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \partial H(\boldsymbol{x}_{k+1}) \\ & = \partial\Phi(\boldsymbol{x}_{k+1}) - \nabla F(\boldsymbol{x}_{k+1}) + \nabla F(\boldsymbol{x}_k) + \frac{1}{\tau}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k), \end{aligned} \tag{B.6}$$

where $\partial H : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ and $\partial\Phi : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ denote the (set-valued) *subdifferential operators* of H and Φ respectively (see Definition 2 of the supplementary material of [109]).

Equation (B.6) can moreover be rewritten as

$$\nabla F(\boldsymbol{x}_{k+1}) - \nabla F(\boldsymbol{x}_k) - \frac{1}{\tau}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \in \partial\Phi(\boldsymbol{x}_{k+1}).$$

Furthermore, from the Lipschitz continuity of F, we have

$$\|\nabla F(\boldsymbol{x}_{k+1}) - \nabla F(\boldsymbol{x}_k) - \frac{1}{\tau}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)\| \leq \left(\beta + \frac{1}{\tau}\right)\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|,$$

and hence from (B.5):

$$\lim_{\|\boldsymbol{x}\|\to+\infty} \left\| \nabla F(\boldsymbol{x}_{k+1}) - \nabla F(\boldsymbol{x}_k) - \frac{1}{\tau}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \right\| = 0. \tag{B.7}$$

Let $\{\boldsymbol{x}_{k_j}\}_{j\in\mathbb{N}}$ be a convergent subsequence of $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$, with limit $\boldsymbol{x}_\star$. Then, we have from (B.7) and Item 2 of Proposition 1 of the supplementary material of [109]:

$$\mathbf{0}_n \in \lim_{j\to+\infty} \partial\Phi(\boldsymbol{x}_{k_j}) = \partial\Phi(\boldsymbol{x}_\star),$$

which completes the proof. ■

We now show Theorems 10.4 and 10.5 by applying Lemma B.1 to the implicit genFRI problem in unconstrained form (B.8)

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \|\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \;+\; \iota_{\mathcal{H}_K}\left(T_P\left(\boldsymbol{x}\right)\right) \;+\; \iota_{\mathbb{B}_\rho^\Gamma}\left(\boldsymbol{x}\right). \tag{B.8}$$

To do so, we must first convert (B.8) into an optimisation problem of the form (B.1), defined over $\mathbb{R}^n$ for some $n \in \mathbb{N}$. We achieve this by proceeding as in [148, Section 7.8] and identifying $\mathbb{C}^N$ with $\mathbb{R}^{2N}$ (respectively $\mathbb{C}^L$ with $\mathbb{R}^{2L}$) in the canonical way

$$\boldsymbol{x} \in \mathbb{C}^N \leftrightarrow \hat{\boldsymbol{x}} := \begin{bmatrix} \Re(\boldsymbol{x}) \\ \Im(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{2N},$$

where $\Re$ and $\Im$ denote the real and imaginary parts respectively. Such an identification makes the canonical inner products and norms on $\mathbb{C}^N$ and $\mathbb{R}^{2N}$ (respectively $\mathbb{C}^L$ and $\mathbb{R}^{2L}$) consistent with one another, i.e. for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$, we have

$$\langle \boldsymbol{x}, \boldsymbol{z}\rangle_{\mathbb{C}^N} = \boldsymbol{z}^H\boldsymbol{x} = \Re(\boldsymbol{z})^T\Re(\boldsymbol{x}) + \Im(\boldsymbol{z})^T\Im(\boldsymbol{x}) = \hat{\boldsymbol{z}}^T\hat{\boldsymbol{x}} = \langle \hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}\rangle_{\mathbb{R}^{2N}},$$

and

$$\|\boldsymbol{x}\|_{\mathbb{C}^N} = \sqrt{\boldsymbol{x}^H\boldsymbol{x}} = \sqrt{\|\Re(\boldsymbol{x})\|_{\mathbb{R}^N}^2 + \|\Im(\boldsymbol{x})\|_{\mathbb{R}^N}^2} = \|\hat{\boldsymbol{x}}\|_{\mathbb{R}^{2N}}.$$

Still following [148, Section 7.8], we moreover identify the linear map $\boldsymbol{G} : \mathbb{C}^N \to \mathbb{C}^L$ with a linear map $\hat{\boldsymbol{G}} : \mathbb{R}^{2N} \to \mathbb{R}^{2L}$ with matrix representation:

$$\hat{\boldsymbol{G}} := \begin{bmatrix} \Re(\boldsymbol{G}) & -\Im(\boldsymbol{G}) \\ \Im(\boldsymbol{G}) & \Re(\boldsymbol{G}) \end{bmatrix} \in \mathbb{R}^{2L\times 2N}.$$

Again, it is easy to show that the two operators are consistent, in the sense that

$$\widehat{\boldsymbol{G}\boldsymbol{x}} = \hat{\boldsymbol{G}}\hat{\boldsymbol{x}}, \quad \text{and} \quad \widehat{\boldsymbol{G}^H\boldsymbol{x}} = \hat{\boldsymbol{G}}^T\hat{\boldsymbol{x}}, \quad \forall \boldsymbol{x} \in \mathbb{C}^N.$$

Similarly, the Toeplitzification operator $T_P : \mathbb{C}^N \to \mathbb{C}^{(N-P)\times(P+1)}$ is identified with the linear operator $\hat{T}_P : \mathbb{R}^{2N} \to \mathbb{C}^{(N-P)\times(P+1)}$ defined as

$$\hat{T}_P(\hat{\boldsymbol{x}}) := T_P(\Re(\boldsymbol{x})) + jT_P(\Im(\boldsymbol{x})), \quad \forall \boldsymbol{x} \in \mathbb{C}^N,$$

where j is the complex 2-root of unity. From the linearity of T_P, this definition yields indeed $T_P(\boldsymbol{x}) = \hat{T}_P(\hat{\boldsymbol{x}})$. Finally, the $\boldsymbol{\Gamma}$-ball $\mathbb{B}_\rho^{\boldsymbol{\Gamma}} \subset \mathbb{C}^N$ is identified with

$$\mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}} := \left\{\hat{\boldsymbol{x}} \in \mathbb{R}^{2N} : \|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \leq \rho\right\},$$

where $\hat{\boldsymbol{\Gamma}} \in \mathbb{R}^{2N\times 2N}$ is a positive definite and diagonal matrix defined as

$$\hat{\boldsymbol{\Gamma}} := \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \boldsymbol{\Gamma} \end{bmatrix}.$$

Again, we trivially have $\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} = \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}$ and hence $\boldsymbol{x} \in \mathbb{B}_\rho^{\boldsymbol{\Gamma}} \Leftrightarrow \hat{\boldsymbol{x}} \in \mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}}$ for all $\boldsymbol{x} \in \mathbb{C}^N$.
In summary, the optimisation problem (B.8) is hence equivalent to the following optimisation problem with search space $\mathbb{R}^{2N}$:

$$\min_{\hat{\boldsymbol{x}}\in\mathbb{R}^{2N}} \left\|\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\right\|_{\mathbb{R}^{2L}}^2 + \iota_{\mathcal{H}_K}\left(\hat{T}_P\left(\hat{\boldsymbol{x}}\right)\right) + \iota_{\mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}}}\left(\hat{\boldsymbol{x}}\right). \tag{B.9}$$

Letting $\hat{F}(\hat{\boldsymbol{x}}) := \left\|\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\right\|_{\mathbb{R}^{2L}}^2$ and $\hat{H}(\hat{\boldsymbol{x}}) := \iota_{\mathcal{H}_K}\left(\hat{T}_P\left(\hat{\boldsymbol{x}}\right)\right) + \iota_{\mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}}}\left(\hat{\boldsymbol{x}}\right)$ we have $\hat{F} : \mathbb{R}^{2N} \to \mathbb{R}_+$ and $\hat{H} : \mathbb{R}^{2N} \to \{0, +\infty\}$, so that (B.9) is indeed of the form (B.1). We must now verify assumptions 1, 2 and 3 of Lemma B.1:

1. $\hat{F}$ is proper, differentiable and $\nabla\hat{F}$ Lipschitz continuous. $\hat{F}$ is proper since

$$\hat{F}(\mathbf{0}_{2N}) = \|\hat{\boldsymbol{y}}\|_{\mathbb{R}^{2L}}^2 = \|\boldsymbol{y}\|_{\mathbb{C}^L}^2 < +\infty.$$

It is differentiable, with gradient given by

$$\nabla\hat{F}(\hat{\boldsymbol{x}}) = 2\hat{\boldsymbol{G}}^T(\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}) = \widehat{\nabla F(\boldsymbol{x})}, \quad \hat{\boldsymbol{x}} \in \mathbb{R}^{2N}. \tag{B.10}$$

The gradient (B.10) is moreover $\hat{\beta}$-*Lipschitz continuous* with respect to the norm $\|\cdot\|_{\hat{\boldsymbol{\Gamma}}}$ on $\mathbb{R}^{2N}$, and its Lipschitz constant is given by:

$$\begin{aligned}
\hat{\beta} &= 2\|\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}\|_{\hat{\boldsymbol{\Gamma}}} \\
&= \sup\left\{2\left\|\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}\hat{\boldsymbol{x}}\right\|_{\hat{\boldsymbol{\Gamma}}} : \hat{\boldsymbol{x}} \in \mathbb{R}^{2N}, \|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} = 1\right\} \\
&= \sup\left\{2\left\|\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{x}\right\|_{\boldsymbol{\Gamma}} : \boldsymbol{x} \in \mathbb{C}^N, \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}} = 1\right\} \\
&= \beta < +\infty.
\end{aligned} \tag{B.11}$$

■

2. $\hat{H}$ is proper and lower semi-continuous. $\hat{H}$ is proper since for all $\rho > 0$, and K,

$$\hat{H}(\mathbf{0}_{2N}) = \iota_{\mathcal{H}_K}\left(\mathbf{0}_{(N-P)\times(P+1)}\right) + \iota_{\mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}}}\left(\mathbf{0}_{2N}\right) = 0 < +\infty.$$

The indicator functions are moreover lower semi-continuous since the sets $\mathcal{H}_K$ and $\mathbb{B}_\rho^{\hat{\boldsymbol{\Gamma}}}$ are both closed. Since T_P is a bounded linear operator, it is contin-

uous and hence $\hat{H}$ is indeed lower semi-continuous as composition between continuous and lower semi-continuous functions. ■

3. $\hat{\Phi} = \hat{F} + \hat{H}$ is coercive. It is easy to see that $\hat{\Phi} = \hat{F} + \hat{H} \geq 0$. To show that $\hat{\Phi}$ is coercive, it is hence sufficient to show that

$$\lim_{\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \to +\infty} \hat{\Phi}(\hat{\boldsymbol{x}}) = +\infty.$$

To this end, we distinguish two cases, which correspond respectively to the assumptions of Theorems 10.4 and 10.5:

1. $\rho \in]0, +\infty[$: in this case $\hat{\Phi}$ is trivially coercive since

$$\iota_{\mathbb{B}^{\hat{\boldsymbol{\Gamma}}}_{\rho}}(\hat{\boldsymbol{x}}) = +\infty, \qquad \forall \|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \geq \rho.$$

2. $\rho = +\infty$ and $\boldsymbol{G}$ injective: When $\rho = +\infty$, the term $\iota_{\mathbb{B}^{\hat{\boldsymbol{\Gamma}}}_{\rho}}$ is always null and $\hat{\Phi}$ simplifies to

$$\hat{\Phi}(\hat{\boldsymbol{x}}) = \left\| \hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \right\|^2_{\mathbb{R}^{2L}} + \iota_{\mathcal{H}_K}\left(\hat{T}_P(\hat{\boldsymbol{x}})\right), \quad \hat{\boldsymbol{x}} \in \mathbb{R}^{2N}.$$

From [148, Section 7.8], we have moreover that

$$\det\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}\right) = |\det(\boldsymbol{G}^H\boldsymbol{G})|^2 \neq 0, \tag{B.12}$$

since $\boldsymbol{G}$ is injective by assumption. From the reverse triangle inequality, we have hence

$$\left\| \hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \right\|_{\mathbb{R}^{2L}} \geq \sigma_{\min}\|\hat{\boldsymbol{x}}\|_{\mathbb{R}^{2N}} - \|\hat{\boldsymbol{y}}\|_{\mathbb{R}^{2L}}, \quad \forall \hat{\boldsymbol{x}} \in \mathbb{R}^{2N},$$

where $\sigma_{\min} = \sqrt{\lambda_{\min}(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}})} > 0$ is the square root of the eigenvalue of $\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}$ with lowest magnitude, which is non-null from (B.12). From the equivalence of norms in finite dimensions, there exist moreover $c_1, c_2 > 0$ such that

$$c_1\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \leq \|\hat{\boldsymbol{x}}\|_{\mathbb{R}^{2N}} \leq c_2\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}}, \quad \forall \hat{\boldsymbol{x}} \in \mathbb{R}^{2N}.$$

This yields

$$\left\| \hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \right\|_{\mathbb{R}^{2L}} \geq \sigma_{\min}c_1\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} - \|\hat{\boldsymbol{y}}\|_{\mathbb{R}^{2L}}, \quad \forall \hat{\boldsymbol{x}} \in \mathbb{R}^{2N},$$

and hence

$$\lim_{\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \to +\infty} \left\| \hat{\boldsymbol{G}}\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \right\|_{\mathbb{R}^{2L}} \geq \lim_{\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} \to +\infty} \sigma_{\min}c_1\|\hat{\boldsymbol{x}}\|_{\hat{\boldsymbol{\Gamma}}} = +\infty,$$

which shows that $\hat{\Phi}$ is indeed coercive.

■

We can hence apply Lemma B.1, to show that the iterates $\{\hat{\boldsymbol{x}}_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^{2N}$ generated by PGD applied to (B.9):

$$\hat{\boldsymbol{x}}_{k+1} \in \operatorname{prox}^{\hat{\boldsymbol{\Gamma}}}_{\tau\hat{H}}\left(\hat{\boldsymbol{x}}_k - \tau\nabla\hat{F}\left(\hat{\boldsymbol{x}}_k\right)\right), \tag{B.13}$$

with $\tau < 1/\hat{\beta}$ and $\hat{\boldsymbol{x}}_0 \in \mathbb{R}^{2N}$, are bounded. Moreover, any limit point $\hat{\boldsymbol{x}}_\star$ of $\{\hat{\boldsymbol{x}}_k\}_{k\in\mathbb{N}}$ is a critical point of (B.9).

Observe finally, that the iterations (B.13) can be rewritten in complex form as

$$\boldsymbol{x}_{k+1} \in \operatorname{prox}^{\boldsymbol{\Gamma}}_{\tau H}\left(\boldsymbol{x}_k - \tau\nabla F\left(\boldsymbol{x}_k\right)\right), \tag{B.14}$$

with $\tau < 1/\beta$ and $\boldsymbol{x}_0 \in \mathbb{C}^N$, and where we have used (B.10), (B.11) and

$$\begin{aligned}\operatorname{prox}^{\hat{\boldsymbol{\Gamma}}}_{\tau\hat{H}}(\hat{\boldsymbol{x}}) &= \arg\min_{\hat{\boldsymbol{z}}\in\mathbb{R}^{2N}} \frac{1}{2\tau}\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{z}}\|^2_{\hat{\boldsymbol{\Gamma}}} + \hat{H}(\boldsymbol{z}) \\ &= \widehat{\operatorname{prox}^{\boldsymbol{\Gamma}}_{\tau H}(\boldsymbol{x})}, \quad \forall\hat{\boldsymbol{x}} \in \mathbb{R}^{2N},\end{aligned}$$

which follows trivially from the previous identifications. By identification and equivalence between the real and complex optimisation problems (B.9) and (B.8), we can hence conclude that limit points of the iterates $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}} \subset \mathbb{C}^N$ generated by (B.14) are critical points of (B.8), which achieves the proof. ■

3 Proof of Theorem 10.8

The proof of Theorem 10.8 relies on the four lemmas hereafter. The first lemma shows that gradient descent is Lipschitz continuous, and exhibits step size ranges for which it is also a contraction. This is a famous result in optimisation [88, 176], traditionally stated in terms of the ℓ_2 canonical norm. Lemma B.2 in contrast assumes the $\boldsymbol{\Gamma}$-norm as underlying norm, since the latter is more natural for our particular problem.

Lemma B.2 — *Contractive Gradient Descent.* *Let $\boldsymbol{G} \in \mathbb{C}^{L\times N}$ be injective, and $\boldsymbol{\Gamma}$ be the diagonal and definite positive matrix defined in (10.10). Define*

$$\alpha := 2\lambda_{min}\left(\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right), \tag{B.15}$$

$$\beta := 2\lambda_{max}\left(\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right), \tag{B.16}$$

where $\lambda_{min}(\boldsymbol{M})$ and $\lambda_{max}(\boldsymbol{M})$ denote the minimum and maximum eigenvalue of a matrix $\boldsymbol{M}$ respectively. Let $\tau \in \mathbb{R}$ be a positive constant and consider the linear map

$$D_\tau : \begin{cases}\mathbb{C}^N \to \mathbb{C}^N, \\ \boldsymbol{x} \mapsto \boldsymbol{x} - 2\tau\boldsymbol{G}^H\left(\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}\right),\end{cases} \tag{B.17}$$

for some $\boldsymbol{y} \in \mathbb{C}^L$. Then, D_τ is Lipschitz continuous with respect to the norm induced by $\boldsymbol{\Gamma}$:

$$\|D_\tau(\boldsymbol{x}) - D_\tau(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} \leq L_\tau \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}}, \qquad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N,$$

with Lipschitz contant:

$$L_\tau = \max\{|1 - \tau\alpha|, |1 - \tau\beta|\}. \tag{B.18}$$

Moreover, D_τ is contractive, i.e. $0 < L_\tau < 1$, for $0 < \tau < 2/\beta$, and minimised for $\tau = 2/(\alpha + \beta)$.

Proof. We have

$$\begin{aligned}
\|D_\tau(\boldsymbol{x}) - D_\tau(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} &= \left\|(\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G})(\boldsymbol{x} - \boldsymbol{z})\right\|_{\boldsymbol{\Gamma}} \\
&\leq \left\|\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right\|_{\boldsymbol{\Gamma}} \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}}, \\
&= L_\tau \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}}
\end{aligned}$$

where the Lipschitz constant $L_\tau := \left\|\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right\|_{\boldsymbol{\Gamma}} > 0$ is the operator norm of $\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}$ induced by the $\boldsymbol{\Gamma}$-norm on $\mathbb{C}^N$:

$$\begin{aligned}
\left\|\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right\|_{\boldsymbol{\Gamma}} &= \sup_{\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}=1} \| \left(\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right)\boldsymbol{x}\|_{\boldsymbol{\Gamma}} \\
&= \sup_{\|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}=1} \left\|\boldsymbol{\Gamma}^{1/2}\left(\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right)\boldsymbol{x}\right\|_2 \\
&= \sup_{\|\tilde{\boldsymbol{x}}\|_2=1} \left\|\boldsymbol{\Gamma}^{1/2}\left(\boldsymbol{I}_N - 2\tau\boldsymbol{G}^H\boldsymbol{G}\right)\boldsymbol{\Gamma}^{-1/2}\tilde{\boldsymbol{x}}\right\|_2 \\
&= \left\|\boldsymbol{I}_N - 2\tau\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right\|_2.
\end{aligned} \tag{B.19}$$

Note that since $\boldsymbol{G}$ is injective, $\boldsymbol{G}^H\boldsymbol{G}$ is positive definite and hence we easily get [88] that the eigenvalues of $\boldsymbol{I}_N - 2\tau\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}$ are contained in the interval $[1 - \tau\beta, 1 - \tau\alpha]$, where $\beta \geq \alpha > 0$ are defined in (B.15) and (B.16) respectively. Its spectral norm is hence given by:

$$\left\|\boldsymbol{I}_N - 2\tau\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\Gamma}^{-1/2}\right\|_2 = \max\{|1 - \tau\alpha|, |1 - \tau\beta|\},$$

which proves (B.18). Finally, the restriction on τ for L_τ to be smaller than one follows from basic algebra, and is discussed in [176]. ∎

The second lemma states that in a Hilbert space, orthogonal projection maps onto closed convex sets are non-expansive. This is a known result of approximation theory [50, 118].

Lemma B.3 — *Non-Expansiveness of Closed Convex Projections.* *Let $\mathscr{H}$ be some Hilbert space with some inner-product norm $\|\cdot\|$ and $\mathcal{C} \subset \mathscr{H}$ a* closed,

convex *set. Then the orthogonal projection map onto $\mathcal{C}$, defined as*

$$\Pi_{\mathcal{C}}(x) = \arg\min_{z\in\mathcal{C}} \|x - z\|, \qquad \forall x \in \mathscr{H},$$

is non-expansive*, i.e.*

$$\|\Pi_{\mathcal{C}}(x) - \Pi_{\mathcal{C}}(z)\| \leq \|x - z\|, \quad \forall x, z \in \mathscr{H}.$$

Proof. Lemma B.3 is proven in [50, Theorem 5.5]. ■

The third lemma states that the singular value projection map $\Pi_{\mathcal{H}_k}$ is locally non-expansive in every neighbourhood of the manifold of matrices with rank exactly k.

Lemma B.4 — *Local Non-Expansiveness of the Singular Value Projection.* *Let $\mathbb{C}^{m\times n}$ be the space of complex-valued rectangular matrices of size $m \times n$, and $\mathcal{H}_k \subset \mathbb{C}^{m\times n}$, $\mathcal{R}_k \subset \mathbb{C}^{m\times n}$ the sets of matrices with rank at most and exactly $k \leq \max\{m, n\}$ respectively. Denote further by $\Pi_{\mathcal{H}_k}$ the orthogonal projection map onto $\mathcal{H}_k$ given in (10.21). Then, for every $\boldsymbol{R} \in \mathcal{R}_k$, the map $\Pi_{\mathcal{H}_k}$ is well-defined (single-valued) and* locally non-expansive

$$\|\Pi_{\mathcal{H}_k}(\boldsymbol{X}) - \Pi_{\mathcal{H}_k}(\boldsymbol{Z})\|_F \leq \|\boldsymbol{X} - \boldsymbol{Z}\|_F, \quad \forall \boldsymbol{X}, \boldsymbol{Y} \in \mathcal{U},$$

for some neighbourhood $\mathcal{U} \ni \boldsymbol{R}$.

Proof. Since $\mathcal{R}_k$ is dense in $\mathcal{H}_k$ [5, Proposition 2.1], we have $\Pi_{\mathcal{H}_k} = \Pi_{\mathcal{R}_k}$ in a neighbourhood $\mathcal{W}$ of every $\boldsymbol{R} \in \mathcal{R}_k$ (see [108, Example 2.3] for a detailed proof of this fact). Moreover, [100, Lemma 3] tells us that, for every $\boldsymbol{R} \in \mathcal{R}_k$, $\Pi_{\mathcal{R}_k}$ is, in a neighbourhood $\mathcal{U} \ni \boldsymbol{R}$ such that $\mathcal{U} \subset \mathcal{W}$, well-defined (single-valued), continuous and differentiable, with gradient given by: $\nabla\Pi_{\mathcal{R}_k} = \Pi_{\mathbb{T}_{\mathcal{R}_k}(\boldsymbol{R})}$ where $\mathbb{T}_{\mathcal{R}_k}(\boldsymbol{R}) \subset \mathbb{C}^{m\times n}$ is the tangent plane of the manifold $\mathcal{R}_k$ in $\boldsymbol{R}$ (see [108, Example 2.2]). Since $\mathbb{T}_{\mathcal{R}_k}(\boldsymbol{R})$ is by definition a linear subspace of $\mathbb{C}^{m\times n}$, the orthogonal projection operator $\Pi_{\mathbb{T}_{\mathcal{R}_k}(\boldsymbol{R})}$ is bounded with unit spectral norm. The map $\Pi_{\mathcal{R}_k} = \Pi_{\mathcal{H}_k}$ is consequently 1-Lipschitz continuous (i.e. non-expansive) with respect to the Frobenius norm in the neighbourhood $\mathcal{U}$ of $\boldsymbol{R} \in \mathcal{R}_k$. ■

The last lemma finally, makes use of Lemmas B.3 and B.4 to show that the denoising operator $H_n(\boldsymbol{x}) = T_P^{\dagger}[\Pi_{\mathbb{T}_P}\Pi_{\mathcal{H}_K}\Pi_{\mathbb{B}_\rho}]^n T_P(\boldsymbol{x})$ is locally non-expansive with respect to the $\boldsymbol{\Gamma}$-norm:

Lemma B.5 — *Local Non-Expansiveness of Denoiser.* *Let $\mathbb{C}^{(N-P)\times(P+1)}$ be the space of complex-valued rectangular matrices of size $(N - P) \times (P + 1)$, $P \leq \lfloor N/2 \rfloor$, and $\mathcal{H}_K \subset \mathbb{C}^{(N-P)\times(P+1)}$, $\mathcal{R}_K \subset \mathbb{C}^{(N-P)\times(P+1)}$ the sets of*

matrices with rank at most and exactly $K \leq P$ respectively. Let

$$H_n(\boldsymbol{x}) := T_P^\dagger [\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}]^n T_P(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{C}^N,$$

be the approximate proximal operator (10.40). Then, H_n is locally well-defined (single-valued) and non-expansive with respect to the $\boldsymbol{\Gamma}$-norm

$$\|H_n(\boldsymbol{x}) - H_n(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} \leq \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}},$$

for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$ such that $T_P(\boldsymbol{x})$, $T_P(\boldsymbol{z})$ are in some neighbourhood of some matrix $\boldsymbol{R} \in \mathcal{R}_K$.

Proof. First, we have, for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$:

$$\|H_n(\boldsymbol{x}) - H_n(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} = \left\| T_P^\dagger (D_n T_P(\boldsymbol{x}) - D_n T_P(\boldsymbol{z})) \right\|_{\boldsymbol{\Gamma}}, \tag{B.20}$$

where $D_n = [\Pi_{\mathbb{T}_P} \Pi_{\mathcal{H}_K} \Pi_{\mathbb{B}_\rho}]^n$. Notice that for $\boldsymbol{X} \in \mathbb{T}_P$, we have

$$\begin{aligned}
\left\| T_P^\dagger(\boldsymbol{X}) \right\|_{\boldsymbol{\Gamma}}^2 &= \langle \boldsymbol{\Gamma} T_P^\dagger(\boldsymbol{X}), T_P^\dagger(\boldsymbol{X}) \rangle_2 \\
&= \langle \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1} T_P^*(\boldsymbol{X}), T_P^\dagger(\boldsymbol{X}) \rangle_2 \\
&= \langle \boldsymbol{X}, T_P T_P^\dagger(\boldsymbol{X}) \rangle_F \\
&= \langle \boldsymbol{X}, \Pi_{\mathbb{T}_P} \boldsymbol{X} \rangle_F \\
&= \langle \boldsymbol{X}, \boldsymbol{X} \rangle_F \\
&= \|\boldsymbol{X}\|_F^2.
\end{aligned}$$

Since the range of D_n is $\mathbb{T}_P$, (B.20) becomes:

$$\|H_n(\boldsymbol{x}) - H_n(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} = \|D_n T_P(\boldsymbol{x}) - D_n T_P(\boldsymbol{z})\|_F .$$

Assuming now that $T_P(\boldsymbol{x})$ and $T_P(\boldsymbol{z})$ are in some neighbourhood of some point $\boldsymbol{R} \in \mathcal{R}_K$, we can invoke Lemmas B.3 and B.4 recursively to obtain:

$$\begin{aligned}
\|D_n T_P(\boldsymbol{x}) - D_n T_P(\boldsymbol{z})\|_F &\leq \|T_P(\boldsymbol{x}) - T_P(\boldsymbol{z})\|_F \\
&= \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}} ,
\end{aligned}$$

where we have used:

$$\|T_P(\boldsymbol{x})\|_F^2 = \langle T_P(\boldsymbol{x}), T_P(\boldsymbol{x}) \rangle_F = \langle T_P^* T_P(\boldsymbol{x}), \boldsymbol{x} \rangle_2 = \|\boldsymbol{x}\|_{\boldsymbol{\Gamma}}, \ \forall \boldsymbol{x} \in \mathbb{C}^N.$$

We finally get

$$\|H_n(\boldsymbol{x}) - H_n(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} \leq \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}} ,$$

for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$ such that $T_P(\boldsymbol{x}), T_P(\boldsymbol{z})$ are in some neighbourhood of some matrix $\boldsymbol{R} \in \mathcal{R}_K$. ∎

We are now ready to show Theorem 10.8. Let

$$U_{\tau,n}(\boldsymbol{x}) := H_n\left(\boldsymbol{x} - \tau \nabla F(\boldsymbol{x})\right) = H_n(D_\tau(\boldsymbol{x})), \qquad \boldsymbol{x} \in \mathbb{C}^N.$$

Then, for every $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$ such that $T_P(\boldsymbol{x}), T_P(\boldsymbol{z})$ are in some neighbourhood of some matrix $\boldsymbol{R} \in \mathcal{R}_K$, $U_{\tau,n}$ is locally Lipschitz continuous as composition between two (locally) Lipschitz continuous functions H_n and D_τ, see Lemmas B.5 and B.2 respectively. Moreover, the Lipschitz constant is the product of the Lipschitz constants of H_n and D_τ, 1 and L_τ in (B.18) respectively. We have therefore

$$\|U_{\tau,n}(\boldsymbol{x}) - U_{\tau,n}(\boldsymbol{z})\|_{\boldsymbol{\Gamma}} \leq L_\tau \|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\Gamma}},$$

for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$ such that $T_P(\boldsymbol{x}), T_P(\boldsymbol{z})$ are in some neighbourhood of some matrix $\boldsymbol{R} \in \mathcal{R}_K$. Finally, the restriction on τ for L_τ to be smaller than one results from Lemma B.2.

■

C

Supplementary Material to Chapter 11

1 Linear Algebra Tools

Chapter 11 makes heavy use of the *Kronecker product* and related operators. To ease the user's understanding, we provide a short description of these operators along with proofs of common transforms used throughout the text. Useful references for this section are [84, 116].

1.1 Conventions

In this chapter, we adopt the following conventions:

- Vectors are denoted with bold lowercase letters: $\boldsymbol{y}$.
- Matrices are denoted with bold uppercase letters: $\boldsymbol{A}$.
- If $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{a}_k \in \mathbb{C}^M$ denotes the k-th column of $\boldsymbol{A}$.
- The i-th entry of vector $\boldsymbol{y}$ is denoted $[\boldsymbol{y}]_i$.
- The (i,j)-th entry of matrix $\boldsymbol{A}$ is denoted $[\boldsymbol{A}]_{ij}$.
- The conjugation operation is denoted by overlining a vector or a matrix respectively: $\overline{\boldsymbol{a}}$, $\overline{\boldsymbol{A}}$.
- The modulus of a complex number $z \in \mathbb{C}$ is denoted by $|z|$.

1.2 Hadamard, Kronecker and Khatri-Rao products

The Hadamard product is the element-wise multiplication operator:

Definition C.1 — *Hadamard Product.* *Let $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ and $\boldsymbol{B} \in \mathbb{C}^{M \times N}$. The* Hadamard product $\boldsymbol{A} \odot \boldsymbol{B} \in \mathbb{C}^{M \times N}$ *is defined as*

$$[\boldsymbol{A} \odot \boldsymbol{B}]_{ij} = [\boldsymbol{A}]_{ij} [\boldsymbol{B}]_{ij} .$$

Moreover, we denote by $\boldsymbol{A}^{\odot 2}$ the Hadamard square of a matrix: $\boldsymbol{A} \odot \boldsymbol{A}$.

The Kronecker product generalises the vector outer product to matrices, and represents the tensor product between two finite-dimensional linear maps:

Definition C.2 — *Kronecker Product.* *Let $\boldsymbol{A} \in \mathbb{C}^{M_1 \times N_1}$ and $\boldsymbol{B} \in \mathbb{C}^{M_2 \times N_2}$.*

The Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{C}^{M_1 M_2 \times N_1 N_2}$ *is defined as*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} [\boldsymbol{A}]_{11} \boldsymbol{B} & \cdots & [\boldsymbol{A}]_{1N_1} \boldsymbol{B} \\ \vdots & \ddots & \vdots \\ [\boldsymbol{A}]_{M_1 1} \boldsymbol{B} & \cdots & [\boldsymbol{A}]_{M_1 N_1} \boldsymbol{B} \end{bmatrix}.$$

The main properties of the Kronecker product are [116]:

$$(\boldsymbol{A} \otimes \boldsymbol{B})^H = \boldsymbol{A}^H \otimes \boldsymbol{B}^H, \tag{C.1}$$
$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD}), \tag{C.2}$$
$$(\boldsymbol{A} \otimes \boldsymbol{B}) \odot (\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A} \odot \boldsymbol{C}) \otimes (\boldsymbol{B} \odot \boldsymbol{D}). \tag{C.3}$$

The Khatri-Rao product finally, is a column-wise Kronecker product:

Definition C.3 — *Khatri-Rao Product.* *Let* $\boldsymbol{A} \in \mathbb{C}^{M_1 \times N}$ *and* $\boldsymbol{B} \in \mathbb{C}^{M_2 \times N}$. *The* Khatri-Rao product $\boldsymbol{A} \circ \boldsymbol{B} \in \mathbb{C}^{M_1 M_2 \times N}$ *is defined as*

$$\boldsymbol{A} \circ \boldsymbol{B} = [\boldsymbol{a}_1 \otimes \boldsymbol{b}_1, \ldots, \boldsymbol{a}_N \otimes \boldsymbol{b}_N].$$

1.3 Matrix identities

In imaging problems, $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{A} \circ \boldsymbol{B}$ are often too large to be stored in memory. However it is not the matrix itself that is of interest in many circumstances, but rather the effect of a linear map such as $f(\boldsymbol{x}) = (\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{x}$. The matrix identities below allow us to evaluate $f(\boldsymbol{x})$ without ever having to compute large intermediate arrays. They make use of the vectorisation operator, defined hereafter:

Definition C.4 — *Vectorisation.* *Let* $\boldsymbol{A} \in \mathbb{C}^{M \times N}$. *The* vectorisation operator $\mathrm{vec}(\cdot)$ *reshapes a matrix into a vector by stacking its columns:*

$$[\mathrm{vec}(\boldsymbol{A})]_{M(j-1)+i} = [\boldsymbol{A}]_{ij}.$$

Conversely, the matricisation operator $\mathrm{mat}_{M,N}(\cdot)$ *reshapes a vector into a matrix:*

$$[\mathrm{mat}_{M,N}(\boldsymbol{a})]_{ij} = [\boldsymbol{a}]_{M(j-1)+i}.$$

Commonly used matrix identities are the following [84, 181]:

$$\mathrm{vec}(\boldsymbol{ABC}) = \left(\boldsymbol{C}^T \otimes \boldsymbol{A}\right) \mathrm{vec}(\boldsymbol{B}) \tag{C.4}$$
$$\mathrm{vec}(\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{b})\boldsymbol{C}) = \left(\boldsymbol{C}^T \circ \boldsymbol{A}\right) \boldsymbol{b} \tag{C.5}$$
$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F = \mathrm{tr}\left(\boldsymbol{A}^H \boldsymbol{B}\right) = \mathrm{vec}(\boldsymbol{A})^H \mathrm{vec}(\boldsymbol{B}) \tag{C.6}$$
$$\mathrm{vec}(\boldsymbol{b}\boldsymbol{a}^T) = \boldsymbol{a} \otimes \boldsymbol{b} \tag{C.7}$$

In this work, we furthermore make use of the following nonstandard matrix identities, proven hereafter:

$$(\boldsymbol{A} \circ \boldsymbol{B})^H \operatorname{vec}(\boldsymbol{C}) = \operatorname{diag}\left(\boldsymbol{B}^H \boldsymbol{C} \overline{\boldsymbol{A}}\right) \tag{C.8}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^H (\boldsymbol{A} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{C}) = \operatorname{vec}(\boldsymbol{B}^H \boldsymbol{B} \boldsymbol{C} \boldsymbol{A}^T \overline{\boldsymbol{A}}) \tag{C.9}$$

$$(\boldsymbol{A} \circ \boldsymbol{B})^H (\boldsymbol{A} \circ \boldsymbol{B}) \boldsymbol{c} = \operatorname{diag}(\boldsymbol{B}^H \boldsymbol{B} \operatorname{diag}(\boldsymbol{c}) \boldsymbol{A}^T \overline{\boldsymbol{A}}) \tag{C.10}$$

$$(\boldsymbol{A} \circ \boldsymbol{B})^H (\boldsymbol{A} \circ \boldsymbol{B}) = \boldsymbol{A}^H \boldsymbol{A} \odot \boldsymbol{B}^H \boldsymbol{B}. \tag{C.11}$$

Proof. (C.8)

$$\begin{aligned}\left[(\boldsymbol{A} \circ \boldsymbol{B})^H \operatorname{vec}(\boldsymbol{C})\right]_i &= \langle [\boldsymbol{A} \circ \boldsymbol{B}]_i, \operatorname{vec}(\boldsymbol{C}) \rangle = (\boldsymbol{a}_i \otimes \boldsymbol{b}_i)^H \operatorname{vec}(\boldsymbol{C}) \\ &\overset{(C.7)}{=} \operatorname{vec}(\boldsymbol{b}_i \boldsymbol{a}_i^T)^H \operatorname{vec}(\boldsymbol{C}) \overset{(C.6)}{=} \operatorname{tr}\left(\overline{\boldsymbol{a}}_i \boldsymbol{b}_i^H \boldsymbol{C}\right) \\ &= \operatorname{tr}\left(\boldsymbol{b}_i^H \boldsymbol{C} \overline{\boldsymbol{a}}_i\right) = \left[\boldsymbol{B}^H \boldsymbol{C} \overline{\boldsymbol{A}}\right]_{ii} = \left[\operatorname{diag}\left(\boldsymbol{B}^H \boldsymbol{C} \overline{\boldsymbol{A}}\right)\right]_i\end{aligned}$$

■

Proof. (C.9)

$$\begin{aligned}(\boldsymbol{A} \otimes \boldsymbol{B})^H (\boldsymbol{A} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{C}) &\overset{(C.1)}{=} \left(\boldsymbol{A}^H \otimes \boldsymbol{B}^H\right) (\boldsymbol{A} \otimes \boldsymbol{B}) \operatorname{vec}(\boldsymbol{C}) \\ &\overset{(C.3)}{=} \left[\left(\boldsymbol{A}^H \boldsymbol{A}\right) \otimes \left(\boldsymbol{B}^H \boldsymbol{B}\right)\right] \operatorname{vec}(\boldsymbol{C}) \\ &\overset{(C.4)}{=} \operatorname{vec}(\boldsymbol{B}^H \boldsymbol{B} \boldsymbol{C} \boldsymbol{A}^T \overline{\boldsymbol{A}})\end{aligned}$$

■

Proof. (C.10)

$$\begin{aligned}(\boldsymbol{A} \circ \boldsymbol{B})^H (\boldsymbol{A} \circ \boldsymbol{B}) \boldsymbol{c} &\overset{(C.5)}{=} (\boldsymbol{A} \circ \boldsymbol{B})^H \operatorname{vec}\left(\boldsymbol{B} \operatorname{diag}(\boldsymbol{c}) \boldsymbol{A}^T\right) \\ &\overset{(C.8)}{=} \operatorname{diag}\left(\boldsymbol{B}^H \boldsymbol{B} \operatorname{diag}(\boldsymbol{c}) \boldsymbol{A}^T \overline{\boldsymbol{A}}\right)\end{aligned}$$

■

Proof. (C.11)

$$\begin{aligned}\left[(\boldsymbol{A} \circ \boldsymbol{B})^H (\boldsymbol{A} \circ \boldsymbol{B})\right]_{ij} &= \langle \boldsymbol{a}_i \otimes \boldsymbol{b}_i, \boldsymbol{a}_j \otimes \boldsymbol{b}_j \rangle \overset{(C.7)}{=} \langle \operatorname{vec}(\boldsymbol{b}_i \boldsymbol{a}_i^T), \operatorname{vec}(\boldsymbol{b}_j \boldsymbol{a}_j^T) \rangle \\ &\overset{(C.6)}{=} \operatorname{tr}\left(\overline{\boldsymbol{a}_i} \boldsymbol{b}_i^H \boldsymbol{b}_j \boldsymbol{a}_j^T\right) = \operatorname{tr}\left(\boldsymbol{b}_i^H \boldsymbol{b}_j \boldsymbol{a}_j^T \overline{\boldsymbol{a}_i}\right) \\ &= \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle\end{aligned}$$

When put in matrix form, the above yields

$$(\boldsymbol{A} \circ \boldsymbol{B})^H (\boldsymbol{A} \circ \boldsymbol{B}) = \boldsymbol{A}^H \boldsymbol{A} \odot \boldsymbol{B}^H \boldsymbol{B}$$

■

2 Derivation: PGD for elastic-net problem 11.3

This section shows how to obtain proximal iteration (11.4) from (11.3).

Recall that the sound intensity map is obtained by solving the convex optimisation problem:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}_+^N}{\arg\min}\, f(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{C.12}$$

$$f(\boldsymbol{x}) = \frac{1}{2}\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{x})\boldsymbol{A}^H\right\|_F^2 \overset{(C.5)}{=} \frac{1}{2}\left\|\mathrm{vec}(\hat{\boldsymbol{\Sigma}}) - \left(\overline{\boldsymbol{A}}\circ\boldsymbol{A}\right)\boldsymbol{x}\right\|_2^2, \tag{C.13}$$

$$g(\boldsymbol{x}) = \lambda\left[\gamma\left\|\boldsymbol{x}\right\|_1 + (1-\gamma)\left\|\boldsymbol{x}\right\|_2^2\right], \tag{C.14}$$

where g is an elastic-net regularizer with $\lambda \geq 0$ and $\gamma \in]0,1[$.

PGD is a fixed-point method to solve problems of the form (C.12) where f, g are closed proper convex with f differentiable. It consists of iterating the proximal update equation until convergence:

$$\boldsymbol{x}^k = \mathbf{prox}_{\alpha g}\left(\boldsymbol{x}^{k-1} - \alpha\boldsymbol{\nabla} f(\boldsymbol{x}^{k-1})\right), \tag{C.15}$$

where $\alpha > 0$ is the step size and $\mathbf{prox}_{\alpha g}$ is the proximal operator associated with (C.14), given by (see proof below):

$$\mathbf{prox}_{\alpha g}(\boldsymbol{x}) = \underset{\boldsymbol{u}\in\mathbb{R}_+^N}{\arg\min}\, g(\boldsymbol{u}) + \frac{1}{2\alpha}\left\|\boldsymbol{u}-\boldsymbol{x}\right\|_2^2, \tag{C.16}$$

$$= \mathbf{ReLu}\left(\frac{\boldsymbol{x} - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1}\right), \qquad \forall \boldsymbol{x}\in\mathbb{R}^N. \tag{C.17}$$

The quantity $\boldsymbol{\nabla} f \in \mathbb{R}^N$ finally is obtained using the rules of vector calculus [140]:

$$\begin{aligned}\boldsymbol{\nabla} f(\boldsymbol{x}) &= \left\{\frac{\partial}{\partial \boldsymbol{x}}\left[\mathrm{vec}(\hat{\boldsymbol{\Sigma}}) - \left(\overline{\boldsymbol{A}}\circ\boldsymbol{A}\right)\boldsymbol{x}\right]\right\}\cdot\left[\mathrm{vec}(\hat{\boldsymbol{\Sigma}}) - \left(\overline{\boldsymbol{A}}\circ\boldsymbol{A}\right)\boldsymbol{x}\right] \\ &= \left(\overline{\boldsymbol{A}}\circ\boldsymbol{A}\right)^H\left[\left(\overline{\boldsymbol{A}}\circ\boldsymbol{A}\right)\boldsymbol{x} - \mathrm{vec}(\hat{\boldsymbol{\Sigma}})\right].\end{aligned} \tag{C.18}$$

Combining (C.15), (C.17) and (C.18) leads to (11.4).

Proof: (Analytic expression for $\mathbf{prox}_{\alpha g}$). Replacing (C.14) in (C.16), we get for $\boldsymbol{x}\in\mathbb{R}^N$:

$$\begin{aligned}\mathbf{prox}_{\alpha g}(\boldsymbol{x}) &= \underset{\boldsymbol{u}\in\mathbb{R}_+^N}{\arg\min}\,\lambda\left[\gamma\left\|\boldsymbol{u}\right\|_1 + (1-\gamma)\left\|\boldsymbol{u}\right\|_2^2\right] + \frac{1}{2\alpha}\left\|\boldsymbol{u}-\boldsymbol{x}\right\|_2^2 \\ &= \underset{(u_1,\dots,u_N)\in\mathbb{R}_+^N}{\arg\min}\sum_{n=1}^N \lambda\left[\gamma|u_n| + (1-\gamma)u_n^2\right] + \frac{1}{2\alpha}(u_n - x_n)^2 \\ &= \underset{(u_1,\dots,u_N)\in\mathbb{R}_+^N}{\arg\min}\sum_{n=1}^N \lambda\left[\gamma u_n + (1-\gamma)u_n^2\right] + \frac{1}{2\alpha}\left[u_n^2 + x_n^2 - 2u_n x_n\right]\end{aligned}$$

$$= \underset{(u_1,\dots,u_N)\in\mathbb{R}_+^N}{\arg\min} \sum_{n=1}^{N} \varphi_n(u_n). \tag{C.19}$$

Notice that (C.19) is the sum of N independent objective functionals, hence each can be independently minimised. (We drop the subscript of φ_n below for simplicity.) Let $\hat{u}$ be the minimiser:[1]

[1] Which exists since the optimisation problem is convex.

$$\hat{u} = \underset{u\geq 0}{\arg\min}\, \varphi(u) = \underset{u\geq 0}{\arg\min}\, \lambda\left[\gamma u + (1-\gamma)u^2\right] + \frac{1}{2\alpha}\left[u^2 + x^2 - 2ux\right], \tag{C.20}$$

for some fixed $x \in \mathbb{R}$. Then two cases can occur:

- $x \leq 0$: the objective functional being composed of positive terms only, any $\hat{u} > 0$ will increase the objective. Therefore $\hat{u} = 0$.
- $x > 0$: In this case the Karush Kuhn Tucker (KKT) conditions [27, 161] tell us that $\hat{u}$ is a minimizer of (C.20) if

$$\begin{aligned} \hat{u}\varphi'(\hat{u}) &= 0 \\ \varphi'(\hat{u}) &\geq 0 \quad \text{if } \hat{u} = 0. \end{aligned}$$

  Plugging $\varphi'(u) = \lambda\gamma + \left(2\lambda(1-\gamma) + \alpha^{-1}\right)u - \alpha^{-1}x$ and solving the above yields

$$\hat{u} = \begin{cases} \frac{x-\lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1} & x > \lambda\alpha\gamma, \\ 0 & x \leq \lambda\alpha\gamma \end{cases}.$$

Both cases can be written in short as

$$\hat{u} = \underset{u\geq 0}{\arg\min}\, \varphi(u) = \left[\frac{x - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1}\right]_+, \qquad \forall x \in \mathbb{R},$$

leading to an element-wise proximal operator of the form

$$\mathbf{prox}_{\alpha g}(\boldsymbol{x}) = \left[\frac{\boldsymbol{x} - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1}\right]_+ = \mathbf{ReLu}\left(\frac{\boldsymbol{x} - \lambda\alpha\gamma}{2\lambda\alpha(1-\gamma)+1}\right), \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

■

3 Network gradient evaluation

This section shows how to obtain derivatives of data-fidelity term $\mathcal{L}_t$ from eq. (11.14) w.r.t. network parameters $\boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{\tau}$.[2]

[2] For notational simplicity, this section drops the subscript in $\mathcal{L}_t$.

3.1 Problem statement

Recall that

$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{\Omega}) = \left\{ \frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}} \in \mathbb{R}^{K+1},\ \frac{\partial\mathcal{L}}{\partial\boldsymbol{B}} \in \mathbb{C}^{M\times N},\ \frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}} \in \mathbb{R}^N \right\},$$

$$\mathcal{L}(\boldsymbol{\Omega}) = \frac{1}{2}\frac{\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}^L(\boldsymbol{\Omega})\right\|_2^2}{\|\hat{\boldsymbol{x}}\|_2^2}, \tag{C.21}$$

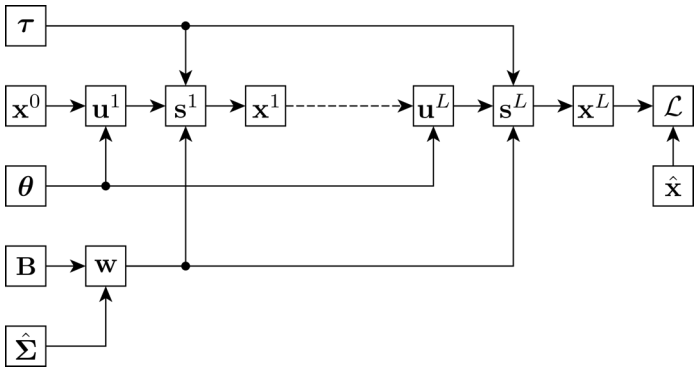


Figure C.1: L-layer computational graph of $\mathcal{L}$.

where $\boldsymbol{x}^L(\boldsymbol{\Omega}) \in \mathbb{R}_+^N$ is given by recurrence relation (11.1):

$$\boldsymbol{x}^l(\boldsymbol{\Omega}) = \sigma\left[P_{\boldsymbol{\theta}}(\boldsymbol{L})\boldsymbol{x}^{l-1} + \left(\overline{\boldsymbol{B}} \circ \boldsymbol{B}\right)^H \mathrm{vec}(\hat{\boldsymbol{\Sigma}}) - \boldsymbol{\tau}\right] \tag{C.22}$$

$$= \sigma\left[\boldsymbol{u}^l + \boldsymbol{w} - \boldsymbol{\tau}\right] \tag{C.23}$$

$$= \sigma\left[\boldsymbol{s}^l\right], \qquad l = 1, \ldots, L \tag{C.24}$$

with $\boldsymbol{x}^0 \in \mathbb{R}_+^N$ some arbitrary constant, $\sigma : \mathbb{R} \to \mathbb{R}$ a point-wise non-linearity, and $P_{\boldsymbol{\theta}}(\boldsymbol{L}) = \sum_{k=0}^K \theta_k T_k(\boldsymbol{L})$ a polynomial filter of order K expressed in terms of Chebychev polynomials.

$\boldsymbol{\nabla}\mathcal{L}$ can be efficiently evaluated using *reverse-mode algorithmic differentiation*[15] in a two-stage process:

- Forward pass: evaluate eq. (C.21) while storing all intermediate values $\boldsymbol{w}, \boldsymbol{\tau}, \left\{\boldsymbol{s}^l\right\}_{l=1,\ldots,L}$;
- Backward pass: walk the computational graph (Fig. C.1) backwards to evaluate derivatives w.r.t. $\boldsymbol{\theta}, \boldsymbol{B}, \boldsymbol{\tau}$.

3.2 Conventions

- If $\boldsymbol{u} \in \mathbb{R}^N$, $\boldsymbol{v} \in \mathbb{R}^M$, the *Jacobian matrix* $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{v}} \in \mathbb{R}^{N \times M}$ is defined as

$$\left[\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{v}}\right]_{ij} = \frac{\partial \left[\boldsymbol{u}\right]_i}{\partial \left[\boldsymbol{v}\right]_j}.$$

  Gradients of scalar-valued functions are therefore row vectors.
- If $\boldsymbol{u} \in \mathbb{R}^N$, $\boldsymbol{V} \in \mathbb{R}^{M \times Q}$, the *Jacobian tensor* $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{V}} \in \mathbb{R}^{N \times M \times Q}$ is defined as

$$\left[\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{V}}\right]_{ijk} = \frac{\partial \left[\boldsymbol{u}\right]_i}{\partial \left[\boldsymbol{V}\right]_{jk}}.$$

3.3 Common intermediate gradients

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}^L}\right]_i = \frac{\partial \mathcal{L}}{\partial \left[\boldsymbol{x}^L\right]_i} = \left[\frac{\boldsymbol{x}^L - \hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2^2}\right]_i \tag{C.25}$$

$$\left[\frac{\partial \boldsymbol{x}^l}{\partial \boldsymbol{s}^l}\right]_{ij} = \frac{\partial \left[\boldsymbol{x}^l\right]_i}{\partial \left[\boldsymbol{s}^l\right]_j} = \delta_{i-j}\sigma'\left(\left[\boldsymbol{s}^l\right]_j\right) = \left[\operatorname{diag}\left(\sigma'\left(\boldsymbol{s}^l\right)\right)\right]_{ij}, \qquad l = 1, \dots, L \tag{C.26}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}^l}\frac{\partial \boldsymbol{x}^l}{\partial \boldsymbol{s}^l} \overset{(C.26)}{=} \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}^l}\operatorname{diag}\left(\sigma'\left(\boldsymbol{s}^l\right)\right), \qquad l = 1, \dots, L \tag{C.27}$$

$$\left[\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{u}^l}\right]_{ij} = \frac{\partial \left[\boldsymbol{s}^l\right]_i}{\partial \left[\boldsymbol{u}^l\right]_j} = \frac{\partial}{\partial \left[\boldsymbol{u}^l\right]_j}\left[\boldsymbol{u}^l + \boldsymbol{w} - \boldsymbol{\tau}\right]_i = \delta_{i-j} = \left[\boldsymbol{I}_N\right]_{ij} \tag{C.28}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}^l} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{u}^l} \overset{(C.28)}{=} \frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}, \qquad l = 1, \dots, L \tag{C.29}$$

$$\left[\frac{\partial \boldsymbol{u}^l}{\partial \boldsymbol{x}^{l-1}}\right]_{ij} = \left[\frac{\partial}{\partial \boldsymbol{x}^{l-1}} P_{\boldsymbol{\theta}}(\boldsymbol{L})\boldsymbol{x}^{l-1}\right]_{ij} = \left[P_{\boldsymbol{\theta}}(\boldsymbol{L})\right]_{ij}, \qquad l = 1, \dots, L \tag{C.30}$$

$$\left[\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{w}}\right]_{ij} = \frac{\partial \left[\boldsymbol{s}^l\right]_i}{\partial \left[\boldsymbol{w}\right]_j} = \frac{\partial}{\partial \left[\boldsymbol{w}\right]_j}\left[\boldsymbol{u}^l + \boldsymbol{w} - \boldsymbol{\tau}\right]_i = \delta_{i-j} = \left[\boldsymbol{I}_N\right]_{ij} \tag{C.31}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{w}} \overset{(C.31)}{=} \sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l} \tag{C.32}$$

3.4 $\partial\mathcal{L}/\partial\theta$

$$\left[\frac{\partial \boldsymbol{u}^l}{\partial \boldsymbol{\theta}}\right]_{ij} = \frac{\partial}{\partial \left[\boldsymbol{\theta}\right]_j}\sum_{k=0}^{K}\left[\boldsymbol{\theta}\right]_k\left[T_k(\boldsymbol{L})\boldsymbol{x}^{l-1}\right]_i = \left[T_j(\boldsymbol{L})\boldsymbol{x}^{l-1}\right]_i, \qquad l = 1, \dots, L \tag{C.33}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\right]_i = \sum_{l=1}^{L}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}^l}\frac{\partial \boldsymbol{u}^l}{\partial \boldsymbol{\theta}}\right]_i \overset{(C.29)}{\underset{(C.33)}{=}} \sum_{i=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}T_i(\boldsymbol{L})\boldsymbol{x}^{l-1}, \qquad i = 0, \dots, K \tag{C.34}$$

3.5 $\partial\mathcal{L}/\partial\mathbf{B}$

$\frac{\partial\mathcal{L}}{\partial\boldsymbol{B}}$ can be obtained by evaluating $\frac{\partial\mathcal{L}}{\partial\boldsymbol{w}}\frac{\partial\boldsymbol{w}}{\partial\boldsymbol{B}}$, but $\frac{\partial\boldsymbol{w}}{\partial\boldsymbol{B}} \in \mathbb{C}^{N\times M\times N}$ is difficult to obtain directly. We therefore proceed in multiple steps:

1. Decompose $\boldsymbol{w}$ as $(\boldsymbol{w}_1+\boldsymbol{w}_2+\boldsymbol{w}_3)$ and express $\{\boldsymbol{w}_k\}_{k=1,2,3}$ explicitly in terms of $\hat{\boldsymbol{\Sigma}}_R, \hat{\boldsymbol{\Sigma}}_I, \boldsymbol{B}_R, \boldsymbol{B}_I$:

$$
\begin{aligned}
\boldsymbol{w} &= \left(\overline{\boldsymbol{B}} \circ \boldsymbol{B}\right)^H \operatorname{vec}(\hat{\boldsymbol{\Sigma}}) \overset{(C.8)}{=} \operatorname{diag}\left(\boldsymbol{B}^H \hat{\boldsymbol{\Sigma}} \boldsymbol{B}\right) \qquad (C.35)\\
&= \operatorname{diag}\left(\left[\boldsymbol{B}_R + j\boldsymbol{B}_I\right]^H \left[\hat{\boldsymbol{\Sigma}}_R + j\hat{\boldsymbol{\Sigma}}_I\right] \left[\boldsymbol{B}_R + j\boldsymbol{B}_I\right]\right)\\
&= \operatorname{diag}\left(\left[\boldsymbol{B}_R^T - j\boldsymbol{B}_I^T\right] \left[\hat{\boldsymbol{\Sigma}}_R + j\hat{\boldsymbol{\Sigma}}_I\right] \left[\boldsymbol{B}_R + j\boldsymbol{B}_I\right]\right)\\
&= \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R - \boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_I\right)\\
&+ j\operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_I + \boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I - \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R\right)\\
&\overset{\boldsymbol{w}\in\mathbb{R}_+^N}{=} \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right) - \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_I\right)\\
&= \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right) - \operatorname{diag}\left(\boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I^T\boldsymbol{B}_R\right)\\
&\overset{\hat{\boldsymbol{\Sigma}}_I=-\hat{\boldsymbol{\Sigma}}_I^T}{=} \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right) + \operatorname{diag}\left(\boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right)\\
&= \operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R + \boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + 2\boldsymbol{B}_I^T\hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right)\\
&\overset{(C.8)}{=} \underbrace{\left(\boldsymbol{B}_R \circ \boldsymbol{B}_R\right)^T \operatorname{vec}(\hat{\boldsymbol{\Sigma}}_R)}_{\boldsymbol{w}_1} + \underbrace{\left(\boldsymbol{B}_I \circ \boldsymbol{B}_I\right)^T \operatorname{vec}(\hat{\boldsymbol{\Sigma}}_R)}_{\boldsymbol{w}_2} + \underbrace{2\left(\boldsymbol{B}_R \circ \boldsymbol{B}_I\right)^T \operatorname{vec}(\hat{\boldsymbol{\Sigma}}_I)}_{\boldsymbol{w}_3}.
\end{aligned}
$$

2. Derive analytic forms for $\left\{\frac{\partial\boldsymbol{w}_k}{\partial\boldsymbol{B}_{R/I}}\right\}_{k=1,2,3}$:

$$
\begin{aligned}
\left[\frac{\partial\boldsymbol{w}_1}{\partial\boldsymbol{B}_R}\right]_{ijk} &= \frac{\partial\left[\boldsymbol{w}_1\right]_i}{\partial\left[\boldsymbol{B}_R\right]_{jk}} \overset{(C.8)}{=} \frac{\partial}{\partial\left[\boldsymbol{B}_R\right]_{jk}}\left[\operatorname{diag}\left(\boldsymbol{B}_R^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R\right)\right]_i \qquad (C.36)\\
&= \frac{\partial}{\partial\left[\boldsymbol{B}_R\right]_{jk}}(\boldsymbol{b}_i^R)^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{b}_i^R\\
&= \delta_{i-k}\frac{\partial}{\partial\left[\boldsymbol{B}_R\right]_{jk}}(\boldsymbol{b}_k^R)^T\hat{\boldsymbol{\Sigma}}_R\boldsymbol{b}_k^R\\
&= \delta_{i-k}\sum_{q=1}^{M}\sum_{g=1}^{M}\left[\hat{\boldsymbol{\Sigma}}_R\right]_{qg}\frac{\partial}{\partial\left[\boldsymbol{B}_R\right]_{jk}}\left\{\left[\boldsymbol{B}_R\right]_{qk}\left[\boldsymbol{B}_R\right]_{gk}\right\}\\
&= \delta_{i-k}\sum_{q=1}^{M}\left(\left[\hat{\boldsymbol{\Sigma}}_R\right]_{jq} + \left[\hat{\boldsymbol{\Sigma}}_R\right]_{qj}\right)\left[\boldsymbol{B}_R\right]_{qk}\\
&\overset{\hat{\boldsymbol{\Sigma}}_R=\hat{\boldsymbol{\Sigma}}_R^T}{=} 2\delta_{i-k}(\boldsymbol{b}_k^R)^T\boldsymbol{\sigma}_j^R
\end{aligned}
$$

$$\begin{aligned}\left[\frac{\partial \boldsymbol{w}_2}{\partial \boldsymbol{B}_I}\right]_{ijk} &= \frac{\partial \left[\boldsymbol{w}_2\right]_i}{\partial \left[\boldsymbol{B}_I\right]_{jk}} \overset{(C.8)}{=} \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} \left[\operatorname{diag}\left(\boldsymbol{B}_I^T \hat{\boldsymbol{\Sigma}}_R \boldsymbol{B}_I\right)\right]_i \\ &= \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} (\boldsymbol{b}_i^I)^T \hat{\boldsymbol{\Sigma}}_R \boldsymbol{b}_i^I \\ &= \delta_{i-k} \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} (\boldsymbol{b}_k^I)^T \hat{\boldsymbol{\Sigma}}_R \boldsymbol{b}_k^I \\ &= \delta_{i-k} \sum_{q=1}^{M} \sum_{g=1}^{M} \left[\hat{\boldsymbol{\Sigma}}_R\right]_{qg} \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} \left\{\left[\boldsymbol{B}_I\right]_{qk} \left[\boldsymbol{B}_I\right]_{gk}\right\} \\ &= \delta_{i-k} \sum_{q=1}^{M} \left(\left[\hat{\boldsymbol{\Sigma}}_R\right]_{jq} + \left[\hat{\boldsymbol{\Sigma}}_R\right]_{qj}\right) \left[\boldsymbol{B}_I\right]_{qk} \\ &\overset{\hat{\boldsymbol{\Sigma}}_R = \hat{\boldsymbol{\Sigma}}_R^T}{=} 2\delta_{i-k} (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^R \end{aligned} \tag{C.37}$$

$$\begin{aligned}\left[\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_R}\right]_{ijk} &= \frac{\partial \left[\boldsymbol{w}_3\right]_i}{\partial \left[\boldsymbol{B}_R\right]_{jk}} \overset{(C.8)}{=} 2 \frac{\partial}{\partial \left[\boldsymbol{B}_R\right]_{jk}} \left[\operatorname{diag}\left(\boldsymbol{B}_I^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{B}_R\right)\right]_i \\ &= 2 \frac{\partial}{\partial \left[\boldsymbol{B}_R\right]_{jk}} (\boldsymbol{b}_i^I)^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{b}_i^R \\ &= 2\delta_{i-k} \frac{\partial}{\partial \left[\boldsymbol{B}_R\right]_{jk}} (\boldsymbol{b}_k^I)^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{b}_k^R \\ &= 2\delta_{i-k} (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^I \end{aligned} \tag{C.38}$$

$$\begin{aligned}\left[\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_I}\right]_{ijk} &= \frac{\partial \left[\boldsymbol{w}_3\right]_i}{\partial \left[\boldsymbol{B}_I\right]_{jk}} \overset{(C.8)}{=} 2 \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} \left[\operatorname{diag}\left(\boldsymbol{B}_I^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{B}_R\right)\right]_i \\ &= 2 \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} (\boldsymbol{b}_i^I)^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{b}_i^R \\ &= 2\delta_{i-k} \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} (\boldsymbol{b}_k^I)^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{b}_k^R \\ &\overset{\hat{\boldsymbol{\Sigma}}_I = -\hat{\boldsymbol{\Sigma}}_I^T}{=} -2\delta_{i-k} \frac{\partial}{\partial \left[\boldsymbol{B}_I\right]_{jk}} (\boldsymbol{b}_k^R)^T \hat{\boldsymbol{\Sigma}}_I \boldsymbol{b}_k^I \\ &= -2\delta_{i-k} (\boldsymbol{b}_k^R)^T \boldsymbol{\sigma}_j^I \end{aligned} \tag{C.39}$$

3. Combine $\left\{\frac{\partial \boldsymbol{w}_k}{\partial \boldsymbol{B}_{R/I}}\right\}_{k=1,2,3}$ with $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}$ to obtain $\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}} \in \mathbb{C}^{M \times N}$:

$$\left[\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{w}_k}\right]_{ij} = \frac{\partial \left[\boldsymbol{w}\right]_i}{\partial \left[\boldsymbol{w}_k\right]_j} = \frac{\partial}{\partial \left[\boldsymbol{w}_k\right]_j} \left[\boldsymbol{w}_1 + \boldsymbol{w}_2 + \boldsymbol{w}_3\right]_i = \delta_{i-j} = \left[\boldsymbol{I}_N\right]_{ij}, \; k = 1,2,3 \tag{C.40}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_k} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{w}_k} \overset{\text{(C.32)}}{\underset{\text{(C.40)}}{=}} \sum_{l=1}^{L} \frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}, \qquad k = 1, 2, 3 \tag{C.41}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1} \frac{\partial \boldsymbol{w}_1}{\partial \boldsymbol{B}_R}\right]_{jk} = \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right]_i \left[\frac{\partial \boldsymbol{w}_1}{\partial \boldsymbol{B}_R}\right]_{ijk} \overset{\text{(C.36)}}{=} 2 \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right]_i \delta_{i-k} (\boldsymbol{b}_k^R)^T \boldsymbol{\sigma}_j^R \tag{C.42}$$

$$= 2 \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right]_k (\boldsymbol{b}_k^R)^T \boldsymbol{\sigma}_j^R = \left[2 \hat{\boldsymbol{\Sigma}}_R^T \boldsymbol{B}_R \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right)\right]_{jk}$$

$$\overset{\hat{\boldsymbol{\Sigma}}_R = \hat{\boldsymbol{\Sigma}}_R^T}{=} \left[2 \hat{\boldsymbol{\Sigma}}_R \boldsymbol{B}_R \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right)\right]_{jk}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2} \frac{\partial \boldsymbol{w}_2}{\partial \boldsymbol{B}_I}\right]_{jk} = \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right]_i \left[\frac{\partial \boldsymbol{w}_2}{\partial \boldsymbol{B}_I}\right]_{ijk} \overset{\text{(C.37)}}{=} 2 \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right]_i \delta_{i-k} (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^R \tag{C.43}$$

$$= 2 \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right]_k (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^R = \left[2 \hat{\boldsymbol{\Sigma}}_R^T \boldsymbol{B}_I \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right)\right]_{jk}$$

$$\overset{\hat{\boldsymbol{\Sigma}}_R = \hat{\boldsymbol{\Sigma}}_R^T}{=} \left[2 \hat{\boldsymbol{\Sigma}}_R \boldsymbol{B}_I \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right)\right]_{jk}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3} \frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_R}\right]_{jk} = \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_i \left[\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_R}\right]_{ijk} \overset{\text{(C.38)}}{=} 2 \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_i \delta_{i-k} (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^I \tag{C.44}$$

$$= 2 \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_k (\boldsymbol{b}_k^I)^T \boldsymbol{\sigma}_j^I = \left[2 \hat{\boldsymbol{\Sigma}}_I^T \boldsymbol{B}_I \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right]_{jk}$$

$$\overset{\hat{\boldsymbol{\Sigma}}_I = -\hat{\boldsymbol{\Sigma}}_I^T}{=} \left[-2 \hat{\boldsymbol{\Sigma}}_I \boldsymbol{B}_I \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right]_{jk}$$

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3} \frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_I}\right]_{jk} = \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_i \left[\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_I}\right]_{ijk} \overset{\text{(C.39)}}{=} -2 \sum_{i=1}^{N} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_i \delta_{i-k} (\boldsymbol{b}_k^R)^T \boldsymbol{\sigma}_j^I \tag{C.45}$$

$$= -2 \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right]_k (\boldsymbol{b}_k^R)^T \boldsymbol{\sigma}_j^I = \left[-2 \hat{\boldsymbol{\Sigma}}_I^T \boldsymbol{B}_R \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right]_{jk}$$

$$\overset{\hat{\boldsymbol{\Sigma}}_I = -\hat{\boldsymbol{\Sigma}}_I^T}{=} \left[2 \hat{\boldsymbol{\Sigma}}_I \boldsymbol{B}_R \operatorname{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right]_{jk}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}_R} &= \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\frac{\partial \boldsymbol{w}_1}{\partial \boldsymbol{B}_R} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_R} \\ &\overset{\substack{(C.42)\\(C.44)}}{=} 2\left\{\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R\,\mathrm{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_1}\right) - \hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_I\,\mathrm{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right\} \\ &\overset{(C.41)}{=} 2\left\{\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_R - \hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_I\right\}\mathrm{diag}\left(\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\right) \\ &= 2\Re\left\{\hat{\boldsymbol{\Sigma}}\boldsymbol{B}\right\}\mathrm{diag}\left(\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\right)\end{aligned} \tag{C.46}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}_I} &= \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\frac{\partial \boldsymbol{w}_2}{\partial \boldsymbol{B}_I} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\frac{\partial \boldsymbol{w}_3}{\partial \boldsymbol{B}_I} \\ &\overset{\substack{(C.43)\\(C.45)}}{=} 2\left\{\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I\,\mathrm{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_2}\right) + \hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\,\mathrm{diag}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_3}\right)\right\} \\ &\overset{(C.41)}{=} 2\left\{\hat{\boldsymbol{\Sigma}}_R\boldsymbol{B}_I + \hat{\boldsymbol{\Sigma}}_I\boldsymbol{B}_R\right\}\mathrm{diag}\left(\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\right) \\ &= 2\Im\left\{\hat{\boldsymbol{\Sigma}}\boldsymbol{B}\right\}\mathrm{diag}\left(\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\right)\end{aligned} \tag{C.47}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{B}_R} + j\frac{\partial \mathcal{L}}{\partial \boldsymbol{B}_I} = 2\hat{\boldsymbol{\Sigma}}\boldsymbol{B}\,\mathrm{diag}\left(\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\right) \tag{C.48}$$

3.6 $\partial\mathcal{L}/\partial\tau$

$$\left[\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{\tau}}\right]_{ij} = \frac{\partial\left[\boldsymbol{s}^l\right]_i}{\partial\left[\boldsymbol{\tau}\right]_j} = \frac{\partial}{\partial\left[\boldsymbol{\tau}\right]_j}\left[\boldsymbol{u}^l + \boldsymbol{w} - \boldsymbol{\tau}\right]_i = -\delta_{i-j} = \left[-\boldsymbol{I}_N\right]_{ij}, \qquad l = 1,\ldots,L \tag{C.49}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\tau}} = \sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l}\frac{\partial \boldsymbol{s}^l}{\partial \boldsymbol{\tau}} \overset{(C.49)}{=} -\sum_{l=1}^{L}\frac{\partial \mathcal{L}}{\partial \boldsymbol{s}^l} \tag{C.50}$$

Combining eqs. (C.34), (C.48) and (C.50) leads to Algorithms 11.13 and 11.14.

4 Real Data Experiments (Supplement)

Results in the main text present a summary of DeepWave's performance on a specific real-world dataset. The goal of this section is to provide a more elaborate description of the dataset, training process, and emphasise interesting observations.

4.1 Dataset Description

Two real-world datasets are considered:

Dataset 1 [131] consists of a series of 92 microphone recordings from the *Pyramic*[155] array (Fig. C.2) taken in an anechoic chamber to evaluate the performance of different direction-of-arrival algorithms [132]. Specifically, the dataset contains a series of 3 second recordings of human speech emitted by loudspeakers positioned around the edge of the chamber and located at the same height. Each recording has one, two, or three speakers active simultaneously. Recordings contain both male and female speech samples to cover a wide audible range.

Dataset 2 [150] consists of a larger collection of microphone recordings from the *Pyramic*[155] array (Fig. C.2) taken in an anechoic chamber. The goal of this dataset is to provide a generic dataset on which to evaluate the performance of array processing algorithms on real-life recordings with all the non-idealities involved. Specifically, the dataset contains 2700 recordings of human speech emitted from every direction of the anechoic chamber at a resolution of 2 degrees in azimuth and three different elevations ({-15, 0, 15} degrees). Recordings contain both male and female speech samples to cover a wide audible range. While the total number of recordings is significant, since each recording contains emissions from a single source, different audio samples can be combined to simulate complex multi-source sound fields. This data-augmentation task therefore allows us to assess the generalizability of DeepWave to such setups. Concretely, we construct a synthetic dataset of 5700 distinct microphone recordings with one, two, or three active speakers simultaneously.

4.2 Data Pre-Processing

The raw time-series are pre-processed to get a suitable training set for DeepWave as follows:

- Instantaneous empirical covariances $\left\{\hat{\boldsymbol{\Sigma}}_t\right\}_t$ are obtained for 9 equi-spaced frequency bands spanning $[1500, 4500]$ Hz every 100 ms using *Short-Time Fourier Transforms (STFT)* [95, 183].
- APGD ground truths $\{\hat{\boldsymbol{x}}_t\}_t$ were estimated by solving eq. (11.3) with $\gamma = 0.5$, step size $\alpha = 1/\left\|\overline{\boldsymbol{A}} \circ \boldsymbol{A}\right\|_2^2$, and $\lambda_t = \max([\boldsymbol{x}_t^1]_1, \ldots, [\boldsymbol{x}_t^1]_N)/(\alpha\gamma)$, where $\boldsymbol{x}_t^1 \in \mathbb{R}^N$ is the APGD estimate obtained after one iteration of eq. (11.5).

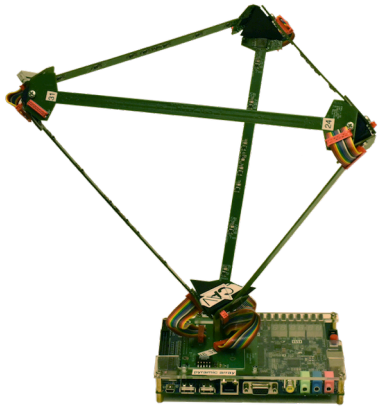


Figure C.2: Pyramic 48-element microphone array [155] used to acquire real-world dataset [131]. Eight microphones are mounted on six PCBs that form the edges of a tetrahedron.

After pre-processing, we obtain 2760 training samples $\mathcal{T} = \left\{\left(\hat{\boldsymbol{\Sigma}}_t, \hat{\boldsymbol{x}}_t\right)\right\}_t$ per frequency band.

4.3 Network Training

DeepWave is trained by solving (11.14) using stochastic gradient descent (SGD) with momentum acceleration [174]. The optimisation problem is initialised as given in eq. (11.15). Training is done on an 80% random subset of $\mathcal{T}$ using mini-batches of size $N_{batch} = 100$, with the remaining 20% serving as a validation set. The learning rate was set to 10^{-8}.

Regularisation parameters were chosen based on a grid search with optimal values $\lambda_{\boldsymbol{\theta}} = \lambda_{\boldsymbol{B}} = \lambda_{\boldsymbol{\tau}} = 0.1$. It was noticed during our experiments that regularising $\boldsymbol{\theta}$ and $\boldsymbol{B}$ provides little benefit to generalisation error and hence can be omitted. Regularisation of $\boldsymbol{\tau}$ is important however to ensure convergence to smooth biases. This is particularly relevant for rich acoustic fields where sources have no spatial constraints (see Section 5).

Training and validation losses converged in less than 10 epochs for the optimal parameterisation, i.e. when $L = 5$ and K ranges from 10 to 23 depending on the frequency band. Total training time for **Dataset 1** was 10 minutes per band on an i7-8550U CPU with 32GB memory. Due to disk space constraints, **Dataset 2** was trained on a dual-socket Intel E5-2680v3 with 256GB memory. Total training time for **Dataset 2** was roughly 3 hours per band.

4.4 Experimental Results

In this section, we provide the supporting plots for the claims made in Section 4 of the main paper:

- Fig. C.3 shows DeepWave's learnt bias parameter on **Dataset 1**. Unlike APGD, the latter is highly nonuniform in space, and slightly stronger in magnitude.
- Fig. C.4 shows the impulse response of DAS and DeepWave trained on Dataset 1 at 3.5 kHz, obtained by simulating the data from a single point-

source in the field. Such plots were used to compute resolution scores of all algorithms across frequency bands.

- Fig. C.5 shows example spherical fields obtained with DeepWave, DAS, APGD and APGD prematurely terminated applied to recordings in the validation set of **Dataset 1**. Resolution and contrast comparisons are moreover carried out. The true colour images displayed in Fig. C.5 were obtained by mapping frequency channels into a colour spectrum (see the color-frequency mapping in Fig. 11.4d).
- A video showing the evolution in time of DeepWave and DAS azimuthal sound fields (as in Figs. 11.4a and 11.4b) is also available online: https://www.youtube.com/watch?v=PwB3CS2rHdI.

5 Further Experiments in Simulation

Results in Chapter 11 present a summary of DeepWave's performance on two real-world datasets. Though the datasets represent real-world scenarios, the downside is that sound emissions are assumed to come from few directions in space. It is therefore challenging to test DeepWave's generalisability on this dataset alone. The goal of this section is to investigate how well DeepWave generalises to richer datasets through simulation.

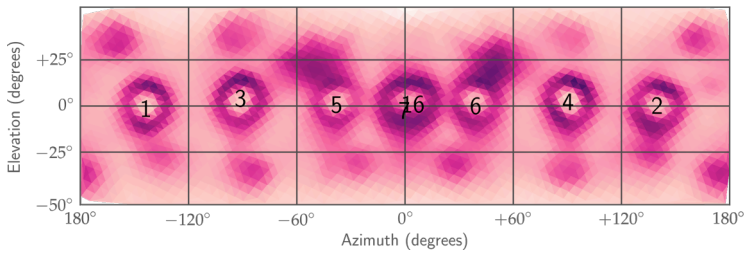


Figure C.3: Bias parameter τ learnt by SGD run on the dataset described in Section 4.1. We observe that the biasing is more prominent at sidelobes and around actual sources. This results in an increased angular resolution with fewer artefacts.

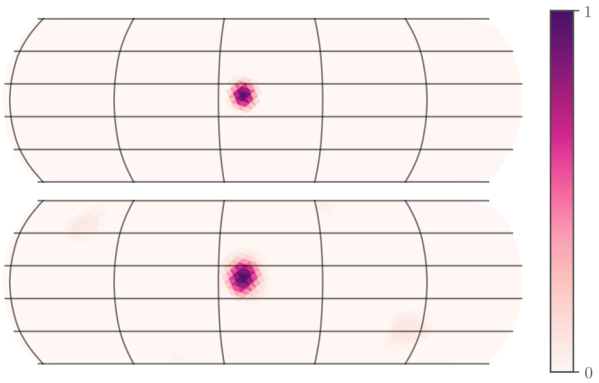
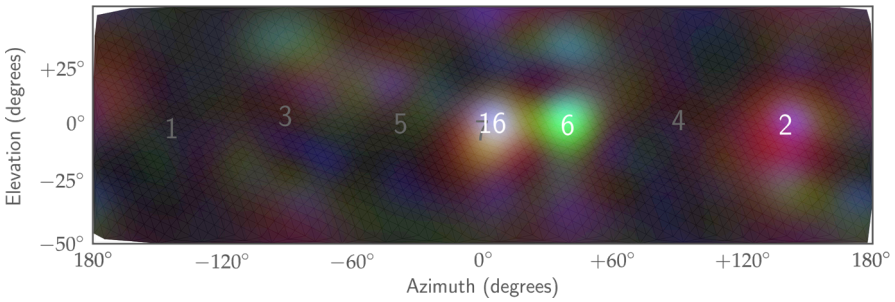
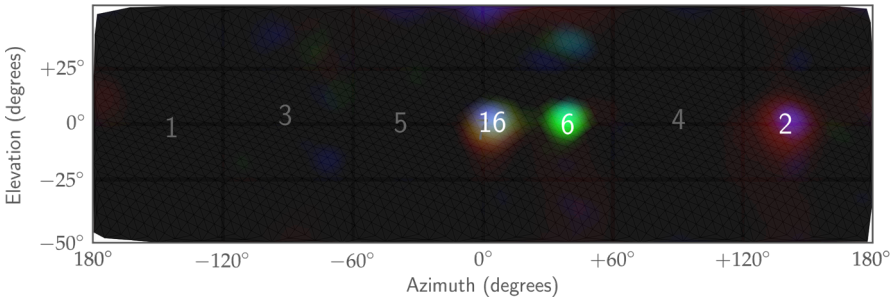


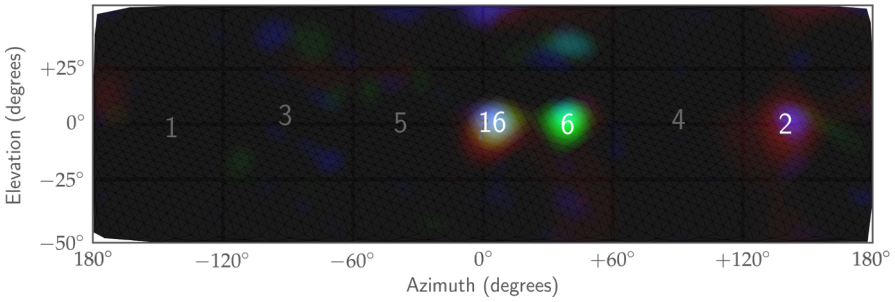
Figure C.4: Impulse response of DeepWave (top) vs. DAS (bottom). We notice a shrinkage of the main lobe, resulting in increased angular resolution.



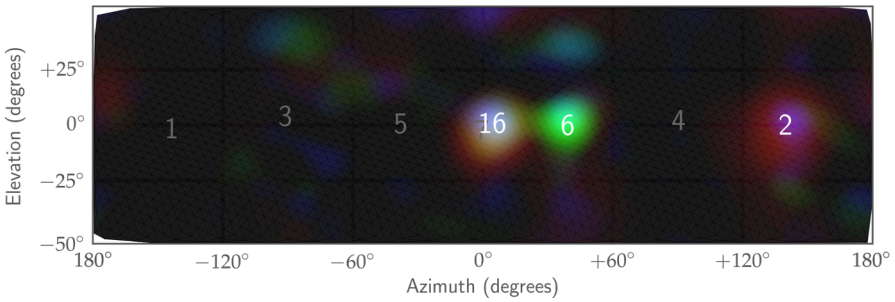
(a) DAS spherical sound field (resolution: 25.3°, contrast: 0.78).



(b) DeepWave spherical sound field (resolution: 18.5°, contrast: 0.97).



(c) APGD spherical sound field (resolution: 13°, contrast: 0.97).



(d) APGD (terminated) spherical sound field (resolution: 21.4°, contrast: 0.94).

Figure C.5: Intensity field reconstruction comparison between DAS, DeepWave ($L = 5$), APGD (converged, $N_{iter} = 17$), APGD (premature termination, $N_{iter} = 5$). In terms of resolution and contrast, DeepWave and APGD have similar performance, outperforming DAS by approximately 27% resolution-wise and 20% contrast-wise across frequency bands. When limited to a number of iterations equal to the depth L of DeepWave, APGD performances degrade considerably.

5.1 Dataset description

The simulated dataset is designed to mimic a key application of acoustic cameras: accurate mapping of the sound field in an open-air setting from a given direction. To this end, the setup is modelled as follows:

- The scene is assumed to be a 120° spherical viewport in which sources are uniformly distributed.
- Source emissions follow a narrow-band point-source model at 2 kHz [95, 181], where their intensities are either uniform or Rayleigh-distributed with rate parameter $r = 1$. All images below show equi-amplitude visualisations only as they are easier to assess through visual inspection.
- Emissions from the scene are captured by a 64-element spherical microphone array of radius $r = 20$ cm.
- Empirical covariances matrices $\hat{\mathbf{\Sigma}} \in \mathbb{C}^{64\times 64}$ are synthesised using the traditional far-field measurement equation [181, eq.(12)] for point sources.
- APGD ground truths are obtained as described in Section 4.2.

The final dataset consists of 20,000 images that contain up to 10 sources in the field. Training the network is identical to Section 4.3, except for the batch-size which increases to 200 and the learning rate that is set to 10^{-7}. In particular training converges in less than 10 epochs, with a total runtime proportionally larger than in Section 4.3 due to the increased dataset size. The

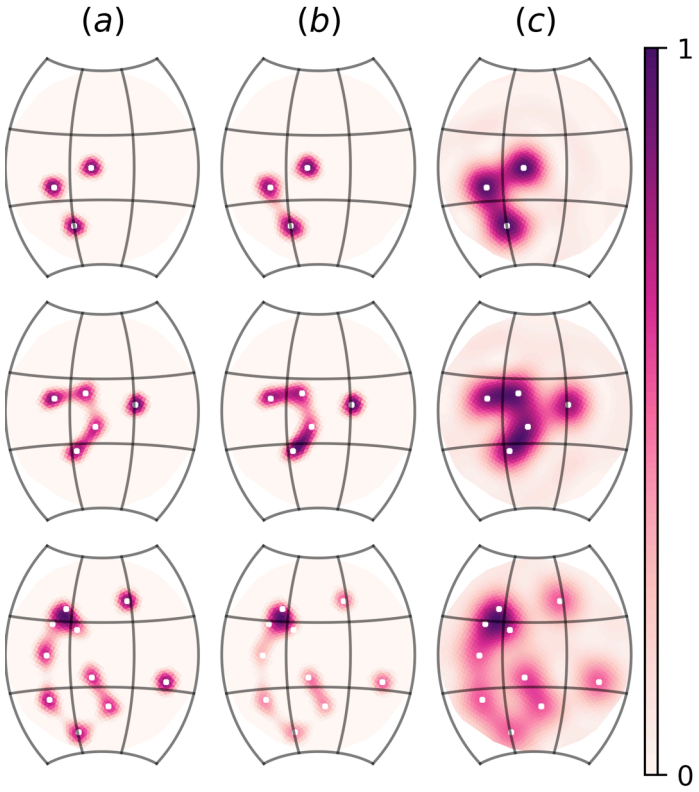


Figure C.6: Intensity field reconstruction comparison between (a) APGD ($N_{\text{iter}} = 48$), (b) DeepWave ($L = 5$), and (c) DAS. The image quality results corroborate with the observations made in Fig. C.5.

optimal parameterisation of the network is achieved with $L = 6$ and $K = 18$.

5.2 Experimental results

- Fig. C.6 shows example spherical fields obtained with DeepWave, DAS and APGD applied to recordings in the validation set of DeepWave.
- Fig. C.7 investigates the influence of DeepWave's depth on the validation loss. Profiles show that 5 or 6 layers are sufficient for the investigated dataset.
- Table C.1 investigates the runtime of APGD, DAS and DeepWave for different depths. DAS and DeepWave execute several orders of magnitude faster than APGD, regardless of network depths. Similar conclusions apply to the real-data setup investigated in Section 4.

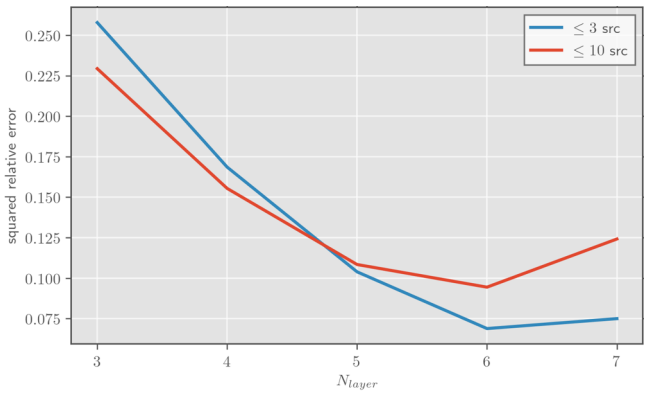


Figure C.7: Influence of network depth on validation loss. The plot shows the relative squared-error on the validation set between APGD ground truth $\hat{\boldsymbol{x}}$ and DeepWave output $\boldsymbol{x}^L$ as a function of network depth L using simulated data. The red curve corresponds to the full unconstrained dataset with up to 10 sources present in the field. The blue curve is obtained by retraining the network on a subset of the dataset where only up to 3 sources are present. Precision loss for small L comes from insufficient sparsification of network output w.r.t. ground truth. On the other hand error increase for L large are due to amplitude mismatches between ground truth and network output. This is presumably caused by the use of the rectified $\tanh$ activation function to avoid gradient explosion during training.

Method	N_{iter}/L	Execution time [s]
APGD (converged)	48	7.26
DeepWave	6	0.0072
DeepWave	5	0.0063
DeepWave	3	0.0059
DeepWave	1	0.0031
DAS		0.0024

Table C.1: Runtime comparison of imaging methods on simulated dataset. Execution times averaged over 50 runs for a specific $\hat{\boldsymbol{\Sigma}} \in \mathbb{C}^{64\times 64}$. DeepWave inference time is comparable to Delay-and-Sum and is adequate to obtain a fluid framerate on an acoustic camera. Runtimes in DeepWave weakly depends on network depth L due to strong sparsity of the deblurring operator $\boldsymbol{\mathcal{D}}$: the main contributor to the former is evaluation of the backprojection term $\boldsymbol{\mathcal{B}}\,\mathrm{vec}(\hat{\boldsymbol{\Sigma}})$. In stark contrast to DeepWave, APGD requires three orders of magnitude more time to reach similar accuracy.

$\boldsymbol{\theta}$	$\boldsymbol{B}$	$\boldsymbol{\tau}$	$\mathcal{L}_{\text{test}}$	rel. improv. [%]	rel. improv. [%]
✗	✗	✗	0.160417		
✗	✗	✓	0.054927	65.76 (✗✗✗)	
✗	✓	✗	0.159698	0.45 (✗✗✗)	
✓	✗	✗	0.159948	0.29 (✗✗✗)	
✓	✗	✓	0.054910	65.77 (✗✗✗)	0.03 (✗✗✓)
✓	✓	✗	0.159234	0.74 (✗✗✗)	0.29 (✗✓✗)
✗	✓	✓	0.054917	65.77 (✗✗✗)	0.02 (✗✗✓)
✓	✓	✓	0.054900	65.78 (✗✗✗)	0.03 (✗✓✓)

Table C.2: DeepWave performance comparison on simulated dataset described in Appendix 5.1 as a function of parameter degrees of freedom. A ✗ in the first three columns means that the associated parameter was frozen during training. In contrast a ✓ means that the parameter is optimized during training. $\mathcal{L}_{\text{test}}$ represents the data-fidelity loss term of eq. (11.14) evaluated over the test set. Finally, the last two columns show the relative improvement of $\mathcal{L}_{\text{test}}$ w.r.t. the baseline parameterisation given in parentheses. The results show that learning the shrinkage operator $\boldsymbol{\tau}$ has the strongest net effect on improving predictive performance, while *for this setup* the deblurring $P_{\boldsymbol{\theta}}(\boldsymbol{L})$ and backprojection operators $\boldsymbol{B}$ provide marginal gains.

6 Ablation study

Results in the main text and above present a summary of DeepWave's performance after optimal tuning of network parameters $\boldsymbol{\theta}$, $\boldsymbol{B}$, $\boldsymbol{\tau}$ during training. Given the physical interpretation of these parameters as deblurring, backprojection and shrinkage operators respectively, we carry out an ablation study to investigate the relative importance of each parameter on DeepWave's ability to reconstruct ground truth APGD images.

Concretely, eight instances of DeepWave with $L = 6$ are trained on the simulated dataset described in Appendix 5.1. Each instance corresponds to a particular combination of free/frozen parameters such that all possible parameter triplets are taken into consideration. Frozen parameters remain at the initialisation point (11.15) of SGD. Network performance is assessed by computing the data-fidelity term $\frac{1}{T}\sum_{t=1}^{T}\mathcal{L}_t$ from (11.14) over the test set. The results are shown in Table C.2.

As expected, freezing all three parameters (✗✗✗) produces the worst reconstructions as the network fails to converge to the ground truth after so few iterations. At the other end of the spectrum, learning all parameters (✓✓✓) leads to the best predictive performance, with a relative improvement of 65.78% over not learning anything. However the contributions of each parameter vary significantly: Learning $\boldsymbol{\tau}$ (✗✗✓) has the strongest net effect (65.76%), whereas learning $\boldsymbol{\theta}$ (✓✗✗), $\boldsymbol{B}$ (✗✓✗) provide minimal gains over no learning (✗✗✗). The second half of Table C.2 shows similar observations hold when training parameter pairs, where learning any parameter in addition to $\boldsymbol{\tau}$ only provides small marginal gains over just learning the latter (✗✗✓). The reason for the marginal gains obtained when learning $\boldsymbol{\theta}$ and $\boldsymbol{B}$ is that the deblurring and backprojection operators are, for the specific experimental conditions investigated (point sources, non-reverberant environments (i.e. anechoic chambre), near-spherical geometries), very well modelled by initialisation scheme (11.15). However, for environments containing reverberation and non-spherical array geometries, the observations above may differ significantly. In these contexts, learning $\boldsymbol{\theta}$ and $\boldsymbol{B}$ may lead to better predictive performance.

Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55, Courier Corporation, 1965. (Cited on page 137)

[2] P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, M. Baker, A. Balbi, A. Banday, R. Barreiro, et al., *Planck early results. i. the planck mission*, Astronomy & Astrophysics, 536 (2011), p. A1. (Cited on pages 11, 157, and 162)

[3] J. Allen, *Sea surface temperature anomaly (1 month - aqua/amsr-e, 2002-11)*. https://neo.sci.gsfc.nasa.gov/view.php?datasetId=AMSRE_SSTAn_M&year=2011. Accessed: 2019-06-26. (Cited on pages 146 and 147)

[4] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Transactions on Signal Processing, 45 (1997), pp. 913–917. (Cited on page 130)

[5] F. Andersson and M. Carlsson, *Alternating projections on low-dimensional manifolds*. (Cited on pages xxvi, 182, 191, and 239)

[6] F. Andersson, M. Carlsson, and P.-A. Ivert, *A fast alternating projection method for complex frequency estimation*, arXiv preprint arXiv:1107.2028, (2011). (Cited on page 191)

[7] Argo, *Argo float data and metadata from global data assembly centre (argo gdac)*, SEANOE, (2000). (Cited on pages 11, 145, and 147)

[8] H. Attouch and J. Peypouquet, *The rate of convergence of nesterov's accelerated forward-backward method is actually faster than 1/k^2*, SIAM Journal on Optimization, 26 (2016), pp. 1824–1834. (Cited on page 119)

[9] S. Aziznejad and M. Unser, *An l1 representer theorem for multiple-kernel regression*, arXiv preprint arXiv:1811.00836, (2018). (Cited on page 223)

[10] A. Badoual, J. Fageot, and M. Unser, *Periodic splines and gaussian processes for the resolution of linear inverse problems*, IEEE Transactions on Signal Processing, (2018). (Cited on pages 14, 18, and 55)

[11] C. B. Barber, D. P. Dobkin, D. P. Dobkin, and H. Huhdanpaa, *The quickhull algorithm for convex hulls*, ACM Transactions on Mathematical Software (TOMS), 22 (1996), pp. 469–483. (Cited on page 105)

[12] H. Bauschke, F. Deutsch, H. Hundal, and S.-H. Park, *Accelerating the convergence of the method of alternating projections*, Transactions of the American Mathematical Society, 355 (2003), pp. 3433–3461. (Cited on page 182)

[13] H. H. Bauschke and J. M. Borwein, *Dykstra's alternating projection algorithm for two sets*, Journal of Approximation Theory, 79 (1994), pp. 418–443. (Cited on page 182)

[14] B. J. Baxter and S. Hubbert, *Radial basis functions for the sphere*, in Recent progress in multivariate approximation, Springer, 2001, pp. 33–47. (Cited on page 5)

[15] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, *Automatic differentiation in machine learning: a survey*, Journal of Marchine Learning Research, 18 (2018), pp. 1–43. (Cited on pages 215 and 248)

[16] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM journal on imaging sciences, 2 (2009), pp. 183–202. (Cited on pages 10, 115, 123, 178, 206, 209, and 210)

[17] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, vol. 1, Springer Science & Business Media, 2008. (Cited on page 205)

[18] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*, Springer Science & Business Media, 2011. (Cited on pages 84 and 96)

[19] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini, *Image deblurring with poisson data: from cells to galaxies*, Inverse Problems, 25 (2009), p. 123006. (Cited on pages 131 and 132)

[20] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for poisson data*, Inverse problems, 26 (2010), p. 105004. (Cited on pages 131 and 132)

[21] A. Besson, L. Roquette, D. Perdios, M. Simeoni, M. Arditi, P. Hurley, Y. Wiaux, and J.-P. Thiran, *A physical model of non-stationary blur in ultrasound imaging*, IEEE Transactions on Computational Imaging, (2019). (Cited on page 206)

[22] E. Bezzam, R. Scheibler, J. Azcarreta, H. Pan, M. Simeoni, R. Beuchat, P. Hurley, B. Bruneau, C. Ferry, and S. Kashani, *Hardware and software for reproducible research in audio array signal processing*, in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Ieee, 2017, pp. 6591–6592. (Cited on page 208)

[23] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, *Iterative methods for image deblurring*, Proceedings of the IEEE, 78 (1990), pp. 856–883. (Cited on page 206)

[24] A. B. Binder, *Lunar prospector: overview*, Science, 281 (1998), pp. 1475–1476. (Cited on page 165)

[25] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, *Sparse Sampling of Signal Innovations*, IEEE Signal Process. Mag., 25 (2008), pp. 31–40. (Cited on pages 10, 175, 176, 177, 178, 179, 183, 184, and 185)

[26] A. K. Boyat and B. K. Joshi, *A review paper: Noise models in digital image processing*, arXiv preprint arXiv:1505.03489, (2015). (Cited on page 131)

[27] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004. (Cited on page 247)

[28] C. Boyer, A. Chambolle, Y. D. Castro, V. Duval, F. De Gournay, and P. Weiss, *On representer theorems and convex regularization*, SIAM Journal on Optimization, 29 (2019), pp. 1260–1281. (Cited on pages 6, 8, 15, 18, 34, 35, 79, and 221)

[29] K. Bredies and M. Carioni, *Sparsity of solutions for variational inverse problems with finite-dimensional data*, arXiv preprint arXiv:1809.05045, (2018). (Cited on pages 6, 15, and 18)

[30] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, *Geometric deep learning: going beyond euclidean data*, IEEE Signal Processing Magazine, 34 (2017), pp. 18–42. (Cited on pages 4, 89, 101, 108, and 111)

[31] T. F. Brooks and W. M. Humphreys, *A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays*, Journal of Sound and Vibration, 294 (2006), pp. 856–879. (Cited on pages 205 and 206)

[32] L. Brusniak, J. Underbrink, and R. Stoker, *Acoustic imaging of aircraft noise sources using large aperture phased arrays*, in 12th AIAA/CEAS Aeroacoustics Conference (27th AIAA Aeroacoustics Conference), 2006, p. 2715. (Cited on page 205)

[33] J. A. Cadzow, *Signal enhancement-a composite property mapping algorithm*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 36 (1988), pp. 49–62. (Cited on page 178)

[34] A. Chambolle and C. Dossal, *On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"*, Journal of Optimization theory and Applications, 166 (2015), pp. 968–982. (Cited on pages 119 and 210)

[35] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of mathematical imaging and vision, 40 (2011), pp. 120–145. (Cited on page 118)

[36] T. F. Chan and C.-K. Wong, *Convergence of the alternating minimization algorithm for blind deconvolution*, Linear Algebra and its Applications, 316 (2000), pp. 259–285. (Cited on page 186)

[37] Y. Chen and T. Pock, *Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration*, IEEE transactions on pattern analysis and machine intelligence, 39 (2017), pp. 1256–1272. (Cited on page 206)

[38] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, John Wiley & Sons, 2013. (Cited on page 155)

[39] Y.-K. Cho, *Compactly supported reproducing kernels for l^2-based sobolev spaces and hankel-schoenberg transforms*, arXiv preprint arXiv:1702.05896, (2017). (Cited on page 140)

[40] N. Chu, J. Picheral, A. Mohammad-Djafari, and N. Gac, *A robust super-resolution approach with sparsity constraint in acoustic imaging*, Applied Acoustics, 76 (2014), pp. 197–208. (Cited on page 206)

[41] Z. Chu and Y. Yang, *Comparison of deconvolution methods for the visualization of acoustic sources based on cross-spectral imaging function beamforming*, Mechanical Systems and Signal Processing, 48 (2014), pp. 404–422. (Cited on page 206)

[42] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, Springer, 2011, pp. 185–212. (Cited on page 178)

[43] L. Condat, *A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479. (Cited on pages xxvi, 10, 115, 116, 117, 118, and 125)

[44] L. Condat, A. Hirabayashi, L. Condat, and A. Hirabayashi, *Cadzow Denoising Upgraded : A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements*, (2015). (Cited on pages 178, 179, 184, and 198)

[45] T. W. Crowther, H. B. Glick, K. R. Covey, C. Bettigole, D. S. Maynard, S. M. Thomas, J. R. Smith, G. Hintler, M. C. Duguid, G. Amatulli, et al., *Mapping tree density at a global scale*, Nature, 525 (2015), p. 201. (Cited on pages 2, 3, and 154)

[46] T. Debarre, J. Fageot, H. Gupta, and M. Unser, *Solving continuous-domain problems exactly with multiresolution b-splines*, in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 5122–5126. (Cited on page 14)

[47] M. Defferrard, X. Bresson, and P. Vandergheynst, *Convolutional neural networks on graphs with fast localized spectral filtering*, in Advances in neural information processing systems, 2016, pp. 3844–3852. (Cited on pages 105, 142, 206, 208, and 213)

[48] M. Defferrard, N. Perraudin, T. Kacprzak, and R. Sgier, *Deepsphere: towards an equivariant graph-based spherical cnn*, arXiv preprint arXiv:1904.05146, (2019). (Cited on pages 4 and 142)

[49] M. Detay, *La lune: un monde d'eau!-etat des connaissances et perspectives hydrotechniques*, La Houille Blanche, (2010), pp. 127–136. (Cited on page 165)

[50] F. R. Deutsch, *Best approximation in inner product spaces*, Springer Science & Business Media, 2012. (Cited on pages 238 and 239)

[51] S. Diamond and S. Boyd, *Matrix-free convex optimization modeling*, in Optimization and its applications in control and data sciences, Springer, 2016, pp. 221–264. (Cited on page 128)

[52] I. Dokmanić and Y. M. Lu, *Sampling sparse signals on the sphere: Algorithms and applications*, IEEE Transactions on Signal Processing, 64 (2015), pp. 189–202. (Cited on page 176)

[53] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218. (Cited on page 185)

[54] ESA, *Planck legacy archive*. https://pla.esac.esa.int, 2018. [Online; accessed 12-July-2019]. (Cited on pages 146 and 162)

[55] R. Escalante and M. Raydan, *Alternating projection methods*, SIAM, 2011. (Cited on pages 179, 181, and 182)

[56] S. Fisher and J. W. Jerome, *Spline solutions to l1 extremal problems in one and several variables*, Journal of Approximation Theory, 13 (1975), pp. 73–83. (Cited on page 14)

[57] A. Flinth and P. Weiss, *Exact solutions of infinite dimensional total-variation regularized problems*, arXiv preprint arXiv:1708.02157, (2017). (Cited on pages 6 and 14)

[58] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Bull. Am. Math, 54 (2017), pp. 151–165. (Cited on page 206)

[59] M. Fréchet, *Sur les ensembles de fonctions et les opérations linéaires*, CR Acad. Sci. Paris, 144 (1907), pp. 1414–1416. (Cited on pages xxv, 24, and 25)

[60] W. Freeden and M. Schreiner, *Spherical functions of mathematical geosciences: a scalar, vectorial, and tensorial setup*, Springer Science & Business Media, 2008. (Cited on pages 1, 53, 60, and 61)

[61] E. Fuselier and G. B. Wright, *Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1753–1776. (Cited on pages 4 and 5)

[62] GFW, *Forest fires and their impact*. https://fires.globalforestwatch.org/about/, 2019. [Online; accessed 04-July-2019]. (Cited on page 154)

[63] ——, *Forests and key themes critical to sustainability and the health of ecosystems*. https://www.globalforestwatch.org/topics/biodiversity, 2019. [Online; accessed 04-July-2019]. (Cited on page 154)

[64] Q. L. Gia, I. H. Sloan, and H. Wendland, *Multiscale analysis in sobolev spaces on the sphere*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 2065–2090. (Cited on pages 136, 137, and 140)

[65] T. Gneiting et al., *Strictly and non-strictly positive definite functions on spheres*, Bernoulli, 19 (2013), pp. 1327–1349. (Cited on pages 4 and 5)

[66] Á. González, *Measurement of areas on a sphere using fibonacci and latitude–longitude lattices*, Mathematical Geosciences, 42 (2010), p. 49. (Cited on pages 4 and 97)

[67] K. M. Gorski, E. Hivon, A. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, *Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere*, The Astrophysical Journal, 622 (2005), p. 759. (Cited on pages 3, 4, 102, and 213)

[68] J. Gray, *The shaping of the riesz representation theorem: A chapter in the history of analysis*, Archive for History of Exact Sciences, 31 (1984), pp. 127–187. (Cited on page 25)

[69] P. J. Green and B. W. Silverman, *Nonparametric regression and generalized linear models: a roughness penalty approach*, Chapman and Hall/CRC, 1993. (Cited on page 13)

[70] K. Gregor and Y. LeCun, *Learning fast approximations of sparse coding*, in Proceedings of the 27th International Conference on International Conference on Machine Learning, Omnipress, 2010, pp. 399–406. (Cited on pages 206, 207, 209, and 211)

[71] B. Gu, D. Wang, Z. Huo, and H. Huang, *Inexact proximal gradient methods for non-convex and non-smooth optimization*, in Thirty-Second AAAI Conference on Artificial Intelligence, 2018. (Cited on pages 178 and 192)

[72] H. Gupta, J. Fageot, and M. Unser, *Continuous-domain solutions of linear inverse problems with tikhonov vs. generalized tv regularization*, arXiv preprint arXiv:1802.01344, (2018). (Cited on pages 3, 6, 8, 14, 18, 34, 35, 74, and 221)

[73] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, *Cnn-based projected gradient descent for consistent ct image reconstruction*, IEEE transactions on medical imaging, 37 (2018), pp. 1440–1453. (Cited on pages 178, 192, and 206)

[74] G. M. Hallegraeff, *Ocean climate change, phytoplankton community responses, and harmful algal blooms: a formidable predictive challenge 1*, Journal of phycology, 46 (2010), pp. 220–235. (Cited on page 146)

[75] I. Halperin, *The product of projection operators*, Acta Sci. Math.(Szeged), 23 (1962), pp. 96–99. (Cited on page 182)

[76] P. Hanjie, T. Blu, M. Vetterli, and D. Frederike, *Unified algorithmic framework for reconstructing signals with finite rate of innovation*. https://github.com/hanjiepan/FRI_pkg. Accessed: 2019-10-18. (Cited on page 196)

[77] R. Hansen and P. Andersen, *A 3d underwater acoustic camera—properties and applications*, in Acoustical Imaging, Springer, 1996, pp. 607–611. (Cited on page 205)

[78] D. P. Hardin, T. Michaels, and E. B. Saff, *A comparison of popular point configurations on s2*, Dolomites Research Notes on Approximation, 9 (2016). (Cited on pages 3, 10, 97, 98, and 102)

[79] S. Hubbert, *Radial basis function interpolation on the sphere*, PhD thesis, University of London, 2002. (Cited on pages 4, 5, 37, 38, 43, 44, and 45)

[80] T. P. Hughes, J. T. Kerry, M. Álvarez-Noriega, J. G. Álvarez-Romero, K. D. Anderson, A. H. Baird, R. C. Babcock, M. Beger, D. R. Bellwood, R. Berkelmans, et al., *Global warming and recurrent mass bleaching of corals*, Nature, 543 (2017), p. 373. (Cited on page 146)

[81] P. Hurley and M. Simeoni, *Flexibeam: analytic spatial filtering by beamforming*, in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Ieee, 2016, pp. 2877–2880. (Cited on pages 162 and 205)

[82] ——, *Flexarray: Random phased array layouts for analytical spatial filtering*, in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Ieee, 2017, pp. 3380–3384. (Cited on pages 162 and 205)

[83] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, *Deep convolutional neural network for inverse problems in imaging*, IEEE Transactions on Image Processing, 26 (2017), pp. 4509–4522. (Cited on page 206)

[84] K. Jinadasa, *Applications of the matrix operators vech and vec*, Linear Algebra and its Applications, 101 (1988), pp. 73–79. (Cited on pages 209, 243, and 244)

[85] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*, PTR Prentice Hall Englewood Cliffs, 1993. (Cited on page 162)

[86] C. H. Jones and G. A. Gilmour, *Acoustic camera apparatus*, Aug. 5 1975. US Patent 3,898,608. (Cited on page 205)

[87] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: Open source scientific tools for Python*. http://www.scipy.org/, 2001–Now. Accessed: 2019-10-18. (Cited on page 198)

[88] A. Jung, *A fixed-point of view on gradient methods for big data*, Frontiers in Applied Mathematics and Statistics, 3 (2017), p. 18. (Cited on pages 237 and 238)

[89] C. Justice, L. Giglio, S. Korontzi, J. Owens, J. Morisette, D. Roy, J. Descloitres, S. Alleaume, F. Petitcolin, and Y. Kaufman, *The modis fire products*, Remote Sensing of Environment, 83 (2002), pp. 244–262. (Cited on page 154)

[90] M. Kamionkowski, A. Kosowsky, and A. Stebbins, *Statistics of cosmic microwave background polarization*, Physical Review D, 55 (1997), p. 7368. (Cited on page 162)

[91] J. Keiner, S. Kunis, and D. Potts, *Fast summation of radial functions on the sphere*, Computing, 78 (2006), pp. 1–15. (Cited on page 4)

[92] K. Kim, N. Neretti, and N. Intrator, *Mosaicing of acoustic camera images*, IEE Proceedings-Radar, Sonar and Navigation, 152 (2005), pp. 263–270. (Cited on page 205)

[93] G. S. Kimeldorf and G. Wahba, *A correspondence between bayesian estimation on stochastic processes and smoothing by splines*, The Annals of Mathematical Statistics, 41 (1970), pp. 495–502. (Cited on pages 13 and 18)

[94] Y. Knyazikhin, *Modis leaf area index (lai) and fraction of photosynthetically active radiation absorbed by vegetation (fpar) product (mod 15) algorithm theoretical basis document*, http://eospso. gsfc. nasa. gov/atbd/modistabls. heml, (1999). (Cited on page 154)

[95] H. Krim and M. Viberg, *Two decades of array signal processing research: the parametric approach*, IEEE signal processing magazine, 13 (1996), pp. 67–94. (Cited on pages 208, 209, 254, and 258)

[96] H. W. Kuhn, *The hungarian method for the assignment problem*, Naval research logistics quarterly, 2 (1955), pp. 83–97. (Cited on page 198)

[97] M. Kuusela, *Preprocessed argo data*. https://github.com/mkuusela/PreprocessedArgoData. Accessed: 2019-06-26. (Cited on pages 146 and 147)

[98] M. Kuusela and M. L. Stein, *Locally stationary spatio-temporal interpolation of argo profiling float data*, Proceedings of the Royal Society A, 474 (2018), p. 20180400. (Cited on pages 11, 145, 146, and 147)

[99] H.-C. Lai and L.-J. Lin, *The fenchel-moreau theorem for set functions*, Proceedings of the American Mathematical Society, (1988), pp. 85–90. (Cited on page 116)

[100] M. J. Lai and A. Varghese, *On convergence of the alternating projection method for matrix completion and sparse recovery problems*, arXiv preprint arXiv:1711.02151, (2017). (Cited on page 239)

[101] D. Lawrence, R. Elphic, W. Feldman, T. Prettyman, O. Gasnault, and S. Maurice, *Small-area thorium features on the lunar surface*, Journal of Geophysical Research: Planets, 108 (2003). (Cited on pages 3, 11, 165, 167, and 168)

[102] D. Lawrence, E. Guinness, and S. Slavney, *Lunar prospector gamma ray spectrometer and neutron spectrometer reduced data*. https://pds-geosciences.wustl.edu/missions/lunarp/reduced_grsns.html, 2005. [Online; accessed 14-October-2019]. (Cited on pages 146 and 167)

[103] D. Lawrence, S. Maurice, and W. Feldman, *Gamma-ray measurements from lunar prospector: Time series data reduction for the gamma-ray spectrometer*, Journal of Geophysical Research: Planets, 109 (2004). (Cited on pages 11, 165, 167, and 170)

[104] Q. Le Gia, I. Sloan, and H. Wendland, *Multiscale approximation for functions in arbitrary sobolev spaces by scaled radial basis functions on the unit sphere*, Applied and Computational Harmonic Analysis, 32 (2012), pp. 401–412. (Cited on pages 136, 137, and 138)

[105] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, nature, 521 (2015), p. 436. (Cited on page 206)

[106] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, et al., *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86 (1998), pp. 2278–2324. (Cited on page 209)

[107] A. S. Lewis, D. R. Luke, and J. Malick, *Local linear convergence for alternating and averaged nonconvex projections*, Foundations of Computational Mathematics, 9 (2009), pp. 485–513. (Cited on page 182)

[108] A. S. Lewis and J. Malick, *Alternating projections on manifolds*, Mathematics of Operations Research, 33 (2008), pp. 216–234. (Cited on pages 182 and 239)

[109] H. Li and Z. Lin, *Accelerated proximal gradient methods for nonconvex programming*, in Advances in neural information processing systems, 2015, pp. 379–387. (Cited on pages 187, 221, 232, 233, and 234)

[110] H. Li, Y. Yang, D. Chen, and Z. Lin, *Optimization algorithm inspired deep neural network structure design*, arXiv preprint arXiv:1810.01638, (2018). (Cited on pages 206 and 209)

[111] J. Liang, J. Fadili, and G. Peyré, *Activity identification and local linear convergence of forward–backward-type methods*, SIAM Journal on Optimization, 27 (2017), pp. 408–437. (Cited on page 119)

[112] J. Liang and C.-B. Schönlieb, *Faster fista*, arXiv preprint arXiv:1807.04005, (2018). (Cited on pages 123, 206, and 210)

[113] ——, *Improving fista: Faster, smarter and greedier*, arXiv preprint arXiv:1811.01430, (2018). (Cited on pages 119 and 123)

[114] M. Lifshits, *Lectures on gaussian processes*, in Lectures on Gaussian Processes, Springer, 2012, pp. 1–117. (Cited on page 221)

[115] M. A. Lifshits, *Gaussian random functions*, vol. 322, Springer Science & Business Media, 2013. (Cited on page 221)

[116] S. Liu and G. Trenkler, *Hadamard, khatri-rao, kronecker and other matrix products*, International Journal of Information and Systems Sciences, 4 (2008), pp. 160–177. (Cited on pages 243 and 244)

[117] P.-L. Loh and M. J. Wainwright, *Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima*, in Advances in Neural Information Processing Systems, 2013, pp. 476–484. (Cited on page 221)

[118] D. R. Luke, *Finding best approximation pairs relative to a convex and prox-regular set in a hilbert space*, SIAM Journal on Optimization, 19 (2008), pp. 714–739. (Cited on page 238)

[119] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195. (Cited on page 206)

[120] D. Marinucci and G. Peccati, *Random fields on the sphere: representation, limit theorems and cosmological applications*, vol. 389, Cambridge University Press, 2011. (Cited on page 3)

[121] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust statistics: theory and methods (with R)*, John Wiley & Sons, 2019. (Cited on page 221)

[122] M. T. McCann, K. H. Jin, and M. Unser, *Convolutional neural networks for inverse problems in imaging: A review*, IEEE Signal Processing Magazine, 34 (2017), pp. 85–95. (Cited on page 206)

[123] S. Mei, Y. Bai, and A. Montanari, *The landscape of empirical risk for non-convex losses*, arXiv preprint arXiv:1607.06534, (2016). (Cited on page 221)

[124] A. Meyer and D. Döbler, *Noise source localization within a car interior using 3d-microphone arrays*, Proceedings of the BeBeC, (2006), pp. 1–7. (Cited on page 206)

[125] V. Michel, *Lectures on constructive approximation*, AMC, 10 (2013), p. 12. (Cited on pages xxvi, 5, 8, 13, 37, 38, 39, 42, 43, 45, 48, 51, 59, 85, 86, 94, 95, 96, 99, 212, and 227)

[126] F. J. Narcowich and J. D. Ward, *Scattered data interpolation on spheres: error estimates and locally supported basis functions*, SIAM Journal on Mathematical Analysis, 33 (2002), pp. 1393–1410. (Cited on pages 4, 5, and 140)

[127] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer Science & Business Media, 2013. (Cited on page 116)

[128] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994. (Cited on page 130)

[129] T. Oliphant, *NumPy: A guide to NumPy*. http://www.numpy.org/, 2006–Now. Accessed: 2019-10-18. (Cited on page 198)

[130] H. Pan, T. Blu, and M. Vetterli, *Towards Generalized FRI Sampling With an Application to Source Resolution in Radioastronomy*, IEEE

Trans. Signal Process., 65 (2017), pp. 821–835. (Cited on pages 10, 175, 176, 177, 178, 185, 186, 195, and 196)

[131] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanić, and M. Vetterli, *Audio speech recordings used in the paper FRIDA: FRI-based DOA Estimation for Arbitrary Array Layout*, Mar. 2017. This work was supported by the Swiss National Science Foundation grant 20FP-1 151073, LABEX WIFI under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL* and by Agence Nationale de la Recherche under reference ANR-13-JS09-0001-01. (Cited on pages 215, 254, and 255)

[132] H. Pan, R. Scheibler, E. F. Bezzam, I. Dokmanic, and M. Vetterli, *Frida: Fri-based doa estimation for arbitrary array layouts*, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2017), pp. 5. 3186–3190. (Cited on page 254)

[133] H. Pan, M. Simeoni, P. Hurley, T. Blu, and M. Vetterli, *Leap: Looking beyond pixels with continuous-space estimation of point sources*, Astronomy & Astrophysics, 608 (2017), p. A136. (Cited on page 3)

[134] N. Parikh, S. Boyd, et al., *Proximal algorithms*, Foundations and Trends® in Optimization, 1 (2014), pp. 127–239. (Cited on pages 10, 115, 116, 122, 126, 130, 131, 132, 178, 186, 187, 188, 206, and 207)

[135] S. C. Park, M. K. Park, and M. G. Kang, *Super-resolution image reconstruction: a technical overview*, IEEE signal processing magazine, 20 (2003), pp. 21–36. (Cited on page 206)

[136] C. L. Parkinson, *Aqua: An earth-observing satellite mission to examine water and other climate variables*, IEEE Transactions on Geoscience and Remote Sensing, 41 (2003), pp. 173–183. (Cited on page 147)

[137] R. Pascanu, T. Mikolov, and Y. Bengio, *On the difficulty of training recurrent neural networks*, in International conference on machine learning, 2013, pp. 1310–1318. (Cited on page 215)

[138] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran, *A deep learning approach to ultrasound image recovery*, in 2017 IEEE International Ultrasonics Symposium (IUS), Ieee, 2017, pp. 1–4. (Cited on page 206)

[139] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, *Deepsphere: Efficient spherical convolutional neural network with healpix sampling for cosmological applications*, Astronomy and Computing, (2019). (Cited on pages 105, 106, 108, 128, 206, 207, and 213)

[140] K. B. Petersen, M. S. Pedersen, et al., *The matrix cookbook*, Technical University of Denmark, 7 (2008), p. 510. (Cited on page 246)

[141] R. Pina and R. Puetter, *Bayesian image reconstruction: The pixon and optimal image modeling*, Publications of the Astronomical Society of the Pacific, 105 (1993), p. 630. (Cited on pages 11, 146, 165, 169, and 171)

[142] B. RAFAELY, *Fundamentals of spherical array processing*, vol. 8, Springer, 2015. (Cited on pages 3, 37, 39, 42, 48, 205, 211, 212, and 213)

[143] C. E. RASMUSSEN AND C. K. WILLIAMS, *Gaussian process for machine learning*, MIT press, 2006. (Cited on pages 137 and 138)

[144] R.-D. REISS, *A course on point processes*, Springer Science & Business Media, 2012. (Cited on pages 2 and 155)

[145] R. W. REYNOLDS AND T. M. SMITH, *Improved global sea surface temperature analyses using optimum interpolation*, Journal of climate, 7 (1994), pp. 929–948. (Cited on page 146)

[146] J. R. RICE AND J. S. WHITE, *Norms for smoothing and estimation*, SIAM review, 6 (1964), pp. 243–256. (Cited on pages 69, 130, and 131)

[147] C. RICHE DE PRONY, *Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastique et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures*, J. de l'École Polytechnique, 1 (1795), pp. 24–76. (Cited on page 183)

[148] B. RIMOLDI, *Principles of Digital Communication: A Top-Down Approach*, Cambridge University Press, 2016. (Cited on pages 115, 177, 234, and 236)

[149] W. J. RIPPLE, C. WOLF, T. M. NEWSOME, M. GALETTI, M. ALAMGIR, E. CRIST, M. I. MAHMOUD, AND W. F. LAURANCE, *World scientists' warning to humanity: A second notice*, BioScience, 67 (2017), pp. 1026–1028. (Cited on page 1)

[150] S. ROBIN, *Pyramic Dataset : 48-Channel Anechoic Audio Recordings of 3D Sources*, Mar. 2018. The author would like to acknowledge Juan Azcarreta Ortiz, Corentin Ferry, and René Beuchat for their help in the design and usage of the Pyramic array. Hanjie Pan, Miranda Kreković, Mihailo Kolundzija, and Dalia El Badawy for lending a hand, or even two, during experiments. Finally, Juan Azcarreta Ortiz, Eric Bezzam, Hanjie Pan and Ivan Dokmanić for feedback on the documentation and dataset organization. (Cited on pages 217 and 254)

[151] J. ROMBERG, *Imaging via compressive sampling*, IEEE Signal Processing Magazine, 25 (2008), pp. 14–20. (Cited on page 206)

[152] O. RONNEBERGER, P. FISCHER, AND T. BROX, *U-net: Convolutional networks for biomedical image segmentation*, in International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241. (Cited on page 206)

[153] W. RUDIN, *Functional analysis, mcgrawhill*, Inc, New York, (1991). (Cited on pages xxv, 23, and 34)

[154] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, *Plug-and-play methods provably converge with properly trained denoisers*, arXiv preprint arXiv:1905.05406, (2019). (Cited on pages 178 and 192)

[155] R. Scheibler, J. Azcarreta, R. Beuchat, and C. Ferry, *Pyramic: Full stack open microphone array architecture and dataset*, in 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), IEEE, 2018, pp. 226–230. (Cited on pages 208, 212, 213, 215, 254, and 255)

[156] B. Scholkopf, R. Herbrich, and A. Smola, *A generalized representer theorem. computational learning theory*, Lecture Notes in Computer Science. 2001a, 2111 (2001), pp. 416–426. (Cited on pages 13, 14, and 18)

[157] R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, 2005. (Cited on page 155)

[158] C. E. Shannon, *A mathematical theory of communication*, ACM SIGMOBILE Mobile Computing and Communications Review, 5 (2001), pp. 3–55. (Cited on page 177)

[159] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains*, IEEE Signal Processing Magazine, 3 (2013), pp. 83–98. (Cited on pages 106, 107, 128, 142, and 213)

[160] P. Sijtsma, *Clean based on spatial source coherence*, International journal of aeroacoustics, 6 (2007), pp. 357–374. (Cited on pages 205 and 206)

[161] M. Simeoni, *Statistical Inference in Positron Emission Tomography*, tech. rep., SWISS FEDERAL INSTITUTE OF TECHNOLOGY LAUSANNE, 2014. (Cited on pages 2, 132, 155, and 247)

[162] M. Simeoni, A. Besson, P. Hurley, and M. Vetterli, *Cadzow plug-and-play gradient descent for generalised FRI*. Under Submission, 2019. (Cited on pages 10 and 175)

[163] M. M. J.-A. Simeoni, *Towards more accurate and efficient beamformed radio interferometry imaging*, tech. rep., 2015. (Cited on page 3)

[164] ——, *Deconvolution of gaussian random fields using the graph fourier transform*, tech. rep., 2016. (Cited on pages 209 and 221)

[165] M. M. J.-A. Simeoni and P. Hurley, *Graph spectral clustering of convolution artefacts in radio interferometric images*, tech. rep., 2019. (Cited on page 205)

[166] M. M. J.-A. Simeoni, S. Kashani, P. Hurley, and M. Vetterli, *Deepwave: A recurrent neural-network for real-time acoustic imaging*, in Advances in neural information processing systems, 2019. (Cited on pages 12 and 205)

[167] E. Solomentsev, *Spherical harmonics*, http1/wwwencyclopediaofinatb. org/indexphp, (2001). (Cited on page 39)

[168] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces (PMS-32)*, vol. 32, Princeton university press, 2016. (Cited on page 212)

[169] R. Stockli, *Leaf Area Index (TERRA/MODIS)*. https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD15A2_M_LAI&year=2016, 1999. [Online; accessed 14-December-2018]. (Cited on pages 2, 146, and 154)

[170] R. Stockli and J. Allen, *ACTIVE FIRES (1 MONTH - TERRA/MODIS)*. https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD14A1_M_FIRE&year=2016, 1999. [Online; accessed 14-December-2018]. (Cited on pages 146 and 154)

[171] R. Stockli, G. Feldman, and N. Kuring, *Land and sea surface temperature (TERRA/MODIS & AQUA/MODIS)*. https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD11C1_M_LSTDA&year=2016, 1999. [Online; accessed 14-December-2018]. (Cited on page 2)

[172] D. Stoyan and A. Penttinen, *Recent applications of point process methods in forestry statistics*, Statistical Science, (2000), pp. 61–78. (Cited on pages 2 and 155)

[173] J. Sun, H. Li, Z. Xu, et al., *Deep admm-net for compressive sensing mri*, in Advances in neural information processing systems, 2016, pp. 10–18. (Cited on pages 206 and 209)

[174] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, *On the importance of initialization and momentum in deep learning*, in International conference on machine learning, 2013, pp. 1139–1147. (Cited on pages 215 and 255)

[175] M. K. Tam, *Alternating projection methods failure in the absence of convexity*. (Cited on page 182)

[176] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, Journal of Optimization Theory and Applications, 178 (2018), pp. 455–476. (Cited on pages 237 and 238)

[177] M. Unser, *A unifying representer theorem for inverse problems and machine learning*, arXiv preprint arXiv:1903.00687, (2019). (Cited on pages xxv, 7, 8, 15, 18, 27, 28, 29, 30, 31, 34, 35, and 79)

[178] M. Unser, J. Fageot, and H. Gupta, *Representer theorems for sparsity-promoting l1 regularization*, IEEE Transactions on Information Theory, 62 (2016), pp. 5167–5180. (Cited on pages 10, 33, 81, 101, and 206)

[179] M. Unser, J. Fageot, and J. P. Ward, *Splines are universal solutions of linear inverse problems with generalized tv regularization*, SIAM Review, 59 (2017), pp. 769–793. (Cited on pages xxvi, 6, 14, 18, 23, 25, 26, 67, 68, and 72)

[180] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*, Cambridge University Press, 2014. (Cited on pages 15, 21, 25, 51, and 221)

[181] A.-J. van der Veen and S. J. Wijnholds, *Signal processing tools for radio astronomy*, in Handbook of Signal Processing Systems, Springer, 2013, pp. 421–463. (Cited on pages 146, 157, 164, 205, 209, 210, 213, 218, 221, 244, and 258)

[182] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, *Plug-and-play priors for model based reconstruction*, in 2013 IEEE Global Conference on Signal and Information Processing, IEEE, 2013, pp. 945–948. (Cited on pages 178 and 192)

[183] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*, Cambridge University Press, 2014. (Cited on pages 42, 48, 68, 69, 90, 91, 96, 177, 184, and 254)

[184] M. Vetterli, P. Marziliano, and T. Blu, *Sampling signals with finite rate of innovation*, IEEE Trans. Signal Process., 50 (2002), pp. 1417–1428. (Cited on pages 175, 176, 177, 179, and 182)

[185] C. R. Vogel and M. E. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE transactions on image processing, 7 (1998), pp. 813–824. (Cited on pages 6 and 26)

[186] J. von Neumann, *The geometry of orthogonal spaces, functional operators-vol. ii*, Annals of Math. Studies, 22 (1950). (Cited on pages 181 and 182)

[187] M. Vono, N. Dobigeon, and P. Chainais, *Bayesian image restoration under poisson noise and log-concave prior*, in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 1712–1716. (Cited on page 157)

[188] A. E. Walker and J. L. Wilkin, *Optimal averaging of noaa/nasa pathfinder satellite sea surface temperature data*, Journal of Geophysical Research: Oceans, 103 (1998), pp. 12869–12883. (Cited on page 147)

[189] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun, *Discrete laplace operators: no free lunch*, in Symposium on Geometry processing, Aire-la-Ville, Switzerland, 2007, pp. 33–37. (Cited on pages 89, 106, and 108)

[190] H. WENDLAND, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Advances in computational Mathematics, 4 (1995), pp. 389–396. (Cited on page 140)

[191] ——, *Scattered data approximation*, vol. 17, Cambridge university press, 2004. (Cited on pages 14 and 33)

[192] Y. WIAUX, L. JACQUES, G. PUY, A. M. SCAIFE, AND P. VANDERGHEYNST, *Compressed sensing imaging techniques for radio interferometry*, Monthly Notices of the Royal Astronomical Society, 395 (2009), pp. 1733–1742. (Cited on page 206)

[193] S. J. WIJNHOLDS AND A.-J. VAN DER VEEN, *Fundamental imaging limits of radio telescope arrays*, IEEE Journal of Selected Topics in Signal Processing, 2 (2008), pp. 613–623. (Cited on page 210)

[194] J. WILSON, V. EKE, R. MASSEY, R. ELPHIC, B. JOLLIFF, D. LAWRENCE, E. LLEWELLIN, J. MCELWAINE, AND L. TEODORO, *Evidence for explosive silicic volcanism on the moon from the extended distribution of thorium near the compton-belkovich volcanic complex*, Journal of Geophysical Research: Planets, 120 (2015), pp. 92–108. (Cited on page 167)

[195] J. T. WILSON, D. J. LAWRENCE, P. N. PEPLOWSKI, J. T. CAHILL, V. R. EKE, R. J. MASSEY, AND L. F. TEODORO, *Image reconstruction techniques in neutron and gamma ray spectroscopy: Improving lunar prospector data*, Journal of Geophysical Research: Planets, 123 (2018), pp. 1804–1822. (Cited on pages 3, 11, 146, 165, 169, 170, and 171)

[196] WORLD METEOROLOGICAL ORGANIZATION, *Oscar surface*. https://oscar.wmo.int/surface, 2018. [Online; accessed 14-December-2018]. (Cited on page 2)

[197] X. XIONG, K. CHIANG, J. SUN, W. BARNES, B. GUENTHER, AND V. SALOMONSON, *Nasa eos terra and aqua modis on-orbit performance*, Advances in Space Research, 43 (2009), pp. 413–422. (Cited on page 155)

[198] L. XU, J. S. REN, C. LIU, AND J. JIA, *Deep convolutional neural network for image deconvolution*, in Advances in Neural Information Processing Systems, 2014, pp. 1790–1798. (Cited on page 206)

[199] Q. YAO, J. T. KWOK, F. GAO, W. CHEN, AND T.-Y. LIU, *Efficient inexact proximal gradient algorithm for nonconvex problems*, arXiv preprint arXiv:1612.09069, (2016). (Cited on page 192)

[200] C. ZALINESCU, *Convex analysis in general vector spaces*, World scientific, 2002. (Cited on page 116)

[201] S. ZHU, *Compactly supported radial basis functions: how and why?*, SIAM Review, (2012). (Cited on pages 4, 139, and 140)

[202] A. ZONCA, L. SINGER, D. LENZ, M. REINECKE, C. ROSSET, E. HIVON, AND K. GORSKI, *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python*, Journal of Open Source Software, 4 (2019), p. 1298. (Cited on page 163)

[203] H. ZOU AND T. HASTIE, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2005), pp. 301–320. (Cited on page 209)

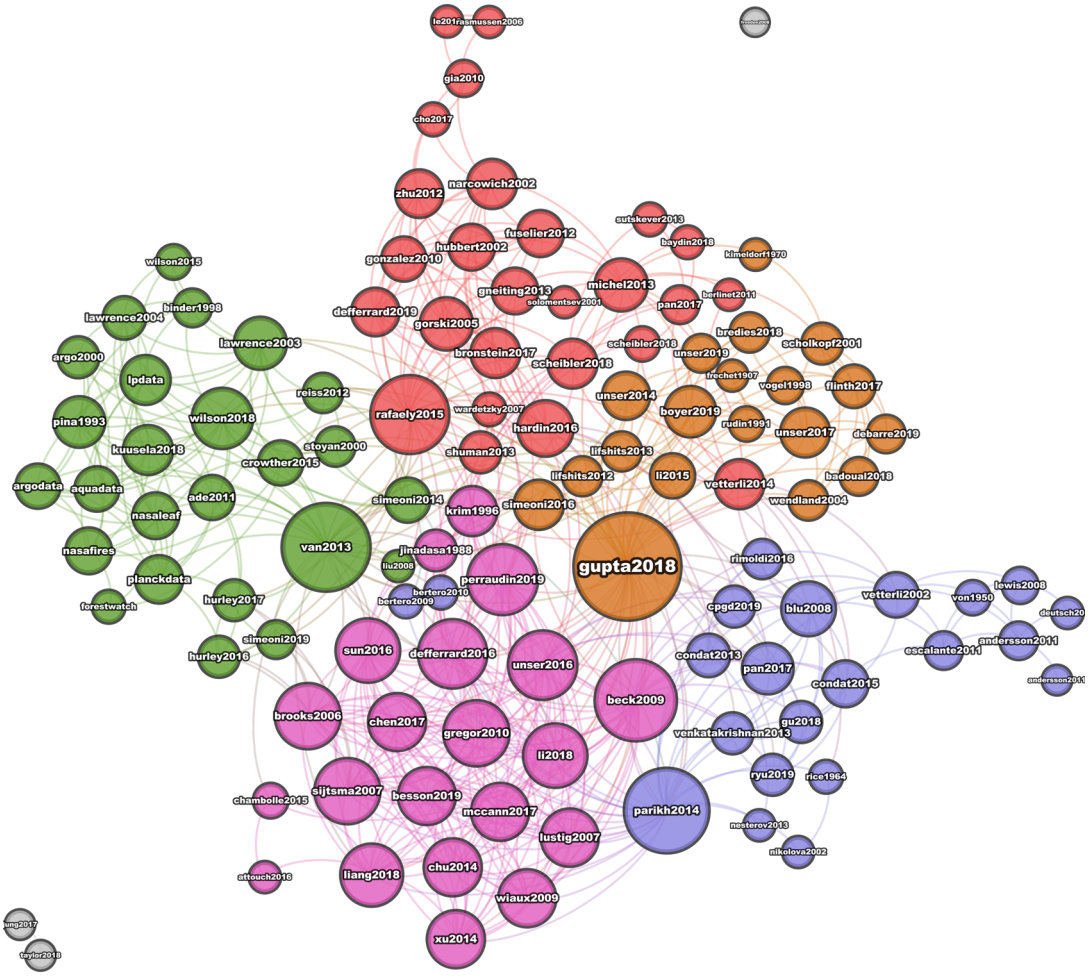


Figure C.8: Graph of all references cited strictly more than once in this thesis. Each publication is represented by a node, sized according to the number of citations within this thesis. Publications are connected with one another if they are cited in a common page. Colors correspond to communities of closely related publications. We observe that each community corresponds to a different thematic of this thesis (green=applications, red=spherical approximation, orange=representer theorems, blue=finite rate of innovation sampling, yellow=neural-networks).

Index

Acronyms

AC acoustic camera.
APGD accelerated proximal gradient descent.

CBP constrained basis pursuit.
cf. confere.
CMB cosmic microwave radiation background.
CNN convolutional neural-network.
CNNs convolutional neural-network.
CPGD Cadzow plug-and-play gradient descent.
CPU central processing unit.

DAS delay and sum.
df degrees of freedom.
DFT discrete Fourier transform.
DOA direction of arrival.

e.g. exempli gratia.
et al. et alii.
etc. et cætera.

FCBP Functional Constrained Basis Pursuit.
FISTA fast iterative soft-thresholding algorithm.
FNSH fully normalised spherical harmonics.
FPBP functional penalised basis pursuit.
FPT functional penalised Tikhonov.
FRI finite rate of innovation.
FWHM full width at half maximum.

genFRI generalised finite rate of innovation.
GFT graph Fourier transform.
GPU graphical processing unit.
GRS gamma-ray spectroscopy.
gSHT generalised spherical harmonic transform.
gTikhonov generalised Tikhonov.
gTV generalised total variation.

HEALPix Hierarchical Equal Area isoLatitude Pixelization.

i.e. id est.
i.f.f. if and only if.
IBM International Business Machines.
ISTA iterative soft-thresholding algorithm.

KL Kullback-Leibler.
KREEP KREEP (potassium (K), rare earth elements (REE), phosphorus (P)).

LAI Leaf Area Index.
LISTA learned iterative soft-thresholding algorithm.
LP Lunar Prospector.
lwsc lower semi-continuous.

MAP method of alternating projections.
MODIS Moderate Resolution Imaging Spectroradiometer.

NASA National Aeronautics and Space Administration.

PBP penalised basis pursuit.
PDS Planetary Data System.
PDS primal-dual splitting method.
PGD proximal gradient descent.
PnP plug-and-play.
PSF point spread function.
PSNR peak signal to noise ratio.

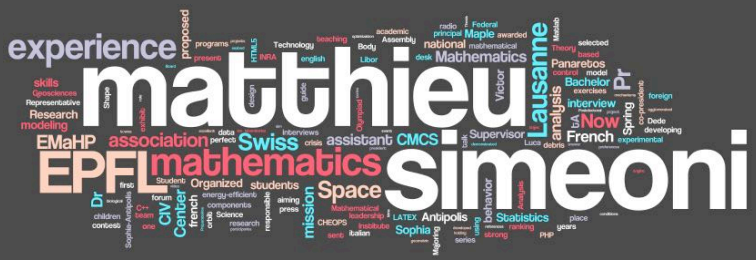
RKHS reproducing kernel Hilbert space.
RNN recurrent neural-network.

s.t. such that.
SGD stochastic gradient descent.
SHT spherical harmonic transform.
SLRA structured low-rank approximation.
SNR signal to noise ratio.
SVD singular value decomposition.

Th Thorium.
TV total variation.

vs. versus.

w.r.t. with respect to.
WMO World Meteorological Organisation.



about

Birth date : 26.08.1991
Nationality : French

✉ Chemin des Brandards 66,
2000 Neuchâtel, Switzerland.

✆ (+41) 78 742 68 25

@ matthieu.simeoni@epfl.ch

languages

	beginner	advanced	fluent
french			●
english			●
italian		●	

programming

	average	good	skilled
Python			●
Matlab			●
R			●
SQL		●	
PHP		●	
HTML5		●	
LaTeX			●

skills

✓ strong experimental design and analysis skills,
✓ strong practical experience with large data sets from sensor networks,
✓ deep knowledge of image and signal processing,
✓ experience in Time Series modeling & Statistical Learning,
✓ experience in extremes and rare events modeling,
✓demonstrated leadership and self-direction,
✓ excellent written and presentation skills.

professional experience

2015-Now **Smart Sensing with Sensor Arrays** — IBM Research, Zurich
Research Scientist (Predoc in the Cognitive Signal Processing group)
Supervisor: Pr. P. Hurley, IBM Zurich; Manager: Dr. C. Bekas, IBM Zurich.
I developed a generic framework called Bluebild for smart sensor arrays, permitting better sensing, estimation and machine learning. Applications include medical imagery, acoustics and radio-astronomy. My work lead to 14 publications and 8 patents. I also gave a press conference during the IBM Research press day. Finally, I helped design a demo on sound localisation showcased in the Zurich IBM Client Center.

Spring 2015 **A New Imager for the Square Kilometre Array (SKA)** — IBM Research, Zurich
Research Scientist (Master Thesis)
Supervisors: Dr. P. Hurley, IBM Zurich; Pr. V. Panaretos, EPFL; Pr. M. Vetterli, EPFL.
As part of my master thesis, I proposed a new imaging algorithm, efficiently exploiting the structure of the hierarchical compression schemes envisioned for the SKA to recover sky images. This imager was a precursor to the Bluebild algorithm, one of the main outcomes of the DOME project and bound to impact significantly the field of radio astronomy. My master thesis was awarded in 2015 the *"IBM Research Prize in Computational Science"*.

Fall 2014 **A Matlab Pipeline for Radio Astronomy** — IBM Research, Zurich
Research Internship (6 months)
Supervisor: Dr. Paul Hurley, CCI, IBM Zurich.
As part of the ASTRON-IBM DOME project, I developed a Matlab GUI application to statistically assess and discriminate among different beamforming and imaging strategies proposed by the DOME WP6 team. The application performs end-to-end processing of the data resulting from modern radio telescopes, and provides a variety of metrics to assess the outputs of the simulations.

awards and scholarships

Fall 2015 **IBM Research Prize in Computational Science** — EPFL, Lausanne
My master thesis was awarded in 2015 the *"IBM Research Prize in Computational Science"*. This prize promotes research in computational sciences and recognizes outstanding projects focused on advanced modelling and simulation methods or their applications to important topics in diverse areas of science and engineering.

2015–2016 **EDIC Scholarship** — EPFL, Lausanne
Based on my outstanding academic results, I was awarded a research scholarship covering the expenses of first year of my PhD.

Spring 2014 **Best Poster Award** — EPFL, Lausanne
As part of my SHS master project, I was awarded the *DH101 Gold award* for the quality of my work.

2013–2015 **EPFL Grant for foreign students** — EPFL, Lausanne
Based on my excellent academic results, I was awarded an EPFL grant for foreign students covering the two years of my Masters.

2009 **Olympiads of Geosciences** — Lycée Thierry Maulnier, Nice
National contest "Olympiad of Geosciences" where I ranked second.

freelance projects

On the behaviour of the LIBOR during the past Financial crisis
Discussion (in french) and time series based analysis of the LIBOR to exhibit any irregular behaviour of the latter during the financial crisis of 2007.

Space debris removal
In the context of the Clean Space One mission, I proposed to the Swiss Space Center a model to understand the growth of the space debris cloud and the risk of collisions.

Data Vizualization & Infographics
I am very interested in data visualization, and I have realized many infographics using the semantic version of Wikipedia, DBpedia. I have also been interested in the analysis and construction of social network graphs, to represent friendship connections on Facebook or LinkedIn.

inventions & patents

As part of my work at IBM, I was main inventor on 3 patents and co-inventor on 5 other patents, on a variety of topics ranging from medical imagery to wireless networks or sensor denoising. Some are listed below:

▶ US10424091B1: *Beamforming for Positron Emission Tomography*,

▶ US20190216432A1: *Method for Ultrasound Analytical Point Spread Function Computation*,

▶ US10386452B2: *Determining Positions of Transducers for Receiving and/or Transmitting Wave Signals*,

▶ US10283834B2: *Methods and apparatuses for versatile beamforming.*,

education

2015–Now **PhD** Candidate — Swiss Federal Institute of Technology (EPFL), Lausanne
Advisors : Pr. Martin Vetterli; Pr. Victor Panaretos; Pr. Paul Hurley.
Computer and Communication Sciences doctoral program
My PhD thesis is in collaboration with IBM Research in Zurich.
Research interests: theory and algorithms for spline-based approximation of spherical fields.

2013–2015 **Master** of Science — Swiss Federal Institute of Technology (EPFL), Lausanne
Major in Applied Mathematics
Specialisation in Statistics and Numerical Analysis, GPA of **5.51/6**
Relevant courses : Signal Processing for communications, Multivariate Statistics, Risk and extreme events modelling, Statistical Theory, Robust and Nonparametric Statistics, Computer Algebra, Isogeometric Analysis, Digital Humanities, Microeconomics.

2011–2013 **Bachelor** of Science — Swiss Federal Institute of Technology (EPFL), Lausanne
Major in Mathematics
Bachelor completed in two years instead of the standard three, GPA of **5.2/6**.
Relevant courses : Algorithms, Time Series, Linear Models, Finite Element Method, Stochastic Processes, Mathematical Modelling of Behaviour, Discrete Optimisation, Discrete Mathematics, Numerical Analysis.

2009–2011 **Classes Préparatoires aux Grandes Écoles** — CIV, Sophia Antipolis
Preparation for national competitive entrance exams to leading French "grandes écoles", specialising in mathematics and physics.

technical experience

Spring 2014 **Semi-Automatic Transcription Tool for Ancient Manuscripts** — EPFL, Lausanne
SHS Master Project (Digital Humanities)
Supervisors: Dr. Andrea Mazzei, DHLAB, EPFL; Pr. Frédéric Kaplan, DHLAB, EPFL.
In this work, I investigated various techniques from the fields of shape analysis and machine learning in order to construct a semi-automatic transcription tool for ancient manuscripts. The average classification accuracy achieved with this technique was 86%. I was awarded the DH101 Gold award for the quality of my work, and my poster was presented at the conference I&C Research Days - "Challenges in Big Data".

Spring 2014 **Statistical Inference in Positron Emission Tomography** — EPFL, Lausanne
Master Semester Project
Supervisors: Mikael Kuusela, SMAT, EPFL; Pr. Victor Panaretos, SMAT, EPFL.
In this project, I investigated statistical algorithms for image reconstruction in the context of positron emission tomography (a medical diagnosis technique).

Fall 2013 **HYDROcontest** — EPFL, Lausanne
Scientific staff in the EPFL team
Supervisors: Pr. Luca Dede', CMCS, EPFL; Pr. Alfio Quarteroni, CMCS, EPFL.
International scientific contest launched by the company HYDROS, aiming to build the fastest, most energy-efficient boat. I was part of the EPFL team composed of 9 laboratories, responsible of the numerical study and optimization of the hydrofoils by means of NURBS and Isogeometric Analysis. Our team scored second at the final race.

Spring 2012 **Orbit design for the CHEOPS mission** — Swiss Space Center, Lausanne
Research assistant
Supervisor: Dr. Anton Ivanov, Swiss Space Center, EPFL.
I designed a 4 years long energy-efficient Lyapunov orbit around L2 for the satellite CHEOPS. The project has now been selected by ESA as the first S-class mission.

communication skills

Press Conference
As part of my work in IBM, I was selected twice by the company to give 30 minute press conferences and present the outcome of my team's work to more than 30 international journalists.

EPFL Student Ambassador
I participated to 3 editions of the CIV students' forum where I was holding a stand to present the 22 programs proposed by EPFL.

references

▶ **Pr. Martin Vetterli,**
Full Professor,
Director of EPFL.
@ **: martin.vetterli@epfl.ch**

▶ **Pr. Victor M. Panaretos,**
Associate Professor,
Head of the Chair of
Mathematical Statistics, EPFL.
@ **: victor.panaretos@epfl.ch**

▶ **Pr. Paul Hurley,**
Professor of Data Science at
Western Sydney University,
Parramatta.
@ **:**
p.hurley@westernsydney.edu.au

more about me

▶ LinkedIn profile ,

▶ My personal webpage ,

▶ My IBM webpage ,

▶ List of publications,

▶ A press interview by IBM,

▶ A video interview (in French),

▶ A press interview (in French),

▶ An article on the EMaHP association published on the first page of the EPFL website.

leadership experience

2012–2018 **EMaHP (EPFL Mathematical Humanitarian Project)** Lausanne
Co-founder and president of the EMaHP association
EMaHP is an association aiming to promote the teaching of mathematics in developing countries. As president of this association I :

- Organized a fundraising of more than 100'000 CHF,
- Developed a series of ludic and interactive mathematical activities,
- Organized a mission in South Africa for 24 participants. During this trip, we went to 6 different towns and traveled more than 4'500 km.
- Participated to the 4th edition of the science festival of Oujda.
- Participated to the science festival of Grahamstown in South Africa.

2012–2013 **EPFL Assembly & SB Faculty Council** EPFL, Lausanne
Representative of the Student Body
The EPFL Assembly is entitled to make proposals concerning all the normative acts of the EPF Board. Elected by the Student Body to represent it in those meetings, I was actively involved into the decision making process and influenced it to suit the best the students needs.

selected publications

2019 **Sparse Spline Approximation on the Hypersphere by Generalised Total Variation Basis Pursuit**
SIAM Journal on Imaging Sciences (SIIMS) (Under submission)
Author : M. Simeoni (IBM/EPFL)

2019 **Cadzow Plug-and-Play Gradient Descent for Generalized FRI**
IEEE Transactions on Signal Processing (Under submission)
Authors : M. Simeoni (IBM/EPFL), A. Besson (E-Scopics), P. Hurley (IBM)

2019 **DeepWave: A Recurrent Neural-Network for Real-Time Acoustic Imaging**
Neural Information Processing Systems (NeurIPS)
Authors : M. Simeoni (IBM/EPFL), S. Kashani (EPFL), P. Hurley (IBM), M. Vetterli (EPFL)

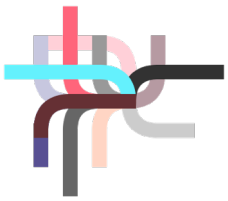
2019 **A Physical Model of Non-stationary Blur in Ultrasound Imaging**
IEEE Transactions on Computational Imaging
Authors : A. Besson (EPFL), L. Roquette (EPFL), D. Perdios (EPFL), M. Simeoni (IBM/EPFL), M. Arditi (EPFL), P. Hurley (IBM), Y. Wiaux (Heriot Watt), J-Thiran (EPFL).

2017 **LEAP: Looking beyond pixels with continuous-space EstimAtion of Point sources.**
Astronomy and Astrophysics
Authors : H. Pan (EPFL), M. Simeoni (IBM/EPFL), P. Hurley (IBM), T. Blu, M.Vetterli (EPFL).

2017 **Flexarray: Random Phased Array Layouts for Analytical Spatial Filtering**
International Conference on Acoustics, Speech, and Signal Processing
Authors : P. Hurley (IBM), M.Simeoni (IBM/EPFL)

2016 **Beamforming towards regions of interest for multi-site mobile networks.**
International Zurich Seminar on Communications
Authors : P. Hurley (IBM), M.Simeoni (IBM/EPFL)

2016 **Flexibeam: Analytic Spatial Filtering by Beamforming.**
International Zurich Seminar on Communications
Authors : P. Hurley (IBM), M.Simeoni (IBM/EPFL)



list of academic publications

Matthieu Simeoni, Lausanne, October 29, 2019

Journal Papers

[1] M. Simeoni, A. Besson, and P. Hurley, "Cadzow plug-and-play gradient descent for generalized FRI," *IEEE Transactions on Signal Processing*, Under Submission.

[2] M. Simeoni, "Sparse spline approximation on the hypersphere by generalised total variation basis pursuit," *SIAM Journal on Imaging Sciences (SIIMS)*, Under Submission.

[3] A. Besson, L. Roquette, D. Perdios, M. Simeoni, M. Arditi, P. Hurley, Y. Wiaux, and J.-P. Thiran, "A physical model of non-stationary blur in ultrasound imaging," *IEEE Transactions on Computational Imaging*, 2019.

[4] H. Pan, M. Simeoni, P. Hurley, T. Blu, and M. Vetterli, "Leap: Looking beyond pixels with continuous-space estimation of point sources," *Astronomy & Astrophysics*, vol. 608, A136, 2017.

Conference Papers

[5] M. M. J.-A. Simeoni, S. Kashani, P. Hurley, and M. Vetterli, "Deepwave: A recurrent neural-network for real-time acoustic imaging," in *Neural Information Processing Systems (NeurIPS), 2019 Thirty-third Conference on*, 2019.

[6] M. Simeoni and P. Hurley, "Graph spectral clustering of convolution artefacts in radio interferometric images," in *Acoustics, Speech and Signal Processing (ICASSP), 2019 IEEE International Conference on*, IEEE, 2019.

[7] L. Roquette, M. Simeoni, and P. Hurley, "A functional framework for ultrasound imaging," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 1837–1841.

[8] L. Roquette, M. M. J.-A. Simeoni, P. Hurley, and A. G. J. Besson, "On an analytical, spatially-varying, point-spread-function," in *2017 IEEE International Ultrasound Symposium (IUS)*, 2017.

[9] M. Gurel, P. Hurley, and M. Simeoni, "Denoising radio interferometric images by subspace clustering," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017.

[10] M. Simeoni and H. Paul, "Sinobeam: Focused Beamforming for PET Scanners," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2017.

[11] P. Hurley and M. Simeoni, "Flexarray: Random phased array layouts for analytical spatial filtering," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, IEEE, 2017.

[12] ——, "On Flexibeam for radio interferometry," in *International BASP Frontiers workshop 2017*, 2017.

[13] M. Simeoni and P. Hurley, "Laplace beamshapes for phased-array imaging," in *International BASP Frontiers workshop 2017*, 2017.

[14] M. Gurel, P. Hurley, and M. Simeoni, "On Denoising Crosstalk in Radio Interferometry," in *International BASP Frontiers workshop 2017*, 2017.

[15] E. Bezzam, R. Scheibler, J. Azcarreta, H. Pan, M. Simeoni, R. Beuchat, P. Hurley, B. Bruneau, C. Ferry, and S. Kashani, "Hardware and software for reproducible research in audio array signal processing," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, IEEE, 2017.

[16] P. Hurley and M. Simeoni, "Flexibeam: Analytic spatial filtering by beamforming," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, IEEE, 2017, pp. 2877–2880.

[17] ——, "Beamforming towards regions of interest for multi-site mobile networks," in *International Zurich Seminar on Communications (IZS 2016)*, 2016.

[18] G. Cherubini, P. Hurley, M. Simeoni, and S. Kazemi, "Iterative image subset scanning for image reconstruction from sensor signals," in *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*, IEEE, 2015, pp. 629–634.

[19] ——, “Imaging in radio interferometry by iterative subset scanning using a modified amp algorithm,” in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, IEEE, 2016, pp. 3326–3330.

[20] M. M. J.-A. Simeoni, “Semi-automatic transcription tool for ancient manuscripts,” in *IC Research Day 2014: Challenges in Big Data*, 2014.

Theses

[21] M. Simeoni, “Towards More Accurate and Efficient Beamformed Radio Interferometry Imaging,” Master Thesis, EPFL, 2015.

[22] ——, “Statistics on manifolds applied to shape theory,” Bachelor Thesis, EPFL, 2013.

Patents

[23] L. Roquette, M. Simeoni, P. Hurley, and K. Sepand, *Efficient computation of spatially varying ultrasound analytical point spread functions*, US Patent App. 15/871,384, 2019.

[24] M. Simeoni, P. Hurley, L. Roquette, and S. Kashani, *Beamforming for tomographic detection systems*, US Patent 10,424,091, 2019.

[25] M. Simeoni, P. Hurley, and G. Merve, *Determining positions of transducers for receiving and/or transmitting wave signals*, US Patent App. 15/480,702, 2018.

[26] G. Merve, M. Simeoni, and P. Hurley, *Reducing noise in phased-array signals from receivers located at different locations*, US Patent App. 15/254,208, 2018.

[27] M. Simeoni, P. Hurley, and G. Cherubini, *Methods and apparatuses for versatile beamforming*, US Patent App. 15/053,662, 2017.

[28] G. Cherubini, P. Hurley, M. Simeoni, and S. Kazemi, *Blind calibration of sensors of sensor arrays*, US Patent App. 14/945,647, 2017.

[29] ——, *Reconstruction using approximate message passing methods*, US Patent App. 14/878,912, 2017.

[30] ——, *Iterative image subset processing for image reconstruction*, US Patent App. 14/878,886, Google Patents, 2017.

Invited Talks

[35] M. Simeoni, “Penalised, sieved maximum likelihood for radio interferometric imaging,” CNRS Workshop GdR 720 ISIS (Information, Signal, Image and ViSion), Paris, November 2018.

[36] ——, “Penalised, sieved maximum likelihood for radio interferometric imaging,” Workshop ”Advanced Statistics for Physics Discovery”, Padova, September 2018.

[37] ——, “Maximum likelihood imaging for radio interferometry,” Seminar, University of Nice, September 2018.

[38] M. Simeoni and P. Hurley, “Bluebild with Flexibeam: A new paradigm for multi-beam observations for the SKA,” Swiss SKA Days, June 2018.

[39] P. Hurley, M. Simeoni, and S. Kashani, “Bluebild: An accurate, flexible imager and its efficient implementation,” Swiss SKA Days, June 2018.

[40] S. Kashani, P. Hurley, and M. Simeoni, “Pypeline : A fast open-source package for advanced imaging with Bluebild,” Swiss SKA Days, June 2018.

[41] M. Simeoni and K. Sepand, “A functional analysis framework for positron emission tomography,” Application Session in EPFL course ”Mathematical Foundations for Signal Processing”, December 2017.

[42] M. Simeoni, “Sieved maximum likelihood for array signal processing,” EPFL Visual Computing Seminar, October 2017.

[43] M. Simeoni and P. Hurley, “Bluebild: A stable, accurate and efficient imager for radio astronomy,” Cosmoclub, CEA, Paris, July 2017.

[44] ——, “Bluebild: A stable, accurate and efficient imager for radio astronomy,” Swiss SKA Days, May 2017.

[45] P. Hurley, M. Simeoni, and S. Kashani, “Bluebild imaging for the SKA and more,” Swiss SKA Days, May 2017.

[46] H. Pan, M. Simeoni, and P. Hurley, “Finding sources in radio astronomy using FRI,” Swiss SKA Days, May 2017.

[47] M. Simeoni and S. Kashani, “Bluebild: A stable, accurate and super-efficient imager for radio astronomy,” Application Session in EPFL course ”Mathematical Foundations for Signal Processing”, December 2016.

[48] M. Simeoni and P. Hurley, “Imaging of Things: From data acquisition to knowledge (and back),” Press Conference at the annual IBM Research Press Day, October 2016.

Supervised Projects

[31] S. Kashani, *See what you hear: A super-resolved, neural network-powered, real-time acoustic camera*, PhD Semester Project, EPFL, 2018.

[32] V. Barros, *Randomized projections for improved sensing and imaging in positron emission tomography*, Master Semester Project, ETH, 2017.

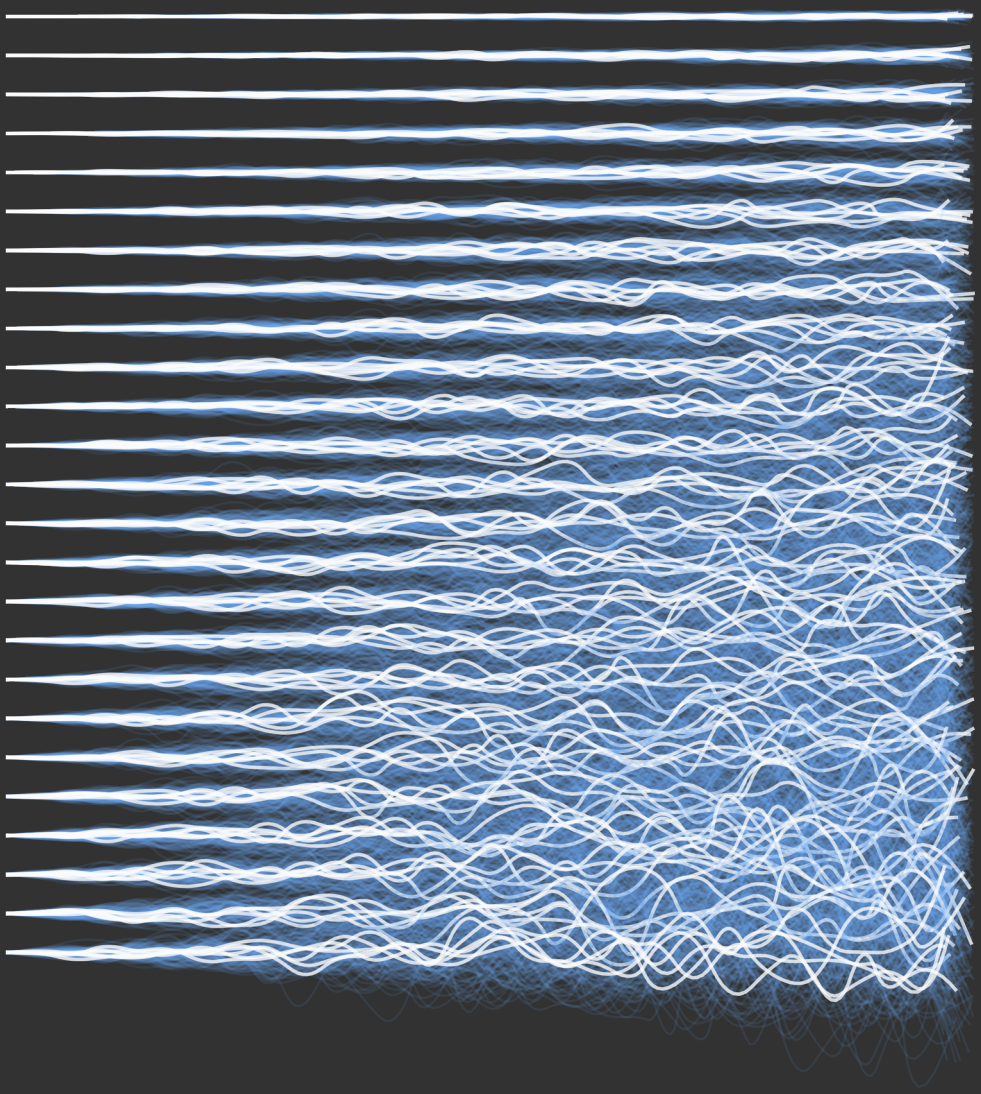
[33] L. Roquette, *A functional framework for enhanced ultrasound imaging*, Master Thesis, EPFL, 2017.

[34] S. Kashani, *Towards real-time high-resolution interferometric imaging with Bluebild*, Master Thesis, EPFL, 2017.

Contributions

[49] V. M. Panaretos, *Statistique pour mathématiciens: Un premier cours rigoureux*. PPUR Presses polytechniques, 2016.

[50] V. M. Panaretos *et al.*, *Statistics for Mathematicians*. Springer, 2016.



*Entia non sunt multiplicanda praeter necessitate**

**More things should not be used than are necessary*