Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies

DISSERTATION
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
Dipl.-Math. Doreen Fischer

eingereicht beim Department Mathematik
der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen
Siegen 2011

1. Gutachter: Prof. Dr. V. Michel
2. Gutachter: Prof. Frederik J. Simons, PhD

Vollzug der Promotion: 21. Juli 2011

Acknowledgements

First of all, I want to thank my supervisor Prof. Dr. Volker Michel for his unquestionable support and his 'open door policy'.
Furthermore, I thank Prof. Dr. Frederik J. Simons, Princeton University, for his interest in my work and his valuable comments.
I want to thank Dr. Arwen Deuss, University of Cambridge, for providing me with the most recent set of splitting function coefficients and for valuable and interesting discussions.
I gratefully acknowledge the financial and scientific support of the Graduate Research Training Program 'Mathematik und Praxis' at the University of Kaiserslautern in the first part of my research and the financial support of the 'Deutsche Forschungsgemeinschaft (DFG)' in the second part of my research.

I am very grateful for the indisputable support and infinite patience of my family.

Abstract

To recover the density of the Earth we invert Newton's gravitational potential. It is a well-known fact that this problem is ill-posed. Thus, we need to develop a regularization method to solve it appropriately.

We apply the idea of a Matching Pursuit (see Mallat and Zhang 1993) to recover a solution stepwise. At step $n+1$, the expansion function d_{n+1} and the weight α_{n+1} are selected to best match the data structure.
One big advantage of this method is that all kinds of different functions may be taken into account to improve the solution stepwise and, thus, the sparsity of the solution may be controlled directly. Moreover, this new approach generates models with a resolution that is adapted to the data density as well as the detail density of the solution.

In the numerical part of this work, we reconstruct the density distribution of the Earth.
For the area of South America, we perform an extensive case study to investigate the performance and behavior of the new algorithm. Furthermore, we research the mass transport in the area of the Amazon where the proposed method shows great potential for further ecological studies, i.e. to reconstruct the mass loss of Greenland or Antarctica.
However, from gravitational data alone it is only possible to recover the harmonic part of the density. To get information about the anharmonic part as well, we need to be able to include other data types, e.g. seismic data in the form of normal mode anomalies. In this work, we will perform such an inversion and present a new model of the density distribution of the whole Earth.

Zusammenfassung

Zur Bestimmung der Dichte der Erde invertieren wir Newtons Gravitationspotential. Da dieses Problem bekanntlich schlecht gestellt ist, entwickeln wir ein geeignetes Regularisierungsverfahren um es zu lösen.

Wir wenden die Idee eines Matching Pursuits (siehe Mallat und Zhang 1993) an. Das heißt, wir bestimmen, der Datenstruktur bestmöglich entsprechend, die Entwicklungsfunktion d und den Koeffizienten α schrittweise. Vorteilhaft ist, dass die unterschiedlichsten Funktionen zur Entwicklung der Lösung beitragen können. Desweiteren erhalten wir Modelle, deren Auflösung an die Datendichte und die Detailstruktur der exakten Lösung angepasst ist.

Wir wenden die Methode auf die Rekonstruktion der Dichteverteilung der Erde (mit Erdinnerem) an.
Für das Gebiet um Südamerika führen wir eine Fallstudie durch, in der wir die Güte und das Verhalten der Methode ausführlich studieren. Desweiteren untersuchen wir den Massentransport im Amazonasgebiet. Die Ergebnisse lassen erwarten, dass die Methode auch für andere ökologisch relevante Problemstellungen, wie die Untersuchung des Massenverlustes in Grönland und der Antarktis, geeignet ist.
Allerdings kann man aus Gravitationsdaten nur den harmonischen Anteil der Dichte rekonstruieren. Das Einbeziehen von seismischen Daten, wie Normal Mode Anomalies, erlaubt es, auch Informationen über den anharmonischen Anteil zu erhalten. In der vorliegenden Arbeit wird das Ergebnis einer solchen Inversion als neues Modell für die gesamte Erde vorgestellt.

Contents

Introduction

Every new satellite mission concerned with the gravitational potential of the Earth allows us to construct more precise models of, e.g. the mass density distribution of the Earth. In a joint inversion with seismic data, e.g. normal mode anomalies or travel times of earthquake waves, we can develop a reasonable model of the density distribution in the interior of the Earth.

Missions like the Gravity Recovery and Climate Experiment (GRACE) (see [85]) allow us to reconstruct temporal variations of the mass density distribution, too. Thus, it is possible to get a more global overview over climate changes than with conventional Earth-bound methods.

Nonetheless, satellite missions can only measure the gravitational potential and not the density distribution itself. Newton's Law states the link between the gravitational potential V and the density distribution ρ as

$$V(x) = \gamma \int_{\mathcal{B}} \frac{\rho(y)}{|x-y|} \mathrm{d}y,$$

where $\mathcal{B}$ is a ball representing the Earth, $x \in \mathbb{R}^3 \setminus \mathcal{B}$ and γ is the gravitational constant. This problem is known as the inverse gravimetric problem and the corresponding data as well as the mentioned seismic data types may be represented as the values of a functional applied to the target function (see, e.g, [18]).

Note that the solution to the inverse gravimetric problem is not unique. In this work, we will use a harmonicity constraint to get a unique solution. Such an a priori condition lacks a physical interpretation but can be motivated by some mathematical arguments. At present, no uniqueness constraint with a satisfactory physical interpretation is known. For further details, see the survey [107].

Needless to say, the inverse gravimetric problem has been solved with a wide variety of methods. Among them are classical methods such as the truncated singular value decompositions (see, e.g. [143, 147]), the domain subdivisions in block-like structures (see, e.g. [74, 93, 96, 145]), and the approximation by point-masses (see, e.g. [74, 127]). Within the last decade, more advanced approximation methods were developed. These are spline methods (see, e.g. [56, 108]) and wavelet methods (see, e.g. [102, 104, 105]). However, most of these methods only allow us to use one kind of predefined basis function on a point grid that is determined

by the given data and, thus, mostly equidistributed. Furthermore, the number of data points is strongly bounded by most methods.

Normal mode anomalies can be observed after major earthquakes as free oscillations of the Earth and give in-depth information about the density distribution of the interior of the Earth. Today, normal modes from earthquakes with a surface wave magnitude larger than 6.5 can be recorded easily such that the number of recordings, that are available for normal mode research, grows steadily. Each normal mode defines a unique splitting function. The coefficients of this splitting function are the given data for the inverse problem. Here, we will use the most recently acquired data by Dr. Arwen Deuss, University of Cambridge.

At the beginning of the century, a number of groups have been concerned with the inversion of normal modes to recover a density model of the deep Earth (see, e.g. [83, 92, 100, 121, 122, 126]). In these works, among others, it was discussed very controversially whether the data situation was sufficient to recover an independent model of the density of the Earth. However, recent advances and the observation of some major events, for example the Sumatra earthquake on December 26, 2004, have greatly increased the number of available normal modes and improved their reliability and the accuracy of the coefficient uncertainties. Thus, using normal mode data in inversions for global models is, again, a promising topic in current research.

Note that a joint inversion of gravitational and seismic data has only rarely been realized. In [60, 81], a combination of gravitational data with normal mode data and travel time data was inverted. However, in recent years, research has apparently been more concerned with the joint inversion of only two out of these three data types, see [40, 43, 90, 129, 130] for a combination of travel times and gravitational data, [17, 18, 19, 82, 95] for a combination of normal mode anomalies and gravitational data, and [84, 87] for joint inversions of normal mode anomalies and travel times. Although, we present here a combination of gravitational and normal mode data the developed method is capable to handle a joint inversion of all three considered data types, too, which will be realized in our further research.

One approach to the inversion of gravitational data seems to be especially promising for our case:
Free-positioned point-mass modeling (see [13, 14]) is a method that positions point-masses shortly below the surface of the Earth stepwise to best match the gravitational potential given by the data. In every step, the nonlinear problem is optimized for the position of the point-masses as well as their magnitudes. Thus, a solution is generated that is adapted to the structure shortly below the surface of the Earth. However, as with most nonlinear optimization methods, this one, too, requires initial values for the positions of the point-masses. Moreover, all ear-

lier positions have to be recomputed when adding another point-mass where the positions of the point-masses in the last step give us an initial value for the new optimization. These are considerable difficulties in regard to the computations. Furthermore, the method becomes instable if the positions of the point-masses are too close to each other or if the point-masses are located deeper than about 100 km below the surface. Additionally, a later study (see [28]) showed that the algorithm is very sensitive with regard to inconsistent gravity data types as well as an irregular distribution of data points. Thus, it is difficult to determine whether one reconstructed a discontinuity in the solution or just the result of data inconsistencies.

Although we favor this idea, we develop an essentially different algorithmic approach due to these problems to realize it. Even though we choose the expansion functions to recover the solution stepwise, too, we use an approach that does not require the determination of some initial value for the position. We will present a new method that allows us to use all different kinds of functions imaginable to expand the solution and to include much more data points as data input than commonly used up to now. In this work, we will use some kind of localized kernel functions that are comparable to point-masses in some characteristics. However, the solution generated by our new algorithm will be adapted to the structure of the signal as well as the data structure. Moreover, the iterative selection of the expansion functions differs with respect to the strategy and the type of available functions, since we use localized basis functions similar to splines or wavelets. The ideas for such a method stem from the field of sparse regularization.

Research in sparse regularization is concerned with solving under-determined or ill-conditioned systems of linear equations with respect to the sparsity of the solution. Of course, there exists a rather large number of different approaches. The main research areas seem to be combinatorial algorithms (see, e.g. [69, 70, 71, 72]) and convex relaxation (see, e.g. [22, 27, 35, 57, 91]), to name a few. However, these methods either need a very large number of measurements or they are very slow. Iterative greedy algorithms (see, e.g. [34, 48, 98, 112, 113, 144]) seem to be efficient with respect to the computational effort and the needed number of measurements (see [112]).

For further details on the general Hilbert space setting as well as sparsity with respect to inverse problems, we refer to, e.g. [12, 33, 36, 42, 142].

However, most of these methods require prior knowledge about the sparsity of the solution. In our case, as well as in most applications, we do not possess this knowledge. Furthermore, the bases display different properties on the ball than in the Euclidean setting. As a consequence, existing methods and algorithms cannot be used directly to solve our problem. In this paper, we introduce a new

approach that allows us to exploit the advantages of iterative greedy algorithms on the ball without prior knowledge of the sparsity.

Approximation on the ball differs strongly from the Euclidean setting. One major difference is the set of available basis functions to expand the solution and their behavior with respect to the reconstruction methods.

The most common approach on the sphere is to use an expansion in terms of spherical harmonics which form an orthogonal basis system to represent the data continuously on the whole sphere. However, because of the global character of these functions small local changes of the data lead to changes in all spherical harmonic coefficients. Furthermore, spherical harmonics are strongly limited by data gaps.

Different groups proposed localized basis systems to remedy the disadvantages of spherical harmonics. Slepian functions, for example, are a locally and globally orthogonal system of functions on the sphere that is optimally suited for a local reconstruction in areas of interest to minimize the effects of data gaps (see [2, 3, 109, 131, 132, 133, 134, 156] for theoretical results and applications). However, this concept is, up to now, limited to the sphere.

The Geomathematics Groups at the University of Kaiserslautern and the University of Siegen developed space localizing kernel functions on the sphere as well as the ball. They, too, allow us to minimize the effects of data gaps or differences in the data density. In this work, we will exploit the localizing character of these functions to reconstruct the detail structure of the density distribution of the Earth. For further theoretical details and applications mostly to geophysical problems, we refer to [7, 18, 55, 56, 63, 64, 66, 88, 102, 108, 114, 148].

This work is divided into four parts: Fundamentals, Matching Pursuit, Numerical Applications and Furthermore.

I. Fundamentals:

In Section 1, we summarize a few important but well-known notations and fundamentals. In Section 2, we give an overview over the involved problems, i.e. the inverse gravimetric problem and the normal mode tomography. Furthermore, we present functional representations for both problems.

II. Matching Pursuit:

Section 3 is concerned with the development of the Functional Matching Pursuit (FMP) as well as some theoretical results regarding the convergence of the algo-

rithm and its convergence rate.

We develop an adaptive and iterative greedy algorithm to solve approximation problems where the data is given directly by a linear and continuous operator, i.e. $\mathcal{F}F = y$ where $y \in \mathbb{R}^l$ is the given data, $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ is the operator corresponding to the inverse problem and $F \in \mathrm{L}^2(\mathcal{B})$ is the unknown target function. Furthermore, this algorithm will not depend on prior knowledge of the sparsity of the target function F.

We intend to approximate the solution F by a (theoretically infinite) expansion $F = \sum_{k=1}^{\infty} \alpha_k d_k$ where each d_k is an element of the dictionary $\mathcal{D}$ and α_k is the associated scalar coefficient. Note that the dictionary is the collection of all functions that may be used in the expansion of the solution. Starting with $F_0 = 0$, we iteratively proceed as follows.

At step $n+1$, the expansion function d_{n+1} and the corresponding coefficient α_{n+1} are selected to best match the signal structure, i.e. to minimize the approximation error $||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l}$ where $R^n := y - \mathcal{F}F_n$ is the residual and $F_n = \sum_{k=1}^{n} \alpha_k d_k$ is the approximation to the target function F that was recovered in the preceding step.

The algorithm can be divided into a preprocessing part that can be parallelized and the main part where, in every step, we just need to search for the optimal dictionary element and update the residual.

Since the inverse gravimetric problem is an ill-posed problem we need to enhance this algorithm to become a regularization method. This process is explained in Section 4 where the Regularized Functional Matching Pursuit (RFMP) is developed and theoretical results with respect to the convergence of the regularized solution are given. Moreover, the main requirements for a regularization method, i.e. the existence and stability of the solution and the convergence, are addressed there as well.

III. Numerical Applications:

Our novel method has the advantage over spline and wavelet methods that different types of functions may be combined to reconstruct the target function. In our numerical applications, we will use the $\mathrm{L}^2(\mathcal{B})$-basis functions $G^{\mathrm{I}}_{m,n,j}, m, n \in \mathbb{N}_0, j = 1, \ldots, 2n+1$, (see [11, 50, 66, 102, 106]) to reconstruct global trends. The kernel functions $K^{\mathrm{I}}_h(x, \cdot), h \in]0, 1[, x \in \mathcal{B}$, (see [5, 7, 17, 19, 105, 106]) which are hat-functions centered at x, where h correlates with the hat-width, are localized functions and will be used to reconstruct detail structures of the solution. Note the similarities of these kernel functions to the point-masses that were used in

free-positioned point-mass modeling. The collection of all functions that may be chosen as functions in the linear expansion of the signal is called a dictionary and will be denoted by $\mathcal{D}$.

Note that the character of the method allows us to use a much denser grid for the centers of the localized kernel functions than in already existing methods to reconstruct the mass density variation of the Earth. This corresponds to the sparsity of the solution. If an equidistributed point grid of centers were chosen with a (global) point density that corresponds to the highest local point density in the irregular grid of centers produced by our method then this grid would have essentially more grid points. This, however, is associated to a very high number of spline basis functions. As a consequence, the system of linear equations required to determine the spline would have a matrix (which is usually ill-conditioned and dense) which cannot be handled numerically due to its size. The new algorithm, however, allows us to gain such a high resolution in regions where this is reasonable without having to solve systems of linear equations.

In Section 5, we will reconstruct the mass density variations of South America out of gravitational data, in particular, and of the whole Earth out of a combination of gravitational and seismic data in the form of normal mode anomalies. We will use the Earth Gravitational Model 2008 (EGM2008) developed by the National Geospatial Intelligence Agency (NGA) (see [117]) as gravitational data input. As data input for the normal mode anomalies we will use splitting function coefficients provided by Dr. Arwen Deuss, University of Cambridge. We discuss the properties of the new method extensively in a case study of the density distribution of South America out of gravitational data only.

In Section 6, we will concern ourselves with the reconstruction of the mass transport in the Amazon area for the year 2008. Here, we will use the monthly solutions collected by the Gravity Recovery and Climate Experiment (GRACE) satellite mission and preprocessed by the Jet Propulsion Laboratory (JPL, see [85]).

IV. Furthermore:

At last we want to summarize this work in Section 7 and give an outlook to further research opportunities in Section 8.

I.

Fundamentals

In this part, we introduce two known L^2-basis systems on the ball as well as a basis system on the ball in form of localized kernel functions for which we show a new property.

Furthermore, we will remind you of some ideas from functional analysis, inverse problem theory, constructive approximation and Sobolev theory.

Moreover, we want to give brief basic introductions to the considered geophysical problems, i.e. inverse gravimetry and normal mode tomography and discuss the considered data.

1. Preliminaries

In this section, we will summarize some basic concepts that are needed to understand the ideas introduced in this work.

1.1. Basic Settings and Spherical Geometry

We denote the set of all positive integers by $\mathbb{N}$, the set of all non-negative integers by $\mathbb{N}_0$, the set of all real numbers by $\mathbb{R}$, and the set of all positive real numbers by $\mathbb{R}^+$.

The three-dimensional Euclidean space is denoted by $\mathbb{R}^3$ with the canonical orthonormal basis $\varepsilon^1 = (1\,,0\,,0)^{\mathrm{T}}$, $\varepsilon^2 = (0\,,1\,,0)^{\mathrm{T}}$ and $\varepsilon^3 = (0\,,0\,,1)^{\mathrm{T}}$. Thus, we can represent every element $x \in \mathbb{R}^3$ in Cartesian coordinates as

$$x = \sum_{i=1}^{3} x_i\, \varepsilon^i = (x_1, x_2, x_3)^{\mathrm{T}} \ .$$

We use the following - usual - definitions in $\mathbb{R}^3$:

$$\begin{aligned}
&\text{inner product} && : && x \cdot y := x^{\mathrm{T}} y = x_1\, y_1 + x_2\, y_2 + x_3\, y_3 \ , \\
& && && x^2 := x \cdot x = x_1^2 + x_2^2 + x_3^2 \ , \\
&\text{norm} && : && |x| := \sqrt{x^2} = \sqrt{x_1^2 + x_2^2 + x_3^2} \ , \\
&\text{vector product} && : && x \wedge y := \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix} , \\
&\text{Laplace operator} && : && \Delta_x := \nabla_x \cdot \nabla_x = \sum_{i=1}^{3} \left(\frac{\partial}{\partial x_i} \right)^2 ,
\end{aligned}$$

where the last two are the representations of the aforementioned operators in Cartesian coordinates of $\mathbb{R}^3$.

Let the unit sphere in $\mathbb{R}^3$ be denoted by $\Omega := \{\xi \in \mathbb{R}^3 |\ |\xi| = 1\}$. Then every element $x \in \mathbb{R}^3$ can be written in the form $x = r\xi$ where $r = |x| \in \mathbb{R}_0^+$ and $\xi \in \Omega$. This representation is unique if $x \neq 0$. Moreover, $x \in \mathbb{R}^3$ can also be represented

in polar coordinates

$$x(r,\varphi,t) = \begin{pmatrix} r\sqrt{1-t^2}\cos\varphi \\ r\sqrt{1-t^2}\sin\varphi \\ rt \end{pmatrix}, \text{ i.e. } x(r,\varphi,\vartheta) = \begin{pmatrix} r\sin\vartheta\cos\varphi \\ r\sin\vartheta\sin\varphi \\ r\cos\vartheta \end{pmatrix},$$

where $r = |x| \in \mathbb{R}_0^+$ denotes the distance to the origin, $\varphi \in [0, 2\pi[$ denotes the longitude, $t = \cos\vartheta \in [-1, 1]$ denotes the polar distance and $\vartheta \in [0, \pi]$ denotes the latitude. Thus, now we may represent the space $\mathbb{R}^3$ in another local basis

$$\varepsilon^r(\varphi,t) = \begin{pmatrix} \sqrt{1-t^2}\cos\varphi \\ \sqrt{1-t^2}\sin\varphi \\ t \end{pmatrix}, \ \varepsilon^\varphi(\varphi) = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix}, \ \varepsilon^t(\varphi,t) = \begin{pmatrix} -t\cos\varphi \\ -t\sin\varphi \\ \sqrt{1-t^2} \end{pmatrix}$$

which is a local tripod on the sphere. Note that $\varepsilon^r \wedge \varepsilon^\varphi = \varepsilon^t$.

With the help of this basis we can give another representation of the gradient ∇_x as well as the Laplace operator Δ_x where these are decomposed into their radial and angular parts, respectively,

$$\begin{aligned} \nabla_x &= \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^* , \\ &\quad \text{where } \nabla^* = \varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} \frac{\partial}{\partial\varphi} + \varepsilon^t \sqrt{1-t^2}\, \frac{\partial}{\partial t} \text{ is the surface gradient and} \\ \Delta_x &= \left(\frac{\partial}{\partial r}\right)^2 + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta^* , \\ &\quad \text{where } \Delta^* = \frac{\partial}{\partial t}(1-t^2)\frac{\partial}{\partial t} + \frac{1}{1-t^2}\left(\frac{\partial}{\partial\varphi}\right)^2 \text{ is the Beltrami operator.} \end{aligned}$$

Let $D \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^m$, $n, m \in \mathbb{N}$. Then $\mathrm{C}^{(k)}(D, W)$ denotes the set of all k-times continuously differentiable functions $F : D \to W$ where $k \in \mathbb{N}_0 \cup \{\infty\}$. If D is additionally compact we write for $F \in \mathrm{C}(D, W)$

$$\|F\|_\infty := \|F\|_{\mathrm{C}(D,W)} := \max_{x\in D} |F(x)|.$$

Then, $(\mathrm{C}(D, W), \|\cdot\|_\infty)$ is a Banach space.

Let now $D \subset \mathbb{R}^m$ be measurable. $\mathrm{L}^p(D, \mathbb{R}^n)$ denotes the set of all equivalence classes of almost everywhere identical functions $F : D \to \mathbb{R}^n$ with $\int_D |F(x)|^p \mathrm{d}x < \infty$ where $1 \le p < \infty$. Note that $\mathrm{L}^p(D, \mathbb{R}^n)$ equipped with the norm

$$\|F\|_{\mathrm{L}^p(D,\mathbb{R}^n)} := \left(\int_D |F(x)|^p \mathrm{d}x\right)^{1/p}$$

is a Banach space.

Let us take a look at $p = 2$ in particular. Then $\mathrm{L}^2(D, \mathbb{R}^n)$ equipped with the scalar product

$$\langle F, G \rangle_{\mathrm{L}^2(D,\mathbb{R}^n)} := \int_D F(x) \cdot G(x) \, \mathrm{d}x, \;\; F, G \in \mathrm{L}^2(D, \mathbb{R}^n),$$

is a Hilbert space.

Let $M \subset X$ where $(X, \|\cdot\|_X)$ is a normed space. Then the closure of M with respect to $\|\cdot\|_X$ is denoted by $\overline{M}^{\|\cdot\|_X}$.

Let a function $G : [-1, 1] \to \mathbb{R}$ and $\xi \in \Omega$ be given. Then a function G_ξ mapping from Ω to $\mathbb{R}$ with $G_\xi(\eta) := G(\xi \cdot \eta)$ for all $\eta \in \Omega$ is called a ξ-zonal function on Ω.

Since $|\xi - \eta|^2 = 2 \, (1 - \xi \cdot \eta)$, $\xi, \eta \in \Omega$, a ξ-zonal function only depends on the spherical distance to ξ.

Moreover, we get that $G \in \mathrm{L}^2([-1, 1])$ satisfies

$$\int_\Omega G_\xi(\eta) \, \mathrm{d}\omega(\eta) = 2\pi \int_{-1}^{1} G(t) \, \mathrm{d}t$$

for all $\xi \in \Omega$.

1.2. Orthogonal Polynomials

In this section, we want to recapitulate the main features of Legendre and Jacobi polynomials. A very detailed introduction to orthogonal polynomials in general as well as to the special cases examined here can be found in [138]. Furthermore, we will introduce orthonormal systems on the sphere as well as on the ball.

1.2.1. Legendre Polynomials

A very special but widely known and used case of orthogonal polynomials are the so-called Legendre polynomials.

Theorem 1.1 (Definition of the Legendre Polynomials)
There exists one and only one system of polynomials $\{P_n\}_{n\in\mathbb{N}_0}$ *called the* Legendre polynomials*, that satisfies the following conditions:*

(i) P_n *is a polynomial of degree* n *defined on* $[-1,1]$*,*

(ii) $\int_{-1}^{1} P_n(t)P_m(t)\ \mathrm{d}t = 0$ *for all* $n, m \in \mathbb{N}_0$ *with* $n \neq m$ *and*

(iii) $P_n(1) = 1$ *for all* $n \in \mathbb{N}_0$*.*

Let us give an alternative definition of the Legendre polynomials.

Definition 1.2 (Rodriguez's Formula)
The Legendre polynomials $\{P_n\}_{n\in\mathbb{N}_0}$ *are given by* Rodriguez's formula

$$P_n(t) = \frac{1}{2^n n!}\ \frac{\mathrm{d}^n}{\mathrm{d}t^n}(t^2-1)^n$$

for $n \in \mathbb{N}_0$ *and* $t \in [-1,1]$*.*

Furthermore, we have certain recurrence formulae

$$\begin{aligned} P'_{n+1}(t) &= tP'_n(t) + (n+1)P_n(t), \\ (t^2-1)P'_n(t) &= ntP_n(t) + nP_{n-1}(t) \text{ and} \\ (n+1)P_{n+1}(t) &= (2n+1)tP_n(t) - nP_{n-1}(t) \end{aligned}$$

for all $n \geq 2$ where $P_0(t) = 1$ and $P_1(t) = t$ for all $t \in [-1,1]$.

Note that we get the following result for the norm.

Lemma 1.3
The $\mathrm{L}^2([-1,1])$*-norm of the Legendre polynomial* P_n *is given by*

$$||P_n||^2_{\mathrm{L}^2([-1,1])} = \int_{-1}^{1} (P_n(t))^2\ \mathrm{d}t = \frac{2}{2n+1}$$

for all $n \in \mathbb{N}_0$*.*

Now let us recapitulate the Clenshaw algorithm, as introduced and proven in [29]. This recursive method is a powerful and efficient feature to numerically evaluate linear combinations of certain polynomials which will be of great use in our computations. Note that the Legendre polynomials fulfill the requirement of the Clenshaw algorithm because of the third recurrence formula.

Theorem 1.4 (Clenshaw Algorithm)
Let a system of one-dimensional functions T_k satisfy the linear second order recurrence relation

$$T_k(x) - a_k(x)\ T_{k-1}(x) - b_k(x)\ T_{k-2}(x) = 0$$

where $k = 2, \ldots, N$, $N \in \mathbb{N}$ with $N \geq 2$. $T_0(x) \neq 0$ and $T_1(x)$ are given. Moreover, let $b_k(x) \neq 0$ for all $k = 2, \ldots, N$.

Then the sum

$$S_N(x) = \sum_{k=0}^{N} A_k(x)\ T_k(x)$$

can be calculated by

$$S_N(x) = (A_0(x) + b_2(x)\ U_2(x))\ T_0(x) + U_1(x)\ T_1(x)$$

where

$$\begin{aligned} U_{N+1}(x) &:= U_{N+2}(x) := 0, \\ U_k(x) &:= a_{k+1}(x)\ U_{k+1}(x) + b_{k+2}(x)\ U_{k+2}(x) + A_k(x) \text{ for } k = N, \ldots, 1. \end{aligned}$$

Moreover, we have explicit representations for particular series of Legendre polynomials.

Lemma 1.5
The identities

$$\sum_{n=0}^{\infty} P_n(t)\ h^n = \frac{1}{\sqrt{1 + h^2 - 2ht}}$$

and

$$\sum_{n=0}^{\infty} (2n+1)\ P_n(t)\ h^n = \frac{1 - h^2}{(1 + h^2 - 2ht)^{3/2}}$$

hold for all $t \in [-1, 1]$ and $h \in]-1, 1[$.

Lastly, let us introduce the associated Legendre function of degree n and order k as

$$P_n^k(t) := (1 - t^2)^{k/2} \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k P_n(t),\ t \in [-1, 1]. \tag{1.1}$$

For further details on this topic we refer to [75] and [78].

1.2.2. Jacobi Polynomials

Let us define the Jacobi polynomials $P_n^{(\alpha,\beta)}$ as follows. Clearly, the Legendre polynomials $P_n = P_n^{(0,0)}$ are a special case hereof.

Theorem 1.6 (Definition of the Jacobi Polynomials)
There exists one and only one system of polynomials $\{P_n^{(\alpha,\beta)}\}_{n\in\mathbb{N}_0}$, $\alpha, \beta > -1$, *called* Jacobi polynomials, *that satisfies the following conditions:*

(i) $P_n^{(\alpha,\beta)}$ *is a polynomial of degree* n *defined on* $[-1, 1]$,

(ii) $\int_{-1}^{1} (1-t)^\alpha (1+t)^\beta P_n^{(\alpha,\beta)}(t) P_m^{(\alpha,\beta)}(t)\ \mathrm{d}t = 0$ *for all* $\alpha, \beta > -1$ *and all* $n, m \in \mathbb{N}_0$ *with* $n \neq m$ *and*

(iii) $P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n} := \dfrac{\Gamma(n+\alpha+1)}{n!\Gamma(\alpha+1)}$ *for all* $\alpha > -1$ *and all* $n \in \mathbb{N}_0$,

where Γ *is the Gamma function.*

For a more detailed description of the Gamma function and its properties we refer to [1].

In the left-hand plot of Figure 1.1, we display the behavior of the Jacobi polynomials $P_m^{(0,\beta)}$ with different parameters β while $\alpha = 0$. The basis systems of $\mathrm{L}^2(\mathcal{B})$ that we want to use in this work depend on Jacobi polynomials of this type only (see Sections 1.2.4 and 1.3). Due to the definition, the extremum at the boundary $x = -1$ changes for different parameters $\beta > -1$ while it stays the same at the boundary $x = 1$.

On the right-hand side of Figure 1.1, we display the behavior of the Jacobi polynomials $P_m^{(0,\beta)}$ for differing parameters m. For increasing m, the number of zeros increases, too. Secondly, the x-value of the largest zero increases.

Evaluating the integral from Theorem 1.6 (ii) we get

$$\begin{aligned} &\int_{-1}^{1} P_m^{(\alpha,\beta)}(t)\ P_n^{(\alpha,\beta)}(t)\ (1-t)^\alpha\ (1+t)^\beta\ \mathrm{d}t \\ = \ &\frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\ \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)}\ \delta_{mn} \end{aligned}$$

where $\delta_{mn} = \begin{cases} 1 & , \ m = n \\ 0 & , \ \text{otherwise} \end{cases}$ is called the Kronecker-Delta.

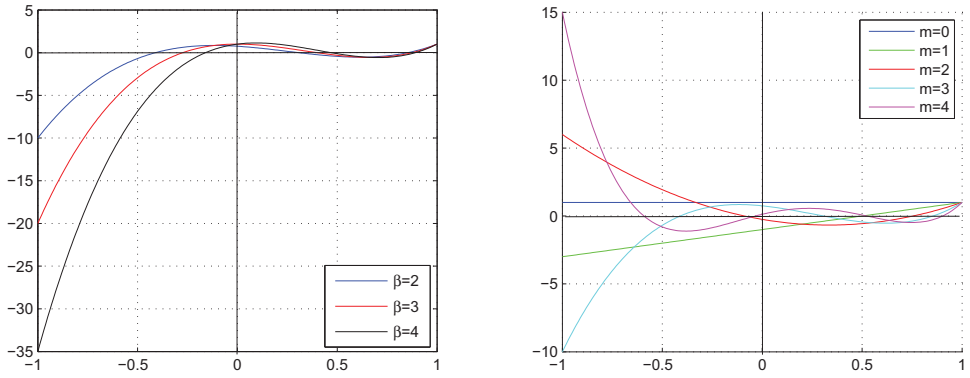

Figure 1.1.: Jacobi polynomials $P_3^{(0,\beta)}$ for $\beta = 2, 3, 4$ (left-hand) and $P_m^{(0,2)}$ for $m = 0, 1, 2, 3, 4$ (right-hand)

In addition, we get the recurrence formulae

$$\begin{aligned} & 2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)P_{n+1}^{(\alpha,\beta)}(t) \\ = \quad & \big[(2n+\alpha+\beta+1)(\alpha^2-\beta^2) \\ & \qquad +(2n+\alpha+\beta+2)(2n+\alpha+\beta+1)(2n+\alpha+\beta)t\big]\, P_n^{(\alpha,\beta)}(t) \\ & -2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2)P_{n-1}^{(\alpha,\beta)}(t) \end{aligned}$$

for all $n \geq 1$ and all $t \in [-1, 1]$ and

$$\begin{aligned} & (2n+\alpha+\beta)(1-t^2)\ \frac{\mathrm{d}}{\mathrm{d}t}\ P_n^{(\alpha,\beta)}(t) \\ = \quad & n\left[\alpha-\beta-(2n+\alpha+\beta)t\right]\ P_n^{(\alpha,\beta)}(t) + 2(n+\alpha)(n+\beta)\ P_{n-1}^{(\alpha,\beta)}(t) \end{aligned}$$

for all $n > 1$ and all $t \in [-1, 1]$, where

$$P_0^{(\alpha,\beta)}(t) = 1 \text{ and } P_1^{(\alpha,\beta)}(t) = \frac{1}{2}(\alpha+\beta+2)t + \frac{1}{2}(\alpha-\beta)$$

for all $t \in [-1, 1]$.

1.2.3. Spherical Harmonics

Spherical harmonics are the restrictions of homogeneous harmonic polynomials to the unit sphere. Thus, let us first explain what a homogeneous harmonic polynomial is and then state some properties of spherical harmonics. A more detailed introduction can be found in [65, 67, 78, 110].

Definition 1.7 (Spherical Harmonic)
Let D be an open and connected subset in $\mathbb{R}^3$.

(a) A function $F \in \mathrm{C}^{(2)}(D)$ is called harmonic *if $\Delta_x F(x) = 0$ for all $x \in D$. The set of all harmonic functions in $\mathrm{C}^{(2)}(D)$ is denoted by* $\mathrm{Harm}(D)$.

(b) A polynomial P on $\mathbb{R}^n$, $n \in \mathbb{N}$, is called homogeneous of degree $m \in \mathbb{N}_0$ *if $P(\lambda x) = \lambda^n P(x)$ for all $x \in \mathbb{R}^3$ and all $\lambda \in \mathbb{R}$. The set of all homogeneous polynomials of degree m on $\mathbb{R}^n$ is denoted by* $\mathrm{Hom}_m(\mathbb{R}^n)$.
The set of their restrictions to a subset $D \subset \mathbb{R}^n$ is defined as

$$\mathrm{Hom}_m(D) := \{P|_D \mid P \in \mathrm{Hom}_m(\mathbb{R}^n)\}.$$

(c) The set of all homogeneous harmonic polynomials on $\mathbb{R}^n$ with degree $m \in \mathbb{N}_0$ is denoted by $\mathrm{Harm}_m(\mathbb{R}^3)$.
Furthermore:

$$\begin{aligned}\mathrm{Harm}_{0\ldots m}(\mathbb{R}^3) &:= \bigoplus_{i=0}^{m} \mathrm{Harm}_i(\mathbb{R}^3),\ m \in \mathbb{N}_0,\\ \mathrm{Harm}_{0\ldots\infty}(\mathbb{R}^3) &:= \bigcup_{i=0}^{\infty} \mathrm{Harm}_{0\ldots i}(\mathbb{R}^3),\end{aligned}$$

and for $D \subset \mathbb{R}^3$

$$\begin{aligned}\mathrm{Harm}_m(D) &= \{P|_D \mid P \in \mathrm{Harm}_m(\mathbb{R}^3)\},\ m \in \mathbb{N}_0,\\ \mathrm{Harm}_{0\ldots m}(D) &= \{P|_D \mid P \in \mathrm{Harm}_{0\ldots m}(\mathbb{R}^3)\},\ m \in \mathbb{N}_0,\\ \mathrm{Harm}_{0\ldots\infty}(D) &= \{P|_D \mid P \in \mathrm{Harm}_{0\ldots\infty}(\mathbb{R}^3)\}.\end{aligned}$$

(d) The elements Y_n of the space $\mathrm{Harm}_n(\Omega)$, $n \in \mathbb{N}_0$, are called the scalar spherical harmonics of degree n.

Note that $\dim \mathrm{Harm}_n(\Omega) = 2n + 1$. Moreover, two scalar spherical harmonics of different degrees are orthogonal in the $\mathrm{L}^2(\Omega)$-sense, i.e.

$$\langle Y_n, Y_m \rangle_{\mathrm{L}^2(\Omega)} = \int_\Omega Y_n(\xi)\ Y_m(\xi)\ \mathrm{d}\omega(\xi) = 0$$

if $n, m \in \mathbb{N}_0$ with $n \neq m$ where $Y_n \in \mathrm{Harm}_n(\Omega)$ and $Y_m \in \mathrm{Harm}_m(\Omega)$.

Moreover, there exist other definitions of spherical harmonics.

Theorem 1.8 (Alternative Definition of Spherical Harmonics)
The spherical harmonics $Y_n \in \mathrm{Harm}_n(\Omega)$, $n \in \mathbb{N}_0$, *are the only infinitely often differentiable eigenfunctions of the Beltrami operator Δ^* corresponding to the eigenvalue $-n(n+1) =: (\Delta^*)^\wedge(n)$, where the sequence $((\Delta^*)^\wedge(n))_{n\in\mathbb{N}_0}$ is called the spherical symbol of Δ^*.*

Now we may define a complete orthonormal system in the space $\mathrm{L}^2(\Omega)$.

Definition 1.9 (Complete Orthonormal System in $\mathbf{Harm}_n(\Omega)$)
For every fixed $n \in \mathbb{N}_0$, the system $\{Y_{n,j}\}_{j=1,\dots,2n+1}$ represents an arbitrary choice of a complete orthonormal system in the space $\mathrm{Harm}_n(\Omega)$ *equipped with the scalar product $\langle \cdot, \cdot \rangle_{\mathrm{L}^2(\Omega)}$, i.e.*

(i) $\langle Y_{n,k}, Y_{n,l} \rangle_{\mathrm{L}^2(\Omega)} = \delta_{kl},\ k, l = 1, \dots, 2n+1$ *and*

(ii) if $F \in \mathrm{Harm}_n(\Omega)$ *satisfies* $\langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)} = 0$ *for all* $j = 1, \dots, 2n+1$ *then* $F = 0$.

We call n the degree and j the order of $Y_{n,j}$.

The Fourier coefficients are given as the scalar product of the function with the basis elements of $\mathrm{L}^2(\Omega)$.

Definition 1.10 (Fourier Coefficients in $\mathbf{L^2(\Omega)}$)
Let $F \in \mathrm{L}^2(\Omega)$. Then

$$F^\wedge(n, j) := \langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)}$$

are the Fourier coefficients *of F with respect to $Y_{n,j}, n \in \mathbb{N}_0, j = 1, \dots, 2n+1$.*

Let us consider the relation between spherical harmonics and Legendre polynomials.

Theorem 1.11 (Addition Theorem of Spherical Harmonics)
Let $\{Y_{n,j}\}_{j=1,\dots,2n+1},\ n \in \mathbb{N}_0$, be an $\mathrm{L}^2(\Omega)$-orthonormal system in $\mathrm{Harm}_n(\Omega)$. *Then*

$$\sum_{j=1}^{2n+1} Y_{n,j}(\xi)\ Y_{n,j}(\eta) = \frac{2n+1}{4\pi} P_n(\xi \cdot \eta) \text{ for all } \xi, \eta \in \Omega.$$

In this work, we will use the following system of real fully normalized spherical harmonics for $\varphi \in [0, 2\pi[$ and $t \in [-1, 1]$ which is commonly used in geoscience:

$$Y_{n,k}(\xi(\varphi, t)) := \sqrt{(2 - \delta_{0k})\ \frac{2n+1}{4\pi}\ \frac{(n-|k|)!}{(n+|k|)!}}\ P_n^{|k|}(t) \begin{cases} \cos(k\varphi) & , k \le 0 \\ \sin(k\varphi) & , k > 0 \end{cases}$$

with $n \in \mathbb{N}_0$ and $k = -n, \dots, n$. Note that we get the index $j = 1, \dots, 2n+1$ from above using the index shift $k = j - n - 1$. Remember the associated Legendre function P_n^k (see Equation (1.1)).

Let us consider the global behavior of the spherical harmonics. We can distinguish three types of behaviors:

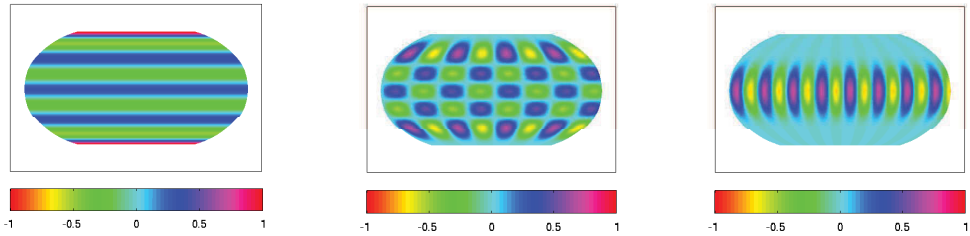

Figure 1.2.: Spherical harmonics $Y_{8,0}$ (left-hand), $Y_{8,4}$ (middle) and $Y_{8,8}$ (right-hand)

In Figure 1.2, we display the fully normalized spherical harmonic $Y_{8,0}$ on the left-hand side. It displays the typical behavior of a zonal function which is a particular feature of all spherical harmonics with $k = 0$. This type of spherical harmonics has n colatitudinal nodes and no longitudinal ones. The spherical harmonics with $0 \neq |k| \neq n$, displayed in the middle of Figure 1.2 in the case of $Y_{8,4}$, are called tesseral. They have $n - |k|$ colatitudinal nodes and $2|k|$ longitudinal ones. The spherical harmonics with $|k| = n$, as for example $Y_{8,8}$ on the right-hand side of Figure 1.2, are called the sectorial functions. They have no colatitudinal nodes and $2n$ longitudinal ones.

Note that all spherical harmonics have a global character. From a theoretical point of view, we can attribute that to the fact that spherical harmonics are based upon polynomials. Thus, these functions are determined to approximate global trends on the sphere and help to construct a basis system of $\mathrm{L}^2(\mathcal{B})$ to approximate global trends on the ball, too (see Section 1.2.4). To approximate detail structures on the ball we will use localized kernel functions which will be introduced in Section 1.3.

1.2.4. An Orthonormal System on the Ball

Let us denote the closed ball with radius $a > 0$ by $\mathcal{B} = \{x \in \mathbb{R}^3 | \ |x| \leq a\}$. From now on, $a := 6371$ km denotes the Earth's radius.

Definition 1.12 (Direct Sum, Orthogonal Complement)
Let H, H_1 and H_2 be Hilbert spaces.

(i) The direct sum *of H_1 and H_2 is defined as $H_1 \oplus H_2$ where for all $x \in H_1 \oplus H_2$ exactly one $y \in H_1$ and exactly one $z \in H_2$ exist such that $x = y + z$.*

(ii) Let $V \subset H$ be closed. Then the orthogonal complement *of V is defined as $V^{\perp_H} := \{x \in H \mid \langle x, v \rangle_H = 0 \text{ for all } v \in V\}$.*

Theorem 1.13 (Decomposition of a Hilbert Space)
Every Hilbert space H can be decomposed orthogonally as $H = V \oplus V^{\perp_H}$ where V is a closed linear subspace of H.

Every function in $\mathrm{L}^2(\mathcal{B})$ can be decomposed uniquely into a harmonic and an anharmonic part (see [102]).

Theorem 1.14 (Decomposition of $\mathrm{L}^2(\mathcal{B})$)
The space $\mathrm{L}^2(\mathcal{B})$ can be decomposed orthogonally as

$$\mathrm{L}^2(\mathcal{B}) = \mathrm{Harm}(\mathcal{B}) \oplus \mathrm{Harm}(\mathcal{B})^{\perp_{\mathrm{L}^2(\mathcal{B})}}$$

where $\mathrm{Harm}(\mathcal{B})^{\perp_{\mathrm{L}^2(\mathcal{B})}}$ is denoted by $\mathrm{Anharm}(\mathcal{B})$. The elements of $\mathrm{Anharm}(\mathcal{B})$ are called anharmonic.

The following complete orthonormal system in $\mathrm{L}^2(\mathcal{B})$ allows such a decomposition easily.

Theorem 1.15 (Complete ONS in $\mathrm{L}^2(\mathcal{B})$: Type I)
A complete orthonormal system *in the space $\mathrm{L}^2(\mathcal{B})$ is given by*

$$G^{\mathrm{I}}_{m,n,j}(x) := \sqrt{\frac{4m+2n+3}{a^3}}\; P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right)\left(\frac{|x|}{a}\right)^n Y_{n,j}\left(\frac{x}{|x|}\right)$$

where $m, n \in \mathbb{N}_0,\ j = 1, \ldots, 2n+1$ and $x \in \mathcal{B}$.

Theorem 1.16 (Complete ONS in $\mathrm{L}^2(\mathcal{B})$: Type II)
Another complete orthonormal system *in the space $\mathrm{L}^2(\mathcal{B})$ is given by*

$$G^{\mathrm{II}}_{m,n,j}(x) := \begin{cases} \sqrt{\frac{2m+3}{a^3}}\; P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) Y_{n,j}\left(\frac{x}{|x|}\right) & , \text{ if } x \neq 0 \\ 1 & , \text{ if } x = 0 \end{cases}$$

where $m, n \in \mathbb{N}_0,\ j = 1, \ldots, 2n+1$ and $x \in \mathcal{B}$.

Let us refer to [11, 50, 66, 102, 106] for the basis system of type I and to [5, 6, 146] for the basis system of type II.

The complete orthonormal system of type I can be divided into harmonic and anharmonic functions where the functions corresponding to $m = 0$ are the harmonic ones. Thus, we can easily reproduce, for example, only the harmonic part of a function with this orthonormal system. Furthermore, every function is a polynomial and well-defined in $x = 0$.

Spline and wavelet bases constructed from the complete orthonormal system of type II can be divided into an angular and a radial part. Thus, this system is

more convenient in numerical applications. As a disadvantage, the functions are not defined in $x = 0$ for $n > 0$ and we have to set the values of the function to an arbitrarily chosen value. However, these functions display a discontinuity at the source $x = 0$.

In Figure 1.4, we display $G^{\mathrm{I}}_{2,4,2}$ and $G^{\mathrm{II}}_{2,4,2}$ on the x_2-x_3-plane of the ball $\mathcal{B}$, i.e. on an equatorial cut through $\mathcal{B}$ (see Figure 1.3). In the plots concerned with the basis function of type II we can clearly see the discontinuity at $x = 0$. The lower right-hand plot in Figure 1.4 displays $G^{\mathrm{II}}_{2,4,2}$ in the area of the ball that we will later consider in our numerical application when reconstructing the density distribution of the Earth in the mantle area based on data given by normal mode anomalies. Clearly, the difficulties near the center of the ball are of no consequence in our applications.

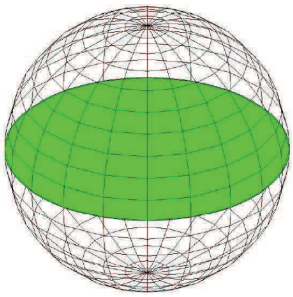

Figure 1.3.: Equatorial cut through the ball $\mathcal{B}$

Let us define the Fourier coefficients with respect to the basis of the space $\mathrm{L}^2(\mathcal{B})$:

Definition 1.17 (Fourier Coefficients in $\mathbf{L^2(\mathcal{B})}$)
Let $F \in \mathrm{L}^2(\mathcal{B})$. Then

$$F^{\wedge}(m,n,j) := \langle F, G_{m,n,j} \rangle_{\mathrm{L}^2(\mathcal{B})}$$

are the Fourier coefficients *of F with respect to $G_{m,n,j}, m, n \in \mathbb{N}_0, j = 1, \cdots, 2n+1$.*

If we write $G_{m,n,j}$ both types I and II can be used. However, the choice has to be fixed for all considerations.

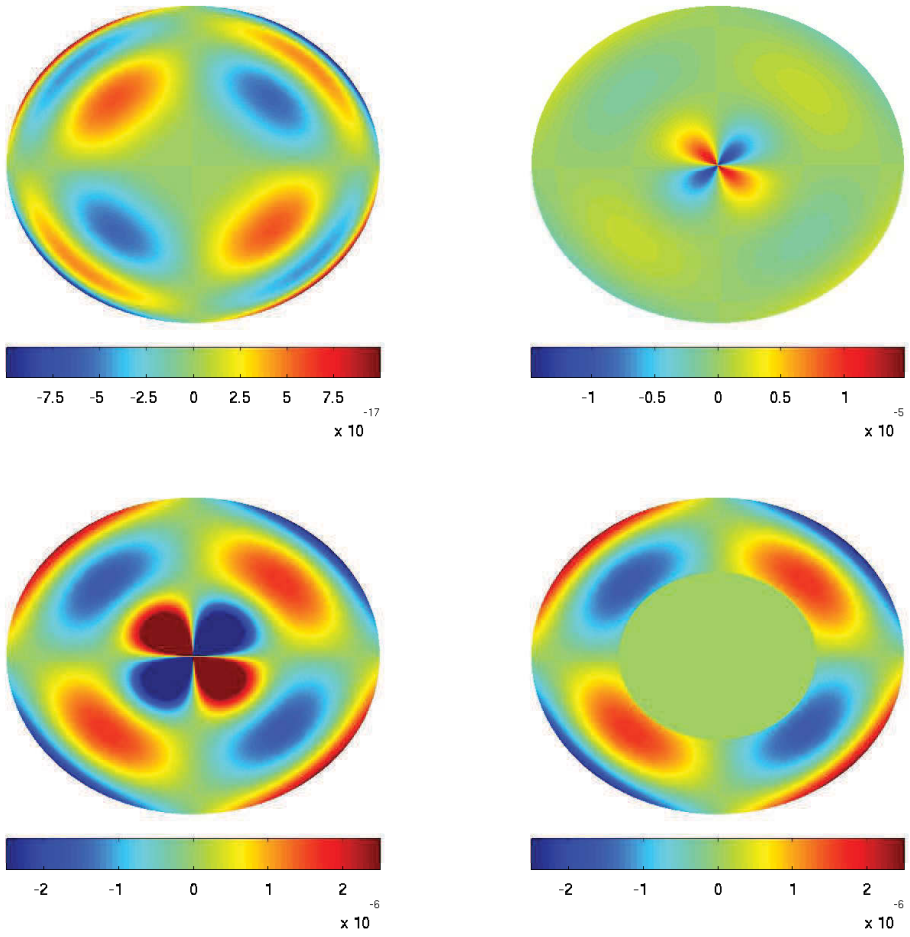

Figure 1.4.: Basis functions $G^{\mathrm{I}}_{2,4,2}$ (upper left-hand) and $G^{\mathrm{II}}_{2,4,2}$ (upper right-hand) in the x_2-x_3-plane (see Figure 1.3); $G^{\mathrm{II}}_{2,4,2}$ with adjusted colorbar (lower left-hand) and plotted in the mantle area only (lower right-hand) in the x_2-x_3-plane

1.3. Kernel Functions

As mentioned before, we use the basis functions of the space $\mathrm{L}^2(\mathcal{B})$, introduced in Section 1.2.4, to reconstruct global trends of the target function. In this section, we introduce localized kernel functions (see [5, 17, 19, 103, 105, 106]) that we will use to reconstruct detail structures of the target function.

Definition 1.18 (Product Kernel)
A function $K \in \mathrm{L}^2(\mathcal{B} \times \mathcal{B})$ is called a product kernel *if it has the form*

$$K(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} K^{\wedge}(m,n)\ G_{m,n,j}(x)\ G_{m,n,j}(y) \tag{1.2}$$

where $x, y \in \mathcal{B}$, provided that all $G_{m,n,j}$ belong to the same type, i.e. type I or type II.

We will mark the kernel functions K^{I} and K^{II} with an index according to the basis functions that we use to construct them. $(K^{\wedge}(m,n))_{m,n\in\mathbb{N}_0}$ is called the symbol of K.

Note that, for numerical reasons, we have to truncate the series, i.e. we truncate the summation over k at E_k. In our applications, we use the Abel-Poisson symbol $K^{\wedge}(m,n) = (A^h_{m,n})^2 = (A^h_m A^h_n)^2$ where we define for fixed $h \in]0,1[$

$$A^h_k = \begin{cases} h^{k/2} & ,\ k \le E_k \\ 0 & ,\ \text{otherwise} \end{cases} ,\ k \in \mathbb{N}.$$

Thus, h is a parameter to influence the localizing character of the kernel function. The hat-width decreases for h getting closer to 1. The peak of $y \mapsto K(x,y)$ is centered at x and the values decrease the larger the distance between x and y is (see Figure 1.5). In this work, we endow the kernel functions with an additional index h to specify the localization property, i.e. we use the notation $K_h(x,\cdot)$ for the kernel function centered at x. For the explicit representation of the kernel functions we refer to Section 3.5.

However, we need to know whether the series in Equation (1.2) converges. Thus, let us state the following result given in [106].

Theorem 1.19 (Convergence of Equation (1.2))
The series in Equation (1.2) converges in the space $\mathrm{L}^2(\mathcal{B} \times \mathcal{B})$ if and only if

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} n\ \left(K^{\wedge}(m,n)\right)^2 < \infty.$$

It converges uniformly on $\mathcal{B} \times \mathcal{B}$ *if*

in the case of type I*:*

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} |K^{\wedge}(m,n)| \ n(2m+n) \ \frac{(n+m+\frac{1}{2})^{2m}}{(m!)^2} < \infty \text{ or}$$

in the case of type II*:*

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} |K^{\wedge}(m,n)| \ nm^5 < \infty.$$

Let us document the localizing character of the kernel functions. In Figure 1.5, we display the localizing character in angular direction for different parameters h for both types K_h^{I} and K_h^{II}. As we mentioned before, the localization property advances with increasing h. However, the kernel function of type I seems to be more localized on the sphere $\partial\mathcal{B}$ than the kernel function of type II.

Furthermore, we can observe a localizing effect in radial direction, too. In Figure 1.6, we display the equatorial plane (i.e. the x_2-x_3-plane, confer Figure 1.3) of the ball $\mathcal{B}$ for $K_h^{\mathrm{I}}(x,\cdot)$ and $K_h^{\mathrm{II}}(x,\cdot)$, $x = (0, 0.5, 0)a$, for different parameters h. Note that the form of the hat of the kernel functions differs.

In Figure 1.7, we display the kernel functions $K_{0.8}^{\mathrm{I}}(x,\cdot)$ and $K_{0.8}^{\mathrm{II}}(x,\cdot)$ at points $x = r\xi$ in the equatorial plane for changing radial distances r. Note the differences of localizing behavior in regard to the local effects near the center x of the kernels.

Using Lemma 1.5, we get another representation of the kernel function of type II:

$$\begin{aligned}
K_h^{\mathrm{II}}(x,y) &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} (A_{m,n}^h)^2 \ G_{m,n,j}^{\mathrm{II}}(x) \ G_{m,n,j}^{\mathrm{II}}(y) \\
&= \left(\sum_{m=0}^{\infty} h^m \ \frac{2m+3}{a^3} \ P_m^{(0,2)}\left(2\frac{|x|}{a} - 1\right) \ P_m^{(0,2)}\left(2\frac{|y|}{a} - 1\right) \right) \\
&\qquad \times \left(\sum_{n=0}^{\infty} h^n \ \frac{2n+1}{4\pi} \ P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right) \right) \\
&= \frac{1-h^2}{4\pi \left(1 + h^2 - 2h\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right)\right)^{3/2}} \\
&\qquad \times \left(\sum_{m=0}^{\infty} h^m \ \frac{2m+3}{a^3} \ P_m^{(0,2)}\left(2\frac{|x|}{a} - 1\right) \ P_m^{(0,2)}\left(2\frac{|y|}{a} - 1\right) \right)
\end{aligned}$$

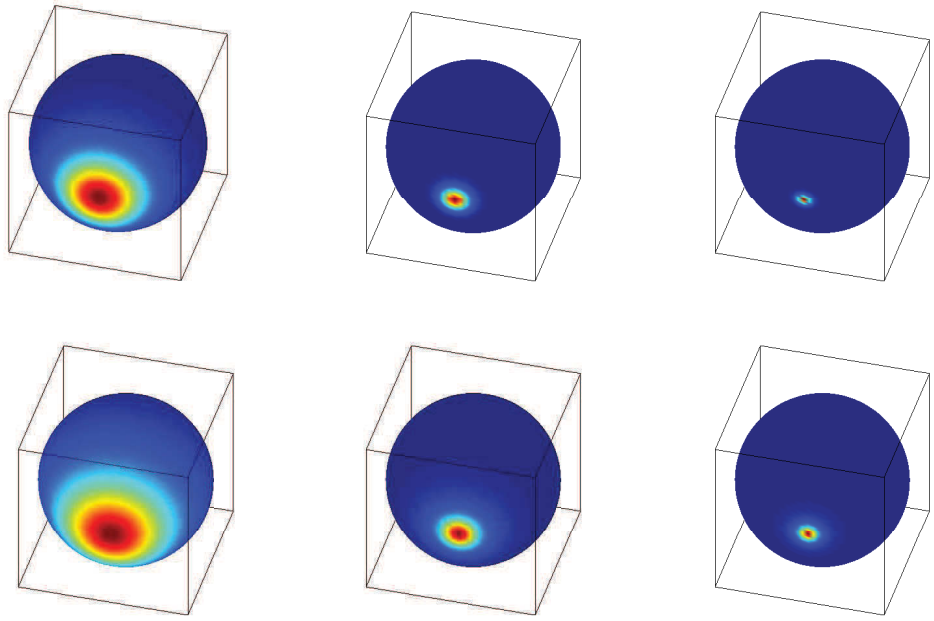

Figure 1.5.: Kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ (top row) and $K_h^{\mathrm{II}}(x,\cdot)$ (bottom row) for $x = (0,-1,0)a$ on the sphere $\partial\mathcal{B}$ with $h = 0.5$ (left-hand), $h = 0.8$ (middle) and $h = 0.9$ (right-hand)

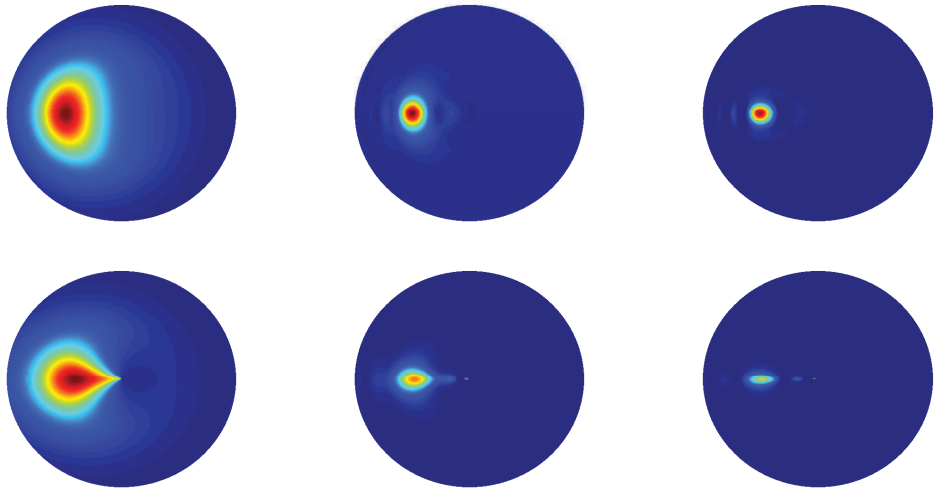

Figure 1.6.: Kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ (top row) and $K_h^{\mathrm{II}}(x,\cdot)$ (bottom row) on the x_2-x_3-plane (see Figure 1.3) for $x = (0,0.5,0)^{\mathrm{T}}a$ with $h = 0.5$ (left-hand), $h = 0.8$ (middle) and $h = 0.9$ (right-hand)

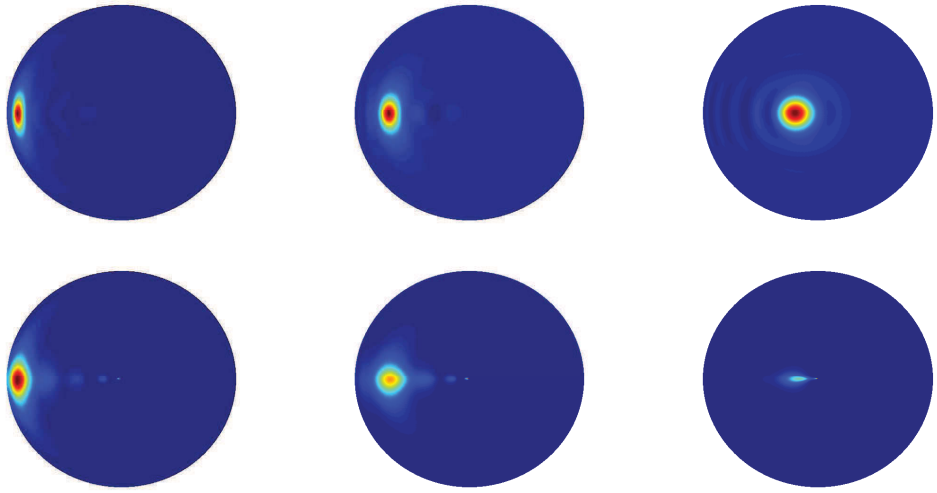

Figure 1.7.: Kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ (top row) and $K_h^{\mathrm{II}}(x,\cdot)$ (bottom row) on the x_2-x_3-plane (see Figure 1.3) with $h = 0.8$ at $x = (0, 0.9, 0)^{\mathrm{T}}a$ (left-hand), $x = (0, 0.7, 0)^{\mathrm{T}}a$ (middle) and $x = (0, 0.2, 0)^{\mathrm{T}}a$ (right-hand)

1.4. Inverse Problems

If we want to obtain the original model (the cause) from observed data (the effect) we deal with an inverse problem. In a mathematical formulation, we get the following definition.

Definition 1.20 (Inverse Problem)
A mathematical model is a mapping $A : X \to Y$ where the set of causes X is mapped to the set of effects Y. If we compute the effect directly from the causes we have a direct problem*, i.e. we compute Ax for $x \in X$. If we try to find the cause from the effects we have an* inverse problem*, i.e. for given $y \in Y$ we find $x \in X$ such that $Ax = y$.*

Hadamard molded the concept of well-posed problems (see [73]).

Definition 1.21 (Well-Posedness (Hadamard))
Let $A : X \to Y$ be a mapping between the topological spaces X and Y. The problem (A, X, Y) is well-posed *if the following three properties are fulfilled:*

(i) The equation $Ax = y$ can be solved for every $y \in Y$.

(ii) This solution is uniquely determined.

(iii) The inverse mapping $A^{-1} : Y \to X$ is continuous, i.e. the solution depends continuously on the data.

If one of these properties is not fulfilled, the problem is ill-posed.

Hadamard believed that mathematical models of physical phenomena should always be well-posed (see [124]). Only incorrect or incomplete models could result in an ill-posed problem. However, this belief is wrong, since most real-world problems are typically ill-posed. Often the third property cannot be fulfilled, for example, if the data is not given exactly but is given by measurements. Even small measurement errors can lead to enormous reconstruction errors if a problem is ill-posed.

If we switch to a Hilbert space setting, we may generalize the concept of the solution of an operator equation such that only item (iii) in Definition 1.21 is critical anymore. Then a problem (A, X, Y) in a Hilbert space setting is ill-posed if the range of A in Y is not closed.

Let us first shortly introduce the more general concept of the Moore-Penrose inverse and list some of its properties. For a more detailed introduction on this topic, we refer to [53, 124].

Definition 1.22 (Moore-Penrose Inverse)
Consider the problem (A, X, Y) between Hilbert spaces.

(i) Let $\mathcal{R}(A) \subset Y$ be the range *of A.*

(ii) $A^*Ax = A^*y$ is called normal equation *where A^* is the* adjoint operator *of A, i.e. $\langle Ax, y\rangle_Y = \langle x, A^*y\rangle_X$ for all $x \in X$ and $y \in Y$.*

(iii) The operator $A^+ : \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp \to X$ mapping each element $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp \subset Y$ to the unique element $x^+ \in \{x \in X \mid A^*Ax = A^*y\}$ with minimimal norm is called generalized inverse *or* Moore-Penrose inverse *of A.*

Theorem 1.23 (Properties of A^+)
The Moore-Penrose inverse fulfills the following properties:

(a) If $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ then $x^+ = A^+y$ is the unique solution of the normal equation in $\mathcal{N}(A)^\perp := \{x \in X \mid Ax = 0\}^\perp$.

(b) $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp = Y$ if and only if the range $\mathcal{R}(A)$ of A is closed.

(c) $\mathcal{R}(A^+) = \mathcal{N}(A)^\perp$

(d) A^+ is continuous if and only if the range $\mathcal{R}(A)$ of A is closed.

Remember the definition of well-posedness by Hadamard (see Definition 1.21): the problem (A, X, Y) is well-posed if the inverse of A exists, is unique and continuous. All three are fulfilled by A^+ if and only if $\mathcal{R}(A) = \overline{\mathcal{R}(A)}$. This justifies an alternative definition of well-posedness by Nashed (see [111]) for the special case of a Hilbert space setting.

Definition 1.24 (Well-Posedness (Nashed))
Let $A : X \to Y$ be a mapping between the Hilbert spaces X and Y. The problem (A, X, Y) is well-posed *if the range of the operator A is closed. Otherwise, it is called* ill-posed.

In order to solve an ill-posed problem we can use a regularization technique to include additional information. Usually, this additional information is a penalty for complexity. In this work, we use a restriction on the smoothness of the solution. A simple form of regularization applied to integral equations, generally termed the Tikhonov regularization, is essentially a trade-off between fitting the data, i.e. reducing the approximation error, and reducing a norm of the solution, i.e. the penalty term.

For further details we refer to, e.g. [53, 124].

1.5. Foundations of Functional Analysis

In this section, we recall some well-known definitions and results from the fields of analysis and functional analysis. We will only introduce those that are needed in later sections. For a more detailed introduction we refer to [4, 76, 157].

First let us state a very important inequality.

Theorem 1.25 (Cauchy-Schwarz Inequality)
Let X be an inner product space. For all $x, y \in X$ the following inequality holds

$$|\langle x, y\rangle_X| \leq ||x||_X ||y||_X$$

where $||\cdot||_X$ is the induced norm of the scalar product $\langle\cdot,\cdot\rangle_X$.

To prove the convergence of the regularized versions of the algorithms that we will develop in this work we need the concepts of weak convergence, strong continuity, weak sequential closedness, and minimum-norm solutions.

Definition 1.26
Let X and Y be Hilbert spaces and $\mathcal{F} : D(\mathcal{F}) \subset X \to Y$ be an operator.

(a) A sequence $(x_n)_n \subset X$ converges weakly *to $x \in X$ if*

$$\langle x_n, y\rangle_X \to \langle x, y\rangle_X \text{ as } n \text{ tends to infinity for all } y \in X.$$

We use the usual notation $x_n \rightharpoonup x$ for weak convergence.

(b) $\mathcal{F}$ is strongly continuous *if $x_n \rightharpoonup x \in X$ implies $\mathcal{F}(x_n) \to \mathcal{F}(x) \in Y$ for n tending to infinity.*

(c) $\mathcal{F}$ is weakly sequentially closed *if we can conclude that $x \in D(\mathcal{F})$ and $\mathcal{F}(x) = y$ for all sequences $(x_n)_n \subset D(\mathcal{F})$ where x_n converges weakly to $x \in X$ and $\mathcal{F}(x_n)$ converges weakly to $y \in Y$ for n tending to infinity.*

(d) F^+ is called a minimum-norm solution *of $\mathcal{F}F = y$ if*

$$||F^+||_X = \min_{\substack{\mathcal{F}F=y \\ F\in D(\mathcal{F})}} ||F||_X.$$

Note that every strongly continuous operator is weakly sequentially closed, too. Moreover, x^+ in Definition 1.22 is, obviously, a minimum-norm solution in the sense of this definition.

Furthermore, let us state that any continuous function on a closed and bounded interval can be uniformly approximated on that interval by polynomials to any degree of accuracy (see [58]).

Theorem 1.27 (Weierstraß Approximation Theorem)
Let F be a continuous real-valued function on the interval $[a, b]$. For every $\varepsilon > 0$ there exists a polynomial P on $[a, b]$ such that $|F(x) - P(x)| < \varepsilon$ for all $x \in [a, b]$.

Moreover, the following two results will be needed (see, e.g. [77, 153]).

Theorem 1.28 (Bolzano-Weierstraß Theorem)
Every bounded sequence in a Hilbert space has a weakly convergent subsequence.

Theorem 1.29
Let $(x_n)_n \subset H$ converge weakly to $x \in H$ for n tending to infinity where H is a Hilbert space. Then

$$||x||_H \leq \liminf_{n\to\infty} ||x_n||_H.$$

The following result is given in [158].

Theorem 1.30
Every linear and compact operator is strongly continuous.

Furthermore, let us recapitulate the requirements for an exchange of a derivative and an integral, i.e. we want to use

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_E f(x,t)\ \mathrm{d}x = \int_E \frac{\partial}{\partial t} f(x,t)\ \mathrm{d}x$$

where $f : E \times I \to \mathbb{R}$ is integrable and I is an interval. For further details on this topic, we refer to [101].

Theorem 1.31 (Differentiation under the Integral Sign)
Let $f : E \times I \to \mathbb{R}$ be summable over the set E for each fixed t in the interval I and let $t_0 \in I$ be a value of t. Let the following conditions be fulfilled:

(i) There exists an integrable function g such that for all $t \in I$ and all $x \in E$

$$\left|\frac{f(x,t) - f(x,t_0)}{t - t_0}\right| \leq g(x).$$

(ii) The partial derivative $\frac{\partial}{\partial t} f(x,t)|_{t=t_0}$ exists for all $x \in E \setminus E_0$ where E_0 is a set of measure 0.

Then the derivative of the integral corresponding to t_0 exists and

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\int_E f(x,t)\ \mathrm{d}x\right)\Big|_{t=t_0} = \int_{E\setminus E_0} \frac{\partial}{\partial t} f(x,t)\Big|_{t=t_0}\ \mathrm{d}x.$$

Note that all integrals in this work are Lebesgue integrals.

1.6. Sobolev Spaces

In this section, we want to define Hilbert spaces of functions on the ball $\mathcal{B}$ such that they have a reproducing kernel and we recover a basis system of the space $\mathrm{L}^2(\mathcal{B})$ consisting of such reproducing kernels (see [106]). For theoretical results on the related version of such spaces defined on the sphere (Sobolev spaces), see [15, 61, 62, 63, 64, 68]. Please remember the definition of a product kernel (see Definition 1.18).

First, let us introduce reproducing kernels.

Definition 1.32 (Reproducing Kernel)
Let H be a Hilbert space of real functions on the subset $D \subset \mathbb{R}^n$. $K_H : D \times D \to \mathbb{R}$ is a reproducing kernel *if*

(a) $K_H(x, \cdot) \in H$ for all $x \in D$ and

(b) $\langle K_H(x, \cdot), F \rangle_H = F(x)$ for all $x \in D$ and all $F \in H$.

For further details on reproducing kernels in general we refer to [39]. For details on the setting of a three-dimensional ball see, e.g. [5, 6, 17].

Definition 1.33 (I/II-Summability)
A real sequence $(K^\wedge(m,n))_{m,n\in\mathbb{N}_0}$ is I-summable *if*

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \left(K^\wedge(m,n)\right)^2 \; n(2m+n) \; \frac{(n+m+\frac{1}{2})^{2m}}{(m!)^2} < \infty.$$

It is called II-summable *if*

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \left(K^\wedge(m,n)\right)^2 \; nm^5 < \infty.$$

Note the similarity to the requirements in Theorem 1.19. From now on, all sequences $(K^\wedge(m,n))_{m,n\in\mathbb{N}_0}$ are assumed to be summable with respect to the corresponding type I or II.

Definition and Theorem 1.34
Let the given (non-trivial) sequence $(K^\wedge(m,n))_{m,n\in\mathbb{N}_0}$ satisfy the respective summability condition. The function F is an element of the Hilbert (Sobolev) space $\mathcal{H} = \mathcal{H}((K^\wedge(m,n)), \mathcal{B})$ if and only if F is function in $\mathrm{L}^2(\mathcal{B})$ with

(i) $\langle F, G_{m,n,j} \rangle_{\mathrm{L}^2(\mathcal{B})} = 0$ for all $m, n \in \mathbb{N}_0$ and $j = 1, \ldots, 2n+1$ with $K^\wedge(m,n) = 0$,

(ii) $$\sum_{\substack{m,n=0 \\ K^\wedge(m,n)\neq 0}}^{\infty} \left(K^\wedge(m,n)\right)^{-2} \sum_{j=1}^{2n+1} \langle F, G_{m,n,j} \rangle^2_{\mathrm{L}^2(\mathcal{B})} < \infty,$$

where the type of the summability condition matches the considered type of basis functions $G_{m,n,j}$.
The scalar product is defined as

$$\langle F_1, F_2 \rangle_{\mathcal{H}} := \sum_{\substack{m,n=0 \\ K^\wedge(m,n)\neq 0}}^{\infty} \left(K^\wedge(m,n)\right)^{-2} \sum_{j=1}^{2n+1} \langle F_1, G_{m,n,j} \rangle_{\mathrm{L}^2(\mathcal{B})} \langle F_2, G_{m,n,j} \rangle_{\mathrm{L}^2(\mathcal{B})}$$

where $F_1, F_2 \in \mathcal{H}$.

In accordance to the spherical case, every element of the Hilbert space $\mathcal{H}$ can be related to a continuous bounded function.

Theorem 1.35 (Sobolev Lemma)
Every space $\mathcal{H}$ in Definition 1.34 is a subspace of $\mathrm{C}(\mathcal{B} \setminus \{0\})$. If we use the basis system of type I*, $\mathcal{H}$ is also a subspace of $\mathrm{C}(\mathcal{B})$.*

As a result of the proof to this theorem, we get for $F \in \mathcal{H}$ the estimate

$$\begin{aligned} & ||F||^2_{\mathrm{C}(\mathcal{B})} \\ \leq & \left(\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (K^{\wedge}(m,n))^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3} \binom{m+n+\frac{1}{2}}{m}^2 \right) ||F||^2_{\mathcal{H}} \end{aligned} \quad (1.3)$$

for type I.

Theorem 1.36 (Reproducing Kernel)
Every $\mathcal{H}((K^{\wedge}(m,n)), \mathcal{B})$ in Definition 1.34 is a reproducing kernel Hilbert space. The corresponding reproducing kernel is the product series associated to the sequence $((K^{\wedge}(m,n))^2)_{m,n \in \mathbb{N}_0}$, i.e.

$$K_{\mathcal{H}}(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \left(K^{\wedge}(m,n)\right)^2 G_{m,n,j}(x) G_{m,n,j}(y)$$

for $x, y \in \mathcal{B}$.

At last, let us state a few density results for the kernel functions of type I. Ultimately, we will show in the third result that the reproducing kernels of type I build a basis in $\mathrm{L}^2(\mathcal{B})$. The ideas of the proofs follow those of the spherical setting. Nonetheless, as far as we know, these results are new.

Theorem 1.37 (Denseness in $\mathcal{H}$)
Let $X \subset \mathcal{B}$ be countable and dense. Then the system $\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}$ is closed (in the sense of the approximation theory) in $(\mathcal{H}, ||\cdot||_{\mathcal{H}})$, i.e.

$$\overline{\mathrm{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}} = \mathcal{H}.$$

Proof.
We know that $\overline{\mathrm{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}}$ is a closed subset of $\mathcal{H}$. Thus, we can decompose the space $\mathcal{H}$ as follows (see Theorem 1.13):

$$\mathcal{H} = \overline{\mathrm{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}} \oplus \left(\overline{\mathrm{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}} \right)^{\perp_{\mathcal{H}}}.$$

To prove the theorem, we need to show that the space on the right-hand side of this direct sum only contains the function that is identical to 0.

Let us start with an arbitrary function

$$F \in \left(\overline{\operatorname{span}\{K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}}\right)^{\perp_{\mathcal{H}}} \subset \mathcal{H} \subset \mathrm{C}(\mathcal{B}).$$

Obviously, we get for all functions $G \in \overline{\operatorname{span}\{K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}}$ that the scalar product vanishes, i.e.

$$\langle F, G\rangle_{\mathcal{H}} = 0.$$

In particular, it follows that $\langle F, K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot)\rangle_{\mathcal{H}} = 0$ for all $x \in X$. Since $K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot)$ is a reproducing kernel we conclude that $F(x) = 0$ for all $x \in X$.

Let us assume that there exists a point $y \in \mathcal{B}$ where $F(y) \neq 0$. Since F is a continuous function we conclude that there exists an open neighborhood U of y where the function F does not vanish. However, X is a dense subset of $\mathcal{B}$ and, thus, there exists a point $x \in X$ where the function F does not vanish either, i.e. $F(x) \neq 0$, which is a contradiction to the explanations above.

Thus, the function F is identical to 0 and we get $\mathcal{H} = \overline{\operatorname{span}\{K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathcal{H}}}$. □

Note that this result is certainly true for the kernel functions of type II, too.

Theorem 1.38 (Denseness in C($\mathcal{B}$))
Let $X \subset \mathcal{B}$ be countable and dense and let $K^{\wedge}(m,n) \neq 0$ for all $m,n \in \mathbb{N}_0$. Then the system $\{K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot) \mid x \in X\}$ is closed in $(\mathrm{C}(\mathcal{B}), ||\cdot||_{\mathrm{C}(\mathcal{B})})$,i.e.

$$\overline{\operatorname{span}\{K_{\mathcal{H}}^{\mathrm{I}}(x,\cdot) \mid x \in X\}}^{||\cdot||_{\mathrm{C}(\mathcal{B})}} = \mathrm{C}(\mathcal{B}).$$

Proof.
First of all we know that

$$\left\{G_{m,n,j}^{\mathrm{I}}\right\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1} \subset \mathcal{H} \subset \mathrm{C}(\mathcal{B}),$$

since $K^{\wedge}(m,n) \neq 0$ for all $m,n \in \mathbb{N}_0$.

Secondly, $\{G_{m,n,j}^{\mathrm{I}}\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1}$ is a basis of the space of all polynomials on $\mathcal{B}$ (see [102] for the anharmonic part where the results for the harmonic part can be shown analogously). With the help of the Weierstraß Approximation Theorem (see Theorem 1.27), we conclude that

$$\overline{\operatorname{span}\left\{G_{m,n,j}^{\mathrm{I}}\right\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1}}^{||\cdot||_{\mathrm{C}(\mathcal{B})}} = \mathrm{C}(\mathcal{B}) \text{ and, thus, } \overline{\mathcal{H}}^{||\cdot||_{\mathrm{C}(\mathcal{B})}} = \mathrm{C}(\mathcal{B}).$$

Now, we conclude for $F \in \mathrm{C}(\mathcal{B})$ and $\varepsilon > 0$ that there exists a function $G \in \mathcal{H}$ such that $||F - G||_{\mathrm{C}(\mathcal{B})} < \frac{\varepsilon}{2}$.

From Theorem 1.37, we know that there exists a function $H \in \operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x, \cdot) \mid x \in X\}$ such that

$$||G-H||_{\mathcal{H}} < \frac{\varepsilon}{2} \left(\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (K^{\wedge}(m,n))^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3} \binom{m+n+\frac{1}{2}}{m}^2 \right)^{-1/2}.$$

With the Sobolev Lemma (see Theorem 1.35) and Inequality (1.3), we conclude that

$$\begin{aligned} ||G-H||_{\mathrm{C}(\mathcal{B})} &\leq \left(\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (K^{\wedge}(m,n))^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3} \binom{m+n+\frac{1}{2}}{m}^2 \right)^{1/2} \\ &\quad \times ||G-H||_{\mathcal{H}} \\ &< \frac{\varepsilon}{2}. \end{aligned}$$

Let us summarize these results:

$$||F-H||_{\mathrm{C}(\mathcal{B})} \leq ||F-G||_{\mathrm{C}(\mathcal{B})} + ||G-H||_{\mathrm{C}(\mathcal{B})} < \varepsilon.$$

Thus, the system $\{K^{\mathrm{I}}_{\mathcal{H}}(x, \cdot) \mid x \in X\}$ is closed in $(\mathrm{C}(\mathcal{B}), ||\cdot||_{\mathrm{C}(\mathcal{B})})$.

□

Theorem 1.39 (Denseness in $\mathrm{L}^2(\mathcal{B})$)
Let $X \in \mathcal{B}$ be countable and dense and let $K^{\wedge}(m,n) \neq 0$ for all $m, n \in \mathbb{N}_0$. Then the system $\{K^{\mathrm{I}}_{\mathcal{H}}(x, \cdot) \mid x \in X\}$ is closed in $(\mathrm{L}^2(\mathcal{B}), ||\cdot||_{\mathrm{L}^2(\mathcal{B})})$, i.e.

$$\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x, \cdot) \mid x \in X\}}^{||\cdot||_{\mathrm{L}^2(\mathcal{B})}} = \mathrm{L}^2(\mathcal{B}).$$

Proof.
First, let us prove that $\overline{\mathrm{C}(\mathcal{B})}^{||\cdot||_{\mathrm{L}^2(\mathcal{B})}} = \mathrm{L}^2(\mathcal{B})$.

Let $\mathcal{B}'$ be an open ball that includes $\mathcal{B}$. All continuous functions F defined on $\mathcal{B}$ can be extended continuously into $\mathcal{B}'$, i.e. by $F(r\xi) := F(a\xi)$ for all $r > a$. Theorem 3.2.2 of [151] yields that $\{\varphi \in \mathrm{C}^{\infty}(\mathcal{B}') \mid \operatorname{supp} \varphi \subset \mathcal{B}' \text{ is compact}\} \subset \mathrm{C}(\mathcal{B}')$ is dense in $\mathrm{L}^2(\mathcal{B}')$. Thus, $\mathrm{C}(\mathcal{B}')|_{\mathcal{B}}$ is dense in $\mathrm{L}^2(\mathcal{B})$, too, where $\mathrm{C}(\mathcal{B}')|_{\mathcal{B}} := \{F|_{\mathcal{B}} \mid F \in \mathrm{C}(\mathcal{B}')\}$. Since any continuous function on $\mathcal{B}'$ is continuous on the subset $\mathcal{B}$ we conclude that $\mathrm{C}(\mathcal{B})$ is dense in $\mathrm{L}^2(\mathcal{B})$.

For all $F \in \mathrm{C}(\mathcal{B})$ we get that

$$||F||_{\mathrm{L}^2(\mathcal{B})} = \sqrt{\frac{4\pi a^3}{3}} ||F||_{\mathrm{C}(\mathcal{B})}.$$

Let $F \in \mathrm{L}^2(\mathcal{B})$ and $\varepsilon > 0$. Since $\mathrm{C}(\mathcal{B})$ is dense in $\mathrm{L}^2(\mathcal{B})$ there exists a function $G \in \mathrm{C}(\mathcal{B})$ such that

$$||F - G||_{\mathrm{L}^2(\mathcal{B})} < \frac{\varepsilon}{2}.$$

From Theorem 1.38, we know that there exists a function $H \in \operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x, \cdot) \mid x \in X\}$ such that

$$||G - H||_{\mathrm{C}(\mathcal{B})} < \frac{\varepsilon}{2\sqrt{\frac{4\pi a^3}{3}}}.$$

As a summary of the above estimates, we conclude that

$$\begin{aligned} ||F - H||_{\mathrm{L}^2(\mathcal{B})} &\leq ||F - G||_{\mathrm{L}^2(\mathcal{B})} + ||G - H||_{\mathrm{L}^2(\mathcal{B})} \\ &\leq ||F - G||_{\mathrm{L}^2(\mathcal{B})} + \sqrt{\frac{4\pi a^3}{3}}||G - H||_{\mathrm{C}(\mathcal{B})} \\ &< \varepsilon \end{aligned}$$

and, thus, we have a basis system of $\mathrm{L}^2(\mathcal{B})$ consisting of reproducing kernel functions of type I.

□

1.7. Point Grids

In our computations, we will use two different types of point grids on the sphere. We will use the equiangular Driscoll-Healy grid, introduced in [49], and the equidistributed Reuter grid, introduced in [123].

Definition 1.40 (Driscoll-Healy Grid)
A Driscoll-Healy grid *on the unit sphere is given by polar coordinates* (φ_i, ϑ_j). *After choosing a control parameter* $\gamma \in \mathbb{N}$ *we have*

(i) $\varphi_i = i\frac{2\pi}{\gamma+1}$ *and*

(ii) $\vartheta_j = j\frac{\pi}{\gamma+1}$

for $i, j = 0, \ldots, \gamma$.

The Driscoll-Healy grid is an equiangular grid with $\gamma^2 + 1$ points. Note that the density of the points is much higher at the poles than at the equator (see left-hand side of Figure 1.8). Moreover, the Driscoll-Healy grid is based on the commonly used geographical grid of the Earth in terms of longitude and latitude which is why data is often given in terms of this grid.

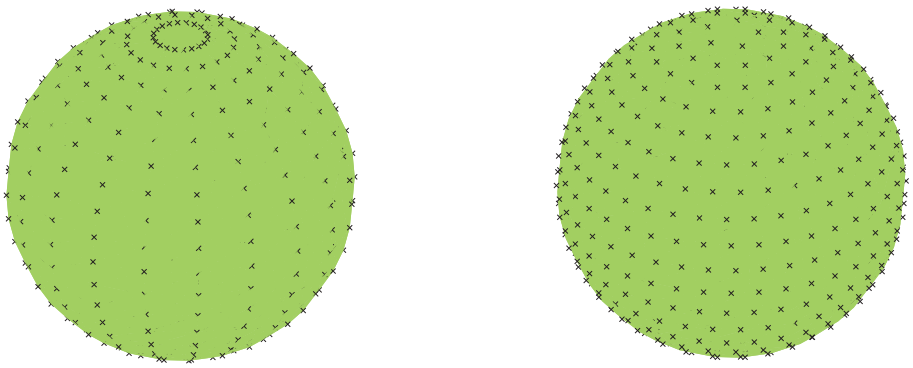

Figure 1.8.: Driscoll-Healy grid (left-hand) and Reuter grid (right-hand) with parameter $\gamma = 20$

Definition 1.41 (Reuter Grid)
A Reuter grid *on the unit sphere is given by the polar coordinates* $(\varphi_{ij}, \vartheta_j)$. *After choosing a control parameter* $\gamma \in \mathbb{N}$ *we have*

(i) $\vartheta_0 = 0$ *and* $\varphi_{01} = 0$ *at the North Pole,*

(ii) $\Delta\vartheta = \frac{\pi}{\gamma}$,

(ii) $\vartheta_i = i\Delta\vartheta$ *for* $1 \leq i \leq \gamma - 1$,

(iv) $\gamma_i = \left\lfloor \frac{2\pi}{\arccos\left(\frac{\cos\Delta\vartheta - \cos^2\vartheta_i}{\sin^2\vartheta_i}\right)} \right\rfloor$, *where the Gauß bracket* $[x]$ *denotes the largest integer less than or equal to* x,

(v) $\varphi_{ij} = \left(j - \frac{1}{2}\right)\frac{2\pi}{\gamma_i}$ *for* $1 \leq j \leq \gamma_i$ *and*

(vi) $\vartheta_\gamma = \pi$ *and* $\varphi_{\gamma 1} = 0$ *at the South Pole.*

The control parameter $\gamma + 1$ gives the number of equidistributed latitudes. The points on each latitude are constructed such that two adjacent points have the same spherical distance as two adjacent latitudes. Thus, we get an equidistributed grid with $N(\gamma) \leq 2 + \frac{4}{\pi}\gamma^2$ points (see right-hand side of Figure 1.8).

2. Basic Problems

We want to reconstruct the mass density distribution of the Earth. It is rather self-evident to consider gravitational data that is assembled by numerous satellite missions. However, we can only reconstruct the harmonic part of the density distribution out of gravitational data (see Section 2.1 and, e.g. [50, 94, 118, 119, 152]). For an approximation of the anharmonic part it is sensible to use seismic data.

However, a joint inversion of gravitational and seismic data has only rarely been realized. In [60, 81], a combination of gravitational data with normal mode data and travel time data was inverted. However, in recent years, research has apparently been more concerned with the joint inversion of only two out of these three data types, see [40, 43, 90, 129, 130] for a combination of travel times and gravitational data, [17, 18, 19, 82, 95] for a combination of normal mode anomalies and gravitational data, and [84, 87] for joint inversions of normal mode anomalies and travel times. Although, we present here a combination of gravitational and normal mode data the presented method is capable to handle a joint inversion of all three considered data types, too, which we will realize in our further research.

In order to fully understand the problem we will give a brief introduction to inverse gravimetry and normal mode tomography. For a more detailed introduction, we refer to [17] and the references therein.

2.1. Inverse Gravimetry

We already mentioned that it is quite obvious to use gravitational data to reconstruct the mass density distribution of the Earth. Newton's Law of Gravitation states the link between the gravitational potential V and the density distribution ρ as

$$V(x) = \gamma \int_{\mathcal{B}} \frac{\rho(y)}{|x-y|} \, \mathrm{d}y, \; x \in \mathbb{R}^3 \setminus \mathcal{B}, \tag{2.1}$$

where γ is the gravitational constant. Note that we can only reconstruct the harmonic part of the mass distribution ρ from the gravitational potential V,

since the operator $T : \mathrm{L}^2(\mathcal{B}) \to T\,(\mathrm{L}^2(\mathcal{B}))$, where

$$(TF)(x) := \int_{\mathcal{B}} \frac{F(x)}{|x-y|}\,\mathrm{d}y \text{ for } x \in \mathbb{R}^3 \setminus \mathcal{B},$$

has the anharmonic subspace of $\mathrm{L}^2(\mathcal{B})$ as its null-space (see [50, 94, 118, 119, 152]).

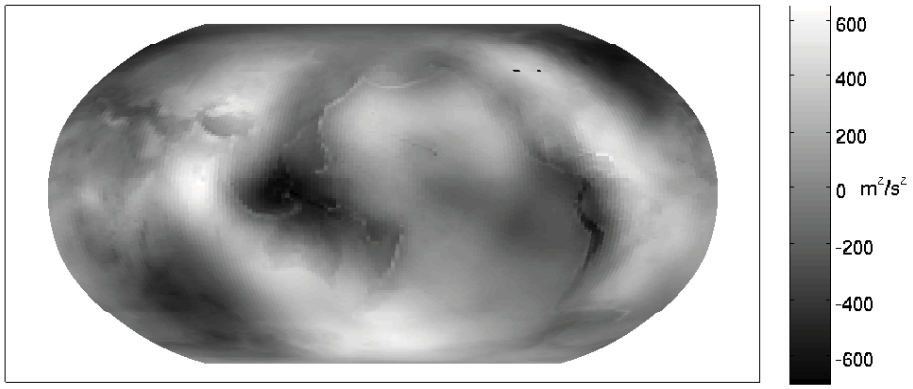

Figure 2.1.: Gravitational potential given by EGM2008 for the whole Earth up to degree 2,190 and order 2,159

In our numerical applications, we use the spherical harmonics model of the Earth's gravitational potential given by the NASA and NIMA joint geopotential model EGM2008 as data input (see [117]). We use EGM2008 from degree 3 up to degree 2,190 and order 2,159 to create a data set that is given pointwise. Out of this data set we want to reconstruct the harmonic part of the density variation at the surface. In our applications, we concentrate on reconstructing the density distribution of South America as a case study. In the Figures 2.1 and 2.2, we display the given data, i.e. the gravitational potential of the whole Earth and South America, in particular, slightly above the surface of the Earth. Clearly, some of the main structures can already be discerned. However, these structures are overlaid with fog that we will eviscerate with our method.

Figure 2.2.: Gravitational potential given by EGM2008 for South America up to degree 2,190 and order 2,159

Let us represent the functionals that map the density to the gravitational potential in terms of both kinds of $\mathrm{L}^2(\mathcal{B})$-basis systems defined in the Definitions 1.15 and 1.16.

In Terms of Type I:

We are interested in the functional for the gravitational potential

$$\mathcal{F}_{\mathrm{G}}^k\ \rho := \int_{\mathcal{B}} \frac{\rho(y)}{|x_k - y|}\ \mathrm{d}y$$

where $\rho \in \mathrm{L}^2(\mathcal{B})$ is mapped to $\mathbb{R}$ and $x_k \in \mathbb{R}^3 \setminus \mathcal{B}$. Furthermore, in [103], a series representation is derived which we use in our considerations:

$$\mathcal{F}_{\mathrm{G}}^k\ \rho = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{4\pi}{2n+1} \sqrt{\frac{a^3}{2n+1}}\ \left\langle \rho, G^{\mathrm{I}}_{0,n,j} \right\rangle_{\mathrm{L}^2(\mathcal{B})}\ \left(\frac{a}{|x_k|}\right)^n\ \frac{1}{|x_k|}\ Y_{n,j}\left(\frac{x_k}{|x_k|}\right).$$

In Terms of Type II:

[19] gives us a representation of the mass density ρ as the unique twice differentiable solution of Equation (2.1) in $\mathcal{B}$ with respect to the constraint that the mass density ρ is harmonic in the interior of $\mathcal{B}$, i.e. $\Delta\rho = 0$:

$$\begin{aligned} \rho \ = \ & \sum_{n=0}^{\infty} \sum_{m=0}^{n} \sum_{j=1}^{2n+1} \frac{2n+1}{4\pi}\ \left(\sum_{l=0}^{n} (2l+3) \left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!} \right)^2 \right)^{-1} \\ & \times \sqrt{\frac{2m+3}{a}}\ \frac{n!(n+2)!}{(n-m)!(n+3+m)!}\ V_{n,j}\ G^{\mathrm{II}}_{m,n,j}. \end{aligned} \tag{2.2}$$

Since the gravitational potential V is harmonic outside the gravitating body and regular at infinity we can expand it in terms of outer harmonics for $x \in \mathbb{R}^3 \setminus \mathcal{B}$:

$$V(x) = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} V_{n,j} \left(\frac{a}{|x|}\right)^{n+1} Y_{n,j}\left(\frac{x}{|x|}\right).$$

If we consider now that the mass density distribution ρ can also be represented in terms of the basis elements

$$\rho = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{j=1}^{2n+1} \langle \rho, G^{\mathrm{II}}_{m,n,j} \rangle_{\mathrm{L}^2(\mathcal{B})} G^{\mathrm{II}}_{m,n,j}$$

we get from Equation (2.2) a representation for $V_{n,j}$, i.e.

$$V_{n,j} = \frac{4\pi}{2n+1}\left(\sum_{l=0}^{n}(2l+3)\left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!}\right)^2\right) \times\sqrt{\frac{a}{2m+3}}\,\frac{(n-m)!(n+3+m)!}{n!(n+2)!}\,\langle\rho, G^{\mathrm{II}}_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}$$

where $n \in \mathbb{N}_0$, $j = 1,\dots,2n+1$ and $m \in \{0,\dots,n\}$.

A harmonic function is given by its values at the surface. Thus, it is for all pairs (n,j) enough to calculate the scalar product $\langle\rho, G^{\mathrm{II}}_{\tilde{m},n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}$ for one $\tilde{m} \in \{0,\dots,n\}$ to obtain $V_{n,j}$ uniquely since the parameter $\tilde{m}$ corresponds exclusively to the radial part of the basis function (see Definition 1.16). This allows us to calculate the other expansion coefficients of the mass density ρ for $m \in \{0,\dots,n\} \setminus \{\tilde{m}\}$ out of $\langle\rho, G^{\mathrm{II}}_{\tilde{m},n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}$.

Let now $\mu(n) \in \{0,\dots,n\}$ be the chosen value for $\tilde{m}$. Then a functional mapping $\rho \in \mathrm{L}^2(\mathcal{B})$ to $V(x_k) \in \mathbb{R}$ can be written as

$$\mathcal{F}^k_{\mathrm{G}}\rho = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\frac{4\pi}{2n+1}\left(\sum_{l=0}^{n}(2l+3)\left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!}\right)^2\right) \times\sqrt{\frac{a^3}{2\mu(n)+3}}\,\frac{(n-\mu(n))!(n+3+\mu(n))!}{n!(n+2)!} \times\langle\rho, G^{\mathrm{II}}_{\mu(n),n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}\,\left(\frac{a}{|x_k|}\right)^n\frac{1}{|x_k|}\,Y_{n,j}\left(\frac{x_k}{|x_k|}\right).$$

Application to Density Deviations:

Since there already exist approximate models of the mass density distribution of the Earth - for example PREM (see [51]) - it is useful to apply the functionals to the deviation $\delta\rho = \rho - \rho_{\mathrm{M}}$ as the difference of the mass density and the density of a reference model.

Thus, we need to calculate the gravitational potential V_δ associated to this deviation. For instance, in a radially symmetric model we have for all $x \in \mathbb{R}^3 \setminus \mathcal{B}$ and the gravitational constant $\gamma = 6.6739 \cdot 10^{-20} \frac{\mathrm{cm}^3}{\mathrm{g} \cdot \mathrm{s}^2}$ that

$$\begin{aligned} V_\delta(x) &= \gamma \int_{\mathcal{B}} \frac{\delta\rho(y)}{|x-y|} \, \mathrm{d}y \\ &= \gamma \int_{\mathcal{B}} \frac{\rho(y) - \rho_{\mathrm{M}}(|y|)}{|x-y|} \, \mathrm{d}y \\ &= V(x) - \gamma \, \frac{4\pi}{|x|} \int_0^a r^2 \rho_{\mathrm{M}}(r) \, \mathrm{d}r. \end{aligned}$$

For PREM as a reference model, we get $V_\delta(x) = V(x) - \gamma \, \frac{4\pi a^3}{3|x|} \, 5.5134 \, \frac{\mathrm{g}}{\mathrm{cm}^3}$ (see [102]).

Note that this is an ill-posed problem.

2.2. Normal Mode Tomography

Free oscillations - or normal modes - of the Earth can be observed after major earthquakes, for example the Sumatra earthquake on December 26, 2004, and give in-depth information about the density structure of the Earth. Most importantly, with this data we may gain information about the anharmonic part of the density as well. Thus, by a joint inversion of normal mode data and gravitational data it is possible to derive an advanced Earth model.

Normal modes are separated into spheroidal modes ${}_kS_l$ and toroidal modes ${}_kT_l$. Here, the index $k \in \mathbb{N}_0$ denotes the overtone of the mode and the index l denotes that the mode is related to the spherical harmonic of degree $l \in \mathbb{N}_0$, i.e. $Y_{l,m}$ where $m = 1, \ldots, 2l+1$. Spheroidal modes produce motion both in the perpendicular direction and parallel to the surface of the Earth. Therefore, they can be seen in both vertical and horizontal components of a seismogram. Only spheroidal modes are both sensitive to the inner core and observable at the surface of the Earth.

In a spherically symmetric, non-rotating, isotropic Earth model (SNREI model) the frequency of the modes is independent of the order m. Thus, we have $2l+1$ modes with the same frequency, i.e. each mode consists of a $(2l+1)$-fold degenerate multiplet, whose singlets all have the same frequency. As a consequence, each mode contributes a single peak to the seismogram.

However, the Earth deviates from the SNREI model, e.g. due to the elliptical shape, lateral heterogeneity or anisotropies within the Earth. These deviations remove the degeneracy of the normal modes such that each mode multiplet is split into a set of $2l+1$ singlets with different frequencies. This means, in reality, we do not observe modes as peaks but as broadened peaks. This effect is called splitting and can be described by a unique splitting function σ. Furthermore, cross-coupling or resonances between different modes change the frequencies of the singlets, too.

For further information on this topic we refer to [32] and [80].

In Figure 2.3, we display some examples of splitting functions (see [41, 80]). We will work with a data set provided by Dr. Arwen Deuss, University of Cambridge. Note that the data is, again, given in spherical harmonic coefficients which we will use as the data input for the inversion in this work.

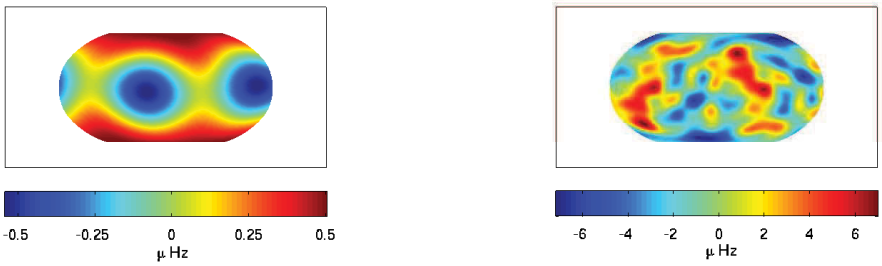

Figure 2.3.: Splitting functions ${}_0S_3$ (left-hand) and ${}_0S_{12}$ (right-hand)

Instead of measuring the splitting function we may measure a shift of the peak frequencies, directly. The representation of splitting functions as frequency shifts has the additional advantage that they can be represented without using spherical harmonics (see, e.g. [86, 135, 154]), i.e. this representation is more useful for local considerations.

As a third option, we may approximate the splitting function coefficients directly and, thus, avoid the problem of locality.

In the following, we present functionals describing these three representations, i.e. the representation by splitting functions, the representation by frequency shifts and the representation by splitting function coefficients. Our computations, however, will be limited to the representation by splitting function coefficients since it is the most time-efficient alternative.

In this work, we will use a spherically symmetric, non-rotating, elastically isotropic Earth model where we have no mode coupling as a reference model, e.g. PREM. Furthermore, we describe the isotropic material properties by the compressional velocity α, the shear velocity β and the mass density ρ.

Splitting Functions

As mentioned above the splitting function only depends on the direction $\xi \in \Omega$. Thus, it can be regarded as a function in $\mathrm{L}^2(\Omega)$, i.e. we can expand an arbitrary splitting function σ as

$$\sigma = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \int_0^a \left(\left(K_n^\alpha, K_n^\beta, K_n^\rho \right) \cdot \delta m^\wedge(n,j) \right)(r) \, \mathrm{d}r \; Y_{n,j}$$

where $\delta m = \left(\frac{\delta\alpha}{\alpha_\mathrm{M}}, \frac{\delta\beta}{\beta_\mathrm{M}}, \frac{\delta\rho}{\rho_\mathrm{M}} \right)$ is the relative deviation and K_n^α, K_n^β and K_n^ρ are sensitivity kernels (see [32]). The reference values α_M, β_M and ρ_M are given in an Earth model, e.g. PREM. Note that

$$\left((\delta m^\wedge(n,j))(r) \right)_i = \langle (\delta m(r\cdot))_i, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)}.$$

Let us derive a functional that maps the deviations to the splitting function:

$$\begin{aligned} &\mathcal{F}_\mathrm{S}^k(\delta\alpha, \delta\beta, \delta\rho) = \sigma(\xi_k) \\ = \; &\sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \int_0^a \left(\left(\frac{K_n^\alpha}{\alpha_\mathrm{M}}, \frac{K_n^\beta}{\beta_\mathrm{M}}, \frac{K_n^\rho}{\rho_\mathrm{M}} \right) \cdot (\delta\alpha, \delta\beta, \delta\rho)^\wedge(n,j) \right)(r) \, \mathrm{d}r \; Y_{n,j}(\xi_k) \end{aligned}$$

where $\xi_k \in \Omega$.

We assume that the material coefficients $(\delta m^\wedge(n,j))(r)$ show a certain decay for n tending to infinity. Thus, the Fourier series is uniformly convergent and we can evaluate it pointwise. Note that, up to now, the data is only given up to degree 12, i.e. the Fourier series is finite anyway.

As a simplification, we assume that the density ρ and the velocities α and β are not independent of each other (see, e.g. [82, 97]):

$$c_\alpha \frac{\delta\alpha}{\alpha_\mathrm{M}} = c_\beta \frac{\delta\beta}{\beta_\mathrm{M}} = \frac{\delta\rho}{\rho_\mathrm{M}}. \tag{2.3}$$

The scaling coefficients c_α and c_β are mostly obtained by empiric studies in the different sciences (see, e.g. [8, 10, 52, 60, 89]). However, using such a simple relation between the density and the velocities is very questionable, for instance, if non-thermal effects result in lateral variations (see [59, 99]). This problem may

be solved by using a vectorial approach (see [17, 18]).

However, using these dependencies, the functional can be simplified to only map the deviation of the density to the splitting function:

$$\mathcal{F}_{\mathrm{S}}^{k}\delta\rho = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\int_{0}^{a}\frac{K_n(r)}{\rho_{\mathrm{M}}(r)}\cdot\left(\delta\rho^{\wedge}(n,j)\right)(r)\ \mathrm{d}r\ Y_{n,j}\left(\xi_k\right)$$

where $K_n = \frac{1}{c_\alpha}K_n^\alpha + \frac{1}{c_\beta}K_n^\beta + K_n^\rho$ and $\xi_k \in \Omega$.

Frequency Shift

Alternatively, we can observe a shift of the peak frequency to avoid the representation in spherical harmonics which is advantageous for a local reconstruction.

Let us denote a frequency shift of a fixed multiplet on the great circle with pole η with

$$\delta\omega(\eta) = \omega(\eta) - \bar{\omega}$$

where $\omega(\eta)$ is the peak frequency and $\bar{\omega}$ is the mean frequency.

If we interpret this shift as a great circle average of the splitting function and use Theorem 2.5 in [17] we get

$$\begin{aligned}\delta\omega(\eta) &= \frac{1}{2\pi}\oint_{\eta}\sigma(\xi)\ \mathrm{d}\Delta(\xi)\\ &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}P_n(0)\int_{0}^{a}\left(\left(K_n^\alpha, K_n^\beta, K_n^\rho\right)\cdot\delta m^{\wedge}(n,j)\right)(r)\ \mathrm{d}r\ Y_{n,j}(\eta).\end{aligned}$$

Note that $P_n(0) = 0$ if and only if n is odd. Thus, only modes with even degree can be used in this approach.

Let us present the functional that maps all deviations $\delta\alpha$, $\delta\beta$ and $\delta\rho$ to the frequency shift function.

$$\begin{aligned}&\mathcal{F}_{\mathrm{F}}^{k}(\delta\alpha,\delta\beta,\delta\rho) = \delta\omega\left(\xi_k\right)\\ =\ &\sum_{\substack{n=0\\ n\ \text{even}}}^{\infty}\sum_{j=1}^{2n+1}P_n(0)\int_{0}^{a}\left(\left(\frac{K_n^\alpha}{\alpha_{\mathrm{M}}},\frac{K_n^\beta}{\beta_{\mathrm{M}}},\frac{K_n^\rho}{\rho_{\mathrm{M}}}\right)\cdot(\delta\alpha,\delta\beta,\delta\rho)^{\wedge}(n,j)\right)(r)\ \mathrm{d}r\ Y_{n,j}\left(\xi_k\right)\end{aligned}$$

With $K_n = \frac{1}{c_\alpha}K_n^\alpha + \frac{1}{c_\beta}K_n^\beta + K_n^\rho$ and $\xi_k \in \Omega$ we may, again, simplify this representation and get a functional where only the density deviation $\delta\rho$ is mapped to

the frequency shift function

$$\mathcal{F}_{\mathrm{F}}^{k}\delta\rho = \sum_{\substack{n=0 \\ n \text{ even}}}^{\infty} \sum_{j=1}^{2n+1} P_n(0) \int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)} \cdot \left(\delta\rho^{\wedge}(n,j)\right)(r)\ \mathrm{d}r\ Y_{n,j}\left(\xi_k\right).$$

Splitting Function Coefficients

As a third option, we may use the splitting function coefficients directly as data input, i.e.

$$\sigma_{kl}^{\wedge}(n,j) = \int_0^a \left(\left(K_n^{(kl,\alpha)}, K_n^{(kl,\beta)}, K_n^{(kl,\rho)}\right) \cdot \delta m^{\wedge}(n,j)\right)(r)\ \mathrm{d}r$$

where again

$$\left(\left(\delta m^{\wedge}(n,j)\right)(r)\right)_i = \langle (\delta m(r\cdot))_i, Y_{n,j}\rangle_{\mathrm{L}^2(\Omega)}.$$

Note that $\sigma_{kl}^{\wedge}(n,j)$ are the splitting function coefficients corresponding to the splitting function σ_{kl} associated to the normal mode ${}_kS_l$, $k,l \in \mathbb{N}_0$. For the sake of readability, we omitted the reference 'kl' to the corresponding normal mode beforehand.

A functional that maps the deviations to the splitting function coefficients may be given as

$$\mathcal{F}_{\mathrm{C}}^{(kl,nj)}(\delta\alpha,\delta\beta,\delta\rho) = \int_0^a \left(\left(\frac{K_n^{(kl,\alpha)}}{\alpha_{\mathrm{M}}}, \frac{K_n^{(kl,\beta)}}{\beta_{\mathrm{M}}}, \frac{K_n^{(kl,\rho)}}{\rho_{\mathrm{M}}}\right) \cdot (\delta\alpha,\delta\beta,\delta\rho)^{\wedge}(n,j)\right)(r)\ \mathrm{d}r.$$

Let us again assume that the velocities α and β as well as the density ρ depend on each other and, thus, $K_n^{kl} = \frac{1}{c_\alpha}K_n^{(kl,\alpha)} + \frac{1}{c_\beta}K_n^{(kl,\beta)} + K_n^{(kl,\rho)}$. Now, let us state the simplified functional that maps the density deviation to the splitting function coefficient:

$$\mathcal{F}_{\mathrm{C}}^{(kl,nj)}\delta\rho = \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} \cdot \left(\delta\rho^{\wedge}(n,j)\right)(r)\ \mathrm{d}r.$$

The choice of the splitting functions to be used

We want to recover the density distribution of the Earth's mantle. Thus, it is appropriate to use splitting functions which are sensitive in that area. The sensitivity of the splitting functions ${}_kS_l$ or ${}_kT_l$, $k, l \in \mathbb{N}_0$, is described by the respective sensitivity kernels $K_n^{(kl,\alpha)}$, $K_n^{(kl,\beta)}$ and $K_n^{(kl,\rho)}$ for the velocities α and β as well as the density ρ.

In Figure 2.4, we display the sensitivity kernels K_2^α, K_2^β and K_2^ρ for the splitting functions ${}_0S_4$ (left-hand side) and ${}_6S_1$ (right-hand side), as an example. Note the abbreviations for the core-mantle boundary (CMB at a radius of $3,486$ km, i.e. a depth of $2,885$ km) and the inner-core boundary (ICB at a radius of $1,216$ km, i.e. a depth of $5,155$ km). Clearly, the sensitivity kernels corresponding to ${}_0S_4$ are mostly sensitive to the mantle, while the sensitivity kernels corresponding to ${}_6S_1$ are sensitive to the whole radius of the Earth. With these possible differences in the behavior of the sensitivity kernels with respect to the radius of the Earth in mind, we should choose the splitting functions, that we will use as a data input, carefully.

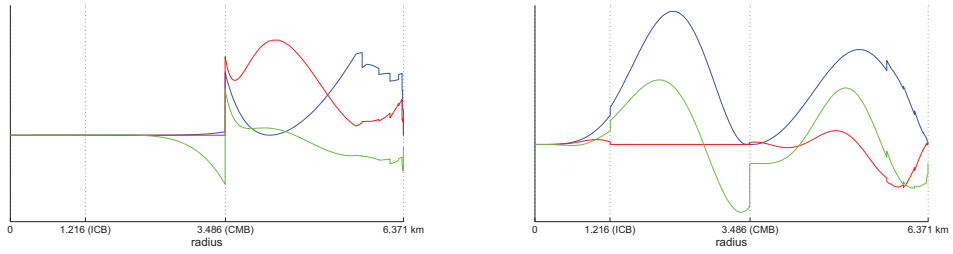

Figure 2.4.: Sensitivity kernels K_2^α (blue), K_2^β (red) and K_2^ρ (green) of the splitting functions ${}_0S_4$ (left-hand) and ${}_6S_1$ (right-hand)

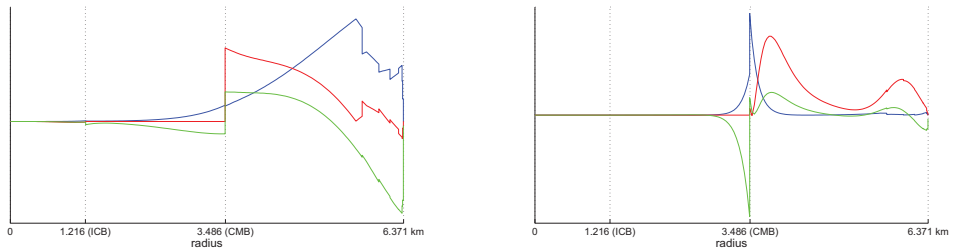

Figure 2.5.: Sensitivity kernels K_2^α (blue), K_2^β (red) and K_2^ρ (green) of the splitting functions ${}_1S_2$ (left-hand) and ${}_1S_{14}$ (right-hand)

Moreover, we have to take into account that the splitting functions display different sensitivities to the velocities α and β as well as the density ρ. For example, the splitting function ${}_1S_2$ depends more on the density than ${}_1S_{14}$ in comparison to the velocities, see the left-hand side and the right-hand side of Figure 2.5, respectively.

In this work, we use the collection of splitting function coefficients given in Table 2.1 where all coefficients corresponding to the given splitting function ${}_kS_l$ up to the given degree n (only even degrees) are used. Current research in the field of normal mode anomalies gives perspectives to the inclusion of odd-degree splitting function coefficients as well.

${}_0S_l$	n	${}_1S_k$	n	${}_2S_k$	n	${}_3S_k$	n	${}_4S_k$	n	${}_5S_k$	n
${}_0S_3$	2	${}_1S_2$	2	${}_2S_4$	6	${}_3S_6$	4	${}_4S_1$	2	${}_5S_3$	6
${}_0S_4$	2	${}_1S_3$	2	${}_2S_5$	6	${}_3S_7$	4	${}_4S_2$	2	${}_5S_4$	6
${}_0S_5$	4	${}_1S_4$	4	${}_2S_6$	6	${}_3S_8$	6	${}_4S_3$	6	${}_5S_5$	6
${}_0S_6$	6	${}_1S_5$	6	${}_2S_7$	6	${}_3S_9$	4	${}_4S_4$	4	${}_5S_6$	8
${}_0S_7$	6	${}_1S_6$	6	${}_2S_8$	6			${}_4S_5$	4		
${}_0S_8$	8	${}_1S_7$	8	${}_2S_9$	6						
${}_0S_9$	8	${}_1S_8$	10	${}_2S_{10}$	6						
${}_0S_{12}$	12	${}_1S_9$	6	${}_2S_{11}$	6						
${}_0S_{13}$	12	${}_1S_{10}$	4	${}_2S_{12}$	6						
${}_0S_{14}$	12	${}_1S_{14}$	2	${}_2S_{13}$	6						
${}_0S_{15}$	12										
${}_0S_{16}$	12										
${}_0S_{17}$	12										
${}_0S_{19}$	12										
${}_0S_{20}$	12										
${}_0S_{21}$	12										

Table 2.1.: List of coefficients used in our computations (see Section 5.3): coefficients of the splitting function ${}_kS_l$ for even degrees up to degree n

II.

Matching Pursuit

In this part, we will develop new methods for a joint inversion of different data types and the corresponding theoretical considerations. The Matching Pursuit (see [98]) decomposes a signal into a linear expansion of functions which are selected stepwise to best match the signal structure such that the difference between the given data and the approximation is minimized. Unfortunately, our data is not given as the values of the target function F. Instead, it is given as the values of a functional applied to the target function, i.e. $y_i = \mathcal{F}^i F,\ i = 1, \ldots, l$, and we cannot apply this algorithm directly.

We will derive an algorithm which we call 'Functional Matching Pursuit' (FMP) that can be applied to this kind of problems. Since we want to solve an ill-posed inverse problem it is very important to develop a regularized version of the algorithm, too. We will present the algorithm ('Regularized Functional Matching Pursuit' (RFMP)) and some theoretical results about the main characteristics of regularization methods - the existence and the stability of the solution as well as the convergence of the regularization.

The iterative character of the method in choosing the expansion functions allows us to directly control the sparsity of the solution. Coupled with the fact that the expansion functions are chosen according to the structure of the solution and the data structure these are the main advantages over already existing methods. Additionally, we may mix different types of functions to best match different structures of the target function.

3. Functional Matching Pursuit

We will start our considerations with an overview over sparse regularization techniques and overcomplete signal representation to motivate our new approach in Section 3.1. Then we will develop the 'Functional Matching Pursuit' (FMP) and give theoretical results with respect to the convergence of the method and its convergence rate. In Section 3.4, we will discuss why we used the original Matching Pursuit as a basis for our research and not the refined version called Orthogonal Matching Pursuit. At last, we will address some issues that arise when implementing the method for our special case, i.e. recovering a tomographic model of the Earth by a joint inversion of gravitational and normal mode data.

3.1. Sparse Regularization and Overcomplete Signal Representation

Research in sparse regularization is, e.g. concerned with finding a sparse or a maximally sparse representation of a signal of an ill-conditioned, linear system of equations. Thus, we minimize an objective function which typically includes a squared error term combined with a penalty term concerned with the sparseness of the solution, e.g. we minimize

$$||Ax - y||_{l^2}^2 + \lambda ||x||_{l^0}$$

where $y \in \mathbb{R}^l$ is the given data, $A \in \mathbb{R}^{m \times l}$ is a matrix and $x \in \mathbb{R}^m$ is the unknown target vector, e.g. wavelet coefficients of a function. However, solving the optimization problem corresponding to

$$||x||_{l^0} := \#\{k \mid x_k \neq 0\} = \sum_{k=1}^{m} |x_k|^0$$

is known to be NP-hard. Thus, we will use a relaxed sparsity measure in the form of $|| \cdot ||_{l^p} := \sum_k |x_k|^p$ where using $p = 1$ can, under certain conditions on the matrix A, again be interpreted as a sparsity constraint (see, e.g. [149]). The research area concerned with the case $p = 1$ is known as compressed sensing (see, e.g. [45, 47] for an overview).

Of course, there exists a rather large number of different approaches to solve this problem. However, all of these methods display disadvantages that disallow us

to use them for our special case. Combinatorial algorithms, e.g. Fourier sampling (see [69, 70]), Chaining Pursuit (see [71]) or HHS Pursuit (see [72]) acquire highly structured samples of the signal that support rapid reconstruction via group testing. Although these methods are very fast, a very high number of data is needed for a proper approximation. The other extreme can be found in the field of convex relaxation, e.g. Basis Pursuit (see [27]), Interior Point methods (see [22, 91]), Projected Gradient Methods (see [57]) and Iterative Thresholding (see [35]). They solve a convex program where the minimizer is known to approximate the target signal. These methods, indeed, require only a small number of data but turn out to be very slow in acquiring the solution.

Iterative greedy algorithms, e.g. Matching Pursuit (MP, see [98]), Orthogonal Matching Pursuit (OMP, see [116, 144]), Regularized Orthogonal Matching Pursuit (ROMP, see [113]), Stagewise Orthogonal Matching Pursuit (StOMP, see [48]), Subspace Pursuit (SP, see [34]) and Compressive Sampling Matching Pursuit (CoSaMP, see [112]), seem to balance running time and efficiency:
With an intermediate running time they display sampling efficiency when building up an approximate solution stepwise. At each iteration, one or more coordinates of the vector x are selected for testing based on the correlation between the columns of A and the regularization measurement vector. While MP, OMP, ROMP and StOMP display a bottom-up approach, i.e. they add a basis function in every step to build up the solution, SP and CoSaMP display a top-down approach, i.e. they iteratively refine the solution x.

Obviously, we decide to use an iterative greedy algorithm. However, because of two main reasons, we cannot use one of the existing methods:

(i) The problem that we want to solve in this work differs from the problem considered in the sparse regularization community, i.e. we want to reconstruct a sparse solution F of $\mathcal{F}F = y$ where $y \in \mathbb{R}^l$ is the possibly noisy data, $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ is a linear and continuous operator and $F \in \mathrm{L}^2(\mathcal{B})$ is the unknown target function. As a consequence, the expansion of the solution F is not build up out of the columns of a matrix A anymore but out of functions in $\mathrm{L}^2(\mathcal{B})$. Furthermore, including a penalty term concerned with the sparsity of the solution requires to compute the L^1-norm of a function which is connected to a very high computational effort in our case.
Our aim is to reconstruct a sparse signal by finding the smoothest expansion to represent the signal, instead of finding a (maximally) sparse representation of a signal as in the sparse regularization community. However, we see a great potential in iterative greedy algorithms to help solve our problem.

(ii) All greedy pursuit algorithms require that the target function is sparse, i.e. that it can be represented exactly as a finite expansion in $\mathrm{L}^2(\mathcal{B})$. Furthermore, the number of elements in this expansion has to be known beforehand. In our case, the target function F is probably not sparse at all. However, we expect that it can be approximated well enough with a finite but unknown number of expansion elements.
Some existing algorithms might be modified to handle this case. The halting condition of the OMP may be changed such that the iteration is stopped when the approximation error is smaller then a certain threshold. However, there are no theoretical results concerning the convergence and stability of such a modification.
The Sparsity Adaptive Matching Pursuit (see [44]) allows signal reconstruction without prior knowledge of the sparsity. However, it involves much more iterations and, thus, a higher computational effort than we can afford in our applications.

Let us give a more detailed overview with regard to overcomplete signal representations and pursuit algorithms (see, e.g. [140, 141, 142]), a particular case of iterative greedy algorithms. These representations offer a wider range of possible elements used in the expansion of the solution than the more traditional approaches and, thus, allow more flexibility and effectiveness in signal reconstruction. Nonetheless, most researches specialize on one kind of basis function at a time, e.g. multiscale Gabor functions (see [98, 120]), amalgams of wavelets and sinusoids (see [27]), collections of windowed cosines were the windows have a range of different widths and locations (see [30]) and systems generated at random (see [46]). In this work, we will attempt to use a combination of localized functions (as in [30]) and a system with a global character.

Applications of recovery out of overcomplete systems can be found in a variety of research areas, e.g. theoretical neurosciences (see [115]), constructive approximation (see [20]), signal processing (see [16, 31]) and image processing (see [136, 137]).

We want to apply the idea of overcompleteness, especially that of the Matching Pursuit (see [98]), to our problem, i.e. we want to recover an approximation to the unknown target function $F \in \mathrm{L}^2(\mathcal{B})$ adaptively and iteratively out of an overcomplete collection of functions were the data $y \in \mathbb{R}^l$ is given explicitly with respect to a linear and continuous operator $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$, i.e. $\mathcal{F}F = y$. However, our algorithms as a combination of ideas from sparse regularization and overcomplete signal representation will not depend on prior knowledge of the sparsity of the target function F.

In the two already known methods, i.e. the Basic Matching Pursuit (see [98])

and the Kernel Matching Pursuit (see [150]), as well as our newly developed ones we get three different versions each:
In the basic version, the optimal expansion functions and the corresponding coefficients are found stepwise. In the version with back-fitting, the expansion functions are found stepwise as well but the optimal set of coefficients is recomputed at every step. This version is more time-consuming but also more accurate. The most accurate one is the version with pre-fitting where all coefficients and the next expansion function are computed at the same time at every step. However, it is easily understood that this is the most time-consuming version as well (see [38]).

Choosing the optimal expansion functions stepwise is advantageous in a number of points. First of all, we can directly control the sparsity of the solution. We will be able to refine solutions or zoom in on interesting structures, since we can reuse earlier results as a starting point in further computations. Furthermore, the expansion functions to represent the solution are chosen according to the structure of the solution as well as the data structure. This allows us to use a much denser grid for the localized kernel functions than in already existing methods to reconstruct the mass density variation of the Earth, e.g. spline and wavelet methods. Moreover, we are able to include much more data points and it is of no importance whether the data grid is well-distributed. Additionally, the algorithm will allow us to mix different types of functions to best match different structures of the target function. In our numerical applications, we will use the $\mathrm{L}^2(\mathcal{B})$-basis system $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\dots,2n+1}$ (see Section 1.2.4) to reconstruct global trends. The localized kernel functions $K^{\mathrm{I}}_h(x,\cdot), h \in]0,1[, x \in \mathcal{B}$, (see Section 1.3) will be used to reconstruct detail structures of the target function.

3.2. The Algorithm

In this section, we will introduce the new algorithm. Let l noisy observations $y_1,\dots,y_l \in \mathbb{R}$ of linear and continuous functionals $\mathcal{F}^1,\dots,\mathcal{F}^l$ applied to the target function F be given, i.e.

$$y_i = \mathcal{F}^i F,\ i = 1,\dots,l.$$

Moreover, let a dictionary $\mathcal{D} \subset \{d \in \mathrm{L}^2(\mathcal{B}) \mid ||d||_{\mathrm{L}^2(\mathcal{B})} = 1\}$ be chosen beforehand. The dictionary is the collection of all functions that may be used in the expansion of the solution. Note that we do not put any further restrictions on the elements of this dictionary. Furthermore, we use normalized functions as dictionary elements since their use reduces the computational effort in the regularized version of the new algorithm. However, this restriction is not imposed by the algorithm itself.

We are looking for an approximation to the target function F as a linear combination of selected dictionary elements $F_n = \sum_{k=1}^{n} \alpha_k d_k$ where n is the number of functions in the expansion, $d_k \in \mathcal{D}$ for all $k = 1, \ldots, n$ are the dictionary elements used in the expansion and $\alpha_k \in \mathbb{R}$ for all $k = 1, \ldots, n$ are the corresponding coefficients of the expansion.

From now on, we will use the notation $\mathcal{F}F := (\mathcal{F}^1 F, \ldots, \mathcal{F}^l F) \in \mathbb{R}^l$, $F \in \mathrm{L}^2(\mathcal{B})$. Note that $\mathcal{F}^k : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}$, $k \in \mathbb{N}_0$, are linear and continuous functionals while $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ is a linear operator with functionals as components. We call the difference between the actual data $y \in \mathbb{R}^l$ and the data corresponding to the approximation the residual $R^n := y - \mathcal{F}F_n \in \mathbb{R}^l$.

The idea of the matching pursuit is to choose a collection of dictionary elements $\{d_1, \ldots, d_n\} \subset \mathcal{D}$ and coefficients $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{R}$ such that they minimize the norm of the residual

$$||R^n||^2_{\mathbb{R}^l} = ||y - \mathcal{F}F_n||^2_{\mathbb{R}^l} = \sum_{i=1}^{l} (y_i - \mathcal{F}^i F_n)^2 \text{ where } F_n = \sum_{k=1}^{n} \alpha_k d_k.$$

We present an iterative method starting with $F_0 = 0$ where we append dictionary elements to the initially empty expansion stepwise while trying to reduce the residual at each stage. Note that no initial information about the solution will be introduced into the algorithm apart from the choice of the dictionary (which involves, for instance, the harmonicity constraint in the case of gravitational data only).

Let us develop the method for the step from n to $n+1$ chosen expansion functions. We assume that F_n is given. Since $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$, we need to look for a combination $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$ that minimizes

$$||y - \mathcal{F}(F_n + \alpha d)||^2_{\mathbb{R}^l} = ||y - \mathcal{F}F_n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l} = ||R^n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l},$$

since the operator $\mathcal{F}$ is linear, i.e.

$$(d_{n+1}, \alpha_{n+1}) = \mathrm{argmin}_{d \in \mathcal{D}, \alpha \in \mathbb{R}} ||R^n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l}.$$

For each choice of $d \in \mathcal{D}$, the corresponding minimizing α has to satisfy

$$\frac{\partial ||R^n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l}}{\partial \alpha} = 0.$$

By computing the partial derivative we get

$$0 = -2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} \text{ and, thus, } \alpha = \frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||^2_{\mathbb{R}^l}}.$$

Let us insert this expression into $||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l}$ to derive an expression from which we can determine the optimal dictionary element d_{n+1} for the step $n+1$.

$$\begin{aligned} ||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} &= ||R^n||^2_{\mathbb{R}^l} - 2\alpha\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + \alpha^2||\mathcal{F}d||^2_{\mathbb{R}^l} \\ &= ||R^n||^2_{\mathbb{R}^l} - 2\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||^2_{\mathbb{R}^l}}\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + \left(\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||^2_{\mathbb{R}^l}}\right)^2 ||\mathcal{F}d||^2_{\mathbb{R}^l} \\ &= ||R^n||^2_{\mathbb{R}^l} - \left(\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||_{\mathbb{R}^l}}\right)^2 \end{aligned}$$

As a consequence, a dictionary element d minimizes $||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l}$ if and only if it maximizes $\left|\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||_{\mathbb{R}^l}}\right|$.

Let d_{n+1} be the dictionary element fulfilling this requirement. Then we conclude that the corresponding coefficient α_{n+1} can be computed by

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}}.$$

Let us state the whole algorithm.

Algorithm 3.1 (Basic Functional Matching Pursuit (FMP))
Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1}d_{n+1}$ *such that*

$$d_{n+1} \text{ maximizes } \left|\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||_{\mathbb{R}^l}}\right| \text{ and } \alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}}.$$

Note that this algorithm does not provide us with the best match to the target function F. Since we determine the expansion functions and the corresponding coefficients at every step iteratively, the expansion with n elements is possibly not optimal at step $n+1$. To remedy this inaccuracy we can do a back-projection in analogy to [38]. That means we choose the dictionary function d_{n+1} as in the original algorithm but recompute the optimal set of coefficients in each step, i.e.

$$(\alpha_1^{n+1}, \ldots, \alpha_{n+1}^{n+1}) = \text{argmin}_{(\alpha_1,\ldots,\alpha_{n+1})\in\mathbb{R}^{n+1}}||y - \mathcal{F}F_{n+1}||^2_{\mathbb{R}^l}.$$

This extension of the algorithm gives us a better approximation while the computation time is increased. To get an even more accurate result we can use pre-fitting, again suggested in [38], where we directly optimize

$$(d_{n+1}, \alpha_1^{n+1}, \ldots, \alpha_{n+1}^{n+1}) = \text{argmin}_{(d,\alpha_1,\ldots,\alpha_{n+1})\in\mathcal{D}\times\mathbb{R}^{n+1}}||y - \mathcal{F}F_{n+1}||^2_{\mathbb{R}^l}.$$

It can easily be seen that this is the most time-consuming version of the three but it will give the best-fitting solution, too. Nonetheless, in this work, we will only use the original version of Algorithm 3.1 (FMP) to reduce the computational effort.

As in the Kernel Matching Pursuit (see [150]), there are no restrictions on the dictionary elements and we may mix different types of functions in the dictionary.

Let us now consider Algorithm 3.1 (FMP) from a computational point of view. We compute $\mathcal{F}d$ and $||\mathcal{F}d||_{\mathbb{R}^l}$ beforehand for all $d \in \mathcal{D}$ and store them. This is an essential part of the algorithm that can be parallelized to reduce the computation time. Furthermore, we can store the solution $\mathcal{F}F_n$ to reduce the costs to compute the residual $R^{n+1} = y - \mathcal{F}F_{n+1}$, since the operator $\mathcal{F}$ is supposed to be linear and, thus, $\mathcal{F}F_{n+1} = \mathcal{F}F_n + \alpha_{n+1}\mathcal{F}d_{n+1}$. Thus, in step $n+1$, we just need to search for the optimal dictionary element d_{n+1} and update the residual R^{n+1}.

3.3. Theoretical Results

By stepping out of the Euclidean setting pursued in [98] we lose some major properties which were exploited in [98] to prove the convergence of the algorithm. However, we sidestep this problem and will now prove the convergence of Algorithm 3.1 (FMP) in this subsection.

In this section, $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ will always be a linear and continuous operator. To treat theoretical questions, let us first rewrite the expression for the residual R^n. For $n = 0$, we get $R^0 = y = \mathcal{F}F$. For $n > 0$, we get with the linearity of the operator $\mathcal{F}$ that

$$\begin{aligned} R^n &= y - \mathcal{F}F_n = y - \mathcal{F}F_{n+1} + \alpha_{n+1}\mathcal{F}d_{n+1} = R^{n+1} + \alpha_{n+1}\mathcal{F}d_{n+1} \\ &= R^{n+1} + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}}\mathcal{F}d_{n+1}. \end{aligned} \tag{3.1}$$

Furthermore, in step $n+1$, we are looking for the dictionary element d_{n+1} such that

$$\frac{|\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}} \geq \sup_{d\in\mathcal{D}} \frac{|\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d\|_{\mathbb{R}^l}}. \tag{3.2}$$

3.3.1. Convergence of Algorithm 3.1 (FMP)

Let us derive some important properties of the residual which we will need to prove the convergence of the algorithm later on.

Theorem 3.2
Let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 3.1 (FMP). Then the $\mathbb{R}^l$-norm of the residual $\left(\|R^n\|_{\mathbb{R}^l}\right)_n$, where the residual R^n is given in Equation (3.1), converges for n tending to infinity. Furthermore, the following equalities hold:

$$\begin{aligned} \|R^n\|_{\mathbb{R}^l}^2 &= \left\|R^{n+1}\right\|_{\mathbb{R}^l}^2 + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} \text{ and} \\ 0 &= \langle R^{n+1}, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}. \end{aligned} \tag{3.3}$$

Proof.
Let us first consider the scalar product $\langle R^{n+1}, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}$ where we use Equation (3.1) on the residual R^{n+1}:

$$\begin{aligned} \langle R^{n+1}, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} &= \left\langle R^n - \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||_{\mathbb{R}^l}^2}\mathcal{F}d_{n+1}, \mathcal{F}d_{n+1}\right\rangle_{\mathbb{R}^l} \\ &= \langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||_{\mathbb{R}^l}^2}||\mathcal{F}d_{n+1}||_{\mathbb{R}^l}^2 \\ &= \langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} \\ &= 0 \end{aligned}$$

Considering the norm of the residual, we get with Equation (3.1)

$$\|R^n\|_{\mathbb{R}^l}^2 = \left\|R^{n+1}\right\|_{\mathbb{R}^l}^2 + \left(\frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||_{\mathbb{R}^l}}\right)^2$$

which proves that Equation (3.3) holds.

Since the second summand of Equation (3.3) is always non-negative we get that

$$\|R^n\|_{\mathbb{R}^l}^2 \geq \left\|R^{n+1}\right\|_{\mathbb{R}^l}^2.$$

Thus, the sequence $\left(\|R^n\|_{\mathbb{R}^l}\right)_n$ is monotonically decreasing. Obviously, it is bounded from below by 0. Consequently, the sequence $\left(\|R^n\|_{\mathbb{R}^l}\right)_n$ converges for n tending to infinity.

□

Now let us do some preliminary work to prove that the residual in itself tends to 0.

Lemma 3.3
Let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 3.1 (FMP). Then

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle^2_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}} = 0.$$

Proof.
Let us display the norm of R^0 by means of a telescoping sum. Due to Theorem 3.2, we get

$$\begin{aligned} ||R^0||^2_{\mathbb{R}^l} &= \sum_{n=0}^{m-1} \left(||R^n||^2_{\mathbb{R}^l} - ||R^{n+1}||^2_{\mathbb{R}^l}\right) + ||R^m||^2_{\mathbb{R}^l} \\ &= \sum_{n=0}^{m-1} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle^2_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}} + ||R^m||^2_{\mathbb{R}^l}. \end{aligned}$$

For m tending to infinity, we get with the convergence of $||R^m||_{\mathbb{R}^l}$ (see Theorem 3.2) that

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle^2_{\mathbb{R}^l}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}} = 0.$$

□

With Theorem 3.2 and Lemma 3.3 we are now able to prove that Algorithm 3.1 (FMP) provides us with the right solution.

Theorem 3.4 (Convergence of Algorithm 3.1 (FMP))
Let the dictionary $\mathcal{D}$ be large enough such that $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$. *Furthermore, let all dictionary elements and their corresponding coefficients be chosen according to Algorithm 3.1 (FMP). Then, the residual R^n converges to* 0 *for n tending to infinity.*

Proof.
We already know from Theorem 3.2 that the sequence of the norm of the residuals $(||R^n||_{\mathbb{R}^l})_n$ converges. Furthermore, we know that it is a monotonically decreasing sequence starting with the element $\|R^0\|_{\mathbb{R}^l} = \|y\|_{\mathbb{R}^l}$, i.e. it is bounded from above. Of course, it is bounded from below by 0. Thus, the sequence of residuals $(R^n)_n$ is bounded, too. That means that there exists a convergent subsequence $(R^{n_j})_j$ in $\mathbb{R}^l$ with limit R^∞ for j tending to infinity.

Due to Lemma 3.3, we already know that

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle^2_{\mathbb{R}^l}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l}} = 0$$

and, consequently,

$$\lim_{j\to\infty} \frac{\langle R^{n_j}, \mathcal{F}d_{n_j+1}\rangle_{\mathbb{R}^l}}{\left\|\mathcal{F}d_{n_j+1}\right\|_{\mathbb{R}^l}} = 0.$$

In combination with Inequality (3.2), we conclude for all dictionary elements $d \in \mathcal{D}$ that

$$\lim_{j\to\infty} \frac{\langle R^{n_j}, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = 0.$$

Putting the limit into the scalar product, we get for all $d \in \mathcal{D}$ that

$$\frac{\langle \lim_{j\to\infty} R^{n_j}, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = \frac{\langle R^\infty, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = 0$$

and, since we chose the dictionary $\mathcal{D}$ large enough such that $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$, we conclude that $R^\infty = 0$. Now we know that the subsequence of the residual $(R^{n_j})_j$ tends to 0 for j tending to infinity. Due to the monotonicity, we get

$$\lim_{n\to\infty} \|R^n\|_{\mathbb{R}^l} = 0.$$

□

To prove that the sequence $(F_n)_n$ converges, we need a particular condition on the dictionary.

Theorem 3.5
Let $y \in \mathbb{R}^l$ be the given data. Moreover, let the dictionary satisfy

1) *$\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$ and*

2) *'semi-frame condition':*
There exists a constant $c > 0$ such that for all $\mathrm{L}^2(\mathcal{B})$-convergent expansions $H = \sum_{k=0}^{\infty} \beta_k d_k$, $\beta_k \in \mathbb{R}$, $d_k \in \mathcal{D}$, (not necessarily pairwise distinct) the following inequality holds:

$$c\|H\|^2_{\mathrm{L}^2(\mathcal{B})} \leq \sum_{k=0}^{\infty} \beta_k^2.$$

3) $C := \inf_{d\in\mathcal{D}} ||\mathcal{F}d||_{\mathbb{R}^l} > 0$

Let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 3.1 (FMP). Then the sequence $(F_n)_n$ of the algorithm converges to a function $F \in \mathrm{L}^2(\mathcal{B})$ with $\mathcal{F}F = y$.

Proof.
Due to Algorithm 3.1 (FMP), Equation (3.3) and Theorem 3.4, we obtain

$$\begin{aligned}\sum_{k=N}^{\infty} \alpha_k^2 &\leq \frac{1}{C^2} \sum_{k=N}^{\infty} \left(\|R^{k-1}\|_{\mathbb{R}^l}^2 - \|R^k\|_{\mathbb{R}^l}^2 \right) = \frac{1}{C^2} \left(\|R^{N-1}\|_{\mathbb{R}^l}^2 - \lim_{k\to\infty} \|R^k\|_{\mathbb{R}^l}^2 \right) \\ &= \frac{1}{C^2} \|R^{N-1}\|_{\mathbb{R}^l}^2 \overset{N\to\infty}{\longrightarrow} 0.\end{aligned}$$

Hence, the conditions on the dictionary imply that

$$\|F - F_{N-1}\|_{\mathrm{L}^2(\mathcal{B})} = \left\| \sum_{k=N}^{\infty} \alpha_k d_k \right\|_{\mathrm{L}^2(\mathcal{B})} \overset{N\to\infty}{\longrightarrow} 0.$$

Since $\mathcal{F}$ is continuous, we finally get $\mathcal{F}F = \lim_{n\to\infty} \mathcal{F}F_n = y$ from Theorem 3.4.

□

Note that Condition 3) of Theorem 3.5 is no constraint in practice, since the dictionary will always be finite in numerical implementations. Furthermore, Condition 1) is obviously necessary, since a data vector $y \in \mathbb{R}^l \setminus \{0\}$ with $\langle y, \mathcal{F}d \rangle_{\mathbb{R}^l} = 0$ for all $d \in \mathcal{D}$ would cause a failure of Algorithm 3.1 (FMP) already in the initial step.

3.3.2. Convergence Rate of Algorithm 3.1 (FMP)

In this subsection, we prove that the norm of the residuals decays exponentially. We use the main ideas from [98] but we have to deal with some difficulties that arise since we do not operate in the Euclidean setting.

The decay of the norm of the residuals depends upon the correlation between the residual and the dictionary elements. We will denote the correlation ratio of a vector $v \in \mathbb{R}^l \setminus \{0\}$ with respect to a dictionary $\mathcal{D}$ with

$$\tau(v) = \sup_{\substack{d\in\mathcal{D} \\ \mathcal{F}d\neq 0}} \frac{|\langle v, \mathcal{F}d \rangle_{\mathbb{R}^l}|}{||v||_{\mathbb{R}^l} ||\mathcal{F}d||_{\mathbb{R}^l}}.$$

Because of the Cauchy-Schwarz inequality (see Theorem 1.25), $\tau(v)$ cannot exceed 1. The maximal value 1 is obtained if there exists a sequence $(d_k)_k \subset \mathcal{D}$ such that $\mathcal{F}d_k$ converges to a vector in $\mathbb{R}^l$ which is collinear with respect to v. Thus, $\tau(v)$ describes how well the data v can be matched by the dictionary

$\mathcal{D}$. Remember the criterion for selecting the optimal dictionary element in Algorithm 3.1 (FMP). Clearly, the dictionary element that maximizes $\tau(R^n)$ is chosen.

Furthermore, let us denote the infimum of the correlation ratio with

$$I(\tau) = \inf_{v \in \mathbb{R}^l \setminus \{0\}} \tau(v)$$

to quantify the 'worst case'.

First, let us guarantee that the norm of the residuals decays exponentially with a rate proportional to $I^2(\tau)$.

Theorem 3.6 (Exponential decay of $||R^n||_{\mathbb{R}^l}$)
Let $F \in \mathrm{L}^2(\mathcal{B})$ with $\mathcal{F}F = y \in \mathbb{R}^l$. Furthermore, let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 3.1 (FMP). Then we get that

$$||R^m||_{\mathbb{R}^l} \leq ||y||_{\mathbb{R}^l} \left(1 - I^2(\tau)\right)^{m/2}$$

for all $m \in \mathbb{N}$.

Proof.
Algorithm 3.1 (FMP) chooses the next dictionary element d_{n+1} according to Inequality (3.2), i.e.

$$\frac{|\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}} \geq \sup_{d \in \mathcal{D}} \frac{|\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = \tau(R^n)||R^n||_{\mathbb{R}^l}.$$

Substituting this result in Equation (3.3), we get that

$$\begin{aligned} \left\|R^{n+1}\right\|_{\mathbb{R}^l}^2 &= \|R^n\|_{\mathbb{R}^l}^2 - \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} \\ &\leq \|R^n\|_{\mathbb{R}^l}^2 - \sup_{d \in \mathcal{D}} \frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2} \\ &= ||R^n||_{\mathbb{R}^l}^2 - \tau^2(R^n)||R^n||_{\mathbb{R}^l}^2 \end{aligned}$$

and we conclude that

$$\left\|R^{n+1}\right\|_{\mathbb{R}^l}^2 \leq \|R^n\|_{\mathbb{R}^l}^2 \left(1 - \tau^2(R^n)\right).$$

Following this line of thought, we get for all $m \in \mathbb{N}$ the inequality

$$\|R^m\|_{\mathbb{R}^l} \leq \left\|R^0\right\|_{\mathbb{R}^l} \prod_{n=0}^{m-1} \left(1 - \tau^2(R^n)\right)^{1/2} \leq \|y\|_{\mathbb{R}^l} \left(1 - I^2(\tau)\right)^{m/2}.$$

□

Note that this result is not very useful if $I(\tau) = 0$. Thus, we show in the following theorem that this is not the case.

Theorem 3.7
Let the dictionary $\mathcal{D}$ be large enough such that $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$. *Then $\tau(v)$ is larger than a strictly positive constant for any $v \in \mathbb{R}^l \setminus \{0\}$, i.e. $I(\tau) > 0$.*

Proof.
Let us assume that the statement of the theorem is wrong, i.e. $I(\tau) = 0$. Then there exists a sequence $(v_n)_n$ of unit vectors in $\mathbb{R}^l$ and a monotonically decreasing sequence $(\tau_n)_n \subset \mathbb{R}$ that converges to 0 for n tending to infinity such that for all $n \in \mathbb{N}_0$

$$\tau(v_n) = \sup_{\substack{d\in\mathcal{D}\\ \mathcal{F}d\neq 0}} \frac{|\langle v_n, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{||v_n||_{\mathbb{R}^l}||\mathcal{F}d||_{\mathbb{R}^l}} = \sup_{\substack{d\in\mathcal{D}\\ \mathcal{F}d\neq 0}} \frac{|\langle v_n, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{||\mathcal{F}d||_{\mathbb{R}^l}} \leq \tau_n. \tag{3.4}$$

We know that the unit sphere $\mathbb{S}^{l-1}$ of $\mathbb{R}^l$ is compact and, thus, there exists a subsequence $(v_{n_p})_p$ that converges to an element v of the unit sphere $\mathbb{S}^{l-1}$ for p tending to infinity.

Let us consider the correlation ratio

$$\begin{aligned}\tau(v) &= \sup_{\substack{d\in\mathcal{D}\\ \mathcal{F}d\neq 0}} \frac{|\langle v, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{||v||_{\mathbb{R}^l}||\mathcal{F}d||_{\mathbb{R}^l}} = \sup_{\substack{d\in\mathcal{D}\\ \mathcal{F}d\neq 0}} \frac{|\langle v, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{||\mathcal{F}d||_{\mathbb{R}^l}} \\ &= \lim_{p\to\infty} \sup_{\substack{d\in\mathcal{D}\\ \mathcal{F}d\neq 0}} \frac{|\langle v_{n_p}, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{||\mathcal{F}d||_{\mathbb{R}^l}} \leq \lim_{p\to\infty} \tau_{n_p} = 0\end{aligned}$$

where we use Inequality (3.4).

Hence, $\langle v, \mathcal{F}d\rangle_{\mathbb{R}^l} = 0$ for all $d \in \mathcal{D}$ and, since $\mathcal{F}$ is linear, also for all $d \in \operatorname{span}\mathcal{D}$. Due to the assumption of the theorem, this implies that $v = 0$ which is a contradiction, since $v \in \mathbb{S}^{l-1}$.

□

The requirement that the dictionary $\mathcal{D}$ is large enough such that $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$ means that all possible combinations of real data $y_1, \ldots, y_l$ can be achieved by taking appropriate linear combinations of dictionary elements. In other words, for given $y \in \mathbb{R}^l$, an algorithm could (theoretically) find a finite linear combination

$$F_n = \sum_{k=1}^{n} \alpha_k d_k,$$

where $n = l$ is possible, such that $\mathcal{F}F_n = y$.

Note that the decay rate decreases if the correlation ratio $\tau(v)$ decreases. Moreover, if the data includes coherent structures with respect to $\mathcal{F}d,\ d \in \mathcal{D}$, then

$\tau(R^n)$ is high for all n, i.e. we get a faster decay of the norm of the residuals. Thus, it is an important step in the preprocessing to choose an appropriate dictionary with respect to the structures of the solution. If we have some idea about the signal structure we may impose this information on the choice of the dictionary functions. Otherwise, we recommend to use a dictionary with more general functions of different kinds to get a faster convergence of the algorithm.

3.4. Outlook: Orthogonal Matching Pursuit

In the Euclidean setting, the Orthogonal Matching Pursuit (OMP) (see [37, 116]) – a refined version of the Matching Pursuit – was developed to increase the quality of the approximation. It can be shown that, in a finite-dimensional space, OMP converges with a finite number of iterations. However, adapting this idea to our setting seems to be problematic. In this section, we want to discuss the arising problems.

In [116], it was proposed to ensure full backward orthogonality of the error to increase the approximation quality. That means we describe every expansion function in terms of the previously chosen ones

$$d_{n+1} = \sum_{k=1}^{n} b_k^n d_k + \zeta_n \text{ with } \langle \zeta_n, d_k \rangle_{\mathrm{L}^2(\mathcal{B})} = 0 \text{ for all } k = 1, \dots, n$$

where ζ_n denotes the 'new' information that cannot be described by the previously chosen expansion functions, i.e. ζ_n is the orthogonal projection of d_{n+1} onto $\overline{\operatorname{span}\{d_1, \dots, d_n\}}^{\perp}$. However, applying this to our current problem seems to be difficult since we do not know how to determine ζ_n with respect to the dictionary elements used in this work.

In [37], another method of orthogonalizing, that allows us to include the setting of inverse problems, was proposed. In step $n+1$ of Algorithm 3.1 (FMP), we choose the dictionary element $d_{n+1} \in \mathcal{D}$ such that Inequality (3.2) holds. Then we project R^n on $\mathcal{F}d_{n+1}$ in Equation (3.1). The idea of the Orthogonal Matching Pursuit is to orthogonalize the directions of the projection, e.g. with a Gram-Schmidt procedure.

For our setting as an inverse problem, this means to orthogonalize $\mathcal{F}d_{n+1}$ with respect to $\{\mathcal{F}d_1, \dots, \mathcal{F}d_n\}$. With $u_1 := \mathcal{F}d_1$ we get with a Gram-Schmidt procedure

$$u_k := \mathcal{F}d_k - \sum_{p=1}^{k-1} \frac{\langle \mathcal{F}d_k, u_p \rangle_{\mathbb{R}^l}}{\|u_p\|_{\mathbb{R}^l}^2} u_p \tag{3.5}$$

for $k = 2, \ldots, n+1$ which provides us with a new approximation

$$\mathcal{F}F_{n+1} = \sum_{k=1}^{n+1} \frac{\langle R^{k+1}, \mathcal{F}d_k \rangle_{\mathbb{R}^l}}{\|u_k\|^2_{\mathbb{R}^l}} u_k \tag{3.6}$$

in orthogonalized elements.

In contrast to the usual setting, we aim to recover an approximation F_n to the target function F itself and not to the data $y = \mathcal{F}F$. In the case of Algorithm 3.1 (FMP) which is based on the Matching Pursuit we may exploit the linearity of the operator $\mathcal{F}$ to recover the approximation F_n. However, this is not possible when using the Orthogonal Matching Pursuit as a basis for our considerations. Although we may resubstitute $\mathcal{F}d_k$ for u_k in Equation (3.6) to, again, use the linearity to get the approximation F_n this seems to be very inconvenient since all u_k iteratively depend on the orthogonalized choices beforehand (see Equation (3.5)).

Thus, in this work, we do not follow the idea of the Orthogonal Matching Pursuit. Furthermore, it is still questionable if the OMP performs faster when applied to real data. Numerical examples showed that the OMP may only be advantageous for very redundant dictionaries. However, we would not call the dictionaries used in this work *very* redundant.

Nonetheless, we see potential to improve the performance of our algorithms by adapting some of the ideas introduced in the Euclidean setting, e.g. the Orthogonal Matching Pursuit, in further research.

3.5. Implementation of the Basic Functional Matching Pursuit

We implement Algorithm 3.1 (FMP) as described in Section 3.2. As a first step, we have to decide which dictionary functions we will use. Secondly, we compute the expressions for $\mathcal{F}d$ as is required in the algorithm. In the following, E_n denotes the truncation index of a series over n.

In this work, we put the $\mathrm{L}^2(\mathcal{B})$-basis system $\{G_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1}$ (see Section 1.2.4) into the dictionary $\mathcal{D}$, to reconstruct global trends of the target function. Furthermore, we use localized kernel functions $K_h(x,\cdot)$, $h \in]0,1[$, $x \in \mathcal{B}$, (see Section 1.3) corresponding to the $\mathrm{L}^2(\mathcal{B})$-basis system $\{G_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1}$ to reconstruct detail structures. Thus, for now, the dictionary $\mathcal{D}$ is given as

$$\begin{aligned}\mathcal{D} \;=\; & \left\{ K_h^{\mathrm{I}}(x,\cdot),\; K_h^{\mathrm{II}}(x,\cdot) \;\middle|\; h \in [0,1], x \in \mathcal{B} \right\} \\ & \cup \left\{ G_{m,n,j}^{\mathrm{I}},\; G_{m,n,j}^{\mathrm{II}} \;\middle|\; m \in \mathbb{N}_0, n \in \mathbb{N}_0, j = 1, \ldots, 2n+1 \right\}.\end{aligned}$$

Using the Dictionary Elements of Type I

First let us consider the localized kernel functions corresponding to the basis of type I. Since the normalized functions are denoted with $K_h(x,\cdot)$ we denote the original kernel functions with $\tilde{K}_h(x,\cdot)$ where

$$K_h(x,\cdot) := \frac{\tilde{K}_h(x,\cdot)}{||\tilde{K}_h(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}.$$

We use this notation for kernel functions corresponding to both types of basis systems. In the following, let us derive the expressions for the kernel function and its norm.

$$\begin{aligned}
\tilde{K}_h^{\mathrm{I}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^2\; G_{m,n,j}^{\mathrm{I}}(x)\; G_{m,n,j}^{\mathrm{I}}(\cdot) \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\frac{4m+2n+3}{a^3}\; P_m^{(0,n+1/2)}\left(2\frac{|\,\cdot\,|^2}{a^2}-1\right) P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right) \\
&\qquad \times\; h^{m+n}\;\frac{2n+1}{4\pi}\; P_n\left(\frac{\cdot}{|\,\cdot\,|}\cdot\frac{x}{|x|}\right)\left(\frac{|\,\cdot\,|}{a}\right)^n\left(\frac{|x|}{a}\right)^n,
\end{aligned}$$

since $x \in \mathcal{B}$ and, in our case, $A_{m,n}^h := h^{\frac{m+n}{2}}$ where $h \in]0,1[$.

The Parseval identity (see, e.g. [65]) yields

$$\begin{aligned}
||\tilde{K}_h^{\mathrm{I}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}^2 &= \int_{\mathcal{B}}\left(\tilde{K}_h^{\mathrm{I}}(x,y)\right)^2 \mathrm{d}y \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{n,m}^h)^4\;\left(G_{m,n,j}^{\mathrm{I}}(x)\right)^2\;\int_{\mathcal{B}}\left(G_{m,n,j}^{\mathrm{I}}(y)\right)^2\;\mathrm{d}y.
\end{aligned}$$

Since $\{G_{m,n,j}^{\mathrm{I}}\}_{m,n\in\mathbb{N}_0,j=1,\ldots,2n+1}$ is an orthonormal system in the space $\mathrm{L}^2(\mathcal{B})$, we conclude that

$$\begin{aligned}
||\tilde{K}_h^{\mathrm{I}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}^2 &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^4\;\left(G_{m,n,j}^{\mathrm{I}}(x)\right)^2 \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^4\;\frac{4m+2n+3}{a^3}\;\left(P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right)\right)^2 \\
&\qquad \times\;\left(Y_{n,j}\left(\frac{x}{|x|}\right)\right)^2\left(\frac{|x|}{a}\right)^{2n}.
\end{aligned}$$

If we now use the addition theorem for spherical harmonics (see Theorem 1.11) we get the identity

$$\begin{aligned} ||\tilde{K}_h^{\mathrm{I}}(x,\cdot)||^2_{\mathrm{L}^2(\mathcal{B})} &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}(A_{m,n}^h)^4\ \frac{4m+2n+3}{a^3}\ \frac{2n+1}{4\pi} \\ &\quad\times\ \left(P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right)\right)^2\ \left(\frac{|x|}{a}\right)^{2n} \end{aligned}$$

where we use that

$$P_n\left(\frac{x}{|x|}\cdot\frac{x}{|x|}\right) = P_n(1) = 1$$

for all $n \in \mathbb{N}_0$ and $x \in \mathcal{B}$. Note that the norm depends on h and the radial distance $|x|$ only.

Remember that these localized kernel functions only influence a part of the ball $\mathcal{B}$ strongly. Thus, they seem to be very promising to reconstruct detail structures of the target function. However, if we used these functions to reconstruct global trends they would be by far too expensive from a computational point of view. Thus, it is advantageous to include a function type with a more global character to reconstruct these global trends. The previously introduced orthonormal basis systems fulfill this characteristic.

Thus, we will consider here the basis functions of type I, i.e.

$$G_{m,n,j}^{\mathrm{I}} = \sqrt{\frac{4m+2n+3}{a^3}}\ P_m^{(0,n+1/2)}\left(2\frac{|\cdot|^2}{a^2}-1\right)\left(\frac{|\cdot|}{a}\right)^n Y_{n,j}\left(\frac{\cdot}{|\cdot|}\right).$$

As derived in Section 3.2, the algorithm chooses the next expansion function to be the dictionary element that maximizes

$$\left|\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{||\mathcal{F}d||_{\mathbb{R}^l}}\right|$$

which translates into

$$\left|\frac{\langle R^n, \mathcal{F}K_h^{\mathrm{I}}(x,\cdot)\rangle_{\mathbb{R}^l}}{||\mathcal{F}K_h^{\mathrm{I}}(x,\cdot)||_{\mathbb{R}^l}}\right| \text{ and } \left|\frac{\langle R^n, \mathcal{F}G_{m,n,j}^{\mathrm{I}}\rangle_{\mathbb{R}^l}}{||\mathcal{F}G_{m,n,j}^{\mathrm{I}}||_{\mathbb{R}^l}}\right|, \text{ respectively.}$$

Thus, we need to compute $\mathcal{F}K_h^{\mathrm{I}}(x,\cdot)$ and $\mathcal{F}G_{m,n,j}^{\mathrm{I}}$ for the operators $\mathcal{F}_{\mathrm{G}}$, $\mathcal{F}_{\mathrm{S}}$, $\mathcal{F}_{\mathrm{F}}$ and $\mathcal{F}_{\mathrm{C}}$ corresponding to gravitational data, data given as splitting functions, as frequency shifts or as splitting function coefficients, respectively (see Section 2).

Operator $\mathcal{F}_{\mathrm{G}}$ corresponding to the gravitational potential

For the $\mathrm{L}^2(\mathcal{B})$-basis $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\dots,2n+1}$, we get

$$\mathcal{F}^k_{\mathrm{G}} G^{\mathrm{I}}_{m,n,j} \;=\; \delta_{m0}\,\frac{4\pi}{2n+1}\sqrt{\frac{a^3}{2n+3}}\,\left(\frac{a}{|x_k|}\right)^n\,\frac{1}{|x_k|}\,Y_{n,j}\left(\frac{x_k}{|x_k|}\right).$$

When introducing this basis system in Section 1.2.4, we mentioned that the basis system of type I can be partitioned into a harmonic and an anharmonic part where the harmonic part corresponds to the basis functions with $m = 0$. Remember that we can only recover information about the harmonic part of the density distribution when we consider gravitational data only. Thus, we will use the harmonic basis functions $G^{\mathrm{I}}_{0,n,j}$ only, when gravitational data is the only data type being considered, since the other functions are elements of the null-space of $\mathcal{F}^k_G$. However, we use all basis functions when we consider different data types to recover information about the anharmonic part, too.

Subsequently, for $x \in \mathcal{B}$ and with the help of the addition theorem for spherical harmonics (see Theorem 1.11) we get

$$\begin{aligned}
\mathcal{F}^k_{\mathrm{G}}\tilde{K}^{\mathrm{I}}_h(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A^h_{m,n})^2\;\mathcal{F}^k_{\mathrm{G}}(G^{\mathrm{I}}_{m,n,j})\;G^{\mathrm{I}}_{m,n,j}(x)\\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A^h_{m,n})^2\;\delta_{m0}\;\frac{4\pi}{2n+1}\sqrt{\frac{a^3}{2n+3}}\\
&\qquad\times\left(\frac{a}{|x_k|}\right)^n\;\frac{1}{|x_k|}\;Y_{n,j}\left(\frac{x_k}{|x_k|}\right)\;G^{\mathrm{I}}_{m,n,j}(x)\\
&= \sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A^h_{0,n})^2\;\frac{4\pi}{2n+1}\sqrt{\frac{a^3}{2n+3}}\left(\frac{a}{|x_k|}\right)^n\frac{1}{|x_k|}\;Y_{n,j}\left(\frac{x_k}{|x_k|}\right)G^{\mathrm{I}}_{0,n,j}(x)\\
&= \sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A^h_{0,n})^2\;\frac{4\pi}{2n+1}\;\left(\frac{|x|}{|x_k|}\right)^n\;\frac{1}{|x_k|}\;Y_{n,j}\left(\frac{x_k}{|x_k|}\right)\;Y_{n,j}\left(\frac{x}{|x|}\right)\\
&= \sum_{n=0}^{E_n}(A^h_{0,n})^2\;\left(\frac{|x|}{|x_k|}\right)^n\;\frac{1}{|x_k|}\;P_n\left(\frac{x_k}{|x_k|}\cdot\frac{x}{|x|}\right)
\end{aligned}$$

where we use the fact that $P_0^{(0,n+1/2)} \equiv 1$. Moreover, since $\mathcal{F}_{\mathrm{G}}$ is a linear operator we get for the normalized kernel function

$$\mathcal{F}^k_{\mathrm{G}}K^{\mathrm{I}}_h(x,\cdot) = \frac{\mathcal{F}^k_{\mathrm{G}}\tilde{K}^{\mathrm{I}}_h(x,\cdot)}{||\tilde{K}^{\mathrm{I}}_h(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}.$$

If we consider $\mathcal{F}_{\mathrm{G}}$ only, we use (as explained above) the dictionary

$$\begin{aligned}\mathcal{D} &= \left\{ K_h^{\mathrm{I}}(x,\cdot) \;\middle|\; h \in \{h_1,\dots,h_s\} \subset [0,1], x \in \mathrm{grid}(\mathcal{B}) \right\} \\ &\quad \cup \left\{ G_{0,n,j}^{\mathrm{I}} \;\middle|\; n \in \mathbb{N}_0, j = 1,\dots,2n+1 \right\},\end{aligned}$$

$s \in \mathbb{N}$, where the point grid $\mathrm{grid}(\mathcal{B})$ was chosen in advance and $A_m^h = \delta_{m0}$.

Operator $\mathcal{F}_{\mathrm{S}}$ corresponding to the splitting functions

The data corresponding to splitting functions is given at points on the surface of the Earth. Thus, we consider locations $\xi_k \in \Omega$ and get for the basis functions

$$\begin{aligned}\mathcal{F}_{\mathrm{S}}^k G_{m,n,j}^{\mathrm{I}} &= \sum_{p=0}^{E_p} \sum_{q=1}^{2p+1} \int_0^a \frac{K_p(r)}{\rho_{\mathrm{M}}(r)} (G_{m,n,j}^{\mathrm{I}}(r\cdot))^\wedge(p,q)\ \mathrm{d}r\ Y_{p,q}(\xi_k) \\ &= \sqrt{\frac{4m+2n+3}{a^3}} \int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r\ Y_{n,j}(\xi_k),\end{aligned}$$

since

$$\begin{aligned}&(G_{m,n,j}^{\mathrm{I}}(r\cdot))^\wedge(p,q) \\ =\ & \int_\Omega G_{m,n,j}^{\mathrm{I}}(r\xi) Y_{p,q}(\xi)\ \mathrm{d}\omega(\xi) \\ =\ & \sqrt{\frac{4m+2n+3}{a^3}} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n \\ & \quad \times \int_\Omega Y_{n,j}(\xi)\ Y_{p,q}(\xi)\ \mathrm{d}\omega(\xi) \\ =\ & \sqrt{\frac{4m+2n+3}{a^3}} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \delta_{np}\ \delta_{jq}. \qquad (3.7)\end{aligned}$$

For $x = t\eta \in \mathcal{B}$ we get

$$\begin{aligned}\mathcal{F}_{\mathrm{S}}^k \tilde{K}_h^{\mathrm{I}}(x,\cdot) &= \sum_{m=0}^{E_m} \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} (A_{m,n}^h)^2\ \mathcal{F}_{\mathrm{S}}^k(G_{m,n,j}^{\mathrm{I}})\ G_{m,n,j}^{\mathrm{I}}(x) \\ &= \sum_{m=0}^{E_m} \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} (A_{m,n}^h)^2\ \frac{4m+2n+3}{a^3}\ P_m^{(0,n+1/2)}\left(2\frac{t^2}{a^2}-1\right) \\ &\quad \times \int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r \\ &\quad \times \left(\frac{t}{a}\right)^n Y_{n,j}(\eta)\, Y_{n,j}(\xi_k).\end{aligned}$$

Applying the addition theorem (see Theorem 1.11), we get

$$\begin{aligned}\mathcal{F}_{\mathrm{S}}^{k}\tilde{K}_{h}^{\mathrm{I}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}(A_{m,n}^{h})^2\ \frac{2n+1}{4\pi}\ \frac{4m+2n+3}{a^3}\ P_m^{(0,n+1/2)}\left(2\frac{t^2}{a^2}-1\right)\\ &\quad\times\int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r\\ &\quad\times\left(\frac{t}{a}\right)^n\ P_n(\xi_k\cdot\eta)\end{aligned}$$

where, again,

$$\mathcal{F}_{\mathrm{S}}^{k}K_{h}^{\mathrm{I}}(x,\cdot)=\frac{\mathcal{F}_{\mathrm{S}}^{k}\tilde{K}_{h}^{\mathrm{I}}(x,\cdot)}{||\tilde{K}_{h}^{\mathrm{I}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}$$

since $\mathcal{F}_{\mathrm{S}}$ is a linear operator.

Operator $\mathcal{F}_{\mathrm{F}}$ corresponding to frequency shifts

The expressions of $\mathcal{F}_{\mathrm{F}}K_{h}^{\mathrm{I}}(x,\cdot)$ and $\mathcal{F}_{\mathrm{F}}G_{m,n,j}^{\mathrm{I}}$ can be derived similarly to those of $\mathcal{F}_{\mathrm{S}}K_{h}^{\mathrm{I}}(x,\cdot)$ and $\mathcal{F}_{\mathrm{S}}G_{m,n,j}^{\mathrm{I}}$, since the structure of the corresponding functionals, i.e. the components of these operators, is very similar. Thus, we only state the results.

For $\xi_k\in\Omega$ and $x=t\eta\in\mathcal{B}$ we get

$$\begin{aligned}\mathcal{F}_{\mathrm{F}}^{k}G_{m,n,j}^{\mathrm{I}} &= P_n(0)\ \sqrt{\frac{4m+2n+3}{a^3}}\ Y_{n,j}(\xi_k)\\ &\quad\times\int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\mathrm{d}r\end{aligned}$$

and

$$\begin{aligned}\mathcal{F}_{\mathrm{F}}^{k}\tilde{K}_{h}^{\mathrm{I}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}P_n(0)(A_{m,n}^{h})^2\ \frac{2n+1}{4\pi}\ \frac{4m+2n+3}{a^3}\ P_m^{(0,n+1/2)}\left(2\frac{t^2}{a^2}-1\right)\\ &\quad\times\int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r\ \left(\frac{t}{a}\right)^n\ P_n(\xi_k\cdot\eta)\end{aligned}$$

where

$$\mathcal{F}_{\mathrm{F}}^{k}K_{h}^{\mathrm{I}}(x,\cdot)=\frac{\mathcal{F}_{\mathrm{F}}^{k}\tilde{K}_{h}^{\mathrm{I}}(x,\cdot)}{||\tilde{K}_{h}^{\mathrm{I}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}$$

since $\mathcal{F}_{\mathrm{F}}$ is a linear operator, too.

Operator $\mathcal{F}_{\mathrm{C}}$ corresponding to splitting function coefficients

Compared to the operator $\mathcal{F}_{\mathrm{S}}$ we do less work when considering the operator $\mathcal{F}_{\mathrm{C}}$, since we are concerned with the splitting function coefficients only. With Equation (3.7), we get for the coefficient $\sigma_{kl}^{\wedge}(n,j), n,k,l \in \mathbb{N}_0, j = 1,\ldots,n+1$, corresponding to the splitting function σ_{kl} associated to the normal mode ${}_kS_l$

$$\begin{aligned}
\mathcal{F}_{\mathrm{C}}^{(kl,nj)} G_{m,p,q}^{\mathrm{I}} &= \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} (G_{m,p,q}^{\mathrm{I}}(r\cdot))^{\wedge}(n,j)\ \mathrm{d}r \\
&= \sqrt{\frac{4m+2n+3}{a^3}} \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r\ \delta_{np}\ \delta_{jq}.
\end{aligned}$$

For $x \in \mathcal{B}$ we get

$$\begin{aligned}
\mathcal{F}_{\mathrm{C}}^{(kl,nj)} \tilde{K}_h^{\mathrm{I}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{p=0}^{E_p}\sum_{q=1}^{2p+1} \left(A_{m,p}^h\right)^2 \mathcal{F}_{\mathrm{C}}^{(kl,nj)}(G_{m,p,q}^{\mathrm{I}})\ G_{m,p,q}^{\mathrm{I}}(x) \\
&= \sum_{m=0}^{E_m}\sum_{p=0}^{E_p}\sum_{q=1}^{2p+1} \left(A_{m,p}^h\right)^2 \sqrt{\frac{4m+2n+3}{a^3}}\ G_{m,p,q}^{\mathrm{I}}(x) \\
&\qquad \times \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r\ \delta_{np}\ \delta_{jq} \\
&= \sum_{m=0}^{E_m} \left(A_{m,n}^h\right)^2 \sqrt{\frac{4m+2n+3}{a^3}} G_{m,n,j}^{\mathrm{I}}(x) \\
&\qquad \times \int_0^a \frac{K_n^{kl}(r)}{\rho_M(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r \\
&= \sum_{m=0}^{E_m} \left(A_{m,n}^h\right)^2 \frac{4m+2n+3}{a^3}\ P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right) \\
&\qquad \times \left(\frac{|x|}{a}\right)^n Y_{n,j}\left(\frac{x}{|x|}\right) \\
&\qquad \times \int_0^a \frac{K_n^{kl}(r)}{\rho_M(r)} P_m^{(0,n+1/2)}\left(2\frac{r^2}{a^2}-1\right)\left(\frac{r}{a}\right)^n\ \mathrm{d}r
\end{aligned}$$

where again

$$\mathcal{F}_{\mathrm{C}}^{(kl,nj)} K_h^{\mathrm{I}}(x,\cdot) = \frac{\mathcal{F}_{\mathrm{C}}^{(kl,nj)} \tilde{K}_h^{\mathrm{I}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{I}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}$$

since $\mathcal{F}_{\mathrm{C}}$ is a linear operator, too.

Comparing the three operators corresponding to the normal mode anomalies, recovering the splitting function coefficients directly with $\mathcal{F}_{\mathrm{C}}^{(kl,nj)}$ seems to be the most efficient version with respect to the computational effort.

If we consider the operators corresponding to gravitation and normal modes, both, we will use anharmonic basis functions $G^{\mathrm{I}}_{m,n,j}$ with $m \neq 0$ as dictionary elements, too, i.e.

$$\begin{aligned}\mathcal{D} &= \left\{ K_h^{\mathrm{I}}(x,\cdot) \;\middle|\; h \in \{h_1,\ldots,h_s\} \subset [0,1], x \in \mathrm{grid}(\mathcal{B}) \right\} \\ &\quad \cup \left\{ G^{\mathrm{I}}_{m,n,j} \;\middle|\; m \in \mathbb{N}_0, n \in \mathbb{N}_0, j = 1,\ldots,2n+1 \right\},\end{aligned}$$

$s \in \mathbb{N}$, where the point grid $\mathrm{grid}(\mathcal{B})$ was chosen in advance.

Using the Dictionary Elements of Type II

If we compare the expressions $\mathcal{F}^k \tilde{K}_h^{\mathrm{I}}(x,\cdot)$ of the different functionals $\mathcal{F}_{\mathrm{G}}^k$, $\mathcal{F}_{\mathrm{S}}^k$, $\mathcal{F}_{\mathrm{F}}^k$ and $\mathcal{F}_{\mathrm{C}}^{(kl,nj)}$, the computational effort will be largest for $\mathcal{F}_{\mathrm{S}}^k$ and $\mathcal{F}_{\mathrm{F}}^k$ corresponding to the normal mode tomography, since we have to compute two sums instead of one as in the case of $\mathcal{F}_{\mathrm{G}}^k \tilde{K}_h^{\mathrm{I}}(x,\cdot)$ and $\mathcal{F}_{\mathrm{C}}^{(kl,nj)} \tilde{K}_h^{\mathrm{I}}(x,\cdot)$. Using the separable basis functions of type II instead of the inseparable basis functions of type I we may expect a reduction of the computation time. However, this is questionable as will be explained below.

Remember the $\mathrm{L}^2(\mathcal{B})$-basis system $\{G^{\mathrm{II}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1\ldots,2n+1}$

$$G^{\mathrm{II}}_{m,n,j}(x) := \begin{cases} \sqrt{\frac{2m+3}{a^3}}\, P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) Y_{n,j}\left(\frac{x}{|x|}\right) & , \text{ if } x \neq 0 \\ 1 & , \text{ if } x = 0 \end{cases}.$$

Because of their global character, we will use these functions to reconstruct global trends of the target function. For the detail structures of the target function we will again use localized kernel functions corresponding to the basis system of type II. As the expressions for $x = 0$ are trivial we only present the results for

$x \in \mathcal{B} \setminus \{0\}$.

$$\begin{aligned}
\tilde{K}_h^{\mathrm{II}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^2\; G_{m,n,j}^{\mathrm{II}}(x)\; G_{m,n,j}^{\mathrm{II}}(\cdot) \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\frac{2m+3}{a^3}\; P_m^{(0,2)}\left(2\frac{|\cdot|}{a}-1\right)P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) \\
&\qquad \times (A_{m,n}^h)^2\;\frac{2n+1}{4\pi}\; P_n\left(\frac{\cdot}{|\cdot|}\cdot\frac{x}{|x|}\right) \\
&= \left(\sum_{m=0}^{E_m}(A_m^h)^2\;\frac{2m+3}{a^3}\; P_m^{(0,2)}\left(2\frac{|\cdot|}{a}-1\right)P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right)\right) \\
&\qquad \times \left(\sum_{n=0}^{E_n}(A_n^h)^2\;\frac{2n+1}{4\pi}\; P_n\left(\frac{\cdot}{|\cdot|}\cdot\frac{x}{|x|}\right)\right),
\end{aligned}$$

since $x \in \mathcal{B} \setminus \{0\}$ and $A_{m,n}^h = h^{\frac{m+n}{2}} = h^{\frac{m}{2}}h^{\frac{n}{2}} = A_m^h A_n^h$ where $h \in]0,1[$. Again, we need to compute the norm of the kernel functions $\tilde{K}_h^{\mathrm{II}}(x,\cdot)$ since we want to use normalized functions in the dictionary.

With the Parseval identity we get

$$\begin{aligned}
&||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}^2 = \int_{\mathcal{B}}\left(\tilde{K}_h^{\mathrm{II}}(x,y)\right)^2 \mathrm{d}y \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^4\;\left(G_{m,n,j}^{\mathrm{II}}(x)\right)^2 \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^4\;\frac{2m+3}{a^3}\;\left(P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right)\right)^2\;\left(Y_{n,j}\left(\frac{x}{|x|}\right)\right)^2 \\
&= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}(A_{m,n}^h)^4\;\frac{2m+3}{a^3}\;\frac{2n+1}{4\pi}\;\left(P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right)\right)^2 \\
&= \left(\sum_{m=0}^{E_m}(A_m^h)^4\;\frac{2m+3}{a^3}\left(P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right)\right)^2\right)\;\left(\sum_{n=0}^{E_n}(A_n^h)^4\;\frac{2n+1}{4\pi}\right).
\end{aligned}$$

Note that this norm, again, only depends on h and the radial distance $|x|$. Let us denote the normalized kernel functions with

$$K_h^{\mathrm{II}}(x,\cdot) := \frac{\tilde{K}_h^{\mathrm{II}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}.$$

As in the case of type I, we need to compute $\mathcal{F}K_h^{\mathrm{II}}(x,\cdot)$ and $\mathcal{F}G_{m,n,j}^{\mathrm{II}}$ for the four different operators $\mathcal{F}_{\mathrm{G}}$, $\mathcal{F}_{\mathrm{S}}$, $\mathcal{F}_{\mathrm{F}}$ and $\mathcal{F}_{\mathrm{C}}$ (see Section 2).

Operator $\mathcal{F}_{\mathrm{G}}$ corresponding to the gravitational potential

Let us present the equations of $\mathcal{F}_{\mathrm{G}} K_h^{\mathrm{II}}(x,\cdot)$ and $\mathcal{F}_{\mathrm{G}} G_{m,n,j}^{\mathrm{II}}$ first.

$$\begin{aligned}
\mathcal{F}_{\mathrm{G}}^k G_{m,n,j}^{\mathrm{II}} &= \delta_{m\mu(n)} \frac{4\pi}{2n+1} \left(\sum_{l=0}^{n} (2l+3) \left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!} \right)^2 \right) \\
&\quad \times \sqrt{\frac{a^3}{2\mu(n)+3}} \, \frac{(n-\mu(n))!(n+3+\mu(n))!}{n!(n+2)!} \\
&\quad \times \left(\frac{a}{|x_k|} \right)^n \frac{1}{|x_k|} \, Y_{n,j} \left(\frac{x_k}{|x_k|} \right)
\end{aligned}$$

and, subsequently, for $x \in \mathcal{B} \setminus \{0\}$ with the help of the addition theorem (see Theorem 1.11)

$$\begin{aligned}
\mathcal{F}_{\mathrm{G}}^k \tilde{K}_h^{\mathrm{II}}(x,\cdot) &= \sum_{n=0}^{E_n} \sum_{m=0}^{\min\{n,E_m\}} \sum_{j=1}^{2n+1} (A_{m,n}^h)^2 \, \mathcal{F}_{\mathrm{G}}^k (G_{m,n,j}^{\mathrm{II}}) \, G_{m,n,j}^{\mathrm{II}}(x) \\
&= \sum_{n=0}^{E_n} \sum_{m=0}^{\min\{n,E_m\}} \sum_{j=1}^{2n+1} (A_{m,n}^h)^2 \sqrt{\frac{2m+3}{a^3}} \, P_m^{(0,2)} \left(2\frac{|x|}{a} - 1 \right) \\
&\qquad \times \delta_{m\mu(n)} \frac{4\pi}{2n+1} \left(\sum_{l=0}^{n} (2l+3) \left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!} \right)^2 \right) \\
&\qquad \times \sqrt{\frac{a^3}{2\mu(n)+3}} \, \frac{(n-\mu(n))!(n+3+\mu(n))!}{n!(n+2)!} \\
&\qquad \times \left(\frac{a}{|x_k|} \right)^n \frac{1}{|x_k|} \, Y_{n,j} \left(\frac{x}{|x|} \right) \, Y_{n,j} \left(\frac{x_k}{|x_k|} \right) \\
&= \sum_{\substack{n=0 \\ \mu(n) \le E_m}}^{E_n} (A_{\mu(n),n}^h)^2 \, P_{\mu(n)}^{(0,2)} \left(2\frac{|x|}{a} - 1 \right) \\
&\qquad \times \left(\sum_{l=0}^{n} (2l+3) \left(\frac{n!(n+2)!}{(n-l)!(n+3+l)!} \right)^2 \right) \\
&\qquad \times \frac{(n-\mu(n))!(n+3+\mu(n))!}{n!(n+2)!} \\
&\qquad \times \left(\frac{a}{|x_k|} \right)^n \frac{1}{|x_k|} \, P_n \left(\frac{x}{|x|} \cdot \frac{x_k}{|x_k|} \right)
\end{aligned}$$

where $\mathcal{F}_{\mathrm{G}}^k K_h^{\mathrm{II}}(x,\cdot) = \dfrac{\mathcal{F}_{\mathrm{G}}^k \tilde{K}_h^{\mathrm{II}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}$, since $\mathcal{F}_{\mathrm{G}}$ is a linear operator.

Operator $\mathcal{F}_{\mathrm{S}}$ corresponding to the splitting functions

For $\xi_k \in \Omega$ we get

$$\begin{aligned}\mathcal{F}_{\mathrm{S}}^k G_{m,n,j}^{\mathrm{II}} &= \sum_{p=0}^{E_p}\sum_{q=1}^{2p+1}\int_0^a \frac{K_p(r)}{\rho_{\mathrm{M}}(r)}(G_{m,n,j}^{\mathrm{II}}(r\cdot)^\wedge(p,q)\ \mathrm{d}r\ Y_{p,q}(\xi_k)\\ &= \sqrt{\frac{2m+3}{a^3}}\int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r\ Y_{n,j}(\xi_k),\end{aligned}$$

since

$$\begin{aligned}(G_{m,n,j}^{\mathrm{II}}(r\cdot)^\wedge(p,q) &= \int_\Omega G_{m,n,j}^{\mathrm{II}}(r\xi)Y_{p,q}(\xi)\ \mathrm{d}\omega(\xi)\\ &= \sqrt{\frac{2m+3}{a^3}}P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \int_\Omega Y_{n,j}(\xi)Y_{p,q}(\xi)\ \mathrm{d}\omega(\xi)\\ &= \sqrt{\frac{2m+3}{a^3}}P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \delta_{np}\ \delta_{jq}. \qquad (3.8)\end{aligned}$$

For $x = t\eta \in \mathcal{B}\setminus\{0\}$ we get

$$\begin{aligned}\mathcal{F}_{\mathrm{S}}^k \tilde{K}_h^{\mathrm{II}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^2\ \mathcal{F}_{\mathrm{S}}^k(G_{m,n,j}^{\mathrm{II}})\ G_{m,n,j}^{\mathrm{II}}(x)\\ &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}\sum_{j=1}^{2n+1}(A_{m,n}^h)^2\ \frac{2m+3}{a^3}\ P_m^{(0,2)}\left(2\frac{t}{a}-1\right)\\ &\qquad\times\int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r\ Y_{n,j}(\eta)\,Y_{n,j}(\xi_k)\\ &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n}(A_{m,n}^h)^2\ \frac{2n+1}{4\pi}\ \frac{2m+3}{a^3}\ P_m^{(0,2)}\left(2\frac{t}{a}-1\right) \qquad (3.9)\\ &\qquad\times\int_0^a \frac{K_n(r)}{\rho_M(r)}P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r\ P_n(\xi_k\cdot\eta)\\ &= \int_0^a\left(\sum_{n=0}^{E_n}\frac{2n+1}{4\pi}\ (A_n^h)^2P_n(\xi_k\cdot\eta)\ \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}\right) \qquad (3.10)\\ &\qquad\times\left(\sum_{m=0}^{E_m}(A_m^h)^2\ \frac{2m+3}{a^3}\ P_m^{(0,2)}\left(2\frac{t}{a}-1\right)\ P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\right)\ \mathrm{d}r\end{aligned}$$

where

$$\mathcal{F}_{\mathrm{S}}^{k} K_h^{\mathrm{II}}(x,\cdot) = \frac{\mathcal{F}_{\mathrm{S}}^{k} \tilde{K}_h^{\mathrm{II}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}},$$

since $\mathcal{F}_{\mathrm{S}}$ is a linear operator. Note that we use the addition theorem (see Theorem 1.11) to get Equation (3.9) and exchange the integration and the summation to get Equation (3.10).

From a computational point of view, it is questionable if the partitioning of the sums in the step from Equation (3.9) to Equation (3.10) is advisable. On the one hand, the partitioning of the sums reduces the computational effort as expected. On the other hand, we have to compute the integral in every step anew. In Equation (3.9) we can compute the integral for the required cases beforehand and store the results which reduces the computational effort on the cost of having to evaluate a double sum.

Operator $\mathcal{F}_{\mathrm{F}}$ corresponding to frequency shifts

Correspondingly, we get for $\xi_k \in \Omega$ and $x = s\eta \in \mathcal{B}$

$$\mathcal{F}_{\mathrm{F}}^{k} G_{m,n,j}^{\mathrm{II}} = P_n(0)\sqrt{\frac{2m+3}{a^3}} \int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\, \mathrm{d}r\; Y_{n,j}(\xi_k)$$

and

$$\begin{aligned}
\mathcal{F}_{\mathrm{F}}^{k} \tilde{K}_h^{\mathrm{II}}(x,\cdot) &= \sum_{m=0}^{E_m}\sum_{n=0}^{E_n} P_n(0)\; (A_{m,n}^h)^2\; \frac{2n+1}{4\pi}\; \frac{2m+3}{a^3}\; P_m^{(0,2)}\left(2\frac{s}{a}-1\right) \\
&\qquad \times \int_0^a \frac{K_n(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\, \mathrm{d}r\; P_n(\xi_k\cdot\eta) \\
&= \int_0^a \left(\sum_{n=0}^{E_n} \frac{2n+1}{4\pi}\; (A_n^h)^2\; P_n(0)\; P_n(\xi_k\cdot\eta)\; \frac{K_n(r)}{\rho_{\mathrm{M}}(r)}\right) \\
&\qquad \times \left(\sum_{m=0}^{E_m} (A_m^h)^2\; \frac{2m+3}{a^3}\; P_m^{(0,2)}\left(2\frac{s}{a}-1\right)\; P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\right)\, \mathrm{d}r
\end{aligned}$$

where, again,

$$\mathcal{F}_{\mathrm{F}}^{k} K_h^{\mathrm{II}}(x,\cdot) = \frac{\mathcal{F}_{\mathrm{F}}^{k} \tilde{K}_h^{\mathrm{II}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}},$$

since $\mathcal{F}_{\mathrm{F}}$ is a linear operator, too.

Operator $\mathcal{F}_{\mathrm{C}}$ corresponding to splitting function coefficients

Compared to the operator $\mathcal{F}_{\mathrm{S}}$ we do less work when considering the operator $\mathcal{F}_{\mathrm{C}}$, since we are concerned with the splitting function coefficients only. With Equation (3.8), we get for the coefficient $\sigma_{kl}^{\wedge}(n,j), n,k,l \in \mathbb{N}_0, j = 1,\ldots,n+1$, corresponding to the splitting function σ_{kl} associated to the normal mode ${}_kS_l$

$$\begin{aligned}
\mathcal{F}_{\mathrm{C}}^{(kl,nj)} G_{m,p,q}^{\mathrm{II}} &= \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} (G_{m,p,q}^{\mathrm{II}}(r\cdot))^{\wedge}(n,j)\ \mathrm{d}r \\
&= \sqrt{\frac{2n+3}{a^3}} \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r\ \delta_{np}\ \delta_{jq}.
\end{aligned}$$

The expression for $\mathcal{F}_{\mathrm{C}}\tilde{K}^{\mathrm{II}}(x,\cdot)$ is

$$\begin{aligned}
\mathcal{F}_{\mathrm{C}}^{(kl,nj)} \tilde{K}_h^{\mathrm{II}}(x,\cdot) &= \sum_{m=0}^{E_m} \sum_{p=0}^{E_p} \sum_{q=1}^{2p+1} \left(A_{m,p}^h\right)^2 \mathcal{F}_{\mathrm{C}}^{(kl,nj)} (G_{m,p,q}^{\mathrm{II}}) G_{m,p,q}^{\mathrm{II}}(x) \\
&= \sum_{m=0}^{E_m} \sum_{p=0}^{E_p} \sum_{q=1}^{2p+1} \left(A_{m,p}^h\right)^2 \sqrt{\frac{2n+3}{a^3}} G_{m,p,q}^{\mathrm{II}}(x) \\
&\qquad \times \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r\ \delta_{np}\ \delta_{jq} \\
&= \sum_{m=0}^{E_m} \left(A_{m,n}^h\right)^2 \frac{2n+3}{a^3}\ P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) Y_{n,j}\left(\frac{x}{|x|}\right) \\
&\qquad \times \int_0^a \frac{K_n^{kl}(r)}{\rho_{\mathrm{M}}(r)} P_m^{(0,2)}\left(2\frac{r}{a}-1\right)\ \mathrm{d}r
\end{aligned}$$

where again

$$\mathcal{F}_{\mathrm{C}}^{(kl,nj)} K_h^{\mathrm{II}}(x,\cdot) = \frac{\mathcal{F}_{\mathrm{C}}^{(kl,nj)} \tilde{K}_h^{\mathrm{II}}(x,\cdot)}{||\tilde{K}_h^{\mathrm{II}}(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}},$$

since $\mathcal{F}_{\mathrm{C}}$ is a linear operator, too.

4. Regularized Functional Matching Pursuit

The ill-posedness of the inverse gravimetric problem requires the use of a regularization technique. Unfortunately, we may doubt the performance of Algorithm 3.1 (FMP) when considering even slightly inaccurate data (see [155] where the ill-posedness that can result was documented for a special case and our considerations regarding noise in our application, see the end of Section 5.1). We use here a Tikhonov regularization, i.e. we try to achieve a trade-off between fitting the data and reducing a norm of the solution. The regularization parameter λ correlates both terms.

In Section 4.1, we introduce a regularized version of Algorithm 3.1 (FMP) that is developed similarly to the unregularized version where the penalty term is concerned with the smoothness of the solution, i.e. its $\mathrm{L}^2(\mathcal{B})$-norm. In Section 4.2, we address the main requirements for regularization methods, i.e. we will give results about the existence and stability of the solution as well as the convergence of the regularized solution in the limit for the regularization parameter for both exact and noisy data. In the last two sections, we will discuss some practical topics when implementing the algorithm and an alternative approach where the penalty term includes the $\mathrm{L}^1(\mathcal{B})$-norm, respectively.

4.1. The Algorithm

Using a Tikhonov regularization, we need to find

$$(d_{n+1}, \alpha_{n+1}) = \mathrm{argmin}_{d\in\mathcal{D},\alpha\in\mathbb{R}} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right)$$

in the step from n to $n+1$ chosen expansion functions where $\lambda > 0$ is the regularization parameter. We derive the optimal dictionary element d_{n+1} and the corresponding coefficient α_{n+1} with the same technique as before.

The minimizing α fulfills

$$\begin{aligned} 0 &= \frac{\partial}{\partial\alpha} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\ &= -2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \frac{\partial}{\partial\alpha} ||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})}. \end{aligned}$$

Since

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\,||F_n+\alpha d||^2_{\mathrm{L}^2(\mathcal{B})}\\
=\;&\frac{\partial}{\partial\alpha}\int_{\mathcal{B}}\left(F_n(y)+\alpha d(y)\right)^2\mathrm{d}y\\
=\;&2\int_{\mathcal{B}}F_n(y)d(y)\;\mathrm{d}y+2\alpha\int_{\mathcal{B}}d(y)^2\mathrm{d}y\\
=\;&2\;\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}+2\alpha||d||^2_{\mathrm{L}^2(\mathcal{B})},
\end{aligned}$$

we get

$$\begin{aligned}
0\;=\;&\frac{\partial}{\partial\alpha}\left(||R^n-\alpha\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||F_n+\alpha d||^2_{\mathrm{L}^2(\mathcal{B})}\right)\\
=\;&-2\;\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}+2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda\left(2\;\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}+2\alpha||d||^2_{\mathrm{L}^2(\mathcal{B})}\right)
\end{aligned}$$

and, thus,

$$\alpha=\frac{\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}-\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}.$$

Again, we insert this into the target function of the optimization:

$$\begin{aligned}
&||R^n-\alpha\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||F_n+\alpha d||^2_{\mathrm{L}^2(\mathcal{B})}\\
=\;&||R^n||^2_{\mathbb{R}^l}+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}+2\alpha\left(-\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}+\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}\right)\\
&+\alpha^2\left(||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}\right)\\
=\;&||R^n||^2_{\mathbb{R}^l}+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}\\
&+2\frac{\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}-\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}\left(-\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}+\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}\right)\\
&+\left(\frac{\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}-\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}\right)^2\left(||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}\right)\\
=\;&||R^n||^2_{\mathbb{R}^l}+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}-\frac{\left(\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}-\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}\right)^2}{||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}.
\end{aligned}\tag{4.1}$$

And, consequently, a dictionary element d minimizes

$$||R^n-\alpha\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||F_n+\alpha d||^2_{\mathrm{L}^2(\mathcal{B})}$$

if and only if it maximizes

$$\left|\frac{\langle R^n,\mathcal{F}d\rangle_{\mathbb{R}^l}-\lambda\langle F_n,d\rangle_{\mathrm{L}^2(\mathcal{B})}}{\sqrt{||\mathcal{F}d||^2_{\mathbb{R}^l}+\lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}}\right|.$$

Thus, we get the following algorithm for the regularized version of the Functional Matching Pursuit:

Algorithm 4.1 (Regularized Basic Functional MP (RFMP))
Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ *such that*

$$d_{n+1} \text{ maximizes } \left| \frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})}}{\sqrt{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda ||d||^2_{\mathrm{L}^2(\mathcal{B})}}} \right| \text{ and}$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1} \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda ||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}}.$$

Here, too, we can improve the approximation quality of the method at the expense of the computation time by introducing back-projection or pre-fitting. Furthermore, we can use the same ideas on preprocessing and parallelization to reduce the computational costs as in the unregularized version (see the end of Section 3.2).

Note that, in this work, all dictionary elements $d \in \mathcal{D}$ are normalized to reduce computational costs, i.e. $||d||_{\mathrm{L}^2(\mathcal{B})} = 1$ for all $d \in \mathcal{D}$ due to our choice of the dictionary.

4.2. Theoretical Results

In this section, we examine the main properties of the regularization method. In the following, let $\mathcal{F} : D(\mathcal{F}) \subset \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ be an arbitrary operator and let $\overline{\text{span } \mathcal{D}} = D(\mathcal{F})$. All other requirements on this operator will be addressed in the lemmata and theorems, individually.

In Section 4.2.1, let us first examine the case where the data is given exactly. We will give results on the existence of a solution F^λ of the regularized optimization problem as well as a convergence result for the regularization, i.e. $F^\lambda \to F^+$ for λ tending to 0 from above, where F^+ is a minimum-norm solution of $\mathcal{F}F = y$, i.e. an exact solution of the unregularized problem with minimal norm. In this section, we do not need to consider the stability of the solution, since the data is given exactly for now.

In Section 4.2.2, we include noisy data. We will give results on the three important topics when dealing with regularization methods - the existence and the

stability of the solution as well as the convergence with respect to the regularization parameter.

The results of the following two subsections follow the ideas of [53, 54, 124, 128].

4.2.1. for Exactly Given Data

Let us now denote the expression that is to be minimized by

$$J_\lambda(F, y) := ||\mathcal{F}F - y||^2_{\mathbb{R}^l} + \lambda ||F||^2_{\mathrm{L}^2(\mathcal{B})}$$

where $\lambda > 0$ is the regularization parameter. Furthermore, let the data $y \in \mathbb{R}^l$ be given exactly. Here, F is a series expansion in dictionary elements, i.e. $F = \sum_{k=1}^{\infty} \alpha_k d_k$.

First, we will prove that there always exists a solution to the optimization problem.

Theorem 4.2 (Existence of a Solution)
Let the operator $\mathcal{F}$ be weakly sequentially closed and let $y \in \mathbb{R}^l$ be arbitrary data. Then there always exists a solution $F^\lambda \in D(\mathcal{F})$ such that

$$J_\lambda(F^\lambda, y) = \min_{F \in D(\mathcal{F})} J_\lambda(F, y), \text{ where } F^\lambda = \sum_{k=1}^{\infty} \alpha_k^\lambda d_k^\lambda.$$

Proof.
Obviously, $J_\lambda(F, y) \geq 0$ for all $F \in D(\mathcal{F})$ and $y \in \mathbb{R}^l$ where $\lambda > 0$. Thus, there exists a sequence of solutions $(F^n)_n \subset D(\mathcal{F})$ such that $J_\lambda(F^n, y)$ monotonically converges from above to $\inf_{F \in D(\mathcal{F})} J_\lambda(F, y)$ for n tending to infinity, where we keep $y \in \mathbb{R}^l$ fixed. Note that each F^n is a series expansion in dictionary elements, i.e.

$$F^n = \sum_{k=1}^{\infty} \alpha_k^n d_k^m.$$

First let us consider the norm of F^n and $\mathcal{F}F^n$. From the definition of $J_\lambda(F, y)$, we get

$$||F^n||_{\mathrm{L}^2(\mathcal{B})} = \sqrt{\frac{1}{\lambda}\left(J_\lambda(F^n, y) - ||\mathcal{F}F^n - y||^2_{\mathbb{R}^l}\right)} \leq \sqrt{\frac{1}{\lambda}\, J_\lambda(F^n, y)} \leq \sqrt{\frac{1}{\lambda}\, J_\lambda(F^1, y)}.$$

Using the Cauchy-Schwarz inequality (see Theorem 1.25), we get

$$\begin{aligned}
\left(||\mathcal{F}F^n||_{\mathbb{R}^l} - ||y||_{\mathbb{R}^l}\right)^2 &= ||\mathcal{F}F^n||^2_{\mathbb{R}^l} - 2||\mathcal{F}F^n||_{\mathbb{R}^l}||y||_{\mathbb{R}^l} + ||y||^2_{\mathbb{R}^l} \\
&\leq ||\mathcal{F}F^n||^2_{\mathbb{R}^l} - 2\langle \mathcal{F}F^n, y\rangle_{\mathbb{R}^l} + ||y||^2_{\mathbb{R}^l} \\
&= ||\mathcal{F}F^n - y||^2_{\mathbb{R}^l} \\
&= J_\lambda(F^n, y) - \lambda ||F^n||^2_{\mathrm{L}^2(\mathcal{B})} \\
&\leq J_\lambda(F^n, y)
\end{aligned}$$

and conclude that

$$||\mathcal{F}F^n||_{\mathbb{R}^l} \leq ||y||_{\mathbb{R}^l} + \sqrt{J_\lambda(F^n, y)} \leq ||y||_{\mathbb{R}^l} + \sqrt{J_\lambda(F^1, y)}.$$

Thus, both these norms $||F^n||_{\mathrm{L}^2(\mathcal{B})}$ and $||\mathcal{F}F^n||_{\mathbb{R}^l}$ are bounded with respect to all n and, consequently, both sequences $(F^n)_n$ and $(\mathcal{F}F^n)_n$ have a weakly convergent subsequence (see Theorem 1.28 (Bolzano-Weierstraß)), i.e. there exists a subsequence $(F^{n_i})_i \subset \mathrm{L}^2(\mathcal{B})$ that converges weakly to $\tilde{F} \in \mathrm{L}^2(\mathcal{B})$ for i tending to infinity such that $(\mathcal{F}F^{n_i})_i \subset \mathbb{R}^l$ converges weakly to $\tilde{y} \in \mathbb{R}^l$ for i tending to infinity.

Since $\mathcal{F}$ was assumed to be weakly sequentially closed (see Definition 1.26), it follows that $\tilde{F} \in D(\mathcal{F})$ and $\mathcal{F}\tilde{F} = \tilde{y}$. Furthermore, with the help of Theorem 1.29, we get

$$\begin{aligned}
\inf_{F \in D(\mathcal{F})} J_\lambda(F, y) &= \lim_{i \to \infty} J_\lambda\left(F^{n_i}, y\right) \\
&= \lim_{i \to \infty} \left(||\mathcal{F}F^{n_i} - y||^2_{\mathbb{R}^l} + \lambda ||F^{n_i}||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\
&\geq ||\mathcal{F}\tilde{F} - y||^2_{\mathbb{R}^l} + \lambda ||\tilde{F}||^2_{\mathrm{L}^2(\mathcal{B})} \\
&= J_\lambda\left(\tilde{F}, y\right)
\end{aligned}$$

and can conclude that

$$J_\lambda\left(\tilde{F}, y\right) = \inf_{F \in D(\mathcal{F})} J_\lambda(F, y)$$

which proves the theorem.

□

We know that a linear operator that maps into a space with finite dimensions is compact, i.e. the operators considered in this work are compact, too. Since every linear and compact operator is strongly continuous (see Theorem 1.30), we can conclude that the operators corresponding to our applications are strongly continuous, too, and therefore weakly sequentially closed. Thus, they fulfill the condition imposed in Theorem 4.2. Note that this is the only serious restriction that will occur in this whole section.

Now we know that there exists a solution F^λ of the regularization problem which fulfills

$$J_\lambda(F^\lambda, y) = \min_{F \in D(\mathcal{F})} J_\lambda(F, y). \tag{4.2}$$

However, we need to prove that Algorithm 4.1 (RFMP) actually converges and provides us with such a solution to the optimization problem. Let us consider the sequence

$$\left(||\mathcal{F}F_n - y||^2_{\mathbb{R}^l} + \lambda ||F_n||^2_{\mathrm{L}^2(\mathcal{B})} \right)_n = \left(||R^n||^2_{\mathbb{R}^l} + \lambda ||F_n||^2_{\mathrm{L}^2(\mathcal{B})} \right)_n$$

and let us prove its convergence. Note that

$$\begin{aligned} R^n &= R^{n+1} + \alpha_{n+1}\mathcal{F}d_{n+1} \\ &= R^{n+1} + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}}\ \mathcal{F}d_{n+1} \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} F_{n+1} &= F_n + \alpha_{n+1}d_{n+1} \\ &= F_n + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}}\ d_{n+1}. \end{aligned} \tag{4.4}$$

Theorem 4.3
Let $\mathcal{F}$ be a linear operator and let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 4.1 (RFMP). Then the sequence $\left(||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}\right)_n$, where the residual R^n is given in Equation (4.3) and F_n is given in Equation (4.4), converges for n tending to infinity.

Proof.
Let us consider the sequence element F_{n+1}. With Equation (4.1) we get

$$\begin{aligned} & ||R^{n+1}||^2_{\mathbb{R}^l} + \lambda||F_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})} \\ = \ & ||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})} - \frac{\left(\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}\right)^2}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}} \\ \leq \ & ||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}. \end{aligned}$$

Thus, the sequence $(||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})})_n$ is monotonically decreasing. Since it is bounded from below by 0, we can conclude that it is convergent, too.

□

Next we will prove that the solution F^λ of the regularization problem converges to a solution of the unregularized Algorithm 3.1 (FMP) for λ tending to 0. Let us denote the regularization with $(\mathcal{R}_\lambda)_{\lambda>0}$, where

$$\mathcal{R}_\lambda : \mathbb{R}^l \to \mathrm{L}^2(\mathcal{B}),\ y \mapsto F^\lambda$$

and F^λ is chosen to fulfill Equation (4.2). Clearly, F^λ depends on the data $y \in \mathbb{R}^l$, too. Since the solution to the regularization is not unique, we now choose F^λ to be one representative of these solutions. From Theorem 4.2 we know that such a solution F^λ exists.

Theorem 4.4 (Convergence of the Regularization)
Let the operator $\mathcal{F}$ be weakly sequentially closed and let y be in the range of $\mathcal{F}$. Furthermore, let F^+ be a minimum-norm solution of $\mathcal{F}F = y$. Then the family $(\mathcal{R}_\lambda(y))_{\lambda>0}$ has at least one convergent subsequence. The limit of each of these subsequences is a minimum-norm solution of $\mathcal{F}F = y$.
If the minimum-norm solution F^+ is additionally unique then the regularization converges to F^+ for λ tending to 0, i.e.

$$\lim_{\lambda\to 0^+} ||\mathcal{R}_\lambda(y) - F^+||_{\mathrm{L}^2(\mathcal{B})} = 0.$$

Proof.
From the definition of the regularization we get

$$\begin{aligned} J_\lambda(F^\lambda, y) &= \min_{F\in D(\mathcal{F})} J_\lambda(F, y) \\ &= \min_{F\in D(\mathcal{F})} \left(||\mathcal{F}F - y||^2_{\mathbb{R}^l} + \lambda ||F||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\ &\leq ||\mathcal{F}F^+ - y||^2_{\mathbb{R}^l} + \lambda ||F^+||^2_{\mathrm{L}^2(\mathcal{B})} \\ &= \lambda ||F^+||^2_{\mathrm{L}^2(\mathcal{B})}, \end{aligned}$$

since F^+ is a minimum-norm solution. Thus,

$$||\mathcal{F}F^\lambda - y||^2_{\mathbb{R}^l} + \lambda ||F^\lambda||^2_{\mathrm{L}^2(\mathcal{B})} \leq \lambda ||F^+||^2_{\mathrm{L}^2(\mathcal{B})}$$

and, consequently,

$$\lim_{\lambda\to 0^+} ||\mathcal{F}F^\lambda - y||^2_{\mathbb{R}^l} = 0. \tag{4.5}$$

Furthermore, we can conclude that

$$||F^\lambda||^2_{\mathrm{L}^2(\mathcal{B})} \leq ||F^+||^2_{\mathrm{L}^2(\mathcal{B})} - \frac{1}{\lambda} ||\mathcal{F}F^\lambda - y||^2_{\mathbb{R}^l} \leq ||F^+||^2_{\mathrm{L}^2(\mathcal{B})}$$

which leads to the assertion that

$$\limsup_{\lambda\to 0^+} ||F^\lambda||^2_{\mathrm{L}^2(\mathcal{B})} \leq ||F^+||^2_{\mathrm{L}^2(\mathcal{B})}. \tag{4.6}$$

Thus, the family $(F^\lambda)_{\lambda>0}$ is bounded and there exists a sequence $(F^{\lambda_k})_k$ that weakly converges to $\tilde{F}$ for k tending to infinity (with λ_k tending to 0 from above). Since the operator $\mathcal{F}$ was assumed to be weakly sequentially closed and $\lim_{k\to\infty} \mathcal{F}F^{\lambda_k} = y$ due to Equation (4.5), we can conclude that $\tilde{F} \in D(\mathcal{F})$ and $\mathcal{F}\tilde{F} = y$.

Since $(F^{\lambda_k})_k$ converges weakly to $\tilde{F}$ for k tending to infinity we know that

$$||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})} \leq \liminf_{k\to\infty} ||F^{\lambda_k}||_{\mathrm{L}^2(\mathcal{B})} \leq ||F^+||_{\mathrm{L}^2(\mathcal{B})} \tag{4.7}$$

where the first inequality stems from Theorem 1.29 while the second inequality stems from Inequality (4.6). Since F^+ is a minimum-norm solution, it follows that

$$||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})} = ||F^+||_{\mathrm{L}^2(\mathcal{B})}$$

and, thus, $\tilde{F}$ is a minimum-norm solution, too.

Moreover, Inequalities (4.6) and (4.7) imply that $\lim_{k\to\infty} ||F^{\lambda_k}||_{\mathrm{L}^2(\mathcal{B})} = ||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})}$ such that the convergence of F^{λ_k} to $\tilde{F}$ for k tending to infinity is strong.

In the second part of the proof, we consider the unique minimum-norm solution F^+. Clearly, each subsequence of $(F_\lambda)_{\lambda>0}$ converges to this unique solution in its norm. Thus, we can conclude that the whole family converges, too, since the limit F^+ is unique.

□

4.2.2. for Noisy Data

Of course, regularization techniques are mostly used to deal with data from measurements, i.e. noisy data y^ε where $||y - y^\varepsilon||_{\mathbb{R}^l} \leq \varepsilon$. Again, we first give a result concerned with the existence of a solution of the regularized algorithm where noisy data is considered.

Theorem 4.5 (Existence of a Solution)
Let the operator $\mathcal{F}$ be weakly sequentially closed. Furthermore, let the noisy data $y^\varepsilon \in \mathbb{R}^l$ be given. Then there always exists a solution $F^{\lambda,\varepsilon} \in D(\mathcal{F})$ such that

$$J_\lambda(F^{\lambda,\varepsilon}, y^\varepsilon) = \min_{F\in D(\mathcal{F})} J_\lambda(F, y^\varepsilon), \text{ where } F^{\lambda,\varepsilon} = \sum_{k=1}^{\infty} \alpha_k^{\lambda,\varepsilon} d_k^{\lambda,\varepsilon}.$$

Proof.
See the proof of Theorem 4.2.

□

Theorem 4.6
Let $\mathcal{F}$ be a linear and continuous operator and let all dictionary elements and the corresponding coefficients be chosen according to Algorithm 4.1 (RFMP). Then $(||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})})_n$, where the residual is given by $R^n := y^\varepsilon - \mathcal{F}F_n$ and F_n is given in Equation (4.4), converges for n tending to infinity.

Proof.
We refer to the proof of Theorem 4.3, since the introduction of noisy data has no influence on the result.

□

The family $\{\mathcal{R}_\lambda\}_{\lambda>0}$, indeed, represents a regularization due to the convergence result in Theorem 4.4 and the continuity of each $\mathcal{R}_\lambda$, which is proven in the following theorem.

Theorem 4.7 (Stability of the Regularized Solution)
Let the operator $\mathcal{F}$ be weakly sequentially closed and let $y^\varepsilon \in \mathbb{R}^l$ be given noisy data. Furthermore, let $(y^{\varepsilon_k})_k \subset \mathbb{R}^l$ be a sequence that converges to y^ε for k tending to infinity. Let $(F^{\lambda,\varepsilon_k})_k$ be a corresponding sequence of minimizing elements of $J_\lambda(\cdot, y^{\varepsilon_k})$. Then there exists a convergent subsequence of $(F^{\lambda,\varepsilon_k})_k$ and every convergent subsequence converges to a minimizing element $F^{\lambda,\varepsilon}$ of $J_\lambda(\cdot, y^\varepsilon)$.
Moreover, if $F^{\lambda,\varepsilon}$ is unique then $(F^{\lambda,\varepsilon_k})_k$ converges to $F^{\lambda,\varepsilon}$ for k tending to infinity.

Proof.
From the definition of $F^{\lambda,\varepsilon_k}$, we get for all $F \in D(\mathcal{F})$ that

$$||\mathcal{F}F^{\lambda,\varepsilon_k} - y^{\varepsilon_k}||^2_{\mathbb{R}^l} + \lambda||F^{\lambda,\varepsilon_k}||^2_{\mathrm{L}^2(\mathcal{B})} = J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k}) \leq J_\lambda(F, y^{\varepsilon_k}). \tag{4.8}$$

Similarly to the proof of Theorem 4.2, we get that

$$||F^{\lambda,\varepsilon_k}||_{\mathrm{L}^2(\mathcal{B})} = \sqrt{\frac{1}{\lambda}\left(J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k}) - ||\mathcal{F}F^{\lambda,\varepsilon_k} - y^{\varepsilon_k}||^2_{\mathbb{R}^l}\right)} \leq \sqrt{\frac{1}{\lambda}J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k})}$$

and

$$||\mathcal{F}F^{\lambda,\varepsilon_k}||_{\mathbb{R}^l} \leq ||y^{\varepsilon_k}||_{\mathbb{R}^l} + \sqrt{J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k})},$$

since, with the Cauchy-Schwarz inequality (see Theorem 1.25),

$$\begin{aligned}
\left(||\mathcal{F}F^{\lambda,\varepsilon_k}||_{\mathbb{R}^l} - ||y^{\varepsilon_k}||_{\mathbb{R}^l}\right)^2 &= ||\mathcal{F}F^{\lambda,\varepsilon_k}||^2_{\mathbb{R}^l} - 2||\mathcal{F}F^{\lambda,\varepsilon_k}||_{\mathbb{R}^l}||y^{\varepsilon_k}||_{\mathbb{R}^l} + ||y^{\varepsilon_k}||^2_{\mathbb{R}^l} \\
&\leq ||\mathcal{F}F^{\lambda,\varepsilon_k}||^2_{\mathbb{R}^l} - 2\langle \mathcal{F}F^{\lambda,\varepsilon_k}, y^{\varepsilon_k}\rangle_{\mathbb{R}^l} + ||y^{\varepsilon_k}||^2_{\mathbb{R}^l} \\
&= ||\mathcal{F}F^{\lambda,\varepsilon_k} - y^{\varepsilon_k}||^2_{\mathbb{R}^l} \\
&= J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k}) - \lambda||F^{\lambda,\varepsilon_k}||^2_{\mathrm{L}^2(\mathcal{B})} \\
&\leq J_\lambda(F^{\lambda,\varepsilon_k}, y^{\varepsilon_k}).
\end{aligned}$$

With Inequality (4.8) we further conclude for all $F \in D(\mathcal{F})$ that

$$\begin{aligned}
||F^{\lambda,\varepsilon_k}||_{\mathrm{L}^2(\mathcal{B})} &\leq \sqrt{\frac{1}{\lambda}J_\lambda(F, y^{\varepsilon_k})} \\
&= \sqrt{\frac{1}{\lambda}||\mathcal{F}F - y^{\varepsilon_k}||^2_{\mathbb{R}^l} + ||F||^2_{\mathrm{L}^2(\mathcal{B})}} \\
&\leq \sqrt{\frac{1}{\lambda}\left(||\mathcal{F}F||_{\mathbb{R}^l} + ||y^{\varepsilon_k}||_{\mathbb{R}^l}\right)^2 + ||F||^2_{\mathrm{L}^2(\mathcal{B})}}.
\end{aligned}$$

Moreover, the definition of $F^{\lambda,\varepsilon_k}$ yields for all $F \in D(\mathcal{F})$ that

$$\begin{aligned} ||\mathcal{F}F^{\lambda,\varepsilon_k}||_{\mathbb{R}^l} &\leq ||y^{\varepsilon_k}||_{\mathbb{R}^l} + \sqrt{J_\lambda(F, y^{\varepsilon_k})} \\ &\leq ||y^{\varepsilon_k}||_{\mathbb{R}^l} + \sqrt{\left(||\mathcal{F}F||_{\mathbb{R}^l} + ||y^{\varepsilon_k}||_{\mathbb{R}^l}\right)^2 + \lambda||F||^2_{\mathrm{L}^2(\mathcal{B})}}. \end{aligned}$$

Thus, both $||F^{\lambda,\varepsilon_k}||_{\mathrm{L}^2(\mathcal{B})}$ and $||\mathcal{F}F^{\lambda,\varepsilon_k}||_{\mathbb{R}^l}$ are bounded for all k since $(y^{\varepsilon_k})_k$ is convergent and, consequently, $||y^{\varepsilon_k}||_{\mathbb{R}^l}$ is bounded, too. With Theorem 1.28 (Bolzano-Weierstraß) and the weak sequential closedness of the operator $\mathcal{F}$ we know that both $(F^{\lambda,\varepsilon_k})_k$ and $(\mathcal{F}F^{\lambda,\varepsilon_k})_k$ have weakly convergent subsequences where $F^{\lambda,\varepsilon_{k_i}}$ converges weakly to $\tilde{F} \in D(\mathcal{F})$ and $\mathcal{F}F^{\lambda,\varepsilon_{k_i}}$ converges weakly to $\mathcal{F}\tilde{F}$.

With Theorem 1.29, we get that

$$||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})} \leq \liminf_{i\to\infty} ||F^{\lambda,\varepsilon_{k_i}}||_{\mathrm{L}^2(\mathcal{B})} \text{ and} \tag{4.9}$$

$$||\mathcal{F}\tilde{F} - y^\varepsilon||_{\mathbb{R}^l} \leq \liminf_{i\to\infty} ||\mathcal{F}F^{\lambda,\varepsilon_{k_i}} - y^{\varepsilon_{k_i}}||_{\mathbb{R}^l} \tag{4.10}$$

and we conclude for all $F \in D(\mathcal{F})$ that

$$\begin{aligned} J_\lambda(\tilde{F}, y^\varepsilon) &= ||\mathcal{F}\tilde{F} - y^\varepsilon||^2_{\mathbb{R}^l} + \lambda||\tilde{F}||^2_{\mathrm{L}^2(\mathcal{B})} \\ &\leq \liminf_{i\to\infty} \left(||\mathcal{F}F^{\lambda,\varepsilon_{k_i}} - y^{\varepsilon_{k_i}}||^2_{\mathbb{R}^l} + \lambda||F^{\lambda,\varepsilon_{k_i}}||^2_{\mathrm{L}^2(\mathcal{B})}\right) \\ &= \liminf_{i\to\infty} J_\lambda(F^{\lambda,\varepsilon_{k_i}}, y^{\varepsilon_{k_i}}) \\ &\leq \limsup_{i\to\infty} J_\lambda(F^{\lambda,\varepsilon_{k_i}}, y^{\varepsilon_{k_i}}) \\ &\leq \lim_{i\to\infty} J_\lambda(F, y^{\varepsilon_{k_i}}) \\ &= J_\lambda(F, y^\varepsilon). \end{aligned}$$

Thus, $\tilde{F}$ minimizes $J_\lambda(\cdot, y^\varepsilon)$. Setting $F = \tilde{F}$ above, we get

$$\lim_{i\to\infty} J_\lambda(F^{\lambda,\varepsilon_{k_i}}, y^{\varepsilon_{k_i}}) = J_\lambda(\tilde{F}, y^\varepsilon). \tag{4.11}$$

Now, let us assume that $(F^{\lambda,\varepsilon_{k_i}})_i$ does not converge (strongly) to $\tilde{F}$ for i tending to infinity. Due to Inequality (4.9), we then get that

$$\limsup_{i\to\infty} ||F^{\lambda,\varepsilon_{k_i}}||_{\mathrm{L}^2(\mathcal{B})} > ||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})}. \tag{4.12}$$

Furthermore, there exists a subsequence $(F^{\lambda,\varepsilon_{k_{i_j}}})_j$ that satisfies

$$\lim_{j\to\infty} ||F^{\lambda,\varepsilon_{k_{i_j}}}||_{\mathrm{L}^2(\mathcal{B})} = \limsup_{i\to\infty} ||F^{\lambda,\varepsilon_{k_i}}||_{\mathrm{L}^2(\mathcal{B})}.$$

With Equation (4.11), we conclude that

$$\begin{aligned}\lim_{j\to\infty}\left(||\mathcal{F}F^{\lambda,\varepsilon_{k_{i_j}}}-y^{\varepsilon_{k_{i_j}}}||_{\mathbb{R}^l}^2+\lambda||F^{\lambda,\varepsilon_{k_{i_j}}}||_{\mathrm{L}^2(\mathcal{B})}^2\right) &= \lim_{j\to\infty}J_\lambda(F^{\lambda,\varepsilon_{k_{i_j}}},y^{\varepsilon_{k_{i_j}}})\\ &= J_\lambda(\tilde{F},y^\varepsilon)\\ &= ||\mathcal{F}\tilde{F}-y^\varepsilon||_{\mathbb{R}^l}^2+\lambda||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})}^2.\end{aligned}$$

Now, with Inequality (4.12) and the convergence of the norm of the subsequence we get that

$$\begin{aligned}\lim_{j\to\infty}||\mathcal{F}F^{\lambda,\varepsilon_{k_{i_j}}}-y^{\varepsilon_{k_{i_j}}}||_{\mathbb{R}^l}^2 &= ||\mathcal{F}\tilde{F}-y^\varepsilon||^2+\lambda\left(||\tilde{F}||_{\mathrm{L}^2(\mathcal{B})}^2-\limsup_{i\to\infty}||F^{\lambda,\varepsilon_{k_i}}||_{\mathrm{L}^2(\mathcal{B})}^2\right)\\ &< ||\mathcal{F}\tilde{F}-y^\varepsilon||_{\mathbb{R}^l}^2\end{aligned}$$

which is a contradiction to Inequality (4.10) and, thus, we proved the assumption wrong and get that $(F^{\lambda,\varepsilon_{k_i}})_i$ (strongly) converges to $\tilde{F}$ for i tending to infinity which proves the first part of the theorem.

If $F^{\lambda,\varepsilon}$ is unique we, again, conclude that each subsequence of $(F^{\lambda,\varepsilon_k})_k$ converges to $F^{\lambda,\varepsilon}$. Therefore, the whole sequence converges to $F^{\lambda,\varepsilon}$, too.

□

For the case of noisy data we will denote the regularization with $(\mathcal{R}^\eta_{\lambda(\varepsilon)})_{\lambda(\varepsilon)>0}$ where

$$\mathcal{R}^\eta_{\lambda(\varepsilon)}:\mathbb{R}^l\to\mathrm{L}^2(\mathcal{B}),\ y^\varepsilon\mapsto F^{\lambda(\varepsilon),\varepsilon,\eta}$$

and $F^{\lambda(\varepsilon),\varepsilon,\eta}$ is chosen such that

$$J_{\lambda(\varepsilon)}(F^{\lambda(\varepsilon),\varepsilon,\eta},y^\varepsilon)\leq\min_{F\in D(\mathcal{F})}J_{\lambda(\varepsilon)}(F,y^\varepsilon)+\eta.$$

Note that the regularization parameter λ was replaced by the function λ that depends on the data error ε. We use $\eta=\eta(\lambda(\varepsilon))$ to describe the error that arises when minimizing numerically. From Theorem 4.5, we already know that there exists a solution $F^{\lambda(\varepsilon),\varepsilon,\eta}$ of the regularized problem with noisy data as input.

The following theorem may be considered as a fusion of Theorems 4.4 and 4.7.

Theorem 4.8 (Convergence of the Regularization)
Let the operator $\mathcal{F}$ be weakly sequentially closed and let y in the range of $\mathcal{F}$ be exactly given data. Furthermore, let F^+ be a minimum-norm solution of $\mathcal{F}F=y$. Let the family $(y^\varepsilon)_{\varepsilon>0}$ fulfill $||y-y^\varepsilon||_{\mathbb{R}^l}\leq\varepsilon$. Let us choose $\lambda:]0,\infty[\to]0,\infty[$ such that $\lambda(\varepsilon)$ and $\frac{\varepsilon^2}{\lambda(\varepsilon)}$ tend to 0 for ε tending to 0 from above. Let the error η satisfy $\eta=\eta(\lambda(\varepsilon))=o(\lambda(\varepsilon))$ for ε tending to 0 from above where o is a Landau symbol.

Then the family $\left(\mathcal{R}_{\lambda(\varepsilon)}^{\eta(\lambda(\varepsilon))}(y^\varepsilon)\right)_{\varepsilon>0}$ *has at least one convergent subsequence and the limit of each of these subsequences is a minimum-norm solution of* $\mathcal{F}F = y$. *If the minimum-norm solution* F^+ *is additionally unique then the regularization converges to* F^+ *in its norm for* ε *tending to* 0 *from above, i.e.*

$$\lim_{\varepsilon\to 0^+} \left|\left|\mathcal{R}_{\lambda(\varepsilon)}^{\eta(\lambda(\varepsilon))}(y^\varepsilon) - F^+\right|\right|_{\mathrm{L}^2(\mathcal{B})} = 0.$$

Proof.
From the definition of the regularization, we get

$$\begin{aligned} J_{\lambda(\varepsilon)}(F^{\lambda(\varepsilon),\varepsilon,\eta}, y^\varepsilon) &\leq \min_{F\in D(\mathcal{F})} J_{\lambda(\varepsilon)}(F, y^\varepsilon) + \eta \\ &= \min_{F\in D(\mathcal{F})} \left(||\mathcal{F}F - y^\varepsilon||_{\mathbb{R}^l}^2 + \lambda(\varepsilon)||F||_{\mathrm{L}^2(\mathcal{B})}^2\right) + \eta \\ &\leq ||\mathcal{F}F^+ - y^\varepsilon||_{\mathbb{R}^l}^2 + \lambda(\varepsilon)||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \eta \\ &= ||y - y^\varepsilon||_{\mathbb{R}^l}^2 + \lambda(\varepsilon)||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \eta \\ &\leq \varepsilon^2 + \lambda(\varepsilon)||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \eta. \end{aligned}$$

Thus, we conclude that

$$||\mathcal{F}F^{\lambda(\varepsilon),\varepsilon,\eta} - y^\varepsilon||_{\mathbb{R}^l}^2 + \lambda(\varepsilon)||F^{\lambda(\varepsilon),\varepsilon,\eta}||_{\mathrm{L}^2(\mathcal{B})}^2 \leq \varepsilon^2 + \lambda(\varepsilon)||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \eta(\lambda(\varepsilon)).$$

Note that, according to our requirements, $\eta(\lambda(\varepsilon)) = o(\lambda(\varepsilon))$ for ε tending to 0 from above, i.e. $\frac{\eta(\lambda(\varepsilon))}{\lambda(\varepsilon)}$ tends to 0 for ε tending to 0 from above and, thus, for $\lambda(\varepsilon)$ tending to 0 from above, too. Furthermore, this implies that η tends to 0 for ε tending to 0 from above. Thus, we get for the limit

$$\lim_{\varepsilon\to 0^+} ||\mathcal{F}F^{\lambda(\varepsilon),\varepsilon,\eta} - y^\varepsilon||_{\mathbb{R}^l}^2 = 0$$

and, consequently,

$$\lim_{\varepsilon\to 0^+} ||\mathcal{F}F^{\lambda(\varepsilon),\varepsilon,\eta} - y||_{\mathbb{R}^l} \leq \lim_{\varepsilon\to 0^+} \left(||\mathcal{F}F^{\lambda(\varepsilon),\varepsilon,\eta} - y^\varepsilon||_{\mathbb{R}^l} + ||y^\varepsilon - y||_{\mathbb{R}^l}\right) = 0 \quad (4.13)$$

Using the assumptions on λ, we conclude that

$$\begin{aligned} ||F^{\lambda(\varepsilon),\varepsilon,\eta}||_{\mathrm{L}^2(\mathcal{B})}^2 &\leq \frac{\varepsilon^2}{\lambda(\varepsilon)} + ||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \frac{\eta(\lambda(\varepsilon))}{\lambda(\varepsilon)} - \frac{1}{\lambda(\varepsilon)}||\mathcal{F}F^{\lambda(\varepsilon),\varepsilon,\eta} - y^\varepsilon||_{\mathbb{R}^l}^2 \\ &\leq \frac{\varepsilon^2}{\lambda(\varepsilon)} + ||F^+||_{\mathrm{L}^2(\mathcal{B})}^2 + \frac{\eta(\lambda(\varepsilon))}{\lambda(\varepsilon)}, \end{aligned}$$

which leads to the assertion that

$$\limsup_{\varepsilon\to 0^+} ||F^{\lambda(\varepsilon),\varepsilon,\eta}||_{\mathrm{L}^2(\mathcal{B})}^2 \leq ||F^+||_{\mathrm{L}^2(\mathcal{B})}^2.$$

Thus, $(F^{\lambda(\varepsilon),\varepsilon,\eta})_{\varepsilon>0}$ is bounded and there exists a subsequence $(F^{\lambda(\varepsilon_k),\varepsilon_k,\eta_k})_k$ with $\eta_k := \eta(\lambda(\varepsilon_k))$ that weakly converges to $\tilde{F}$ for k tending to infinity. Since the operator $\mathcal{F}$ was assumed to be weakly sequentially closed and Equation (4.13) is valid we can conclude that $\tilde{F} \in D(\mathcal{F})$ and $\mathcal{F}\tilde{F} = y$.

Thus, we have the same setup as in the proof to Theorem 4.4 and refer to that for the rest of the proof.

□

4.3. Implementation of the Regularized Basic Functional Matching Pursuit

Remember that, in step $n+1$, we determine the dictionary element and the corresponding coefficient fulfilling

$$(d_{n+1}, \alpha_{n+1}) = \operatorname{argmin}_{d\in\mathcal{D},\alpha\in\mathbb{R}} \left(||R^n + \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right).$$

First let us recapitulate Algorithm 4.1 (RFMP) to see which expressions we need to consider in this section.

Algorithm 4.9 (RFMP)
Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1}d_{n+1}$ *such that*

$$d_{n+1} \text{ maximizes } \left| \frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})}}{\sqrt{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}} \right| \text{ and}$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}}.$$

Note that, in our case, $||d||_{\mathrm{L}^2(\mathcal{B})} = 1$ for all $d \in \mathcal{D}$. Furthermore, we already computed $\mathcal{F}d$ and its norm in Section 3.5. We only need to consider $\langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})}$ for all $d \in \mathcal{D}$.

Using the Dictionary Elements of Type I

Since

$$\langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})} = \left\langle \sum_{i=1}^{n} \alpha_i d_i, d \right\rangle_{\mathrm{L}^2(\mathcal{B})} = \sum_{i=1}^{n} \alpha_i \langle d_i, d\rangle_{\mathrm{L}^2(\mathcal{B})},$$

let us consider the scalar products $\langle d_i, d\rangle_{\mathrm{L}^2(\mathcal{B})}$. We get the following three combinations of dictionary elements where $x, \tilde{x} \in \mathcal{B}$:

$$
\begin{aligned}
\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{\tilde{m},\tilde{n},\tilde{j}} \rangle_{\mathrm{L}^2(\mathcal{B})} &= \delta_{m\tilde{m}}\ \delta_{n\tilde{n}}\ \delta_{j\tilde{j}}, \\
\langle G^{\mathrm{I}}_{m,n,j}, \tilde{K}^{\mathrm{I}}_h(x,\cdot) \rangle_{\mathrm{L}^2(\mathcal{B})} &= \int_{\mathcal{B}} G^{\mathrm{I}}_{m,n,j}(y)\ \tilde{K}^{\mathrm{I}}_h(x,y)\ \mathrm{d}y \\
&= \sum_{\tilde{m}=0}^{E_{\tilde{m}}} \sum_{\tilde{n}=0}^{E_{\tilde{n}}} \sum_{\tilde{j}=1}^{2\tilde{n}+1} (A^h_{\tilde{m},\tilde{n}})^2\ G^{\mathrm{I}}_{\tilde{m},\tilde{n},\tilde{j}}(x) \int_{\mathcal{B}} G^{\mathrm{I}}_{m,n,j}(y)\ G^{\mathrm{I}}_{\tilde{m},\tilde{n},\tilde{j}}(y)\ \mathrm{d}y \\
&= \sum_{\tilde{m}=0}^{E_{\tilde{m}}} \sum_{\tilde{n}=0}^{E_{\tilde{n}}} \sum_{\tilde{j}=1}^{2\tilde{n}+1} (A^h_{\tilde{m},\tilde{n}})^2\ G^{\mathrm{I}}_{\tilde{m},\tilde{n},\tilde{j}}(x)\ \delta_{m\tilde{m}}\ \delta_{n\tilde{n}}\ \delta_{j\tilde{j}} \\
&= (A^h_{m,n})^2 \sqrt{\frac{4m+2n+3}{a^3}}\ P^{(0,n+1/2)}_m\left(2\frac{|x|^2}{a^2}-1\right) \\
&\qquad \times Y_{n,j}\left(\frac{x}{|x|}\right)\left(\frac{|x|}{a}\right)^n,
\end{aligned}
$$

and

$$
\begin{aligned}
\left\langle \tilde{K}^{\mathrm{I}}_h(x,\cdot), \tilde{K}^{\mathrm{I}}_{\tilde{h}}(\tilde{x},\cdot) \right\rangle_{\mathrm{L}^2(\mathcal{B})} &= \int_{\mathcal{B}} \tilde{K}^{\mathrm{I}}_h(x,y)\ \tilde{K}^{\mathrm{I}}_{\tilde{h}}(\tilde{x},y)\ \mathrm{d}y \\
&= \sum_{m=0}^{E_m} \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} (A^h_{m,n})^2\ (A^{\tilde{h}}_{m,n})^2\ G^{\mathrm{I}}_{m,n,j}(x)\ G^{\mathrm{I}}_{m,n,j}(\tilde{x}) \\
&= \sum_{m=0}^{E_m} \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} (A^h_{m,n})^2\ (A^{\tilde{h}}_{m,n})^2\ \frac{4m+2n+3}{a^3} \\
&\qquad \times Y_{n,j}\left(\frac{x}{|x|}\right)\ Y_{n,j}\left(\frac{\tilde{x}}{|\tilde{x}|}\right)\ \left(\frac{|x|\ |\tilde{x}|}{a^2}\right)^n \\
&\qquad \times P^{(0,n+1/2)}_m\left(2\frac{|x|^2}{a^2}-1\right) P^{(0,n+1/2)}_m\left(2\frac{|\tilde{x}|^2}{a^2}-1\right) \\
&= \sum_{m=0}^{E_m} \sum_{n=0}^{E_n} (A^h_{m,n})^2\ (A^{\tilde{h}}_{m,n})^2\ \frac{2n+1}{4\pi}\ \frac{4m+2n+3}{a^3} \\
&\qquad \times P_n\left(\frac{x}{|x|}\cdot\frac{\tilde{x}}{|\tilde{x}|}\right)\ \left(\frac{|x|\ |\tilde{x}|}{a^2}\right)^n \\
&\qquad \times P^{(0,n+1/2)}_m\left(2\frac{|x|^2}{a^2}-1\right) P^{(0,n+1/2)}_m\left(2\frac{|\tilde{x}|^2}{a^2}-1\right).
\end{aligned}
$$

For the normalized dictionary elements we get

$$\begin{aligned}
\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{\tilde{m},\tilde{n},\tilde{j}} \rangle_{\mathrm{L}^2(\mathcal{B})} &= \delta_{m\tilde{m}}\delta_{n\tilde{n}}\ \delta_{j\tilde{j}}, \\
\langle G^{\mathrm{I}}_{m,n,j}, K^{\mathrm{I}}_h(x,\cdot) \rangle_{\mathrm{L}^2(\mathcal{B})} &= \frac{1}{||\tilde{K}^{\mathrm{I}}_h(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}\ \langle G^{\mathrm{I}}_{m,n,j}, \tilde{K}^{\mathrm{I}}_h(x,\cdot) \rangle_{\mathrm{L}^2(\mathcal{B})} \\
&= \frac{1}{||\tilde{K}^{\mathrm{I}}_h(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})}}\ (A^h_{m,n})^2\ \sqrt{\frac{4m+2n+3}{a^3}}\ Y_{n,j}\left(\frac{x}{|x|}\right) \\
&\quad \times\ P^{(0,n+1/2)}_m\left(2\frac{|x|^2}{a^2}-1\right)\ \left(\frac{|x|}{a}\right)^n
\end{aligned}$$

and

$$\begin{aligned}
&\left\langle K^{\mathrm{I}}_h(x,\cdot), K^{\mathrm{I}}_{\tilde{h}}(\tilde{x},\cdot) \right\rangle_{\mathrm{L}^2(\mathcal{B})} \\
= &\ \frac{1}{||\tilde{K}^{\mathrm{I}}_h(x,\cdot)||_{\mathrm{L}^2(\mathcal{B})} ||\tilde{K}^{\mathrm{I}}_{\tilde{h}}(\tilde{x},\cdot)||_{\mathrm{L}^2(\mathcal{B})}} \\
&\times \sum_{m=0}^{E_m}\sum_{n=0}^{E_n} (A^h_{m,n})^2\ (A^{\tilde{h}}_{m,n})^2\ \frac{2n+1}{4\pi}\ \frac{4m+2n+3}{a^3}\ P_n\left(\frac{x}{|x|}\cdot\frac{\tilde{x}}{|\tilde{x}|}\right)\ \left(\frac{|x|\ |\tilde{x}|}{a^2}\right)^n \\
&\qquad \times\ P^{(0,n+1/2)}_m\left(2\frac{|x|^2}{a^2}-1\right) P^{(0,n+1/2)}_m\left(2\frac{|\tilde{x}|^2}{a^2}-1\right).
\end{aligned}$$

Using the Dictionary Elements of Type II

Again we get three combinations of dictionary elements of type II where $x, \tilde{x} \in \mathcal{B}$.

$$\begin{aligned}
\langle G^{\mathrm{II}}_{m,n,j}, G^{\mathrm{II}}_{\tilde{m},\tilde{n},\tilde{j}} \rangle_{\mathrm{L}^2(\mathcal{B})} &= \delta_{m\tilde{m}}\ \delta_{n\tilde{n}}\ \delta_{j\tilde{j}}, \\
\langle G^{\mathrm{II}}_{m,n,j}, \tilde{K}^{\mathrm{II}}_h(x,\cdot) \rangle_{\mathrm{L}^2(\mathcal{B})} &= \int_{\mathcal{B}} G^{\mathrm{II}}_{m,n,j}(y)\ \tilde{K}^{\mathrm{II}}_h(x,y)\ \mathrm{d}y \\
&= (A^h_{m,n})^2\ \sqrt{\frac{2m+3}{a^3}}\ Y_{n,j}\left(\frac{x}{|x|}\right) P^{(0,2)}_m\left(2\frac{|x|}{a}-1\right),
\end{aligned}$$

and

$$\begin{aligned}
&\left\langle \tilde{K}_h^{\mathrm{II}}(x,\cdot), \tilde{K}_{\tilde{h}}^{\mathrm{II}}(\tilde{x},\cdot) \right\rangle_{\mathrm{L}^2(\mathcal{B})} \\
= &\sum_{m=0}^{E_m} \sum_{n=0}^{E_n} \left(A_{m,n}^h\right)^2 \left(A_{m,n}^{\tilde{h}}\right)^2 \frac{2n+1}{4\pi} \frac{2m+3}{a^3} P_n\left(\frac{x}{|x|}\cdot\frac{\tilde{x}}{|\tilde{x}|}\right) \\
&\quad \times P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) P_m^{(0,2)}\left(2\frac{|\tilde{x}|}{a}-1\right) \\
= &\left(\sum_{n=0}^{E_n} \left(A_n^h\right)^2 \left(A_n^{\tilde{h}}\right)^2 \frac{2n+1}{4\pi} P_n\left(\frac{x}{|x|}\cdot\frac{\tilde{x}}{|\tilde{x}|}\right)\right) \\
&\times \left(\sum_{m=0}^{E_m} \left(A_m^h\right)^2 \left(A_m^{\tilde{h}}\right)^2 \frac{2m+3}{a^3} P_m^{(0,2)}\left(2\frac{|x|}{a}-1\right) P_m^{(0,2)}\left(2\frac{|\tilde{x}|}{a}-1\right)\right).
\end{aligned}$$

Of course, these expressions have to be normalized, too, as done above.

4.4. Alternative: The Least Mixed Norms Problem (MRFMP)

Solving the least squares problem to get a viable solution is the classical approach, i.e. we minimize the amount of energy in the system by minimizing the L^2-norm. Although this approach can be mathematically realized pretty easily it often leads to poor results in practical applications.

Algorithm 4.1 (RFMP) already provides us with a solution that displays some properties of a sparse solution. Nonetheless, we may enforce sparseness by minimizing the L^0-norm, i.e. by maximizing the number of zero coefficients. However, for most applications, this is a computationally infeasible problem since it is NP-hard. We may replace the L^0-norm with a L^1-norm (see [21, 22, 23]) to solve an easier problem. If we consider the solution of our problem to be a vector, as usual in the field of compressive sensing, we can find the one with the smallest l^1-norm with already existing, very efficient methods. In our case, however, the solution is still a function existing in $\mathrm{L}^2(\mathcal{B})$ and although it is relatively easy to compute the L^2-norm of the solution, it is not as easy to compute its L^1-norm. Anyway, we cannot expect the behavior corresponding to the l^1-norm to translate to the L^1-setting.

Nonetheless, we will try to derive the algorithm corresponding to the least mixed norms problem.

For practical reasons, let us assume that the set $\mathcal{N}_\alpha := \{y \in \mathcal{B} \mid F_n(y)+\alpha d(y) = 0\}$ is of measure 0 for fixed $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}_0$.

We want to find

$$(d_{n+1}, \alpha_{n+1}) = \text{argmin}_{d\in\mathcal{D},\alpha\in\mathbb{R}} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||_{\mathrm{L}^1(\mathcal{B})} \right),$$

in the step from n to $n+1$ chosen expansion functions. In the following, we will see that we cannot derive the optimal dictionary element d_{n+1} and the corresponding coefficient α_{n+1} with the same technique as before.

The minimizing α fulfills

$$\begin{aligned} 0 &= \frac{\partial}{\partial\alpha} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||_{\mathrm{L}^1(\mathcal{B})} \right) \\ &= -2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda\frac{\partial}{\partial\alpha}||F_n + \alpha d||_{\mathrm{L}^1(\mathcal{B})}. \end{aligned}$$

Since

$$\frac{\partial}{\partial\alpha}||F_n + \alpha d||_{\mathrm{L}^1(\mathcal{B})} = \frac{\partial}{\partial\alpha}\int_{\mathcal{B}} |F_n(y) + \alpha d(y)| \ \mathrm{d}y,$$

we need to know whether we are allowed to exchange the derivative and the integration. Remember the theorem on differentiation under the integral sign (see Theorem 1.31).

Of course, the function $f : \mathcal{B}\times\mathbb{R} \to \mathbb{R},\ (y,\alpha) \mapsto |F_n(y) + \alpha d(y)|$ is integrable on $\mathcal{B}$ for fixed α.

For the fixed value $\alpha_0 \in \mathbb{R}$ we get for all $y \in \mathcal{B}$ and all $\alpha \in \mathbb{R}$ that

$$\begin{aligned} \left|\frac{f(y,\alpha) - f(y,\alpha_0)}{\alpha - \alpha_0}\right| &= \left|\frac{|F_n(y) + \alpha d(y)| - |F_n(y) + \alpha_0 d(y)|}{\alpha - \alpha_0}\right| \\ &\leq \frac{|\alpha d(y) - \alpha_0 d(y)|}{|\alpha - \alpha_0|} = |d(y)| \end{aligned}$$

where we used the inequality $|\ |x| - |y|\ | \leq |x - y|$ for $x, y \in \mathbb{R}$. Note that $|d(y)|$ is an integrable function only depending on y since $d \in \mathrm{L}^1(\mathcal{B})$.

The requirement that the partial derivative exists for all $y \in \mathcal{B} \setminus \mathcal{N}_{\alpha_0}$ is fulfilled, too, since

$$\frac{\partial}{\partial\alpha} f(y,\alpha)\Big|_{\alpha=\alpha_0} = \frac{\partial}{\partial\alpha}|F_n(y) + \alpha d(y)|\Big|_{\alpha=\alpha_0} = \text{sgn}\left(F_n(y) + \alpha_0 d(y)\right) d(y).$$

Thus, all requirements of Theorem 1.31 are fulfilled and we may exchange the derivative and the integration. Consequently, we get that

$$\begin{aligned}
\frac{\partial}{\partial\alpha}||F_n+\alpha d||_{\mathrm{L}^1(\mathcal{B})} &= \frac{\partial}{\partial\alpha}\int_{\mathcal{B}}|F_n(y)+\alpha d(y)|\ \mathrm{d}y \\
&= \int_{\mathcal{B}\setminus\mathcal{N}_\alpha}\frac{\partial}{\partial\alpha}|F_n(y)+\alpha d(y)|\ \mathrm{d}y \\
&= \int_{\mathcal{B}\setminus\mathcal{N}_\alpha}\operatorname{sgn}\left(F_n(y)+\alpha d(y)\right)d(y)\ \mathrm{d}y \\
&= \int_{\mathcal{B}}\operatorname{sgn}\left(F_n(y)+\alpha d(y)\right)d(y)\ \mathrm{d}y.
\end{aligned}$$

Thus, we conclude that

$$0 = 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} - 2\ \langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + \lambda\int_{\mathcal{B}}\operatorname{sgn}\left(F_n(y)+\alpha d(y)\right)d(y)\ \mathrm{d}y.$$

However, we cannot solve this equation for α analytically. Thus, it is not practical to use the $\mathrm{L}^1(\mathcal{B})$-norm approach although it seemed to be promising.

Furthermore, note that Theorem 1.29 is valid in a Hilbert space setting only. Thus, the results regarding the existence, stability and convergence of the regularization problem (see Section 4.2) may not apply here either.

III.

Numerical Applications

In times of sea-level rise, rapid mass-loss in the polar regions and other climate changes of global impact, the processing of data that is collected by the various satellite missions rapidly gains importance. Satellite missions allow a global or supra-regional overview over events that may not be distinguishable from a local point of view or even from the surface of the Earth itself.

In this part, we are concerned with two kinds of problems and data. We will recover the mass density distribution of South America out of the Earth Gravitational Model 2008 (EGM2008) developed by the National Geospatial Intelligence Agency (NGA) (see [117]). Right now, this model is the gravity model with the highest resolution available. Moreover, we recover the mass density distribution of the whole Earth by a joint inversion of gravitational and seismic data in the form of normal mode anomalies. Dr. Arwen Deuss, University of Cambridge, kindly provided us with the most recent model coefficients.

Furthermore, we examine the mass transport in the Amazon area for the year 2008. The satellite mission Gravity Recovery and Climate Experiment (GRACE) has provided us with monthly solutions of the gravitational potential since the mid of 2003. Thus, these solutions are predestined to be used to reflect temporal changes caused, e.g. by large changes of ground water levels or the deglaciation. We will use the solutions provided by the Jet Propulsion Laboratory (JPL, see [85]).

5. Reconstructing the Mass Density Distribution of the Earth

Let us demonstrate the advantages of our new method by reconstructing the mass density distribution of the Earth. As data input, we will use the gravitational potential EGM2008 (see [117]). Its coefficients are given up to spherical harmonic degree $2,190$ and order of $2,159$.

Since it is well-known that the harmonicity constraint in particular and gravitational data in general are only appropriate for the determination of mass anomalies in the uppermost layer of the Earth (see [107]), we reconstruct the density close to the surface when considering gravitational data only, in Section 5.1. Afterwards, we will give a remark on l^1-optimization in Section 5.2. In Section 5.3, we reconstruct the density distribution of the whole Earth by a joint inversion of gravitational and seismic data in the form of normal mode anomalies.

5.1. by Using Gravimetric Data (EGM2008): Example South America

As a case study, let us reconstruct the mass density distribution of South America to analyze the proposed algorithms.

The solution of Algorithm 4.1 (RFMP)

The first step for the application of Algorithm 4.1 (RFMP) is to choose an appropriate dictionary $\mathcal{D}$. Here, we use the dictionary

$$\begin{aligned}\mathcal{D} &= \left\{ K_h^{\mathrm{I}}(x,\cdot) \;\middle|\; h \in \{0.95, 0.97, 0.99\}, x \in \mathrm{grid}(\mathcal{B}) \right\} \\ &\quad \cup \left\{ G_{0,n,j}^{\mathrm{I}} \;\middle|\; n = 3,\dots,8, j = 1,\dots,2n+1 \right\}\end{aligned} \tag{5.1}$$

where $\mathrm{grid}(\mathcal{B})$ is a Driscoll-Healy grid restricted to a spherical rectangle covering South America, i.e. an equiangular grid, with $39,800$ grid points. We stop the summation in the kernel functions at degree $2,190$ in accordance to the degree of the spherical harmonics coefficients used to compute the data. This dictionary suits our needs since we may reconstruct global trends with the basis elements $G_{0,n,j}^{\mathrm{I}}$ while the localized kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ are a very good choice to recover detail structures of the target function. If not specified otherwise, the data

will be given at 25,440 points on a Driscoll-Healy grid slightly above the Earth's surface at 7 km height. The method will be stopped after $N = 20,000$ steps, i.e. 20,000 dictionary functions will be chosen to represent the target function.

Since the inverse gravimetric problem is ill-posed, we use the regularized version of our algorithm, i.e. Algorithm 4.1 (RFMP), to reconstruct the density distribution of South America. However, this means we are confronted with the problem to choose an appropriate regularization parameter λ. This is never an easy task but in our case we are additionally strained with the problem that our computations are too time-consuming to use the common methods. Usually, the parameter is determined by computing the L-curve where for different parameters λ the approximation error $||R^N||^2_{\mathbb{R}^l}$ is plotted against the penalty term $||F_N||^2_{\mathrm{L}^2(\mathcal{B})}$. Another possibility is to choose the regularization parameter by trial and error. However, we want to make an informed choice. Therefore, we adapt the L-curve method to our purpose.

After 100 steps, the weighted norm of the residual $\frac{||R^{100}||^2_{\mathbb{R}^l}}{||y||^2_{\mathbb{R}^l}}$, where $||y||^2_{\mathbb{R}^l} = ||R^0||^2_{\mathbb{R}^l}$ is the initial error, is already very small and we compute an L-curve for different regularization parameters λ after 100 expansion functions were chosen instead of at the end of the computations after 20,000 steps.

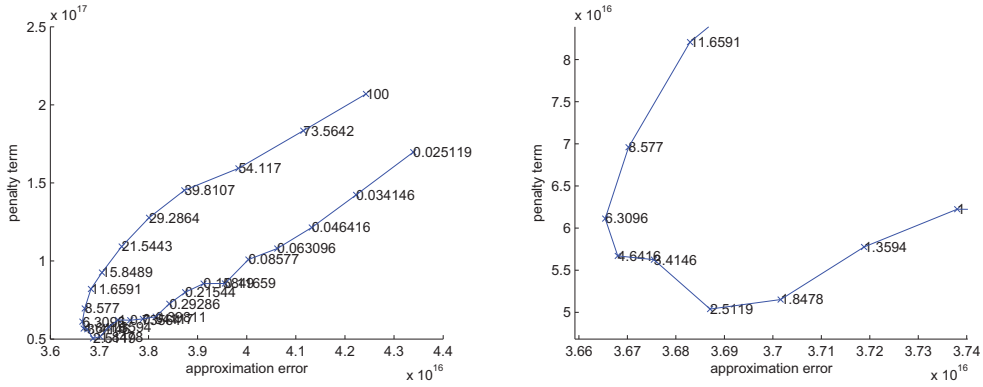

Figure 5.1.: Adapted L-curve for the choices $\lambda_i = 10^{-2+4i/30}$, $i = 3, \ldots, 30$ (left-hand) and zoomed-in on the interesting part (right-hand)

On the left-hand side of Figure 5.1, we display the resulting L-curve for parameters $\lambda_i = 10^{-2+4i/30}$, $i = 3, \ldots, 30$, i.e. $\lambda = 0.025119, \ldots, 100$, where the values of the approximation error are given on the x-axis and the norm of the solution on the y-axis. Furthermore, the value of the regularization parameter λ is given at the corresponding point in the plot itself. On the right-hand side of Figure 5.1, we zoom in on the important part and choose $\lambda = 4.6416$ as an appropriate

regularization parameter.

Using this regularization parameter, we get the solution displayed on the left-hand side of Figure 5.2. We clearly see the outline of the continent and the main topographic structures, i.e. the Andes and the Caribbean. Furthermore, we see parts of Antarctica and the Mid-Atlantic Ridge. In the right-hand side plot of Figure 5.2, we display the center points x of the chosen kernel functions $K_h^{\mathrm{I}}(x, \cdot)$. Note that we artificially include the coast lines of South America in blue as an orientation.

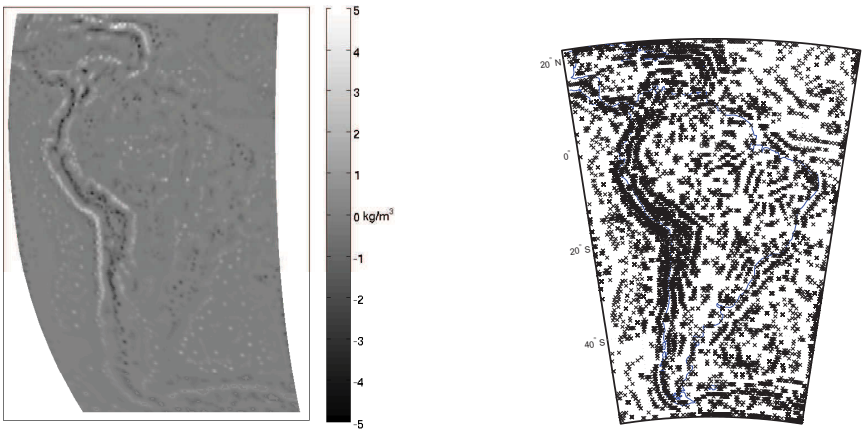

Figure 5.2.: Reconstructed density deviation (left-hand) and center points x of the chosen expansion functions $K_h^{\mathrm{I}}(x, \cdot)$ (right-hand) computed out of $25,440$ data points with $20,000$ selected functions from the dictionary $\mathcal{D}$ (i.e. $F_{20,000}$ is shown), $\lambda = 4.6416$

Remember that we aimed to develop a method that recovers a solution adapted to the structure of the target function. As can be seen in the right-hand side plot of Figure 5.2, the center points of the localized expansion functions are chosen mainly in areas where the detail structure needs to be more accurate, i.e. in the Andes and the Caribbean.

Note that the data was given on a Driscoll-Healy grid which is nearly equidistributed in the area of South-America. Thus, it can be seen clearly that the center points were primarily selected according to the structure of the solution and not just the data structure.

Now imagine a data grid that is as large as the one chosen in this application, i.e. $25,440$ data points. A spline method would not be able to handle this data grid since the corresponding (ill-conditioned and dense) matrix is much too large to be handled in the computations. In spline methods, the chosen data points and the center points x of the kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ used to expand the solution are directly connected. Thus, it is not possible to get a resolution as high as in the Andes with a spline method.

The localized character of the solution

In Figure 5.3, we examine the influence of the solution $F_{20,000}$ as displayed in Figure 5.2, i.e. the influence of the expansion elements, on exactly one point

$$x = (0.3784, -0.8661, -0.3267)^{\mathrm{T}} a$$

situated in the Andes. We mark this point in the left-hand plot of Figure 5.3 with a red dot. On the right-hand side of Figure 5.3, we display the influence of the expansion functions of the solution $F_{20,000}$ on the value at the point x with a logarithmic colorbar, i.e. we display $\log|\alpha_k d_k(x)|$ for all expansion elements. For a localized expansion function, we display $\log|\alpha_k K_{h_k}^{\mathrm{I}}(x,x_k)|$ as a value at the center point x_k. The influence of an expansion function $G_{m_k,n_k,j_k}^{\mathrm{I}}$ is included as an additive at all points in the plot.

Since we use a logarithmic colorbar, the blue color denotes a very small influence of about e^{-15} while the red color denotes a large influence of about e^2 on the value at the point x. Obviously, the value at x is mostly influenced by localized expansion functions. Furthermore, we clearly observe that the impact of the localized expansion functions increases when the distance between x and the center x_k decreases. This behavior was, obviously, expected.

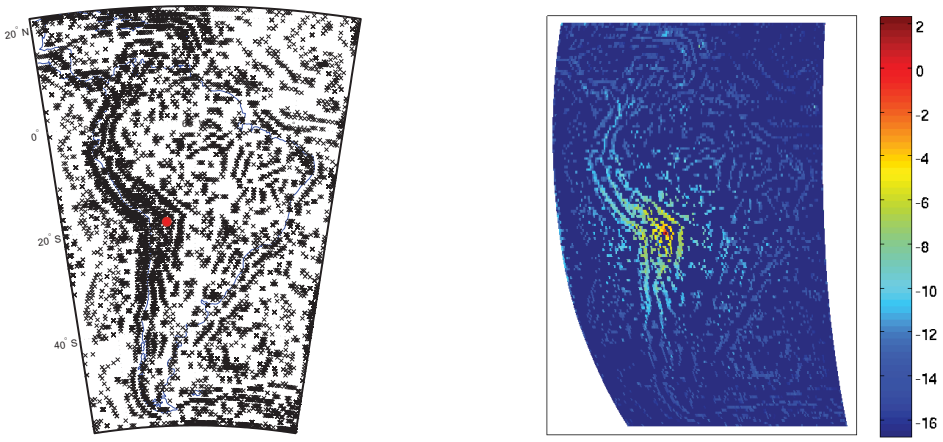

Figure 5.3.: Influence of the expansion functions of the solution $F_{20,000}$ (right-hand) on the point $x = (0.3784, -0.8661, -0.3267)^{\mathrm{T}}a$ (red dot) located in the Andes (left-hand)

The approximation error

We require the method to be adaptive and iterative. Thus, we may expect that the solution is improved when we increase the number of chosen dictionary functions in the expansion. First of all, let us consider the trend of the weighted approximation error, i.e.

$$\frac{||R^n||^2_{\mathbb{R}^l}}{||R^0||^2_{\mathbb{R}^l}} = \frac{||R^n||^2_{\mathbb{R}^l}}{||y||^2_{\mathbb{R}^l}}$$

where $||R^0||^2_{\mathbb{R}^l}$ is the initial error. Thus, we may display the approximation quality in percentages with respect to the initial error where $1 = 100\%$ denotes the starting point.

In Figure 5.4, we display the evolution of the residual for the first 100 steps as well as the development from step 100 to $10,000$ for $\lambda = 4.6416$ (in blue). We can see clearly that the residual decreases rapidly in the beginning. After choosing only 100 expansion functions, the approximation error is already reduced to less than 1% of the initial error. Thus, it is appropriate to terminate the computation after 100 dictionary functions are chosen when computing the adapted L-curve.

Moreover, we see that the approximation error can be reduced to less than 0.5% of the initial error after about 500 steps and after approximately $10,000$ steps we get a weighted approximation error of 0.09% for the optimal choice of the regularization parameter $\lambda = 4.6416$. Let us remark that the approximation error is

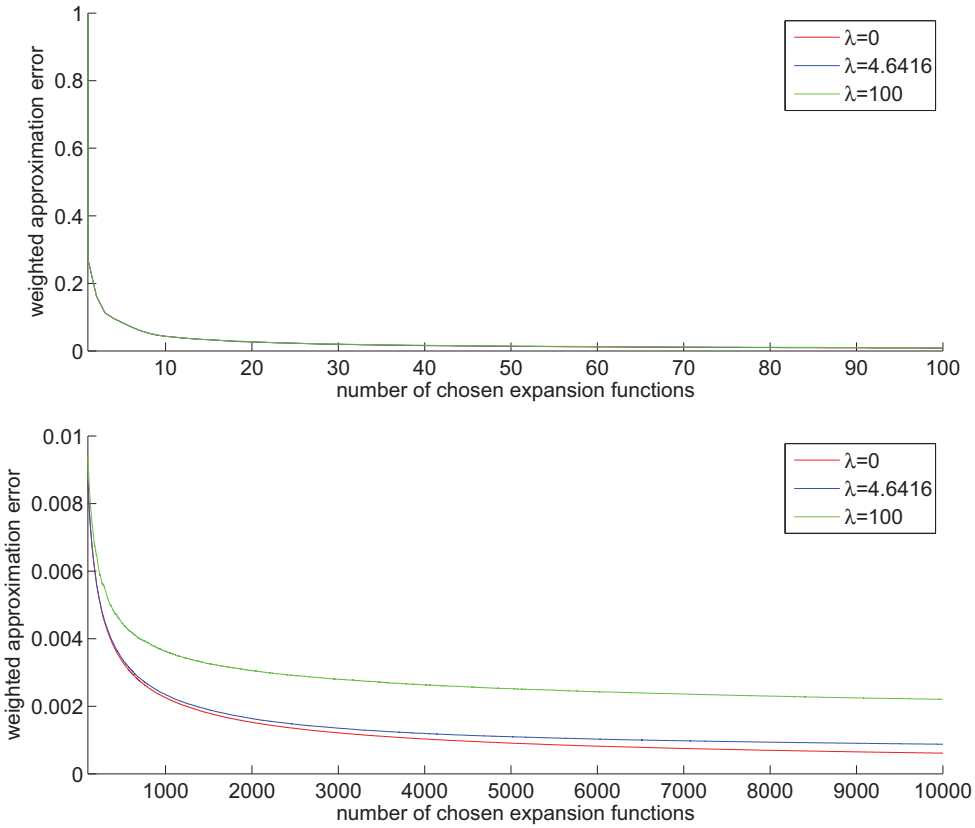

Figure 5.4.: Evolution of the weighted approximation error when reconstructing the density deviation out of 25,440 data points for $n = 1, \ldots, 100$ (top) and $n = 100, \ldots, 10,000$ (bottom) in case of the solution computed with Algorithm 3.1 (FMP), i.e. with regularization parameter $\lambda = 0$ (red), and the solutions computed with Algorithm 4.1 (RFMP) with regularization parameters $\lambda = 4.6416$ (blue) and 100 (green). Note that the evolution of all three is very similar in the upper plot. Only the green line can be seen.

only reduced slightly anymore in the steps from $10,000$ to $20,000$ chosen expansion functions. These results are not unexpected. Remember the regularization functional $||R^n||^2_{\mathbb{R}^l} + \lambda ||F_n||^2_{\mathrm{L}^2(\mathcal{B})}$ that is to be minimized. We may expect that, for large n, the penalty term becomes more important in comparison to the residual. Thus, its influence in the choice of the next dictionary function increases, too. We expect the algorithm to concentrate on the reduction of the approximation error in the beginning and shift its focus to reducing the penalty term, i.e. increasing the smoothness of the solution, in a distinct iteration with respect to the chosen regularization parameter λ.

The convergence based on the evolution of the coefficients α

In Figure 5.5, we display the evolution of the absolute value of the corresponding weights α_n obtained by Algorithm 4.1 for $\lambda = 4.6416$ up to $n = 20,000$.

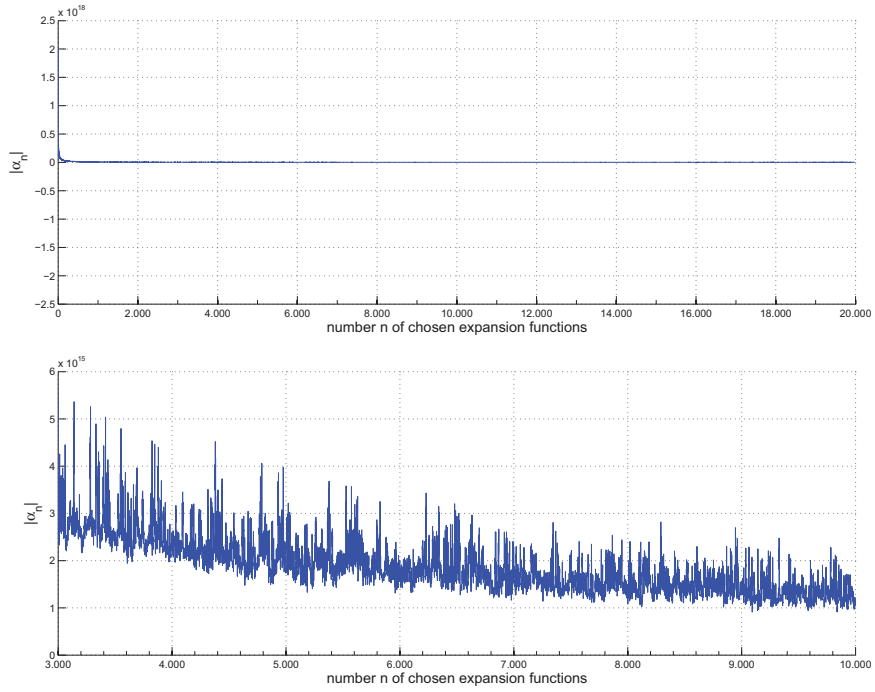

Figure 5.5.: Evolution of $|\alpha|$ when reconstructing the density deviation out of $25,440$ data points for $n = 1, \ldots, 20,000$ (top) and $n = 3,000, \ldots, 10,000$ (bottom) for the solution computed with Algorithm 4.1 (RFMP) with regularization parameter $\lambda = 4.6416$

Note that the absolute values of the weights α_n of the representation of the solution decrease very quickly which is evidence for a fast convergence of the

algorithm. However, they do not decrease monotonically as is the case in the original matching pursuit (see [98]).

The iterative character of the method

We developed an iterative method to reconstruct the target function. Thus, we may expect that the solution improves stepwise. Let us consider the solution after $3,000$ and $10,000$ steps in comparison to the solution after $20,000$ steps (see the left-hand column of Figure 5.6). Clearly, the solution improves if we choose $10,000$ dictionary functions instead of only $3,000$. The main structures of the continent are displayed better. Moreover, we discern connected structures and not just the local peaks of the localized kernel functions. After acquiring a result as $F_{3,000}$, one should obviously decide to refine the solution to get a better resolution. Note that the result $F_{3,000}$ may be used as an initial approximation in the refinement. Thus, we do not lose valuable computation time. We will consider such a refinement later in this section.

At first sight, we do not see much improvement if we increase the number of the chosen expansion functions from $10,000$ to $20,000$ (compare the middle and top plot of the left-hand column of Figure 5.6). However, this is not surprising since the main structures were obviously already reconstructed after choosing only $10,000$ expansion functions (see the display of centers x of the chosen localized expansion functions in the right-hand column of Figure 5.6). However, the additional $10,000$ expansion functions in $F_{20,000}$ clearly refine the resolution of the detail structures of the target function as can be observed, e.g. at the western border of the Andes where the edges are not as rough anymore. Note that these observations match remarkably well with our findings from the approximation error (see Figure 5.4).

Now let us consider the center points x of the chosen kernel functions $K_h^{\mathrm{I}}(x,\cdot)$, again for $F_{3,000}$, $F_{10,000}$ and $F_{20,000}$, as displayed in the right-hand column of Figure 5.6. As suggested by the plot of the solution $F_{3,000}$, the displayed center points again show that the main structures have been identified already this early in the reconstruction process. We clearly see beginning shapes of the Andes and the Caribbean. Moreover, we may even discern the coast lines of the eastern border of the continent. As expected, all these features are much more prominent after $10,000$ expansion functions were chosen.

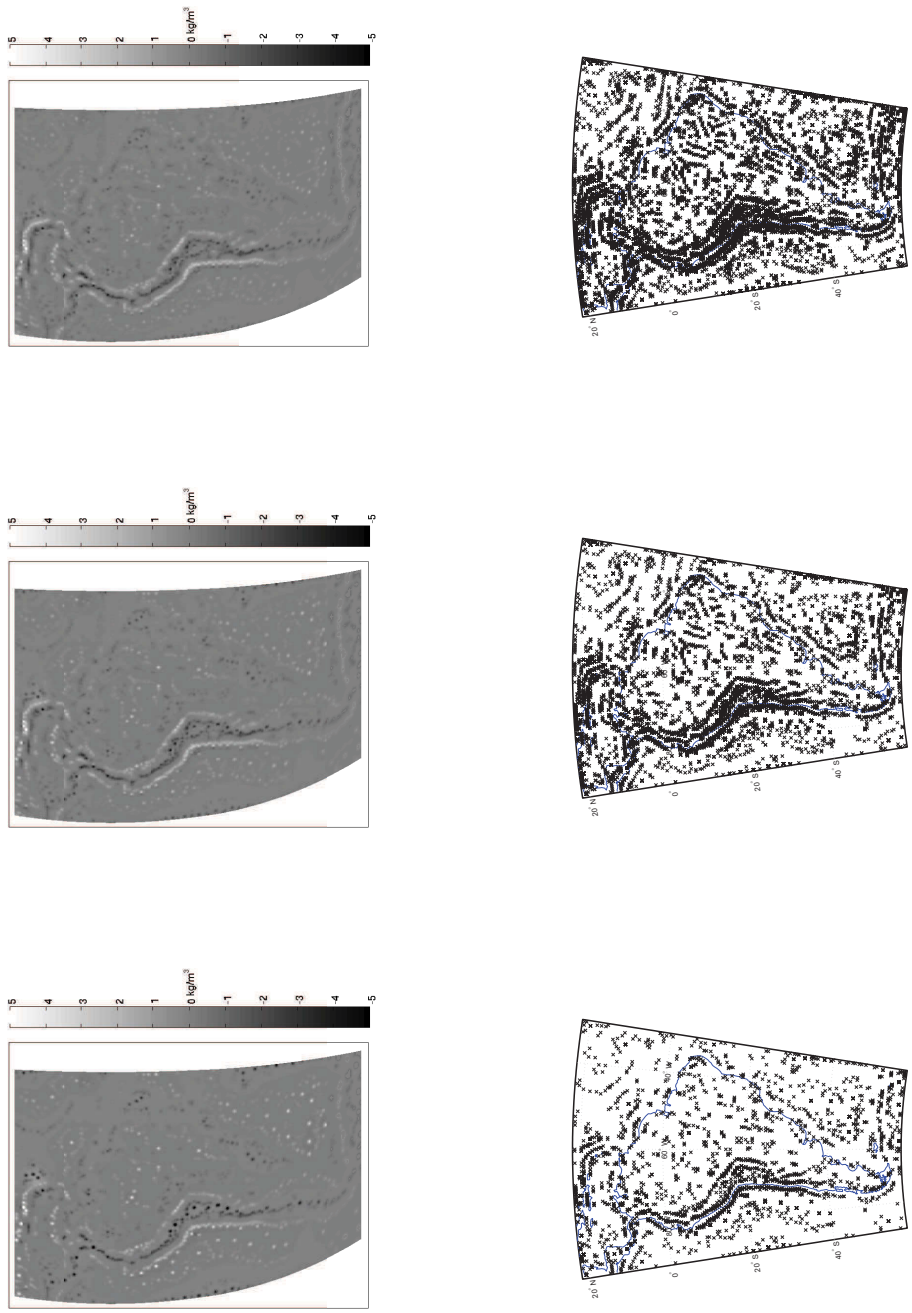

Figure 5.6.: Reconstructed density deviation (left-hand column) and center points x of the chosen dictionary functions $K_h^{\mathrm{I}}(x,\cdot)$ (right-hand column) computed out of $25,440$ data points with $3,000$ (bottom), $10,000$ (middle) and $20,000$ (top) selected expansion functions, $\lambda = 4.6416$

The selection of the center points in the solution $F_{10,000}$ compared to the one in $F_{20,000}$ confirms that the additional $10,000$ expansion functions in $F_{20,000}$ are mainly chosen to refine the detail structure of the approximation since we distinguish a cumulation of centers in the Caribbean, the Andes and part of Antarctica. Note that we again added artificial blue coast lines to the plots of the centers to ease the orientation.

The influence of the regularization parameter λ

In Figure 5.4, we displayed the weighted residuals for the unregularized approximation with Algorithm 3.1 (FMP) ($\lambda = 0$ in red) as well as for the regularized approximations with Algorithm 4.1 (RFMP) for $\lambda = 4.6416$ (in blue) and $\lambda = 100$ (in green). We clearly see that there are no major differences in the development of the residuals up to step 100 (upper plot). However, in later steps, the effects of the choice of the regularization parameter are more obvious. Choosing the regularization parameter $\lambda = 0$, i.e. a reconstruction with no regard for the smoothness of the solution, provides us, as expected, with the smallest weighted approximation error of about 0.06% after $10,000$ expansion functions were chosen. With increasing regularization parameter the importance of the penalty term which forces the smoothness of the solution increases, too, and, thus, the focus shifts from finding the solution with the smallest approximation error to finding a solution that is smooth, too. As a consequence, the approximation error of the smoothest solution ($\lambda = 100$, see Figure 5.7) is with a value of about 0.2% the largest one.

In Figure 5.7, we compare the solution of the unregularized method (Algorithm 3.1 (FMP)) with the solution of the regularized method (Algorithm 4.1 (RFMP)) for the 'optimal' regularization parameter $\lambda = 4.6416$ and the larger parameter $\lambda = 100$. In each case, we reconstruct the density deviation out of $25,440$ data points by selecting $10,000$ expansion functions out of the dictionary $\mathcal{D}$ given in Equation (5.1). The left-hand column of Figure 5.7 displays the solutions with the original colorbar while the colorbars of the same solutions in the right-hand column are adapted for comparison to the values of the 'optimally' regularized solution.

Regarding the plots with adapted colorbar, we clearly see the characteristics of a regularization. With increasing regularization parameter, the influence of the penalty term increases, too. In our case, the penalty term is concerned with the smoothness of the solution, i.e. the solution becomes smoother when we increase the regularization parameter.

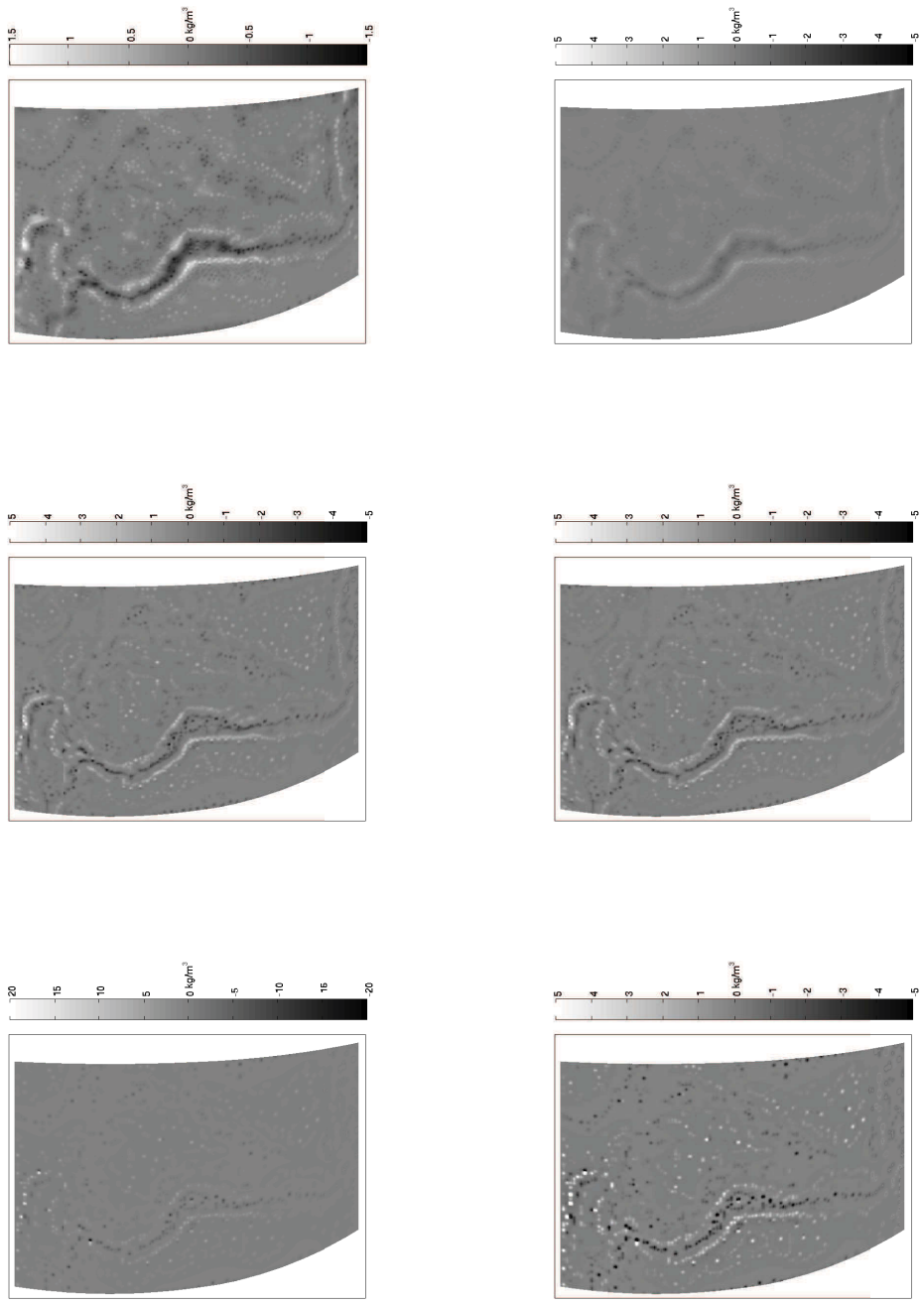

Figure 5.7.: Reconstructed density deviation computed out of 25,440 data points with 10,000 selected expansion functions from the dictionary $\mathcal{D}$ (i.e. $F_{10,000}$ is shown) for $\lambda = 0$ (bottom), $\lambda = 4.6416$ (middle) and $\lambda = 100$ (top) with original (left-hand) and adapted (right-hand) colorbar

If we consider the range of the recovered values for the density deviation we observe another trait of a regularization. Smoothening oftentimes manifests as a change in magnitude of the solution when we have only comparably small variations in the surrounding regions.

Anyway, these considerations demonstrate that the introduced method (Algorithm 4.1 (RFMP)) is, indeed, a regularization method as was proven in Section 4.2.

The evolution of the penalty term

In Figure 5.8, we display the evolution of the penalty term for the unregularized case ($\lambda = 0$ in red) as well as the regularization with different parameters λ for $10,000$ iterations of Algorithms 3.1 (FMP) and 4.1 (RFMP), respectively ($\lambda = 4.6416$ in blue, $\lambda = 10$ in a dashed magenta line style and $\lambda = 100$ in green).

With increasing regularization parameter λ the influence of the penalty term in the choice of the next expansion function increases, too. Thus, we expect the approximation to become smoother with increasing regularization parameter. Remember that, in our case, the smoothness of the approximation is expressed by its $\mathrm{L}^2(\mathcal{B})$-norm, i.e. the penalty term. Thus, we expect the penalty term to decrease with increasing regularization parameter.

Clearly this expectation is fulfilled if we compare the results for $\lambda = 0$, $\lambda = 4.6416$ and $\lambda = 10$. However, the curve for the penalty term corresponding to $\lambda = 100$ displays larger values than the curves corresponding to less regularized computations. That means the values of the $\mathrm{L}^2(\mathcal{B})$-norm are higher, although the approximation is indeed smoother, as we displayed in Figure 5.7 and discussed before.

If we consider the approximation error for the same regularization parameters (see Figure 5.9), we do not observe any unexpected behavior. The approximation error increases for increasing regularization parameter - for all choices of λ.

Unfortunately, we do not understand this behavior, yet. However, for small regularization parameters the evolution of the penalty term is as expected. And for larger regularization parameters the solution becomes smoother although this behavior is not depicted in the evolution of the penalty term.

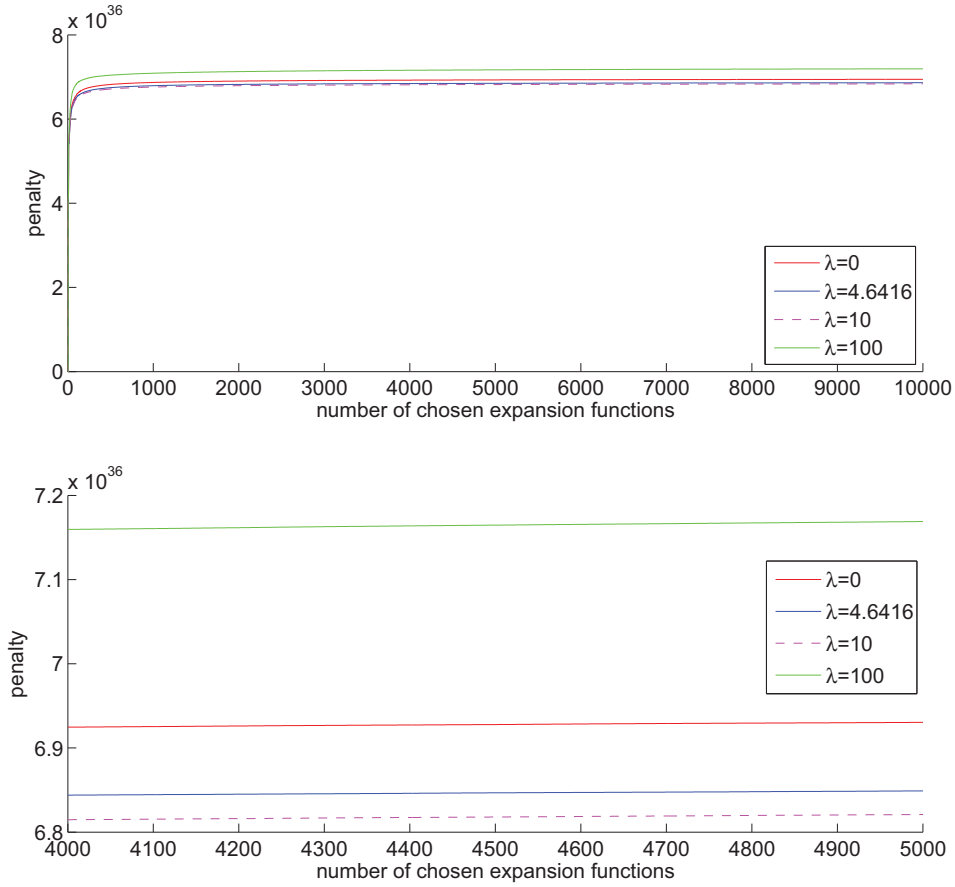

Figure 5.8.: Evolution of the penalty term $||F_n||_{\mathrm{L}^2(\mathcal{B})}$ when reconstructing the density deviation out of $25,440$ data points for $n = 1, \ldots, 10,000$ (top) and $n = 4,000, \ldots, 5,000$ (bottom) for the solution computed with Algorithms 3.1 (FMP) and 4.1 (RFMP) with regularization parameters $\lambda = 0$ (in red), $\lambda = 4.6416$ (in blue), $\lambda = 10$ (in dashed magenta) and $\lambda = 100$ (in green)

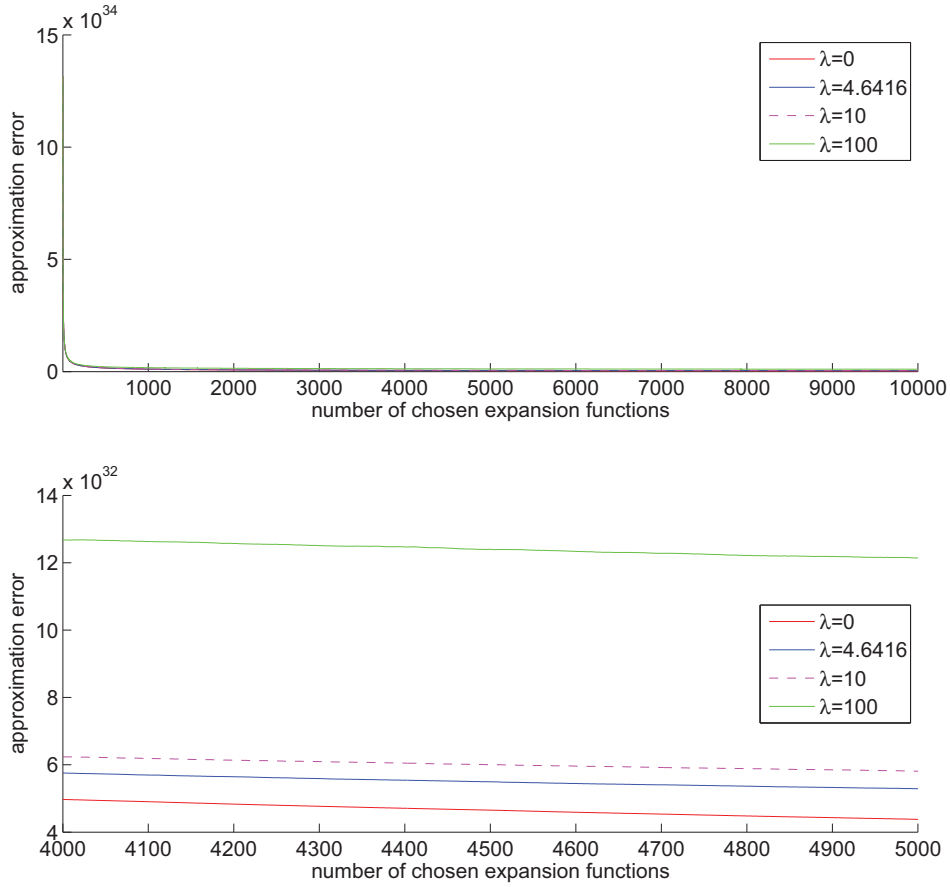

Figure 5.9.: Evolution of the approximation error $||R^n||_{\mathrm{L}^2(\mathcal{B})}$ when reconstructing the density deviation out of $25,440$ data points for $n = 1, \ldots, 10,000$ (top) and $n = 4,000, \ldots, 5,000$ (bottom) for the solution computed with Algorithms 3.1 (FMP) and 4.1 (RFMP) with regularization parameters $\lambda = 0$ (in red), $\lambda = 4.6416$ (in blue), $\lambda = 10$ (in dashed magenta) and $\lambda = 100$ (in green)

The choice of expansion functions, in particular the choice of the localization parameter h

Let us examine the choice of the expansion functions $K_h^{\mathrm{I}}(x,\cdot)$ and $G_{m,n,j}^{\mathrm{I}}$. We already know that the centers x are chosen according to the structure of the solution (see right-hand side of Figure 5.2). Now, we want to consider the choice of the localization parameters h corresponding to the chosen localized expansion functions $K_h^{\mathrm{I}}(x,\cdot)$, displayed in Figure 5.10. Choices of an expansion function $G_{m,n,j}^{\mathrm{I}}$ are denoted with the value 1.

It is not clear whether these results should be ranked as a success. On the one hand, the choice of the localization parameters clearly depends on the choice of the regularization parameter. Remember that the hat-width of a localized kernel function increases if the localization parameter h decreases. With increasing regularization parameter the less localized dictionary elements are preferred, i.e. the algorithm chooses more dictionary elements with lower localization parameter h to build an approximation that is smoother. This behavior was expected.

On the other hand, we expect the algorithm to choose the less localized dictionary elements first to recover main structures of the solution, i.e. we expect it to start with the dictionary elements with global character followed by weakly localized dictionary elements before it chooses strongly localized dictionary elements to recover the detail structures of the target function. Clearly, the algorithm does not fulfill this expectation.

Up to now, we only use localized kernel functions in scaling function form as dictionary elements, i.e.

$$d = K_h(x,\cdot) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}(A_{m,n}^h)^2 G_{m,n,j}(x)G_{m,n,j}(\cdot),\ h\in]0,1[,\ x\in\mathcal{B}.$$

We propose to use localized kernel functions in wavelet form, too, i.e.

$$d = K_h(x,\cdot) - K_{\tilde{h}}(\tilde{x},\cdot),\ h,\tilde{h}\in]0,1[,\ h>\tilde{h},\ x,\tilde{x}\in\mathcal{B}$$

to improve the results.

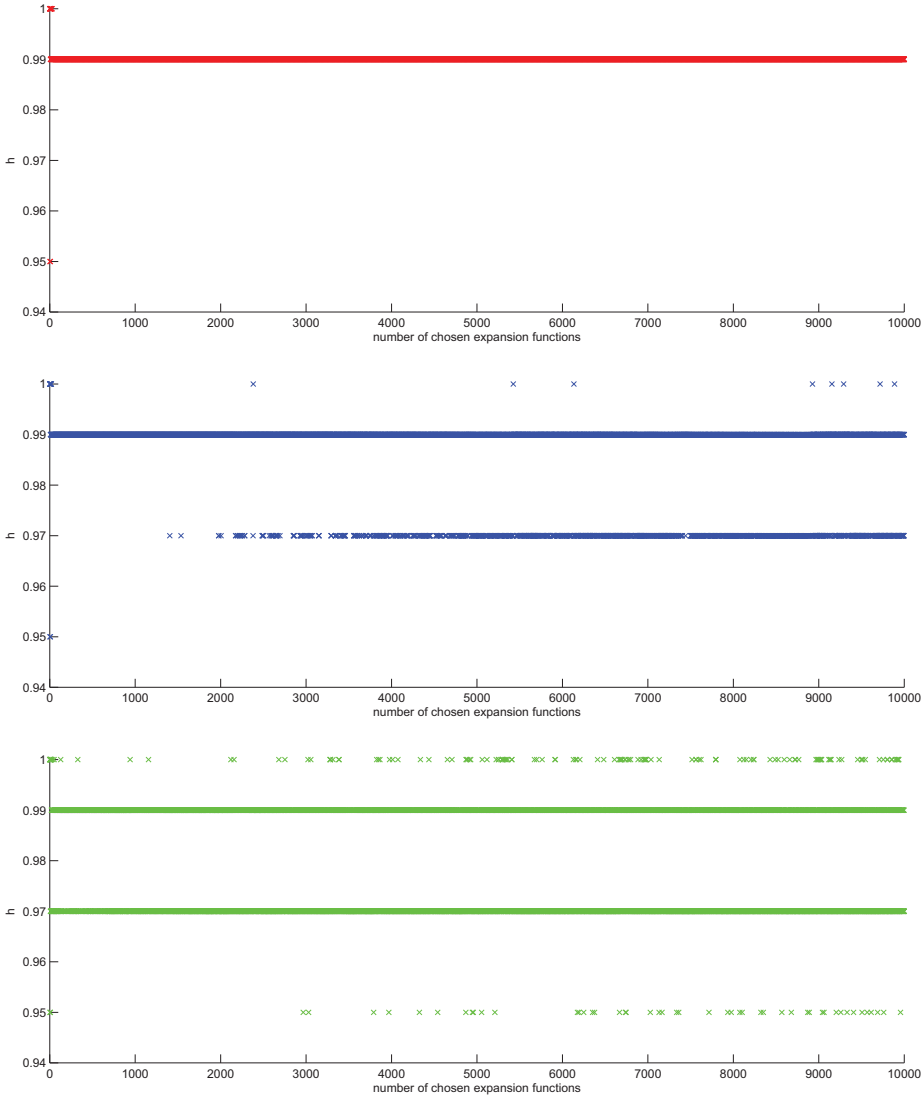

Figure 5.10.: Choice of localization parameters h in the localized expansion elements $K_h^{\mathrm{I}}(x,\cdot)$ when reconstructing the density deviation out of 25,440 data points for $n = 1,\ldots,10,000$ expansion functions chosen by Algorithm 3.1 (FMP), i.e. $\lambda = 0$ (top in red), and Algorithm 4.1 (RFMP) with regularization parameters $\lambda = 4.6416$ (middle in blue) and $\lambda = 100$ (bottom in green). Choices of an expansion element $G_{m,n,j}^{\mathrm{I}}$ are denoted with the value 1.

Refinement

We mentioned that the iterative character of our method allows us to reuse results from previous computations to, e.g. zoom in on certain parts of the solution. We investigate this property by refining Middle-America and the Caribbean.

In the refinement, we want to improve the approximation in a certain area only. Thus, it is reasonable to consider a dictionary that consists of functions with a local character exclusively, e.g. the dictionary

$$\mathcal{D} \;=\; \left\{ K_h^{\mathrm{I}}(x,\cdot) \;\middle|\; h \in \{0.95, 0.97, 0.99\}, x \in \mathrm{grid}(\mathcal{B}) \right\} \tag{5.2}$$

where $\mathrm{grid}(\mathcal{B})$ is a Driscoll-Healy grid restricted to a spherical rectangle covering the area of Middle-America and the Caribbean, i.e. an equiangular grid with $39,800$ grid points. Furthermore, we use a data grid with $25,440$ points on a Driscoll-Healy grid slightly above the Earth's surface at 7 km height for this same area of the refinement. We choose the regularization parameter $\lambda = 10$ with the adapted L-curve method.

In the left-hand column of Figure 5.11, we display the solution and the center points of the chosen localized expansion functions corresponding to Algorithm 4.1 (RFMP) where we use a regularization parameter $\lambda = 4.6416$ and stop the computation after $3,000$ expansion functions were chosen out of the dictionary $\mathcal{D}$ given in Equation (5.1). We considered this solution before when we investigated the iterative character of the method in Figure 5.6. For the refinement, we use this solution as a basis for the new computations. Thus, in this example, we save half of the computation time.

In the middle column of Figure 5.11, we display the refinement and the center points of the $3,000$ additionally chosen expansion functions. Remember that the dictionary $\mathcal{D}$ given in Equation (5.2) for the refinement consists of localized functions only. When comparing the solution without the refinement (upper left-hand plot) and the solution corresponding to the refinement (upper middle plot) we clearly see, that the resolution of Middle-America and the Caribbean is by far improved. If we consider the center points of the chosen expansion functions (lower middle plot) we observe that they are mostly centered in these regions, i.e. Middle-America and the Caribbean, as well. Furthermore, we observe some minor boundary effects. Note that we give some ideas to control these boundary effects in Appendix A.

In the right-hand side column of Figure 5.11, we consider the changes that the refinement introduces. Again we observe, that most changes are located in Middle-America and the Caribbean, i.e. the regions where something happens. This is, of course, expected behavior as we investigated these properties before. However,

it is quite remarkable how well the refinement process works. It allows us to do a coarse investigation of a certain area first. And then, with these results in mind, we may decide where we want to refine this solution. Furthermore, this property allows us to save major computational effort when refining.

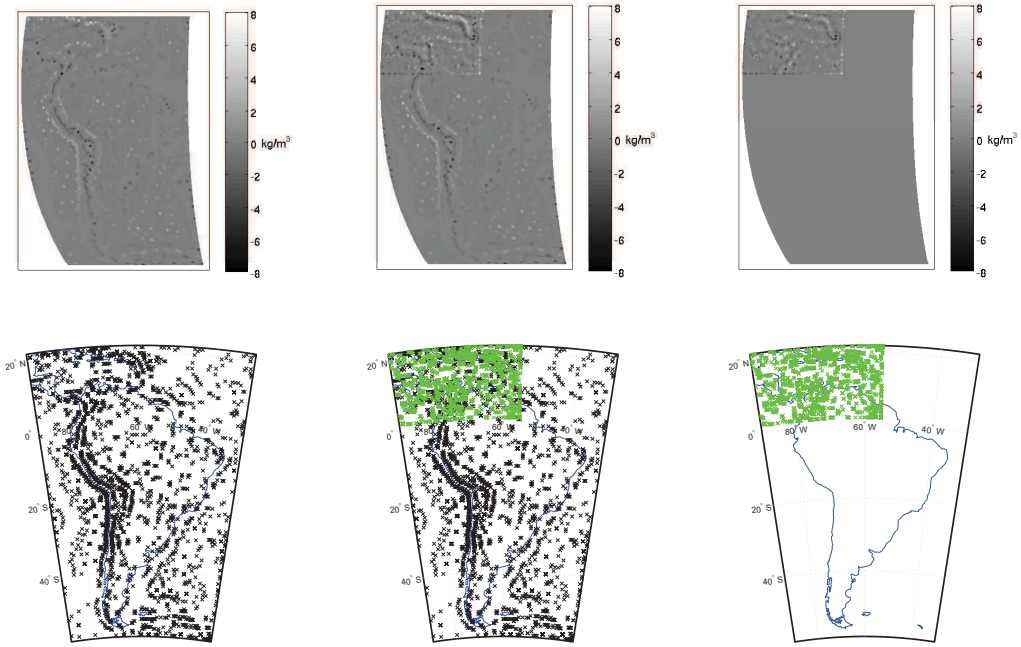

Figure 5.11.: Reconstructed density deviation (top) and the center points x of the chosen localized expansion functions (bottom) computed out of $25,440$ data points with $3,000$ selected expansion functions for $\lambda = 4.6416$ (left-hand column), refined in the North-West out of the set of $25,440$ data points located in the North-West with additional $3,000$ selected expansion functions out of the dictionary given in Equation (5.2) for $\lambda = 10$ (middle column) and the refinement only (right column)

Dealing with noise

In this section, we want to examine the behavior of the regularizing Algorithm 4.1 (RFMP) when applied to noisy data y^ε. Here ε denotes the noise level where a value $\varepsilon = 0.01$ corresponds to a data input that is disturbed with 1% random noise relative to the exact data y, i.e.

$$y_i^\varepsilon = y_i + \varepsilon \ \mathrm{rand}_i \ y_i, \ \ i = 1, \ldots, l$$

where rand_i is a random number in the interval $[0, 1]$.

In this section, we consider the reconstructed density deviations out of $25,440$ data points where we stop the algorithm after $10,000$ expansion functions are chosen out of the dictionary given in Equation (5.1).

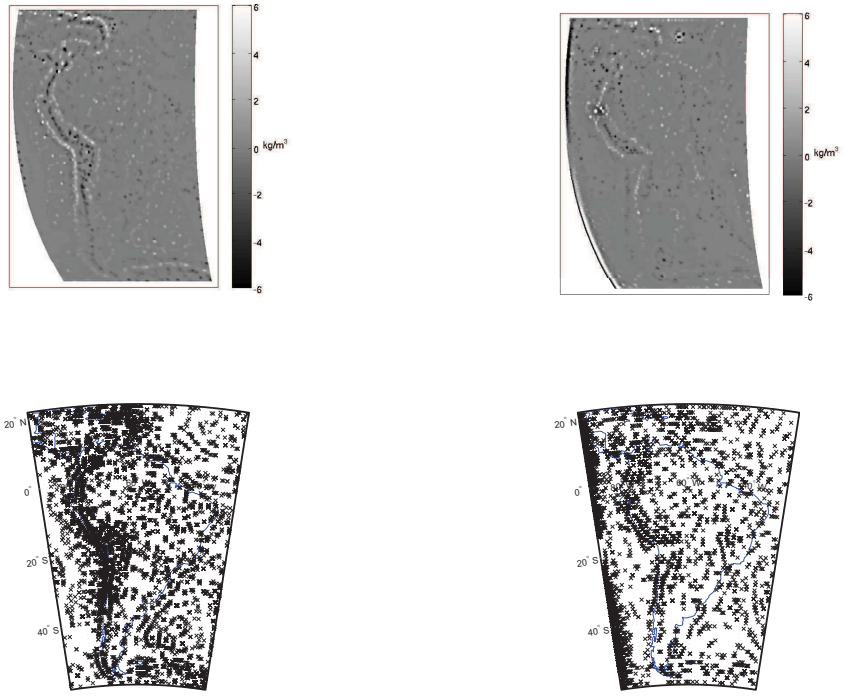

Figure 5.12.: Reconstructed density deviation (top) and center points of the chosen localized expansion functions (bottom) computed out of $25,440$ data points with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ (i.e. $F_{10,000}$ is shown) for $\lambda = 0$ out of exact data ($\varepsilon = 0$, left-hand column) and out of data with 1% noise ($\varepsilon = 0.01$, right-hand column)

In Figure 5.12, we display the solution and the position of the center points of the chosen localized expansion functions of the unregularized Algorithm 3.1 (FMP) when considering exact data and a noise level of $\varepsilon = 0.01$ which corresponds to 1% of noise. We can clearly observe that the introduction of noise has a very negative influence on the reconstruction quality of the algorithm since the main structures are hardly identified anymore. Furthermore, the solution suffers from major boundary effects. Thus, using a regularization method seems to be very important.

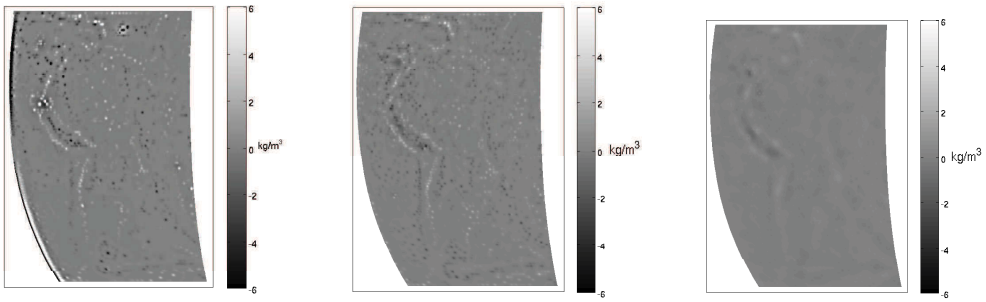

Figure 5.13.: Reconstructed density deviation computed out of $25,440$ data points with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ (i.e. $F_{10,000}$ is shown) for $\lambda = 0$ (left-hand), $\lambda = 10$ (middle) and $\lambda = 100$ (right-hand with adapted colorbar) out of data disturbed with 1% noise ($\varepsilon = 0.01$)

Let us compare $F^{\lambda,\varepsilon} = F^{0,0.01}$, $F^{10,0.01}$ and $F^{100,0.01}$ in Figure 5.13. It is not surprising that the regularized solutions are much smoother than the unregularized one. However, the unregularized solution clearly displays boundary effects that do not appear in the regularized ones. Furthermore, $F^{0,0.01}$ shows some artefacts, e.g. in the Caribbean and the Andes. Again, we do not observe such artefacts in the regularized solutions. As a result, we strongly recommend to use our regularization method when working with noisy data. Of course, this confirms our theoretical results.

In Figure 5.14, we investigate the solutions for the three different noise levels $\varepsilon = 0$, $\varepsilon = 0.01$ and $\varepsilon = 0.1$ where $\varepsilon = 0$ corresponds to the case with exact data that we examined in the prior sections. Moreover, we use the two regularization parameters $\lambda = 10$ (left-hand column) and $\lambda = 100$ (middle column with original colorbar and right-hand column with adapted colorbar) for our considerations.

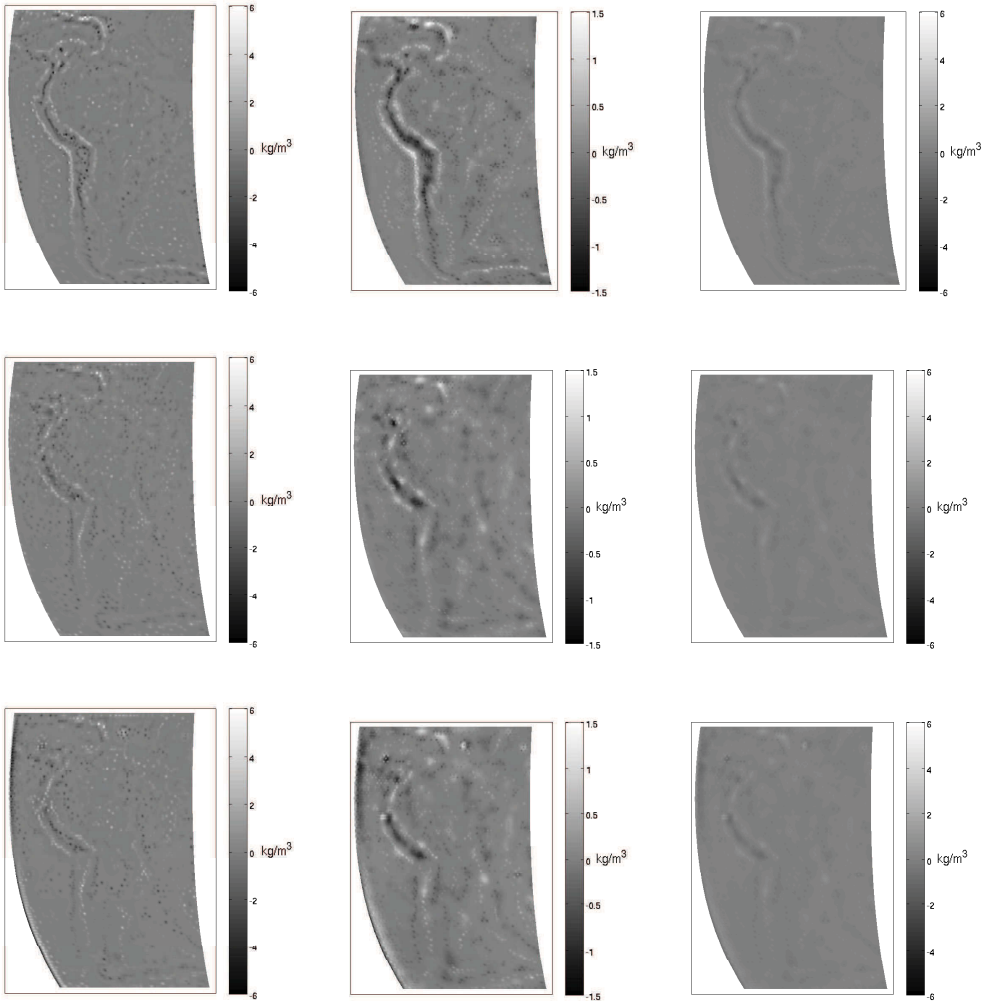

Figure 5.14.: Reconstructed density deviation computed out of $25,440$ data points with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ (i.e. $F_{10,000}$ is shown) for $\lambda = 10$ (left-hand column), $\lambda = 100$ (middle) and $\lambda = 100$ with adapted colorbar (right-hand column) out of exact data ($\varepsilon = 0$, top), data with 1% noise ($\varepsilon = 0.01$, middle row) and data with 10% noise ($\varepsilon = 0.1$, bottom)

Let us first consider the solutions regularized with $\lambda = 10$, i.e. the left-hand column of Figure 5.14. Clearly, some of the important features of South America, namely the Andes, the Caribbean and parts of Antarctica may still be distinguished when we use noisy data as an input to Algorithm 4.1 (RFMP). However, the more shallow structures close to the Mid-Atlantic Ridge and the eastern coast of South America can, obviously, not be recovered correctly anymore. Surprisingly, the reconstruction does not worsen as much as may be expected when raising the noise level from 1% ($\varepsilon = 0.01$) to 10% ($\varepsilon = 0.1$). However, we observe boundary effects in the solution corresponding to $\varepsilon = 0.1$ (see Figure 5.15, too) that are not present in the solution corresponding to $\varepsilon = 0.01$.

When comparing the reconstruction with regularization parameter $\lambda = 10$ to the solutions corresponding to $\lambda = 100$ we clearly observe the regularizing character of the method. All solutions corresponding to $\lambda = 100$ are smoother than the respective ones corresponding to $\lambda = 10$, as expected. However, the solutions corresponding to the noisy data input for $\lambda = 100$ display the same strengths (see Andes, Caribbean and Antarctica) and shortcomings (see the east coast of South America and the Mid-Atlantic Ridge) as the reconstructions corresponding to the regularization parameter $\lambda = 10$. Moreover, the solution corresponding to the noise level $\varepsilon = 0.1$ and the regularization parameter $\lambda = 100$ clearly displays boundary effects as well as some artefacts (see, e.g. the maximum in the east of the Caribbean in white) that are not distinguishable in the solution corresponding to the noise level $\varepsilon = 0.01$ with $\lambda = 100$.

Let us summarize that the reconstruction of the density deviations out of noisy data still produces results where the main structures are distinguishable if we use a regularization method. However, the quality of the reconstruction seems to worsen from North to South. Furthermore, we observe some boundary effects and artefacts for larger noise levels. We will give a short outlook on treating boundary effects in Appendix A.

The choice of the center points x of the localized expansion functions $K_h^{\mathrm{I}}(x, \cdot)$ confirms all these observations (see Figure 5.15 where we display this choice for the regularization parameters $\lambda = 10$ (left-hand column) and $\lambda = 100$ (right-hand column) as well as noise levels $\varepsilon = 0$ (exact data, bottom), $\varepsilon = 0.01$ (middle) and $\varepsilon = 0.1$ (top)). Clearly, the ability of the algorithm to choose the expansion functions according to the structure of the target function is corrupted when we consider noisy data. Furthermore, the main structures of the solution cannot be distinguished in the plot of the center points of the chosen expansion functions, anymore. Nonetheless, an increasing regularization parameter bears a positive impact on this ability since more structures may be detected in the plots corresponding to $\lambda = 100$.

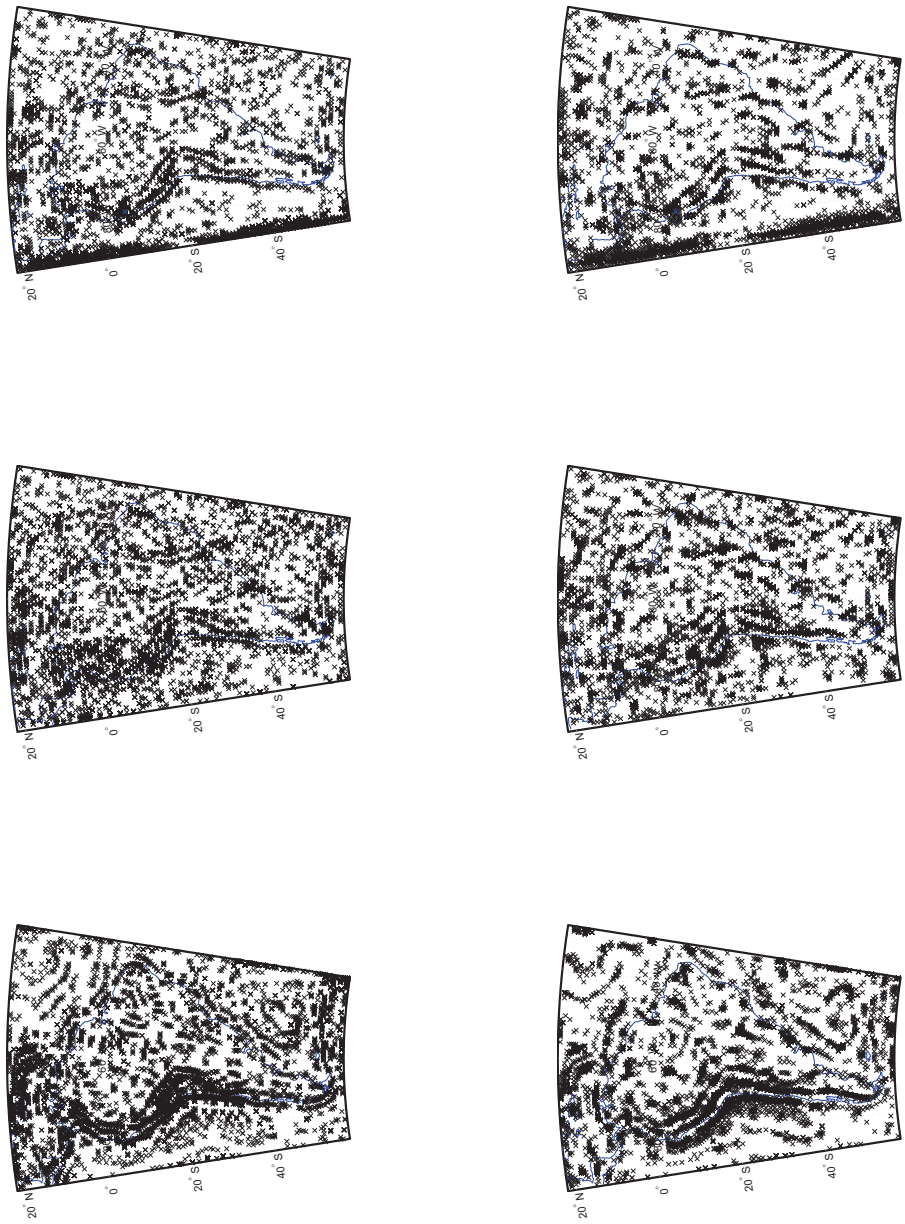

Figure 5.15.: Center points x of the chosen dictionary functions $K_h^{\mathrm{I}}(x,\cdot)$ computed out of $25,440$ data points with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ (i.e. $F_{10,000}$ is shown) out of exact data ($\varepsilon = 0$, bottom), data with 1% noise ($\varepsilon = 0.01$, middle) and data with 10% noise ($\varepsilon = 0.1$, top) for regularization parameters $\lambda = 10$ (left-hand column) and $\lambda = 100$ (right-hand column)

In Figure 5.15, the boundary effects, that we already detected for the noise level $\varepsilon = 0.1$, are more obvious and intensify with increasing regularization parameter λ. Again, we refer to Appendix A for an outlook on treating boundary effects.

At last, let us investigate the behavior of Algorithm 4.1 (RFMP) when considering noisy data by means of the relative error corresponding to the optimal solution for the exact data, i.e. corresponding to $F^{4.6416,0}$. We use a relative error to keep comparability. The optimal solution seems to be an appropriate choice for a reference solution, since we do not know the exact solution.

Let the solution $F_n^{\lambda,\varepsilon}$ be described by the vector $z^{\lambda,\varepsilon} \in \mathbb{R}^M$, $M \in \mathbb{N}$, where $z_i^{\lambda,\varepsilon} := F_n^{\lambda,\varepsilon}(x_i)$, $i = 1, \ldots, M$, and M is the number of grid points x_i for the evaluation of the solution. Then the relative error may be described as

$$\frac{||z^{4.6416,0} - z^{\lambda,\varepsilon}||_{\mathbb{R}^M}^2}{||z^{4.6416,0}||_{\mathbb{R}^M}^2}.$$

For our choice of regularization parameters $\lambda = 0$, $\lambda = 10$ and $\lambda = 100$ and noise levels $\varepsilon = 0$ (corresponding to exact data), $\varepsilon = 0.01$ and $\varepsilon = 0.1$, we get the values displayed in Table 5.1 for the relative error with respect to the solution $F^{4.6416,0}$.

relative	$\lambda = 0$	$\lambda = 10$	$\lambda = 100$
$\varepsilon = 0$	18.6283	1.7455	1.1255
$\varepsilon = 0.01$	10.1487	1.5769	1.0499
$\varepsilon = 0.1$	10.9906	2.4077	1.1169

Table 5.1.: Relative error (corresponding to the optimal solution $F_{10.000}^{4.6416,0}$ with respect to the exact data) of the solutions corresponding to the regularization parameters $\lambda = 0$, $\lambda = 10$ and $\lambda = 100$ and the parameters $\varepsilon = 0$, $\varepsilon = 0.01$ and $\varepsilon = 0.1$ denoting the noise level of the data input

When considering the relative error in Table 5.1, we clearly see that our chosen regularization works. For a certain noise level ε, the relative error corresponding to the unregularized solution $F^{0,\varepsilon}$ is certainly larger than the relative errors corresponding to the regularized solutions $F^{10,\varepsilon}$ and $F^{100,\varepsilon}$. In this example, we reach relative errors for the unregularized solution that are up to 10 times (but at least 4 times) larger than in the regularized cases. Note that the regularization parameter $\lambda = 10$ still corresponds to a weak regularization in our case. Thus, it is very important and appropriate to use a regularization method when confronted with noisy data as we mentioned before.

If we consider the development of the relative error with respect to a certain regularization parameter λ, we observe that, as has to be expected, the value increases when we raise the noise level from $\varepsilon = 0.01$ to $\varepsilon = 0.1$. Surprisingly, the value corresponding to $F^{100,0.1}$ is still rather small although we observe major boundary effects in this case (see Figure 5.14). We may conclude that the approximation quality does not suffer much from these effects.

Note that we are not able to make conclusions about the optimal value of the regularization parameter out of the values given in Table 5.1. One reason is that we compare relative values where the choice of the reference value might play an important role. Secondly, we cannot expect that the behavior with respect to one certain noise level ε is linear, e.g. for $\varepsilon = 0$ we get the values

relative	$\lambda = 0$	$\lambda = 4.6416$	$\lambda = 10$	$\lambda = 100$
$\varepsilon = 0$	18.6283	1	1.7455	1.1255

which is not a linear relation.

In summary, these considerations clearly demonstrate that the proposed regularization method works. Furthermore, this example showed that it is very important to use a regularization method when confronted with noisy data.

5.2. Remark: The Sparse Regularization Point of View (MRFMP_l^1)

As we discussed in Sections 3.1 and 4.4, from a sparse regularization point of view, it seems to be advantageous to use l^1-optimization in the penalty term of the Tikhonov regularization. Since we operate in function spaces we are stuck with the $\mathrm{L}^1(\mathcal{B})$-norm (see Section 4.4) which is rather expensive in computations in comparison to the $\mathrm{L}^2(\mathcal{B})$-norm.

In the Euclidean setting, usually the l^1-norm is minimized (see Section 3.1). Here, we want to show that it is not possible to just adapt methods from the Euclidean setting to our setting of an inverse problem on the ball. Thus, let us consider the l^1-norm of the coefficients α_k of the expansion of the solution as a penalty term in the regularization, i.e. we minimize

$$||R^n||^2_{\mathbb{R}^l} + \lambda ||\mathcal{A}_n||_{l^1}$$

where $\mathcal{A}_n = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}} \in \mathbb{R}^n$ consists of the coefficients of the expansion of

the solution F_n. Note that $||\mathcal{A}_n||_{l^1} = \sum_{k=1}^{n} |\alpha_k|$.

Let us now develop the corresponding algorithm. In step $n+1$, we are looking for the dictionary element d_{n+1} and the corresponding coefficient α_{n+1} that fulfill

$$(d_{n+1}, \alpha_{n+1}) = \text{argmin}_{d\in\mathcal{D},\alpha\in\mathbb{R}} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||\mathcal{A}_{n+1}||_{l^1} \right)$$

where $\mathcal{A}_{n+1} = (\alpha_1, \ldots, \alpha_n, \alpha)^{\mathrm{T}} \in \mathbb{R}^{n+1}$. We derive the optimal dictionary element d_{n+1} and the corresponding coefficient α_{n+1} with the same technique as before.

The minimizing α fulfills

$$\begin{aligned} 0 &= \frac{\partial}{\partial\alpha}\left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda\sum_{k=1}^{n}|\alpha_k| + \lambda|\alpha| \right) \\ &= -2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \text{ sgn } \alpha. \end{aligned}$$

and, thus,

$$\alpha = \frac{2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha}{2||\mathcal{F}d||^2_{\mathbb{R}^l}} \tag{5.3}$$

which is solved by

$$\alpha_{1/2} = \frac{2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} \pm \lambda}{2||\mathcal{F}d||^2_{\mathbb{R}^l}}.$$

Remember that we do not add to the expansion of the solution if $\alpha = 0$. Thus, we do not need to consider this case.

With this approach, we might produce a false solution. Thus, we have to verify whether Equation (5.3) holds true for α_1 and α_2.

Again, we insert the expression for α (see Equation (5.3)) into the target function

of the optimization:

$$\begin{aligned}
& ||R^n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \sum_{k=1}^{n} |\alpha_k| + \lambda|\alpha| \\
= \; & ||R^n||^2_{\mathbb{R}^l} - 2\alpha \langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + \alpha^2 ||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \sum_{k=1}^{n} |\alpha_k| + \lambda|\alpha| \\
= \; & ||R^n||^2_{\mathbb{R}^l} - \frac{2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha}{||\mathcal{F}d||^2_{\mathbb{R}^l}} \langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} \\
& + \left(\frac{2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha}{2||\mathcal{F}d||^2_{\mathbb{R}^l}} \right)^2 ||\mathcal{F}d||^2_{\mathbb{R}^l} \\
& + \lambda \sum_{k=1}^{n} |\alpha_k| + \lambda \frac{|2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha|}{2||\mathcal{F}d||^2_{\mathbb{R}^l}} \\
= \; & ||R^n||^2_{\mathbb{R}^l} + \lambda \sum_{k=1}^{n} |\alpha_k| + \frac{-4\langle R^n, \mathcal{F}d\rangle^2_{\mathbb{R}^l} + \lambda^2 + 2\lambda \, |2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha|}{4||\mathcal{F}d||^2_{\mathbb{R}^l}}.
\end{aligned}$$

And, consequently, a dictionary element d minimizes

$$||R^n - \alpha \mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \sum_{k=1}^{n} |\alpha_k| + \lambda|\alpha|$$

if and only if it minimizes

$$\frac{-4\langle R^n, \mathcal{F}d\rangle^2_{\mathbb{R}^l} + \lambda^2 + 2\lambda|2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha}{4||\mathcal{F}d||^2_{\mathbb{R}^l}}.$$

Thus, we get the following algorithm for the regularized version of the Functional Matching Pursuit with l^1-norm:

Algorithm 5.1 (Basic RFMP - Mixed l^1 (MRFMP_l^1))
Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ *such that*

$$d_{n+1} \textit{ minimizes } \frac{-4\langle R^n, \mathcal{F}d\rangle^2_{\mathbb{R}^l} + \lambda^2 + 2\lambda|2\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha|}{4||\mathcal{F}d||^2_{\mathbb{R}^l}} \textit{ and}$$

$$\alpha_{n+1} = \frac{2\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda \text{ sgn } \alpha_{n+1}}{2||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l}}.$$

In the top row of Figure 5.16, we display the result of the inversion with respect to the proposed l^1-regularization compared to the results we acquired before with

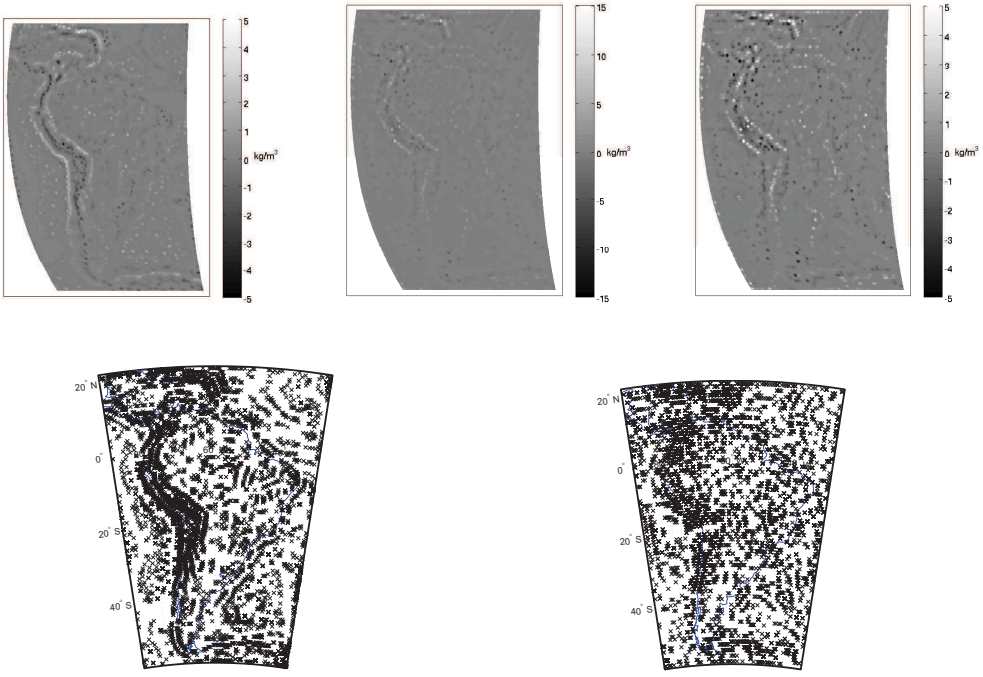

Figure 5.16.: Reconstructed density deviation (top) for the $\mathrm{L}^2(\mathcal{B})$-regularization (left-hand) and the proposed l^1-regularization with original (middle) and adapted colorbar (right-hand) and center points x of the chosen expansion functions $K_h^{\mathrm{I}}(x, \cdot)$ (bottom in the same order) computed out of $25,440$ data points with $10,000$ selected functions from the dictionary $\mathcal{D}$ given in Equation (5.1) for $\lambda = 4.6416$ ($\mathrm{L}^2(\mathcal{B})$-regularization) and $\lambda = 100$ (l^1-regularization)

$\mathrm{L}^2(\mathcal{B})$-regularization.

Clearly, the solution corresponding to the l^1-regularization is less smooth than the original solution. This has to be expected since it was acquired with no regard to this property of the solution. Although the main structures of South America are recovered with this regularization as well we get much more obvious errors in the reconstruction. Furthermore, the amplitudes are too extreme to be a good approximation.

Let us now consider the arrangement of the center points of the chosen expansion functions for both regularizations in the bottom row of Figure 5.16. Clearly and as expected, the expansion functions are not chosen according to the structure of the solution anymore. In the l^1-case, we get a very even arrangement of center points.

Overall, this example shows that ideas and methods from the Euclidean setting cannot just be applied to inverse problems in general. The ideas have to be adjusted carefully, as was done when developing a regularized algorithm in this work.

5.3. by Combining Gravitational and Normal Mode Data

As we mentioned before, we may only reconstruct the harmonic part of the density distribution out of the gravitational potential. For information about the anharmonic part, we need to include seismic data as well. For the joint inversion, we use the gravitational potential EGM2008 up to degree $2,190$ and order $2,159$ as data input. The data are given at $1,560$ data points on a sphere located 7 km above the surface of the Earth. As seismic data, we use $1,738$ splitting function coefficients corresponding to 49 different splitting functions (see Table 2.1) constructed out of the normal mode anomalies of the Earth recovered by Dr. Arwen Deuss, University of Cambridge. For further details on these data types we refer to Section 2.

Since we now operate on the whole Earth, i.e. the interior is included, we need a dictionary that includes anharmonic functions as well. In this section, the dictionary $\mathcal{D}$ is given by

$$\begin{aligned}\mathcal{D} &= \left\{ K_h^{\mathrm{I}}(x,\cdot) \;\middle|\; h \in \{0.9, 0.95, 0.99\}, x \in \mathrm{grid}(\mathcal{B}) \right\} \\ &\quad \cup \left\{ G_{m,n,j}^{\mathrm{I}} \;\middle|\; m,n = 0,\ldots,12, j = 1,\ldots,2n+1 \right\}\end{aligned} \tag{5.4}$$

where $\mathrm{grid}(\mathcal{B})$ is a Driscoll-Healy grid with $9,900$ points on spheres with three different radii $r_1 = 6,071$ km, $r_2 = 6,220.5$ km and $r_3 = 6,370$ km. Thus, $\mathrm{grid}(\mathcal{B})$ consists of $29,700$ points and the dictionary includes more than $89,000$ functions. We choose to reconstruct the upper 300 km of the interior of the Earth as a case study, since the main structures in this region are defined by subduction zones and ridges, i.e. we will be able to discuss the results critically. We stop the summation in the localized kernel functions $K_h^{\mathrm{I}}(x,\cdot)$ at degree $2,190$ in accordance to the degree of the spherical harmonics coefficients used to compute the data.

Algorithm 4.1 (RFMP) will be stopped after $N = 10,000$ expansion functions are chosen to reconstruct the target function.

Normal Mode Anomalies Only

First of all, let us consider the inversion for seismic data only. In Figure 5.17, we display the reconstruction of the density deviation up to a depth of 300 km plotted on an equatorial cut through the Earth (confer Figure 1.3) where we already labeled the main features, i.e. the main subduction zones and ridges, close to the equator. Clearly, the new method reproduces the known features in the upper part of the crust of the Earth.

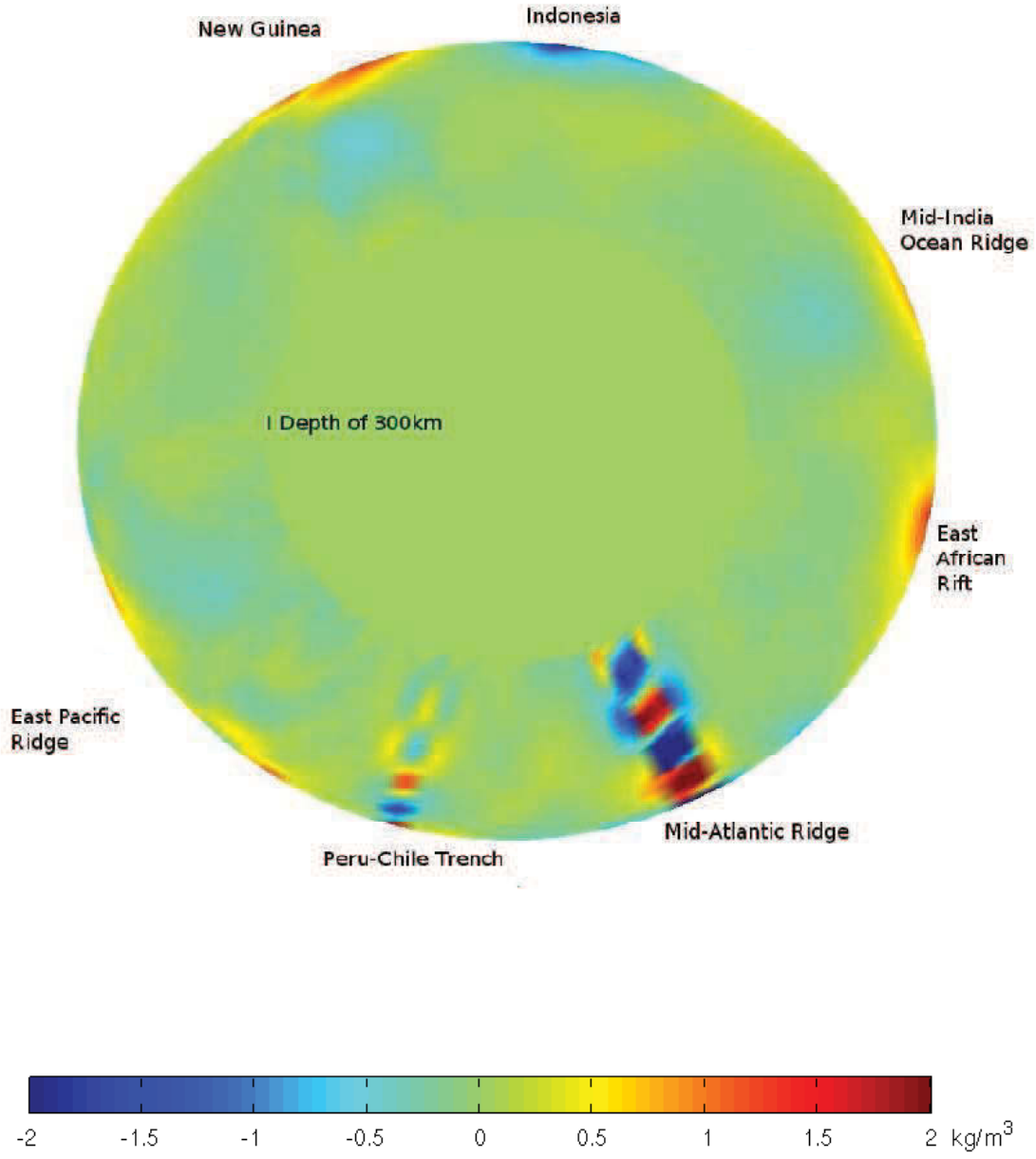

Figure 5.17.: Reconstructed density deviation in the upper 300 km (enlarged) of the x_2-x_3-plane (see Figure 1.3) computed out of $1,738$ splitting function coefficients corresponding to 49 different splitting functions with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ given in Equation (5.4) (i.e. $F_{10,000}$ is shown), $\lambda = 10^{-30}$

Secondly, we will investigate the approximation quality in more detail. Therefore, in the top of Figure 5.18, we compare the data, i.e. the splitting function coefficients corresponding to 49 different splitting functions (see Table 2.1), with the approximation $\mathcal{F}_{\mathrm{C}}F_{10,000}$. We display the values of the coefficients in blue, the values of the approximation in red and the difference in green.

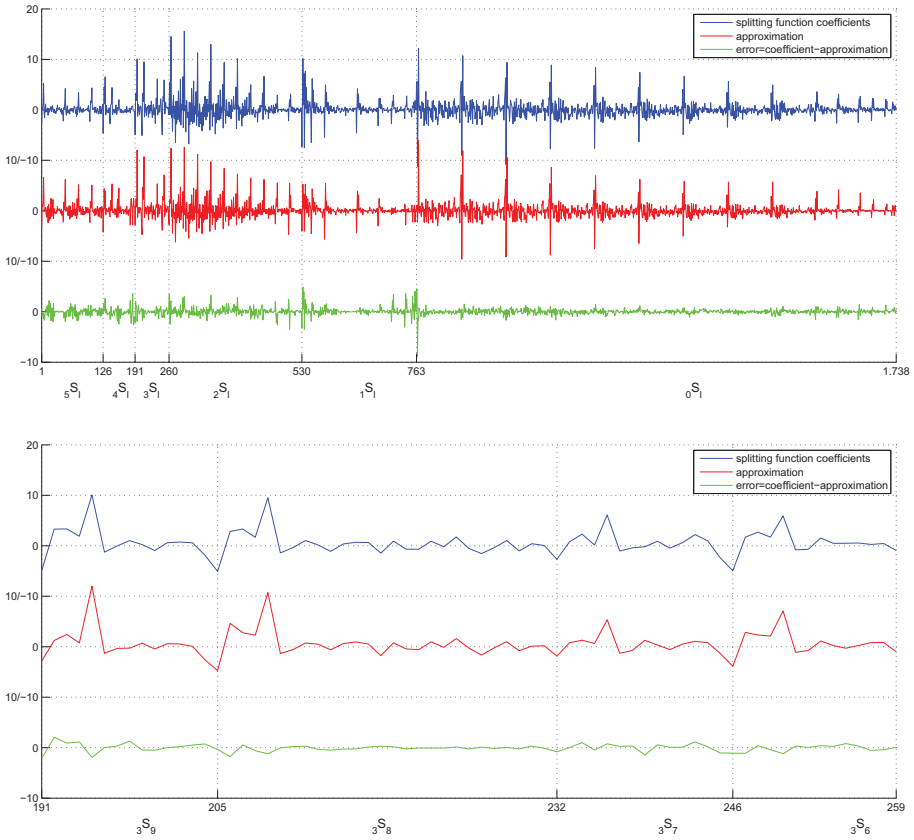

Figure 5.18.: Values of the splitting function coefficients (blue), approximation $\mathcal{F}_{\mathrm{C}}F_{10,000}$ (red) and error (green) for all $1,738$ coefficients (top) and the coefficients corresponding to the splitting functions ${}_3S_l$, $l = 6, \ldots, 9$, with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ given in Equation (5.4), $\lambda = 10^{-30}$

We can clearly see that the main features of the coefficients were reconstructed by our new method. The best match can be found in the coefficients corresponding to the splitting functions ${}_0S_l$, $l = 3, \ldots, 9, 12, \ldots, 17, 19, \ldots, 21$, i.e. the fundamental modes. In the bottom plot of Figure 5.18, we display the value of the coefficients, the approximation and the error for ${}_3S_l$, $l = 6, \ldots, 9$, where the algorithm again delivers a very good approximation to the input data. Most errors

are connected to the amplitude of the approximation and not to the structure itself. Maybe, further research in the direction of finding a fitting regularization parameter will provide us with an even better approximation.

Combined Inversion

If we combine different data types, we get the optimization functional

$$||y-\mathcal{F}F_n||^2_{\mathbb{R}^l}+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}=\sum_{i=1}^{l_\mathrm{a}}(y_i^\mathrm{a}-\mathcal{F}_\mathrm{a}^iF_n)^2+\sum_{i=1}^{l_\mathrm{b}}(y_i^\mathrm{b}-\mathcal{F}_\mathrm{b}^iF_n)^2+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}$$

where $\mathcal{F}_\mathrm{a}$ and $\mathcal{F}_\mathrm{b}$ are two operators corresponding to the two different data types $y^\mathrm{a}\in\mathbb{R}^{l_\mathrm{a}}$ and $y^\mathrm{a}\in\mathbb{R}^{l_\mathrm{b}}$, $l:=l_\mathrm{a}+l_\mathrm{b}$.

In our combination of gravitational data and normal mode anomalies, we expect the inversion to be dominated by the gravitational data (see [17, 18] where a joint inversion of these two data types was considered as well). Thus, we attempt to weight the data appropriately, i.e. we are looking for weights $w_i>0$, $i=1,\dots,l_\mathrm{a}+l_\mathrm{b}$ with respect to the optimization functional

$$||\tilde{y}-\tilde{\mathcal{F}}F_n||^2_{\mathbb{R}^l}+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}=\sum_{i=1}^{l_\mathrm{a}}w_i(y_i^\mathrm{a}-\mathcal{F}_\mathrm{a}^iF_n)^2+\sum_{i=1}^{l_\mathrm{b}}w_{l_{\mathrm{a}+i}}(y_i^\mathrm{b}-\mathcal{F}_\mathrm{b}^iF_n)^2+\lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})}$$

where $\tilde{y}=\mathrm{W}y$ and $\tilde{\mathcal{F}}=\mathrm{W}\mathcal{F}$ for the diagonal matrix $\mathrm{W}=\mathrm{diag}(\sqrt{w_1},\dots,\sqrt{w_{l_\mathrm{a}+l_\mathrm{b}}})\in\mathbb{R}^{(l_\mathrm{a}+l_\mathrm{b})\times(l_\mathrm{a}+l_\mathrm{b})}$. Note that the theoretical results hold for this optimization problem as well.

For our numerical considerations, we weight the data corresponding to the normal mode anomalies with a factor 10^5, i.e. $w_1,\dots,w_{l_\mathrm{G}}=1$ and $w_{l_\mathrm{G}+1},\dots,w_{l_\mathrm{G}+l_\mathrm{C}}=10^5$ where l_G corresponds to the number of gravitational data points and l_C corresponds to the number of splitting function coefficients. Furthermore, we choose the regularization parameter $\lambda=10^{-30}$ as in the case where we consider normal mode anomalies only.

In Figure 5.19, we display the result of this combined inversion. Moreover, in Figure 5.20, we display the difference between the result corresponding to normal mode anomalies only (see Figure 5.17) and the result corresponding to the combined inversion to present the gain of a combined inversion. Note that we use the same coordinate system as in Figure 5.17. Thus, the main features that we labeled in Figure 5.17 are still in the same areas in these figures corresponding to the combined inversion.

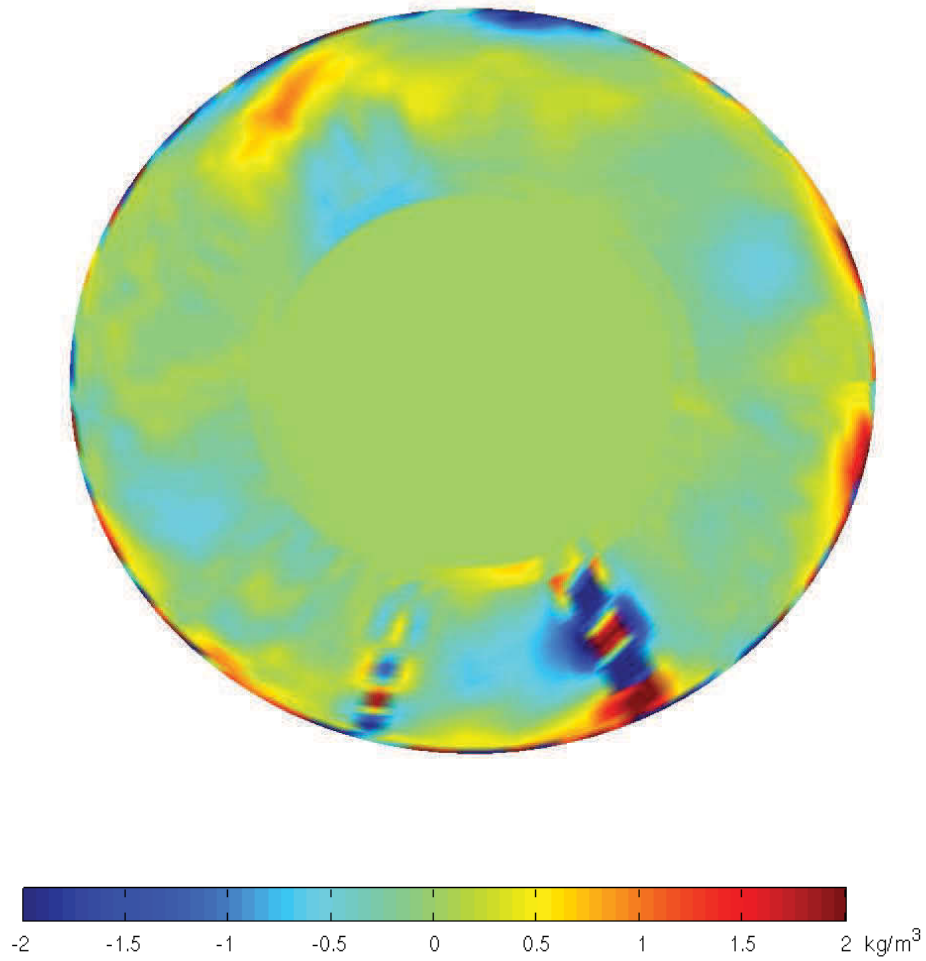

Figure 5.19.: Reconstructed density deviation in the upper 300 km (enlarged) of the x_2-x_3-plane (see Figure 1.3) computed out of $1,738$ splitting function coefficients corresponding to 49 different splitting functions weighted with a factor of 10^5 and $1,560$ gravitational data points with $10,000$ selected expansion functions from the dictionary $\mathcal{D}$ given in Equation (5.4) (i.e. $F_{10,000}$ is shown), $\lambda = 10^{-30}$

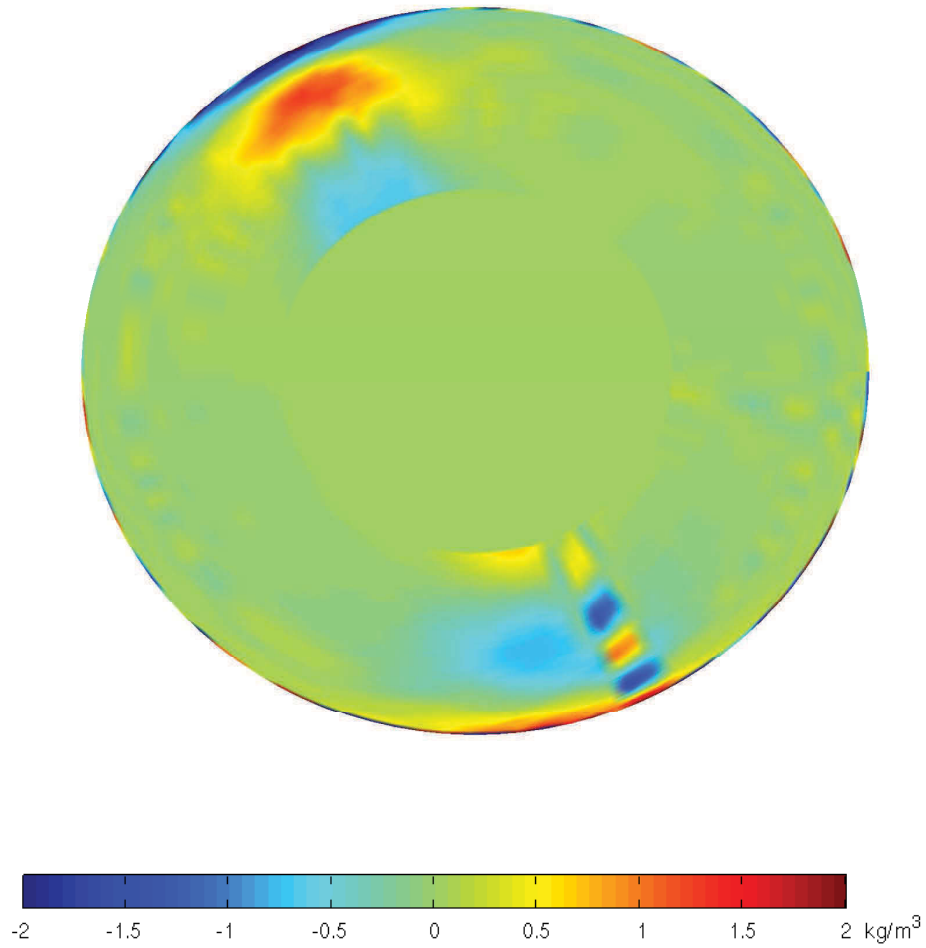

Figure 5.20.: Difference of the result of the combined inversion (see Figure 5.19) and the reconstruction out of normal mode data only (see Figure 5.17) in the upper 300 km (enlarged) of the x_2-x_3-plane (see Figure 1.3)

Clearly and as expected, we gain more information close to the surface of the Earth when we reconstruct the density deviation out of a combination of gravitational data and data corresponding to normal mode anomalies compared to the result corresponding to normal mode anomalies only. Furthermore, we reconstruct different structures in the area of New Guinea and Indonesia, i.e. in the upper part of the plots. Note that the solution is not dominated by the gravitational data as was observed in [17, 18] where the same data types were inverted with a spline method. By weighting, we seem to have reached some kind of balance between both data types.

In [17, 18] a vectorial spline method was developed to invert gravitational data and normal mode anomalies jointly. Comparing the results directly leads us to the conclusion that we get a much more detailed reconstruction with our new method. Next to using more recent and accurate data sets, this may be ascribed mainly to two features of our new method:
In [17, 18], the data corresponding to 10 splitting functions and 500 data points corresponding to the gravitational data were used for the combined inversion which leads to a large ill-conditioned matrix. Our new method allows us to consider the data corresponding to 49 different splitting functions and $1,560$ gravitational data points where we have not reached the limit of data points that may be considered by far.
Secondly, with our new approach, we are able to consider different types of functions in our dictionary to reconstruct the solution. This is, of course, not possible with a spline method.

Thus, our new method seems to be very well matched for the kind of problems considered in this work, since we are able to present a much more detailed model of the interior of the Earth that displays the main features of the upper crust of the Earth, i.e. the subduction zones and ridges.

Nonetheless, we see potential for improvements of these first results. First of all, it seems to be advisable to use point grids that are adapted to the depth, i.e. we might use a denser point grid close to the surface of the Earth. Secondly, we might link the choice of the localization parameter h to the depth, i.e. we might give different choices of localization parameters in different depths, since we expect a more detailed reconstruction close to the surface of the Earth. As a third improvement, we should choose the regularization parameter and the weights more sophisticated.

6. Reconstructing the Mass Transport in the Amazon Area

The satellite mission GRACE (Gravity Recovery And Climate Experiment) was started in 2002 to gain more information about the Earth's gravitational potential which allows us to detect climate phenomena like the water mass transport in the gravitational field (see [24]). The GRACE mission provides us with a monthly global coverage of the gravitational potential such that we are able to investigate temporal variations, too. In this section, we will concentrate on detecting the monthly mass transport, i.e. seasonal changes, in the Amazon area. It is very important to observe the mass transport in the Amazon area regularly, since it is one of the largest watersheds on Earth. Thus, it has been observed by other research groups using different techniques and observation periods (see [25, 26, 55, 139]).

We will use the monthly data provided by the Jet Propulsion Laboratory (JPL, see [85], Release 04). The data, i.e. the spherical harmonics coefficients, are given up to degree and order 120. To analyze the temporal variations of the monthly given gravitational potential of the Amazon area, we will subtract a mean potential from the monthly solutions and use this difference as an input to our algorithm. We use the available coefficients from July 2004 to June 2009 to compute the mean potential.

However, it is well-known that the higher degrees and orders contain noise that needs to be removed from our data. Thus, we have to use some kind of smoothing. Note that smoothing also attenuates the real signal such that we have to expect a change in magnitudes. In [139], a smoothing function with an effective Gaussian radius was suggested. In this work, we prefer wavelets to analyze variations in the gravitational potential of the Earth as suggested in [55], i.e. we consider

$$\sqrt{4\pi}\,\frac{\gamma M}{a}\sum_{n=3}^{E_n}\sum_{j=1}^{2n+1}\psi_J(n)\tilde{V}_{n,j}Y_{n,j}\left(\frac{x}{|x|}\right)$$

to reconstruct the data from the given spherical harmonics coefficients $\tilde{V}_{n,j}$. Note that γM is a constant representing the product of the gravitational constant γ and the Earth's mass M.

Here, we will use the P-wavelet corresponding to the CuP-scaling function as a filter (see also [65]), i.e. the function $\psi_J(n)$ defined as

$$\psi_J(n) = \begin{cases} (1-2^{-J-1}n)^4(1+2^{-J}n)^2 & \\ \quad -(1-2^{-J}n)^4(1+2^{-J+1}n)^2 & , n = 0, \ldots, 2^J - 1 \\ (1-2^{-J-1}n)^4(1+2^{-J}n)^2 & , n = 2^J, \ldots, 2^{J+1} - 1 \\ 0 & , n \geq 2^{J+1} \end{cases} .$$

As shown in Figure 6.1, the filter ψ_J controls up to which degree and to what extent the coefficients of the spherical harmonics are considered. An increasing scale J admits more detail information. However, it bears the risk to include errors or artefacts like satellite tracks as well. Thus, it has to be investigated carefully which filter yields a realistic and useful viable input. We refer to [55] for such an investigation for our case.

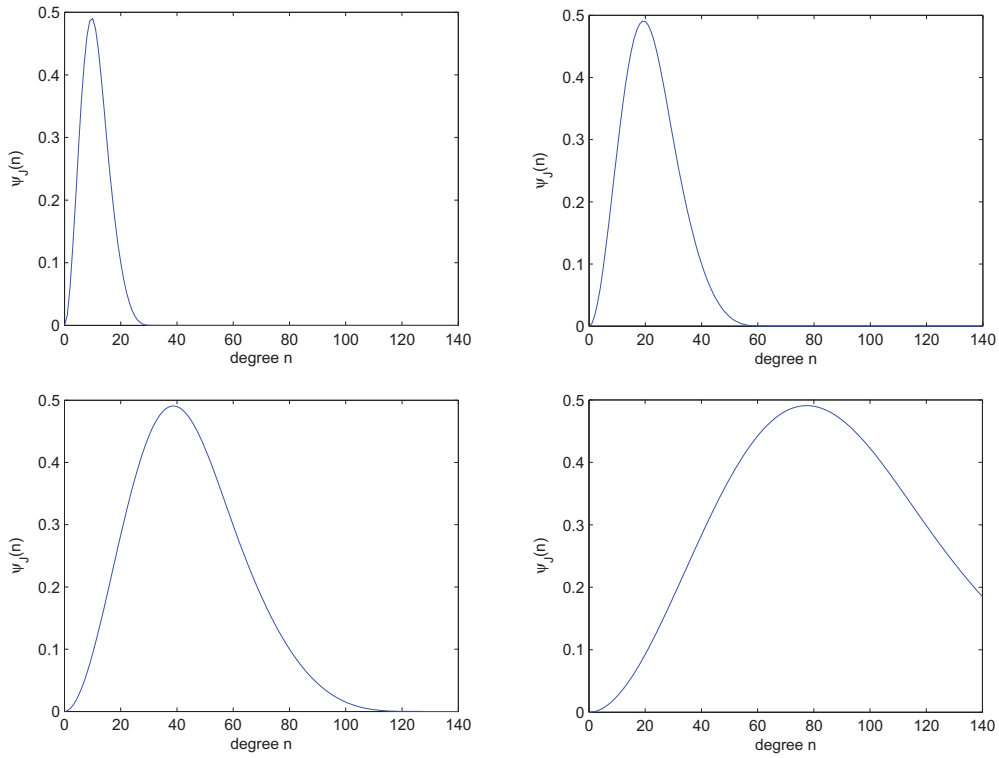

Figure 6.1.: ψ_J for $J = 4$ (top left-hand), $J = 5$ (top right-hand), $J = 6$ (bottom left-hand) and $J = 7$ (bottom right-hand)

In [55], the problem of choosing the right scale was already considered. As a result, it was proposed to use scale $J = 4$, since scale $J = 5$ already includes errors. In Figure 6.2, we display the difference between the mean gravitational potential

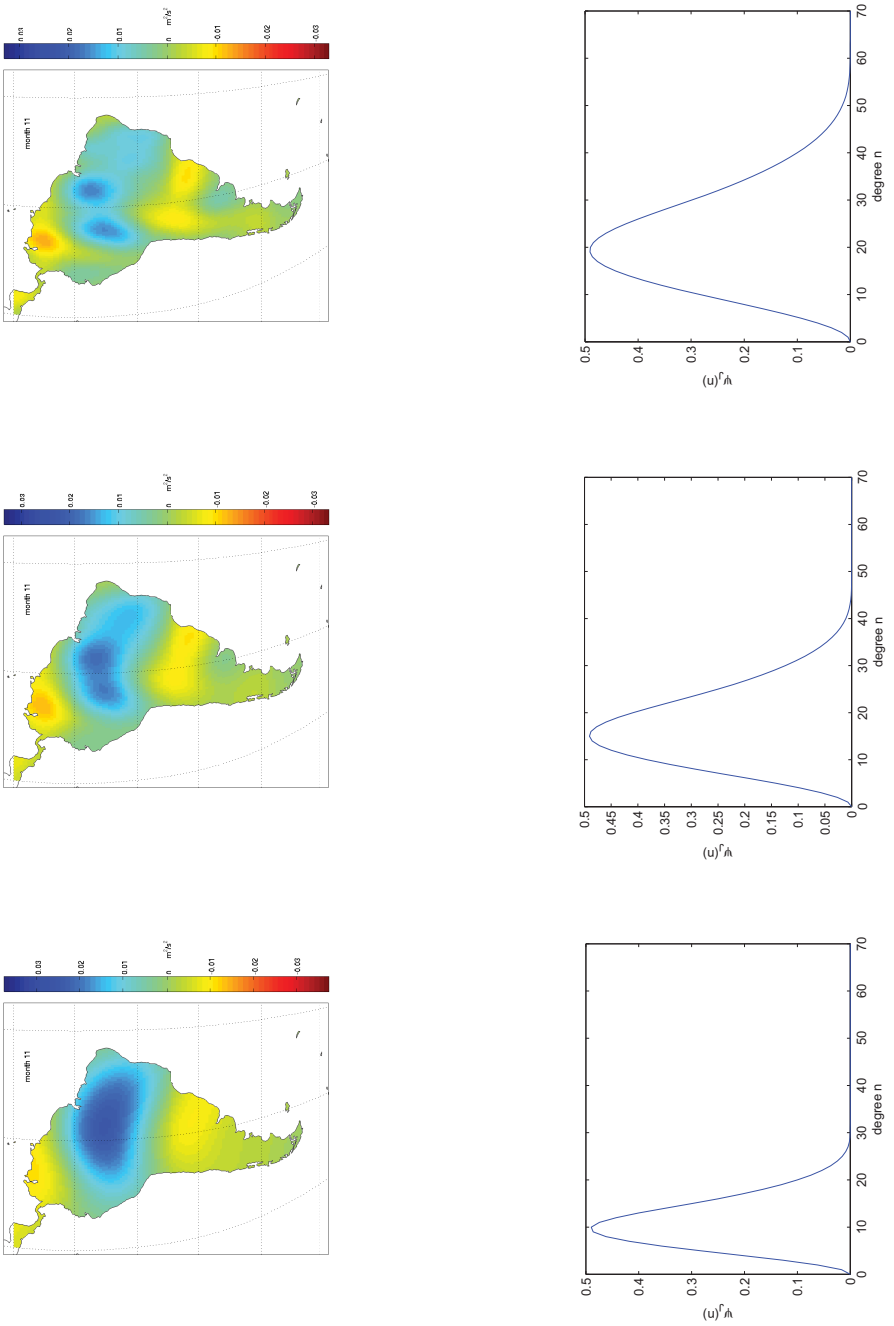

Figure 6.2.: Difference between the mean gravitational potential and the gravitational potential of November 2008 (left-hand column) at scale $J = 4$ (bottom), scale $J = \frac{\ln 25}{\ln 2}$ (middle) and scale $J = 5$ (top) as well as the CuP-filter functions $\psi_J(n)$ for $J = 4$, $J = \frac{\ln 25}{\ln 2}$ and $J = 5$ in the right-hand column (in the same order as above)

and the gravitational potential of November 2008 at scale $J = 4$ and scale $J = 5$ (bottom row and top row of Figure 6.2, respectively). Clearly, the potential at scale $J = 5$ includes too many data errors, e.g. the vertical structures probably indicate satellite tracks. On the other hand, the potential at scale $J = 4$ provides us with a very low scale data input and does not seem promising to recover the desired effects.

At scale $J = 4$, all spherical harmonics coefficients up to degree 31 are considered, while at scale $J = 5$ the degree of considered coefficients is increased up to 63. To get a better input to recover the seasonal changes in the Amazon area we choose a filter where all spherical harmonics coefficients up to degree 49 are considered (see middle row of Figure 6.2, $J = \frac{\ln 25}{\ln 2}$). Now we expect that there are enough details included to recover the desired effects where the contained noise is still suppressed sufficiently.

We will use the dictionary given in Equation (5.1), where the series in the kernel functions is terminated after 100 summations. Moreover, we will stop Algorithm 4.1 (RFMP) after $N = 10,000$ steps. As input we use $11,990$ data points given on a Driscoll-Healy grid computed on a sphere 7 km above the Earth's surface.

Again using the adapted version of the L-curve method to choose the regularization parameter λ, we would ideally choose the monthly parameters given in Table 6.1. In this work, we use the mean value of these regularization parameters, i.e. $\lambda = 8.7128$, for all months to keep comparability.

January	$\lambda = 11.6591$	July	$\lambda = 8.577$
February	$\lambda = 8.577$	August	$\lambda = 8.577$
March	$\lambda = 8.577$	September	$\lambda = 6.3096$
April	$\lambda = 8.577$	October	$\lambda = 8.577$
May	$\lambda = 6.3096$	November	$\lambda = 8.577$
June	$\lambda = 8.577$	December	$\lambda = 11.6591$

Table 6.1.: Ideal choice of regularization parameters λ in 2008

In Figures 6.3 and 6.4, we display the resulting density deviations for January 2008 up to December 2008. Here, the color blue denotes that the humidity is higher than in the mean, i.e. the surface and ground water levels are higher than in the mean, while red denotes that the humidity is lower than in the mean.

Looking at April 2008 (bottom right-hand plot in Figure 6.3), we conclude that we have a rainy season north of the equator and a dry season south of the equator. These findings conform with meteorological and hydrological observations

(see Global Land Data Assimilation System (GLDAS) [79, 125]). In comparison to September 2008 (top left-hand plot in Figure 6.4), we clearly see the seasonal changes, since now there is a dry area north of the equator and a wet region south of the equator.

Moreover, we observe a clear separation of the Amazon watershed and the Orinoco watershed in the north of South America. This separation is a very important feature to be reconstructed, since we do not only have a meteorological separation by the equator but a topographic separation by the Guiana highlands, too.

Overall, the displayed results conform to empiric data from a temporal perspective as well as from a spatial one, i.e. the changes appear in accordance to the seasons in the Amazon area and the equator seems to be a natural interface for the change of conditions.

In Figure 6.5, we display the centers x of the kernel functions $K_h^{\mathrm{I}}(x, \cdot)$ chosen by the algorithm where we artificially included the coast lines of South America for orientation. Again, we clearly see one of the advantages of our new method. The expansion functions are primarily chosen in accordance with the detail density of the solution.

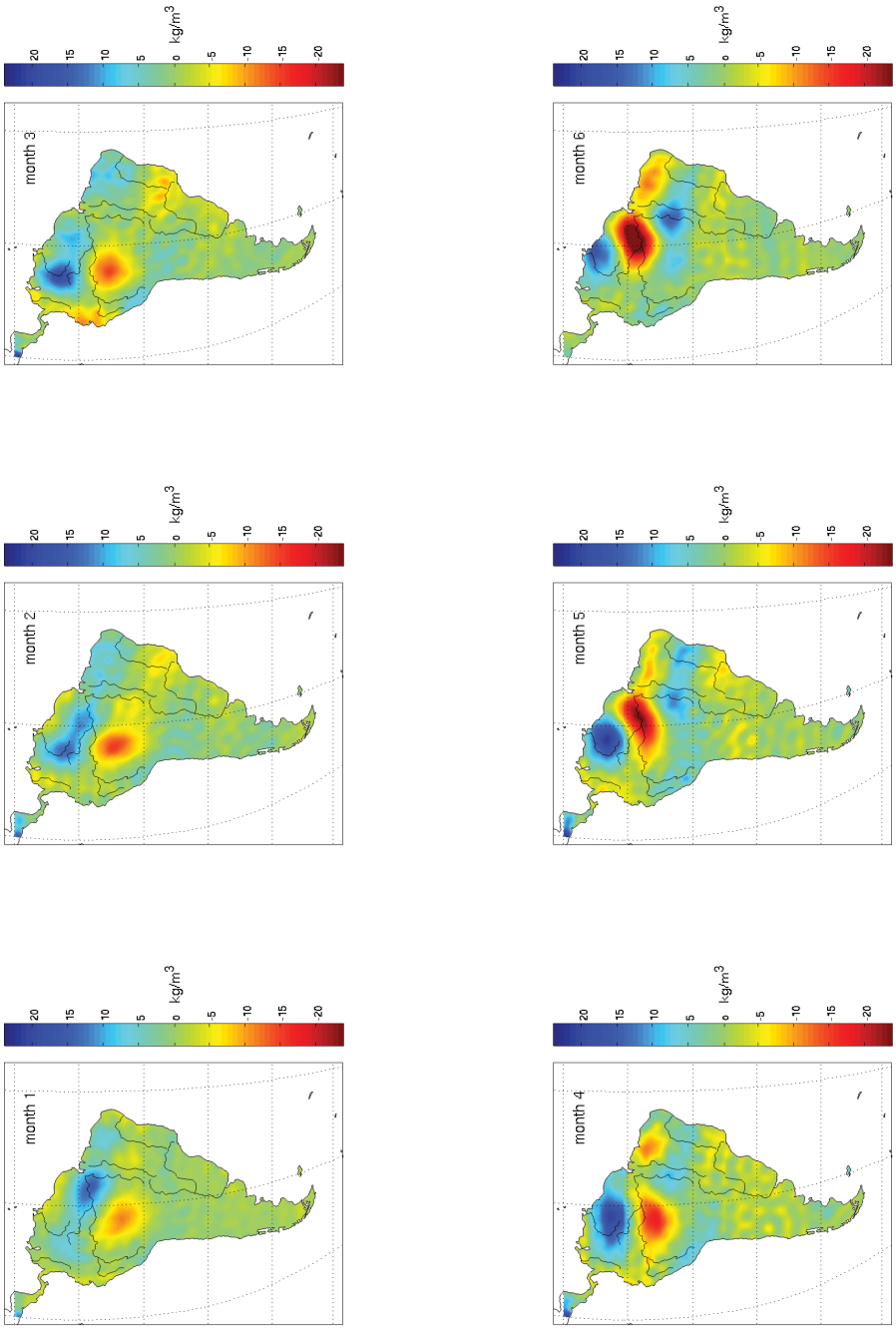

Figure 6.3.: Density deviation from January 2008 (bottom left-hand), February 2008 (middle left-hand) to June 2008 (top right-hand) computed out of 11,990 data points with 10,000 selected expansion functions, $\lambda = 8.7128$

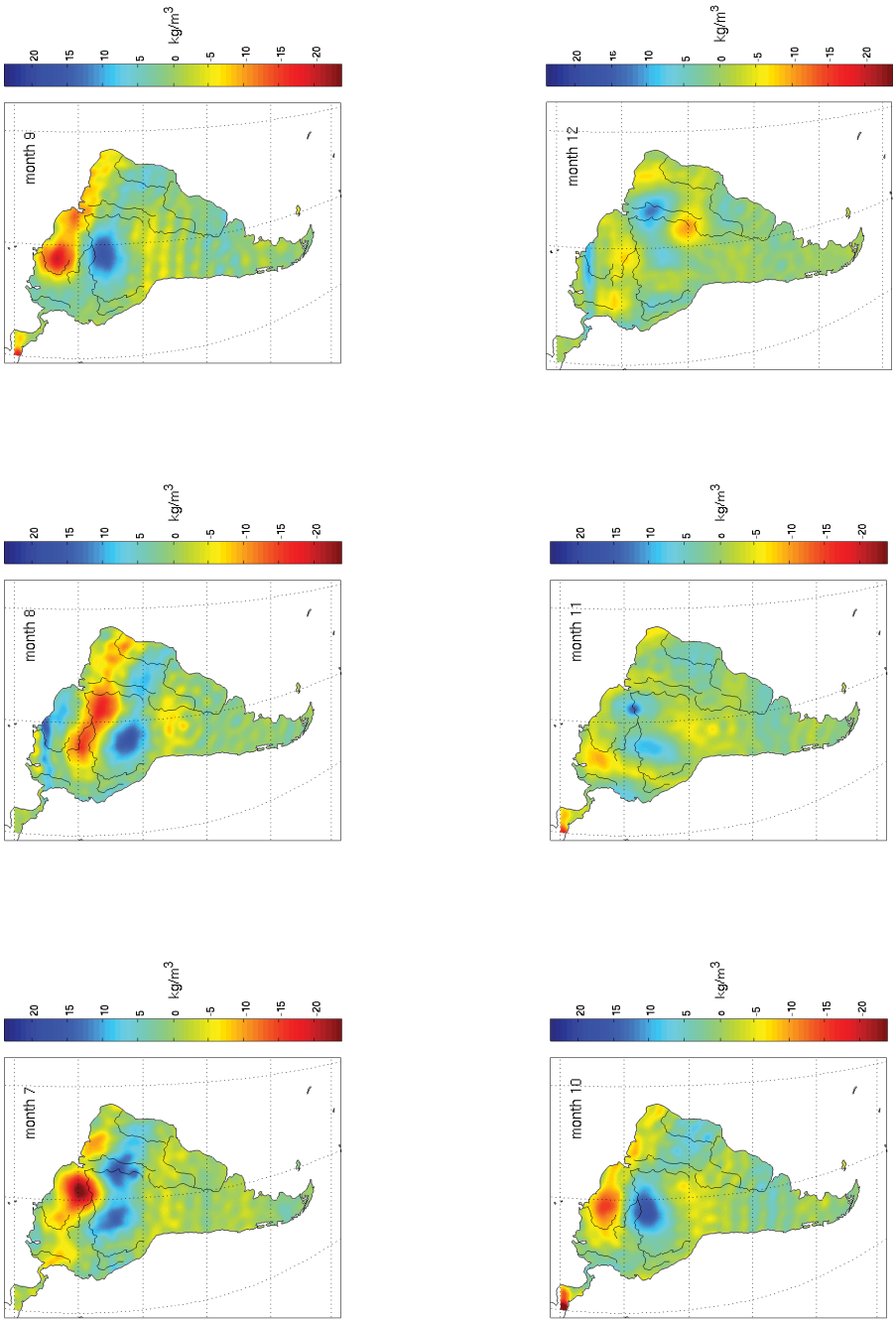

Figure 6.4.: Density deviation from July 2008 (bottom left-hand) to December 2008 (top right-hand) computed out of 11,990 data points with 10,000 selected expansion functions, $\lambda = 8.7128$

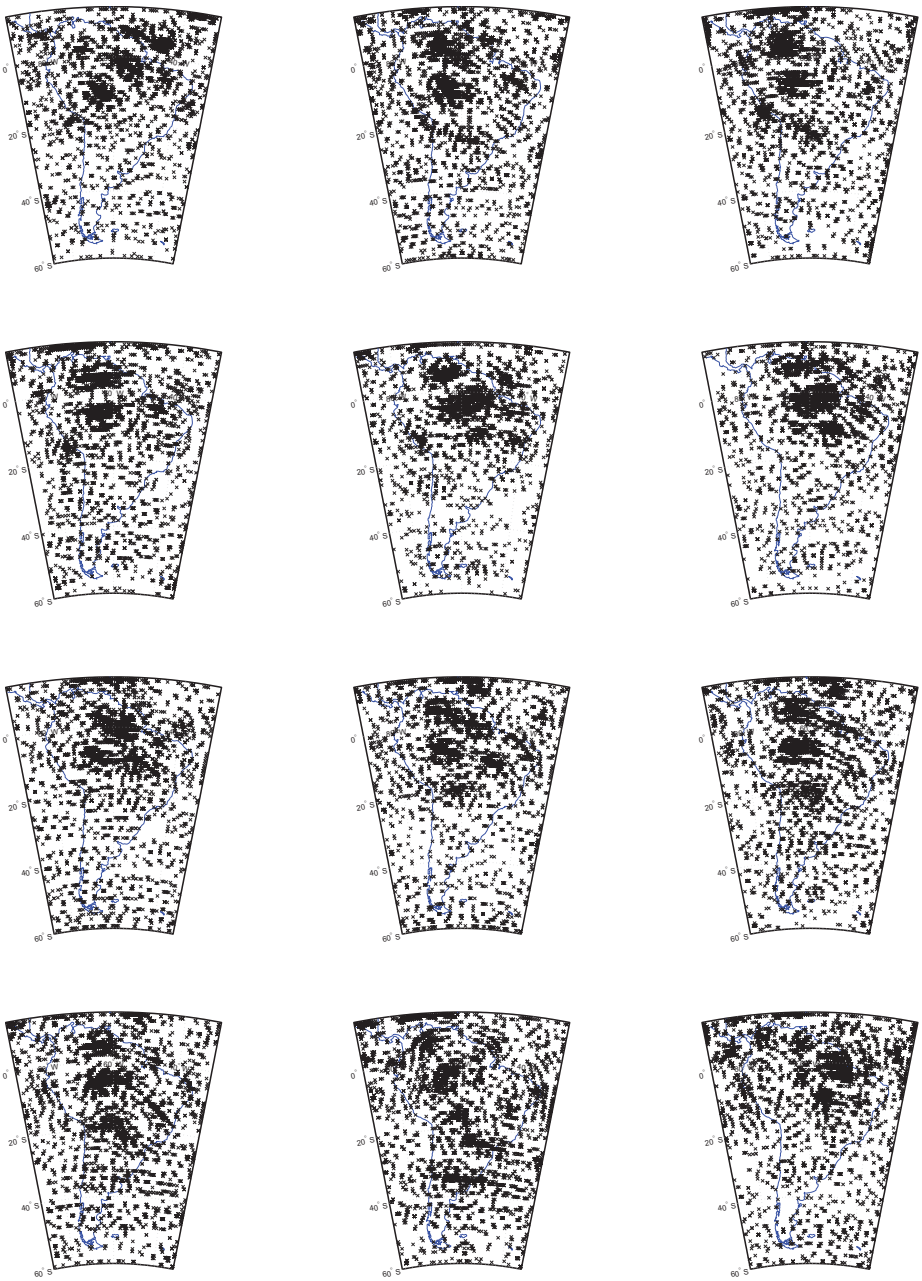

Figure 6.5.: Choice of centers x of the dictionary functions $K_h^{\mathrm{I}}(x, \cdot)$ for the year 2008 (upper left-hand: January, upper right-hand: March) computed out of $11,990$ data points with $10,000$ selected expansion functions, $\lambda = 8.7128$

IV.

In the End

7. Summary

Spline methods are a popular tool when solving ill-posed inverse problems like the inverse gravimetric problem. However, there are a few clear draw-backs of these methods that we were able to rectify in this work.

First of all, when inverting with a spline method we need to solve a linear equation system, whose size corresponds directly to the number of data points used for the inversion. However, systems corresponding to the very high number of data points that we use in this work, i.e. more than $25,000$ data points, cannot be handled numerically anymore. Because of its iterative character, our new method does not know such limitations.

Secondly, spline methods produce a solution that is adapted to the data structure. This, in itself, is a clear advantage over other already existing methods. However, it is even more advantageous to procure a solution that is adapted to the data structure and the detail structure of the solution itself. Again, our new method provides us with such a solution.

A third comment is that we may only use one kind of basis function for the expansion of the solution when using a spline method. However, the reconstruction may be more efficient if, for certain structures, different basis functions were available, e.g. functions with a global character to reconstruct global trends and localized functions to reconstruct detail structures. The new method allows such a mix of expansion functions.

The mystery ingredient is the idea of a Matching Pursuit which is an adaptive and iterative greedy algorithm introduced in signal recovery and subsequently used in the field of sparse regularization. However, we introduce three new concepts into the Matching Pursuit: The data is now given by a linear and continuous operator and not by the target function itself. Secondly, this algorithm will not depend on prior knowledge of the sparsity of the target function. Furthermore, we leave the Euclidean setting to operate on a setting of an inverse problem on the three-dimensional ball.

Since most real-world problems are ill-posed, we construct a regularization method, too, where we introduce different penalty terms, i.e. norms corresponding to the smoothness or the sparsity of the solution as well as the sparsity of the coefficients of the expansion itself. In this work, the best results are obtained with the $L^2(\mathcal{B})$-norm corresponding to the smoothness of the solution.

We prove convergence results for all developed algorithms as well as the convergence rate of the unregularized version of the method. Furthermore, we discuss the main properties of a regularization method, i.e. the existence and the stability of a solution as well as the convergence of the regularization.

As we mentioned before, another main feature of the new methods is that we may collect all different kinds of functions in a so-called dictionary to reconstruct different structures in the solution, accordingly. In our applications, we collect functions with a global character and localized kernel functions in our dictionary. We demonstrate the power of the new methods in a case study with respect to the reconstruction of the density distribution of South America out of satellite data where we have to solve the ill-posed inverse gravimetric problem. The regularized version of the algorithm provides us with solutions that are primarily matched to the structure of the target function and not only the data structure. We discussed the behavior and the results of all methods critically and in detail.

Furthermore, we see great potential in the method to be used in other research areas, too, as we demonstrated by reconstructing the mass transport in the Amazon area for the year 2008 out of satellite data. Of course, the method may be applied to other ecologically relevant problems, too, e.g. the investigation of the deglaciation of Antarctica and Greenland. Furthermore, we consider applications in medical imaging.

Reconstructing the density distribution out of gravitational data only, gives us information about the harmonic part of the density distribution exclusively. To get information about the anharmonic part and, thus, the interior of the Earth, we need to include seismic data, as well. However, the new method is very well matched to the joint inversion of different data types as we demonstrated in the joint inversion of gravitational data and seismic data in the form of normal mode anomalies where we used the most recent data available. We were able to present a new model of the density distribution in the interior of the Earth.

8. Outlook

For Algorithm 3.1 (FMP), i.e. the unregularized version, we were able to show that the speed of the convergence strongly depends on the choice of the dictionary. However, we have not been able to show a similar correlation for the regularized version, Algorithm 4.1 (RFMP), although it surely exists. In further research, we aim to prove the existence of such an interrelation and, in the best case, construct some indicator for the quality of the chosen dictionary depending, e.g. on the coherence of the dictionary elements or the vectors $\mathcal{F}d,\ d \in \mathcal{D}$, as done in related research areas.

Although the performance of the algorithm in our numerical considerations conformed very well with the theoretical expectations, we had to discuss some unexpected results as well. The choice of localized dictionary elements $K_h(x, \cdot)$ involves the choice of the center point x as well as the choice of the parameter h which controls the localizing character of the function. The center points were clearly chosen according to the structure of the target function, as intended. The use of localized kernel functions in the form of scaling functions only, might be a reason for the unexpected choices of the parameter h. In further research, we intend to rectify the problem by using a mixture of localized kernel functions in the form of a scaling function as well as in the form of a wavelet $d = K_h(x, \cdot) - K_{\tilde{h}}(x, \cdot)$ as dictionary elements.

In our further research, we will try to adapt other ideas from signal recovery and sparse regularization to our setting to improve our algorithm. The convergence behavior may be improved by using ideas from the Orthogonal Matching Pursuit, as we discussed before. As an alternative, we will implement the aforementioned back-projection where we use l^1-optimization to compute the (already recovered but sub-optimal) coefficients α_k for the expansion of the solution anew. Furthermore, we will try to upgrade the behavior of the method when confronted with noisy data which, obviously, includes treating boundary effects.

We assumed a rather simple relation between the density and the velocities to simplify our computations. It is, however, questionable whether a description in this way is viable. To solve this problem, we propose to use a vectorial approach as was done in, e.g. [17, 18].

In the applications, we demonstrated that our new method is capable to recon-

struct natural phenomena like the mass transport on the surface of the Earth. However, we have to remark that we were not able to use the full potential of the algorithms since the data that we used was given in spherical harmonics coefficients. We displayed before that spherical harmonics are global functions. Thus, local data errors are distributed to the whole model. Moreover, the detail structure of the data is compromised by this representation. In further research, we intend to include other, less preprocessed data types, too, to increase the performance of the algorithm and produce results that are as detailed and precise as needed in global environment studies.

In this work, we showed that the presented method also allows us to combine different data types. Our main goal is to recover an advanced model for the density distribution of the interior of the Earth. We explained before that gravitational data only gives information about the harmonic part of the density. However, the anharmonic part can be partially recovered by seismic data, e.g. normal mode anomalies or travel times. Here, we presented first results for a joint inversion of gravitational data and normal mode anomalies. In further research, we propose to study a combined inversion of even more different data types with our new method.

V.

Appendix

A. Treating Boundary Effects

If we limit our computations to only a part of the Earth's surface, e.g. the area around South America, and choose the expansion elements and the corresponding coefficients according to Algorithm 3.1 (FMP) or Algorithm 4.1 (RFMP) we observe minor to major boundary effects. These boundary effects can be distinguished clearly when considering noisy data (see the bottom row of Figure 5.14 and the top row of Figure 5.15 corresponding to the noise level $\varepsilon = 0.1$). We want to treat these boundary effects by influencing the choice of the optimal expansion elements, i.e. where the center points of the chosen expansion elements are located, directly with an appropriate weight function.

Remember that the second choice of dictionary functions $d = G^{\mathrm{I}}_{m,n,j}$ is a global function. Thus, it is of no importance when treating boundary effects.

In the next sections, we will introduce both a basic and a regularized algorithm that include a weight function to treat boundary effects. Furthermore, we will present examples for such weight functions. For theoretical results with respect to the algorithms developed in this section we refer to a later work. However, we expect that most results may be transferred.

Weighted Functional Matching Pursuit

To influence the choice of the optimal expansion functions we can, for example, choose to minimize the weighted residual $g(d)^2||R^n||^2_{\mathbb{R}^l}$ where g is the weight function that penalizes proximity of the centers of the localized dictionary elements to the boundaries of the considered area. Its value $g(d)$ will depend only on the center point of the chosen kernel function $d = K_h(x, \cdot)$. If the chosen expansion function is $d = G_{m,n,j}$ the weight function g will not influence this choice. Moreover, the weight function g has to fulfill some more minor requirements:

Assumption A.1
We assume that the weight function g maps continuously onto a codomain $[a, b]$ where $0 < a < b$, i.e. g is strictly positive.

We will give examples for such admissible weight functions later in this section.

To derive the corresponding algorithm, we again assume that, in step $n+1$, F_n is given. Since $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$, we are looking for a combination $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$ that minimizes $g(d)^2 \, ||y - \mathcal{F}(F_n + \alpha d)||^2_{\mathbb{R}^l}$, i.e.

$$(d_{n+1}, \alpha_{n+1}) = \text{argmin}_{d \in \mathcal{D}, \alpha \in \mathbb{R}} \; g(d)^2 ||R^n - \alpha \mathcal{F} d||^2_{\mathbb{R}^l},$$

since the operator $\mathcal{F}$ is assumed to be linear.

For all $d \in \mathcal{D}$, the minimizing α is then given by

$$0 = \frac{\partial}{\partial \alpha} g(d)^2 ||R^n - \alpha \mathcal{F} d||^2_{\mathbb{R}^l} = g(d)^2 \frac{\partial}{\partial \alpha} ||R^n - \alpha \mathcal{F} d||^2_{\mathbb{R}^l}.$$

Clearly, this is the same partial derivative as in the case where we do not weight. Thus, we get

$$0 = g(d)^2 \left(-2 \langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} + 2\alpha ||\mathcal{F} d||^2_{\mathbb{R}^l} \right)$$

and, since $g(d)$ was assumed to be strictly positive for all dictionary elements $d \in \mathcal{D}$ (see Assumption A.1), we get the same formula for α as before

$$\alpha = \frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l}}{||\mathcal{F} d||^2_{\mathbb{R}^l}}.$$

Again, we use this representation of α to derive an expression from which we can determine the optimal dictionary element d_{n+1} at step $n+1$.

Let us insert the expression for α into the weighted residual $g(d)^2 ||R^n - \alpha \mathcal{F} d||^2_{\mathbb{R}^l}$. With the same steps in the calculation as in the case without the weight function g (see Section 3.2) we get

$$g(d)^2 ||R^n - \alpha \mathcal{F} d||^2_{\mathbb{R}^l} = g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} - \left(\frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l}}{||\mathcal{F} d||_{\mathbb{R}^l}} \right)^2 \right).$$

Now let us state the complete algorithm:

Algorithm A.2 (Weighted FMP (WFMP))

Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ *such that*

$$d_{n+1} \text{ minimizes } g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} - \left(\frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l}}{||\mathcal{F} d||_{\mathbb{R}^l}} \right)^2 \right) \text{ and}$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F} d_{n+1} \rangle_{\mathbb{R}^l}}{||\mathcal{F} d_{n+1}||^2_{\mathbb{R}^l}}.$$

Although the expression for the choice of the dictionary function now includes the norm of the residual, we do not expect an increase in the computation time since the norm of the residual was already updated and stored in the preceding step from $n-1$ to n. Note that the weight function directly influences the choice of the optimal dictionary element only, as we required for this version of the algorithm.

Weighted Regularized Functional Matching Pursuit

In this section, we will derive a regularized version of Algorithm A.2 (WFMP) by weighting the target function, i.e. we are looking for the dictionary element and the corresponding coefficient such that

$$(d_{n+1}, \alpha_{n+1}) = \text{argmin}_{d\in\mathcal{D},\alpha\in\mathbb{R}}\ g(d)^2 \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right),$$

in the step from n to $n+1$ chosen expansion functions. Again, the weight function is denoted by g and fulfills Assumption A.1. We can derive the minimizing dictionary element d_{n+1} and the corresponding coefficient α_{n+1} with the same technique as before.

The minimizing α fulfills

$$\begin{aligned} 0 &= \frac{\partial}{\partial\alpha} g(d)^2 \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\ &= g(d)^2 \frac{\partial}{\partial\alpha} \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\ &= g(d)^2 \left(-2\ \langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} + 2\alpha||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda \left(2\ \langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})} + 2\alpha||d||^2_{\mathrm{L}^2(\mathcal{B})} \right) \right) \end{aligned}$$

(see Section 4.1 for more details) and with Assumption A.1, i.e. $g(d) > 0$ for all dictionary elements $d \in \mathcal{D}$, we get

$$\alpha = \frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}}.$$

Again from Section 4.1, we can derive that the dictionary element that is chosen in step $n+1$ minimizes

$$\begin{aligned} & g(d)^2 \left(||R^n - \alpha\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||F_n + \alpha d||^2_{\mathrm{L}^2(\mathcal{B})} \right) \\ = \ & g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} + \lambda||F_n||^2_{\mathrm{L}^2(\mathcal{B})} - \frac{\left(\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})} \right)^2}{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda||d||^2_{\mathrm{L}^2(\mathcal{B})}} \right). \end{aligned}$$

Thus, we get the following algorithm for the weighted regularized version of the Functional Matching Pursuit:

Algorithm A.3 (Weighted RFMP (WRFMP))
Start with $F_0 = 0$.
Given F_n.
Build $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ *such that*

$$d_{n+1} \text{ minimizes } g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} + \lambda ||F_n||^2_{\mathrm{L}^2(\mathcal{B})} - \frac{\left(\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})} \right)^2}{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda ||d||^2_{\mathrm{L}^2(\mathcal{B})}} \right) \text{ and}$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1} \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(\mathcal{B})}}{||\mathcal{F}d_{n+1}||^2_{\mathbb{R}^l} + \lambda ||d_{n+1}||^2_{\mathrm{L}^2(\mathcal{B})}}.$$

Note that this algorithm, too, fulfills our expectation that the weight function only influences the choice of the dictionary element d_{n+1} directly. The computation of the corresponding coefficient depends only indirectly on the weight function.

An Appropriate Weight Function

To treat the boundary effects occurring in our applications, we have to use a weight function that penalizes the choice of a kernel function that is centered near the boundary. Since we minimize, the penalty term should have a large value close to the boundary and a small one in the center of the considered area.

For the sake of clearness, we start with a cosine-roll-off (see Figure A.1) and put on the constraints for the minimization at a later stage, i.e. we choose a weight function that has the value 1 in the middle and decreases towards the boundaries. Let us first give an example of such a weight function on the interval. We will enlarge that definition to a plane and, then, fit it to fulfill Assumption A.1. At last, we will fit it to the requirements of Algorithms A.2 (WFMP) and A.3 (WRFMP).

Let us introduce a cosine-roll-off (see [9]) that can be used as a weight function on the interval $[-1, 1]$. With the help of the parameter $\beta \in [0, 1]$ we will define a cosine flank where β allows us to directly control the length of the flank as well as its position.

Example A.4
On the interval $[-1, 1]$ *the cosine-roll-off can be defined by*

$$f_{\mathrm{1D}}(x) = \begin{cases} 1 & , \ |x| \leq \frac{1-\beta}{2} \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{\beta}\left(|x| - \frac{1-\beta}{2}\right)\right)\right) & , \ \frac{1-\beta}{2} < |x| \leq \frac{1+\beta}{2} \\ 0 & , \ otherwise \end{cases} , x \in [-1, 1]$$

with the parameter $\beta \in [0, 1]$. *In Figure A.1 we display the effect of the parameter choice, i.e. that the slope of the flank steepens with decreasing* β. *Note that the*

codomain of the weight function is $[0, 1]$. *Furthermore, the weight function is continuous.*

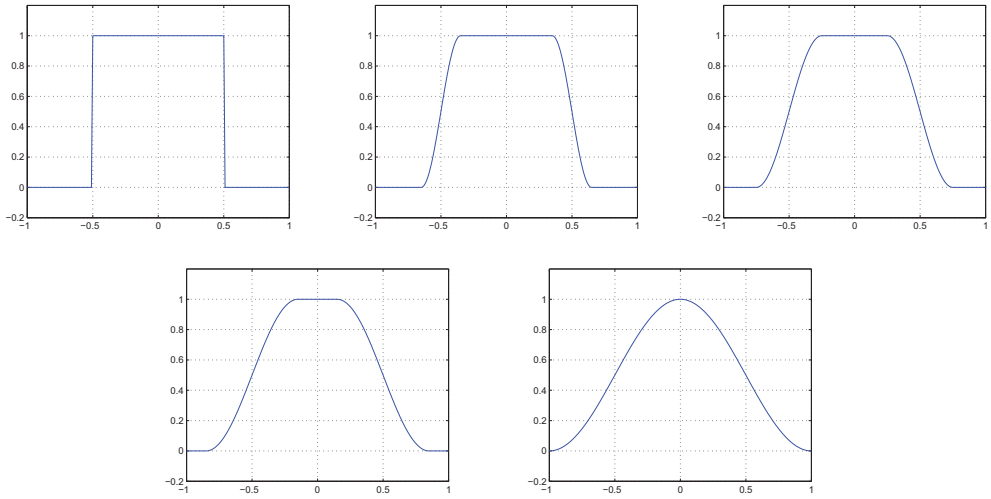

Figure A.1.: Cosine-roll-off f_{1D} on $[-1, 1]$ with parameters $\beta = 0$, $\beta = 0.3$, $\beta = 0.5$, $\beta = 0.7$, $\beta = 1$ (in order of the plots from top left-hand to bottom right-hand)

Now, we will consider two possibilities to expand a one-dimensional weight function on the interval $[-1, 1]$ to a two-dimensional setting on the domain $[-1, 1] \times [-1, 1]$.

Example A.5
We can build a two-dimensional continuous weight-function on $[-1, 1] \times [-1, 1]$

(i) by taking the product of two one-dimensional weight functions on $[-1, 1]$, *i.e.*

$$f_{2D}(x, y) = f_{1D}(x) f_{1D}(y), \ x, y \in [-1, 1],$$

or

(ii) by rotating a one-dimensional weight function on $[-1, 1]$, *i.e.*

$$f_{2D}(x, y) = f_{1D}\left(\sqrt{x^2 + y^2}\right), \ x, y \in [-1, 1].$$

Examples for both possibilities are displayed in Figure A.2.
Clearly, possibility (i) *punishes proximity to the boundaries while possibility* (ii) *punishes distance to the center point* $(0, 0)$.

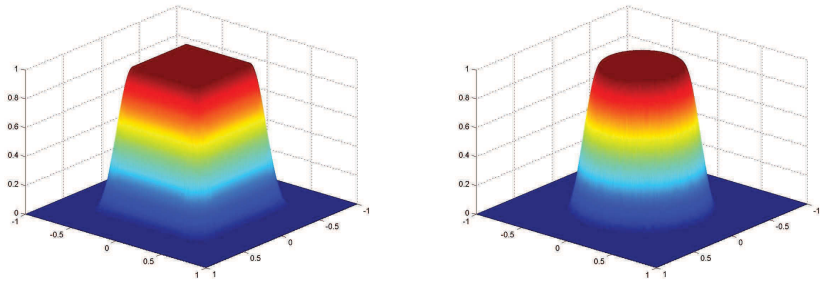

Figure A.2.: Two-dimensional cosine-roll-off f_{2D} on $[-1,1] \times [-1,1]$ with parameter $\beta = 0.3$ as the product of two one-dimensional weight functions on $[-1,1]$ (left-hand) and as a rotation of a one-dimensional weight function on $[-1,1]$ (right-hand)

Looking at the two possibilities to build a two-dimensional weight function on $[-1,1] \times [-1,1]$ in Example A.5, we clearly prefer the first possibility for our purpose since we want to punish proximity to the boundaries.

In Assumption A.1, we require that 0 is not in the codomain of the weight function that we want to use. The weight functions, that we introduced so far, do not fulfill this requirement. However, we can change this fact by scaling accordingly.

Lemma A.6
The codomain of the two-dimensional weight function g_{2D}, which is displayed in Figure A.3 and defined by

$$g_{2D}(x,y) = g_{1D}(x)g_{1D}(y), \;\; x,y \in [-1,1]$$

where

$$\begin{aligned} g_{1D}(x) &= \begin{cases} 1 & , \; |x| \leq 1-\beta \\ \frac{1}{2}\left(1+\cos\left(\frac{\pi}{2\beta}\left(|x|-1+\beta\right)\right)\right) & , 1-\beta < |x| \leq 1+\beta \\ 0 & otherwise \end{cases} \quad , x \in [-1,1], \\ &= \begin{cases} 1 & , \; |x| \leq 1-\beta \\ \frac{1}{2}\left(1+\cos\left(\frac{\pi}{2\beta}\left(|x|-1+\beta\right)\right)\right) & , 1-\beta < |x| \leq 1 \end{cases} , \end{aligned}$$

with $\beta \in [0,1]$ fulfills the requirements of Assumption A.1, i.e.

$$g_{2D}(x,y) \in [a,b], \;\; 0 < a < b, \;\; \textit{for all } x,y \in [-1,1],$$

where g_{2D} is continuous.

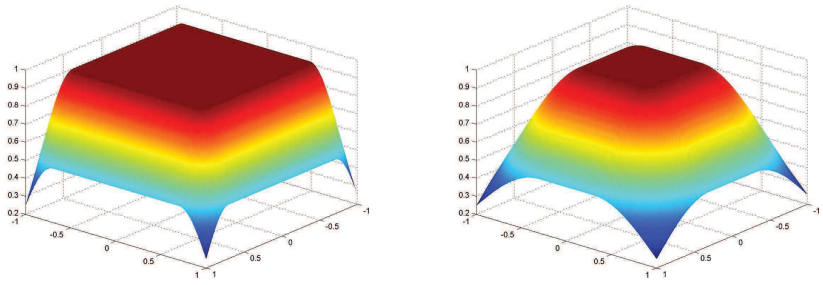

Figure A.3.: Two-dimensional cosine-roll-off on $[-1,1] \times [-1,1]$ with parameters $\beta = 0.3$ (left-hand) and $\beta = 0.7$ (right-hand) as the product of two adapted one-dimensional weight functions on $[-1,1]$

Proof.
First, we will show that g_{1D} fulfills the requirements of Assumption A.1. From the nature of the assumption and our two-dimensional weight function g_{2D} we can then conclude that g_{2D} fulfills the requirements of Assumption A.1, too.

Since we only included continuous functions in the definition of g_{2D}, we can conclude that it is continuous, too.

Clearly, we only need to take a look at the case where $1 - \beta < |x| \leq 1$ and, accordingly, $\beta \in]0,1]$ since otherwise $g_{1D}(x) = 1$.

Since $1 - \beta < |x| \leq 1$, we know that

$$0 < \frac{\pi}{2\beta}\left(|x| - 1 + \beta\right) \leq \frac{\pi}{2}$$

and, consequently,

$$0 \leq \cos\left(\frac{\pi}{2\beta}\left(|x| - 1 + \beta\right)\right) < 1.$$

Thus, we can derive for x with $1 - \beta < |x| \leq 1$ that

$$\frac{1}{2} \leq g_{1D}(x) = \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{2\beta}\left(|x| - 1 + \beta\right)\right)\right) < 1,$$

i.e. we can conclude that $g_{1D}(x) > 0$ for all $x \in [-1,1]$.

□

Note, that for $x, y \in [-1, 1]$

$$\frac{1}{4} \leq g_{2\mathrm{D}}(x, y) = g_{1\mathrm{D}}(x) g_{1\mathrm{D}}(y) < 1,$$

i.e. $g_{2\mathrm{D}}$ maps to the interval $[0.25, 1[$ as may be observed in Figure A.3 where we display the two-dimensional cosine-roll-off for different parameters β.

As a last step, we need to know how to apply the weight function $g_{2\mathrm{D}}$ introduced in Lemma A.6 to our dictionary functions. Furthermore, we want to introduce intervals different to $[-1, 1]$ since we need to characterize the part of the Earth's surface that we limit our computation to, i.e. we want to operate on the interval $[r_1, r_2], 0 < r_1 < r_2$, which is, typically for our case, a subinterval of $[0, 2\pi[$ (longitude) or $[0, \pi]$ (latitude). Therefore, we need to introduce new parameters $0 < r_1 < r_2$ to the definition of $g_{1\mathrm{D}}$ and adapt it from the interval $[-1, 1]$ to the new interval $[r_1, r_2]$ by the transformation

$$x = 2\,\frac{r - r_1}{r_2 - r_1} - 1, x \in [-1, 1], r \in [r_1, r_2].$$

Thus, we get a new one-dimensional weight function on the interval $[r_1, r_2]$

$$g_{1\mathrm{D}}(r, r_1, r_2) = \begin{cases} 1 & , \ |x| \leq 1 - \beta \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{2\beta}\left(|x| - 1 + \beta\right)\right)\right) & , \ 1 - \beta < |x| \leq 1 \end{cases},$$

$$x = 2\,\frac{r - r_1}{r_2 - r_1} - 1, \ r \in [r_1, r_2].$$

Let us define a two dimensional weight function on an arbitrary area that is characterized by the two intervals $[\varphi_1, \varphi_2]$ and $[\vartheta_1, \vartheta_2]$ and a dictionary element as input

$$g_{2\mathrm{D}}(d) = \begin{cases} 1 & , \ d = G_{m,n,j} \\ g_{1\mathrm{D}}(\varphi, \varphi_1, \varphi_2) g_{1\mathrm{D}}(\vartheta, \vartheta_1, \vartheta_2) & , \ d = K_h(x(\varphi, \vartheta, r), \cdot) \end{cases}. \tag{A.1}$$

This function, too, fulfills Assumption A.1 and now maps to the interval $[0.25, 1]$.

Let us consider the role of the weight function in Algorithm A.2 (WFMP). It is a factor of the expression that is to be minimized, i.e.

$$d_{n+1} \text{ minimizes } g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} - \left(\frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l}}{||\mathcal{F}d||_{\mathbb{R}^l}} \right)^2 \right).$$

We want to penalize the choice of a dictionary function which is centered close to the boundaries and, thus, the weight function should have larger values close

to the boundaries than in the middle of the considered area. The aforementioned weight functions do not fulfill this requirement. Thus, we will consider

$$g(d) = (1.25 - g_{2\mathrm{D}}(d)) \in [a, b],$$

where $g_{2\mathrm{D}}(d)$ maps to the interval $[0.25, 1]$ as in Equation (A.1), as a weight function for the weighted version of the Functional Matching Pursuit. Note that $g(d)$ now maps to the interval $[0.25, 1]$, too.

In the regularized case (see Algorithm A.3 (WRFMP)) we, too, minimize the expression where the weight function appears, i.e.

$$d_{n+1} \text{ minimizes } g(d)^2 \left(||R^n||^2_{\mathbb{R}^l} + \lambda ||F_n||^2_{\mathrm{L}^2(\mathcal{B})} - \frac{\left(\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d\rangle_{\mathrm{L}^2(\mathcal{B})}\right)^2}{||\mathcal{F}d||^2_{\mathbb{R}^l} + \lambda ||d||^2_{\mathrm{L}^2(\mathcal{B})}} \right)$$

and, thus, we can choose the same weight function as in the unregularized case, i.e.

$$g(d) = (1.25 - g_{2\mathrm{D}}(d)).$$

Now let us formally define the weight function g which we will use in our further considerations to treat boundary effects.

Definition A.7
The weight function $g : \mathcal{D} \to [0.25, 1]$ fulfills Assumption A.1. It is defined by

$$g(d) = \begin{cases} 1 & , \ d = G_{m,n,j} \\ (1.25 - g_{1\mathrm{D}}(\varphi, \varphi_1, \varphi_2) g_{1\mathrm{D}}(\vartheta, \vartheta_1, \vartheta_2)) & , \ d = K_h(x(\varphi, \vartheta, r), \cdot) \end{cases},$$

where $g_{1\mathrm{D}}$ is given by

$$g_{1\mathrm{D}}(r, r_1, r_2) = \begin{cases} 1 & , \ |x| \leq 1 - \beta \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{2\beta}(|x| - 1 + \beta)\right)\right) & , \ 1 - \beta < |x| \leq 1 \end{cases},$$

$x = 2\,\dfrac{r - r_1}{r_2 - r_1} - 1, \ r \in [r_1, r_2]$.
Here, $[\varphi_1, \varphi_2]$ and $[\vartheta_1, \vartheta_2]$ are subintervals of $[0, 2\pi[$ and $[0, \pi]$, respectively. They characterize the part of the Earth's surface that we limit our computations to.

For some applications, weighting with a function that maps to the interval $[0.25, 1]$ might be too strong, i.e. the influence is so large that no dictionary elements with center points on the flank of the weight function are chosen anymore. To remedy this, we concatenate the weight function with, e.g. a linear function to change the range of the weight functions. We may, for example, use

$$f(x) = 0.1x + 0.9, \ x \in \mathbb{R}.$$

Then, $g(d) = f(g(d)), \ d \in \mathcal{D}$, has the range $[0.925, 1]$ and is potentially better fitted for the application.

Bibliography

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, New York, 1972.

[2] A. Albertella, F. Sansò, and N. Sneeuw. Band-Limited Functions on a Bounded Spherical Domain: The Slepian Problem on the Sphere. *Journal of Geodesy*, 73:436–447, 1999.

[3] A. Albertella and N. Sneeuw. The Analysis of Radiometric Data with Slepian Functions. *Physics and Chemistry of the Earth*, 25:667–672, 2000.

[4] H.W. Alt. *Lineare Funktionalanalysis*. Springer, Berlin, Heidelberg, New York, Tokyo, 1985.

[5] A. Amirbekyan. *The Application of Reproducing Kernel Based Spline Approximation to Seismic Surface and Body Wave Tomography: Theoretical Aspects and Numerical Results*. PhD thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, kluedo.ub.uni-kl.de/volltexte/2007/2103/pdf/ThesisAbel.pdf, 2007.

[6] A. Amirbekyan and V. Michel. Splines on the Three-Dimensional Ball and their Application to Seismic Body Wave Tomography. *Inverse Problems*, 24:1–25, 2008.

[7] A. Amirbekyan, V. Michel, and F.J. Simons. Parameterizing Surface-Wave Tomographic Models with Harmonic Spherical Splines. *Geophysical Journal International*, 174:617–628, 2008.

[8] D.L. Anderson. *Theory of the Earth*. Blackwell Scientific Publications, Oxford, 1989.

[9] J.B. Anderson. *Digital Transmission Engineering*. Wiley Interscience, Lund, 2005.

[10] G. Arfken. *Mathematical Methods for Physicists*. Academic Press, New York, London, 2nd edition, 1970.

[11] L. Ballani, J. Engels, and E.W. Grafarend. Global Base Functions for the Mass Density in the Interior of a Massive Body (Earth). *Manuscripta Geodaetica*, 18:99–114, 1993.

[12] A.R. Barron, A. Cohen, W. Dahmen, and R.A. DeVore. Approximation and Learning by Greedy Algorithms. *The Annals of Statistics*, 36:64–94, 2008.

[13] F. Barthelmes. *Untersuchungen zur Approximation des äußeren Schwerefeldes der Erde durch Punktmassen mit optimierten Positionen.* Veröffentlichungen des Zentralinstituts Physik der Erde, Potsdam, 1986.

[14] F. Barthelmes and R. Dietrich. Use of Point Masses on Optimized Positions for the Approximation of the Gravity Field. In *Determination of the Geoid: Present and Future*, pages 484–493, Berlin, 1991. Springer.

[15] M. Bayer, S. Beth, and W. Freeden. Geophysical Field Modelling by Multiresolution Analysis. *Acta Geodaetica et Geophysica Hungarica*, 33:289–319, 1998.

[16] A.P. Berg and W.B. Mikhael. A Survey of Mixed Transform Techniques for Speech and Image Coding. In *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, volume 4, 1999.

[17] P. Berkel. *Multiscale Methods for the Combined Inversion of Normal Mode and Gravity Variations.* PhD thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, Shaker Verlag, Aachen, 2009.

[18] P. Berkel, D. Fischer, and V. Michel. Spline Multiresolution and Numerical Results for Joint Gravitation and Normal Mode Inversion with an Outlook on Sparse Regularisation. *International Journal on Geomathematics,* accepted for publication*, www.uni-siegen.de/fb6/geomathe/preprints/spg_1.pdf*, 2010.

[19] P. Berkel and V. Michel. On Mathematical Aspects of a Combined Inversion of Gravity and Normal Mode Variations by a Spline Method. *Mathematical Geosciences*, 42:795–816, 2010.

[20] E.J. Candès and D.L. Donoho. New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities. *Communications on Pure and Applied Mathematics*, 57:219–266, 2002.

[21] E.J. Candès, R. Mark, T. Tao, and R. Vershynin. Error Correction via Linear Programming. *IEEE Symposium on Foundations of Computer Science*, pages 295–308, 2005.

[22] E.J. Candès, J. Romberg, and T. Tao. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Fourier Information. *IEEE Transactions on Information Theory*, 52:489–509, 2006.

[23] E.J. Candès and T. Tao. Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51:4203–4215, 2005.

[24] Center for Space Research. University of Texas, Austin. www.csr.utexas.edu/ grace/overview.html.

[25] J.L. Chen, C.R. Wilson, and B.D. Tapley. Satellite Gravity Measurements Confirm Accelerated Melting of Greenland Ice Sheet. *Science*, 313:1958–1960, 2006.

[26] J.L. Chen, C.R. Wilson, B.D. Tapley, Z.L. Yang, and G.Y. Niu. 2005 Drought Event in the Amazon River Basin as Measured by GRACE and Estimated by Climate Models. *Journal of Geophysical Research*, 114:B05404, 2009.

[27] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Review*, 43:129–159, 2001.

[28] S.J. Claessens, W.E. Featherstone, and F. Barthelmes. Experiences with Point-Mass Gravity Field Modelling in the Perth Region, Western Australia. *Geomatics Research Australasia*, 75:53–86, 2001.

[29] C.W. Clenshaw. A Note on the Summation of Chebyshev Series. *Mathematical Tables and Other Aids to Computation*, 9:118–120, 1955.

[30] R. Coifman, Y. Meyer, and V. Wickerhauser. Adapted Wave Form Analysis, Wavelet-Packets and Applications. In *Proceedings of the Second International Conference on Industrial and Applied Mathematics*, 1992.

[31] S.F. Cotter and B.D. Rao. Sparse Channel Estimation via Matching Pursuit with Application to Equalization. *IEEE Transactions on Communications*, 50:374–377, 2002.

[32] F.A. Dahlen and J. Tromp. *Theoretical Global Seismology*. Princeton University Press, Princeton, New Jersey, 1998.

[33] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning and Adaptive Sparse Solution of Inverse Problems. *Math. Comp.*, 2011. S 0025-5718(2011)02507-X.

[34] W. Dai and O. Milenkovic. Subspace Pursuit for Compressive Sensing: Closing the Gap between Performance and Complexity. *IEEE Transactions on Information Theory*, 55:2230–2249, 2009.

[35] I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1457, 2004.

[36] I. Daubechies, M. Fornasier, and I. Loris. Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints. *Journal of Fourier Analysis and Applications*, 14:764–792, 2008.

[37] G. Davis, S.G. Mallat, and M. Avellaneda. Adaptive Greedy Approximations. *Constructive Approximation*, 13:57–98, 1997.

[38] G. Davis, S.G. Mallat, and Z. Zhang. Adaptive Time-Frequency Approximations with Matching Pursuits. *SPIE Journal of Optical Engineering*, 33:2183–2191, 1994.

[39] P.J. Davis. *Interpolation and Approximation.* Dover Publications, New York, 1975.

[40] M. De Stefano and D. Colombo. Geophysical Modeling through Simultaneous Joint Inversion of Seismic, Gravity and Magnetotelluric Data. *SEG 76th Annual Meeting*, 2006.

[41] A. Deuss, J.C.E. Irving, and J.H. Woodhouse. Regional Variation of Inner Core Anisotropy from Seismic Normal Mode Observation. *Science*, 328:1018–1020, 2010.

[42] R.A. DeVore. Nonlinear Approximation. *Acta Numerica*, 7:51–150, 1998.

[43] P. Ditmar. Finding the Shape of a Local Heterogeneity by Means of a Structural Inversion with Constraints. *Geophysical Prospecting*, 50:209–223, 2002.

[44] T.T. Do, L. Gan, N. Nguyen, and T.D. Tran. Sparsity Adaptive Matching Pursuit Algorithm for Practical Compressed Sensing. *Asilomar Conference in Signals, Systems, and Computers*, 2008.

[45] D.L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[46] D.L. Donoho and M. Elad. Optimally Sparse Representation in General (Non-orthogonal) Dictionaries via l^1 Minimization. *Proceedings of the National Academy of Sciences*, 100:2197–2202, 2002.

[47] D.L. Donoho and Y. Tsaig. Extensions of Compressed Sensing. *Signal Processing*, 86:549–571, 2006.

[48] D.L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit. Preprint, 2006.

[49] J.R. Driscoll and R.M. Healy. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.

[50] H.M. Dufour. Fonctions Orthogonales dans la Sphère. Résolution Théorique du Problème du Potentiel Terrestre. *Bulletin Géodésique*, 51:227–237, 1977.

[51] A. Dziewonski and D.L. Anderson. The Preliminary Reference Earth Model. *Physics of the Earth and Planetary Interiors*, 25:297–356, 1981.

[52] A. Dziewonski and J.H. Woodhouse. Global Images of the Earth's Interior. *Science*, 236:37–48, 1987.

[53] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[54] H.W. Engl, K. Kunisch, and A. Neubauer. Convergence Rates for Tikhonov Regularization of Nonlinear Ill-Posed Problems. *Inverse Problems*, 5:523–540, 1989.

[55] M.J. Fengler, W. Freeden, A. Kohlhaas, V. Michel, and T. Peters. Wavelet Modeling of Regional and Temporal Variations of the Earth's Gravitational Potential Observed by GRACE. *Journal of Geodesy*, 81:5–15, 2007.

[56] M.J. Fengler, D. Michel, and V. Michel. Harmonic Spline-Wavelets on the 3-Dimensional Ball and their Application to the Reconstruction of the Earth's Density Distribution from Gravitational Data at Arbitrarily Shaped Satellite Orbits. *Zeitschrift für Angewandte Mathematik und Mechanik*, 86:856–873, 2006.

[57] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE Journal of Selected Topics in Signal Processing*, 1:586–598, 2007.

[58] O. Forster. *Analysis 1*. Vieweg, Braunschweig, Wiesbaden, 2008.

[59] A.M. Forte, A.M. Dziewonski, and R.J. O'Connell. Continent-Ocean Chemical Heterogeneity in the Mantle Based on Seismic Tomography. *Science*, 268:386–388, 1995.

[60] A.M. Forte, R.L. Woodward, and A.M. Dziewonski. Joint Inversions of Seismic and Geodynamic Data for Models of Three-Dimensional Mantle Heterogeneity. *Journal of Geophysical Research*, 99:21857–21877, 1994.

[61] W. Freeden. Stokes' Boundary Value Problem of Physical Geodesy and its Numerical Computation. *Mitteilungen aus dem Institut für Theoretische Geodäsie in Bonn*, 56, 1978.

[62] W. Freeden. On Integral Formulas of the (Unit) Sphere and Their Application to Numerical Computation of Integrals. *Computing*, 25:131–146, 1980.

[63] W. Freeden. On Approximation by Harmonic Splines. *Manuscripta Geodaetica*, 6:193–244, 1981.

[64] W. Freeden. On Spherical Spline Interpolation and Approximation. *Mathematical Methods in the Applied Sciences*, 3:551–575, 1981.

[65] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere (With Applications to Geomathematics)*. Oxford University Press, Oxford, 1998.

[66] W. Freeden and V. Michel. *Multiscale Potential Theory (with Applications to Geoscience)*. Birkhäuser, Boston, 2004.

[67] W. Freeden and M. Schreiner. *Spherical Functions of Mathematical Geosciences. A Scalar, Vectorial, and Tensorial Setup*. Springer, Berlin, Heidelberg, 2009.

[68] W. Freeden, M. Schreiner, and R. Franke. A Survey on Spherical Spline Approximation. *Surveys on Mathematics for Industry*, 7:29–85, 1997.

[69] A.C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M.J. Strauss. Near-Optimal Sparse Fourier Representations via Sampling. In *Proceedings of the 2002 ACM Symposium on Theory of Computing*, 2002.

[70] A.C. Gilbert, S. Muthukrishnan, and M.J. Strauss. Improved Time Bounds for Near-Optimal Sparse Fourier Representation via Sampling. In *Proceedings of SPIE Wavelets*, volume 11, 2005.

[71] A.C. Gilbert, M.J. Strauss, J.A. Tropp, and R. Vershynin. Algorithmic Linear Dimension Reduction in the l^1-Norm for Sparse Vectors. In *Allerton 2006 (44th Annual Allerton Conference on Communication, Control, and Computing)*, 2006.

[72] A.C. Gilbert, M.J. Strauss, J.A. Tropp, and R. Vershynin. One Sketch for All: Fast Algorithms for Compressed Sensing. In *San Diego 2007 (39th ACM Symposium on Theory of Computing)*, 2007.

[73] J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, 1923.

[74] G. Hein, F. Sansò, G. Strykowsky, and C.C. Tscherning. On the Choice of Norm and Base Functions for the Solution of the Inverse Gravimetric Problem. *Ricerche di Geodesia Topografia Fotogrammetria CLUP*, 5:121–138, 1989.

[75] W.A. Heiskanen and H. Moritz. *Physical Geodesy*. W.H. Freeman and Company, San Francisco, 1967.

[76] H. Heuser. *Funktionalanalysis*. B.G. Teubner, Stuttgart, 1992.

[77] H. Heuser. *Funktionalanalysis: Theorie und Anwendung. Lehrbuch*. B.G. Teubner, Wiesbaden, 2006.

[78] E.W. Hobson. *The Theory of Spherical and Ellipsoidal Harmonics*. (Second Reprint), Chelsea Publishing Company, New York, 1965.

[79] Hydrological Sciences Branch. NASA/Goddard Space Flight Center, Greenbelt. mirador.gsfc.nasa.gov.

[80] J.C.E. Irving, A. Deuss, and J.H. Woodhouse. Normal Mode Coupling due to Hemispherical Anisotropic Structure in Earth's Inner Core. *Geophysical Journal International*, 178:962–975, 2009.

[81] M. Ishii and J. Tromp. Normal-Mode and Free-Air Gravity Constraints on Lateral Variations in Velocity and Density of Earth's Mantle. *Science*, 285:1231–1236, 1999.

[82] M. Ishii and J. Tromp. Even-Degree Lateral Variations in the Earth's Mantle Constrained by Free Oscillations and the Free-Air Gravity Anomaly. *Geophysical Journal International*, 145:77–96, 2001.

[83] M. Ishii and J. Tromp. Constraining Large-Scale Mantle Heterogeneity Using Mantle and Inner-Core Sensitive Normal Modes. *Physics of the Earth and Planetary Interiors*, 146:113–124, 2004.

[84] M. Ishii, J. Tromp, A.M. Dziewonski, and G. Ekström. Joint Inversion of Normal Mode and Body Wave Data for Inner Core Anisotropy 1. Laterally Homogeneous Anisotropy. *Journal of Geophysical Research*, 107:2379, 2002.

[85] Jet Propulsion Laboratory. California Institute of Technology, Pasadena. http://podaac.jpl.nasa.gov/grace/index.html.

[86] T.H. Jordan. A Procedure for Estimating Lateral Variations from Low-Frequency Eigenspectra Data. *Geophysical Journal of the Royal Astronomical Society*, 52:441–455, 1978.

[87] T.H. Jordan and D.L. Anderson. Earth Structure from Free Oscillations and Travel Times. *Geophysical Journal of the Royal Astronomical Society*, 36:541–576, 1974.

[88] P. Kammann and V. Michel. Time-Dependent Cauchy-Navier Splines and their Application to Seismic Wave Front Propagation. *Zeitschrift für Angewandte Mathematik und Mechanik (ZAMM)*, 88:155–178, 2008.

[89] S. Karato. Importance of Anelasticity in the Interpretation of Seismic Tomography. *Geophysical Research Letters*, 20:1623–1626, 1993.

[90] A. Khan, J.A.D. Connolly, J. Maclennan, and K. Mosegaard. Joint Inversion of Seismic and Gravity Data for Lunar Composition and Thermal State. *Geophysical Journal International*, 168:243–258, 2007.

[91] S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale l^1-Regularized Least Squares. *IEEE Journal on Special Topics in Signal Processing*, 1:606–617, 2007.

[92] C. Kuo and B. Romanowicz. On the Resolution of Density Anomalies in the Earth's Mantle Using Spectral Fitting of Normal Mode Data. *Geophysical Journal International*, 150:162–179, 2002.

[93] B.J. Last and K. Kubik. Compact Gravity Inversion. *Geophysics*, 48:713–721, 1983.

[94] G. Lauricella. Sulla Distribuzione della Massa nell'Interno dei Pianeti. *Rendiconti dell'Accademia Nazionale dei Lincei*, 21:18–26, 1912.

[95] J.M. Lees and J.C. VanDecar. Seismic Tomography Constrained by Bouguer Gravity Anomalies: Applications in Western Washington. *Pure and Applied Geophysics*, 135:31–52, 1991.

[96] E. Lewi. *Modelling and Inversion of High Precision Gravity Data*. PhD thesis, University of Darmstadt, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Munich, 1997.

[97] X.-D. Li, D. Giardini, and J.H. Woodhouse. Large-Scale Even-Degree Structure of the Earth from Splitting of Long-Period Normal Modes. *Journal of Geophysical Research*, 96:551–577, 1991.

[98] S.G. Mallat and Z. Zhang. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.

[99] G. Masters, G. Laske, H. Bolton, and A. Dziewonski. The Relative Behavior of Shear Velocity, Bulk Sound Speed, and Compressional Velocity in the Mantle: Implications for Chemical and Thermal Structure. In S. Karato,

A.M. Forte, R.C. Liebermann, G. Masters, and L. Stixrude, editors, *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, volume 117, pages 63–87. AGU, Washington D.C., 2000.

[100] G. Masters, G. Laske, and F. Gilbert. Matrix Autoregressive Analysis of Free-Oscillation Coupling and Splitting. *Geophysical Journal International*, 143:478–489, 2000.

[101] E.J. McShane. *Integration*. Princeton University Press, Princeton, New Jersey, 1944.

[102] V. Michel. *A Multiscale Method for the Gravimetry Problem: Theoretical and Numerical Aspects of Harmonic and Anharmonic Modelling*. PhD thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, Shaker Verlag, Aachen, 1999.

[103] V. Michel. *A Multiscale Approximation for Operator Equations in Separable Hilbert Spaces – Case Study: Reconstruction and Description of the Earth's Interior*. Shaker Verlag, Aachen, Habilitation thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, 2002.

[104] V. Michel. Scale Continuous, Scale Discretized and Scale Discrete Harmonic Wavelets for the Outer and the Inner Space of a Sphere and Their Application to an Inverse Problem in Geomathematics. *Applied and Computational Harmonic Analysis*, 12:77–99, 2002.

[105] V. Michel. Regularized Wavelet-Based Multiresolution Recovery of the Harmonic Mass Density Distribution from Data of the Earth's Gravitational Field at Satellite Height. *Inverse Problems*, 21:997–1025, 2005.

[106] V. Michel. Tomography: Problems and Multiscale Solutions. In *Handbook of Geomathematics*, pages 949–972, 2010.

[107] V. Michel and A.S. Fokas. A Unified Approach to Various Techniques for the Non-Uniqueness of the Inverse Gravimetric Problem and Wavelet-Based Methods. *Inverse Problems*, 24:045019, 2008.

[108] V. Michel and K. Wolf. Numerical Aspects of a Spline-Based Multiresolution Recovery of the Harmonic Mass Density out of Gravity Functionals. *Geophysical Journal International*, 173:1–16, 2008.

[109] L. Miranian. Slepian Functions on the Sphere, Generalized Gaussian Quadrature Rule. *Inverse Problems*, 20:877–892, 2004.

[110] C. Müller. *Spherical Harmonics*. Springer, Berlin, Heidelberg, New York, 1966.

[111] M.Z. Nashed. A new Approach to Classification and Regularization of Ill-Posed Operator Equations. In *Inverse and Ill-Posed Problems*, volume 4 of Notes and Reports in Mathematics in Science and Engineering, 1974.

[112] D. Needell and J.A. Tropp. CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples. *Applied and Computational Harmonic Analysis*, 26:301–321, 2009.

[113] D. Needell and R. Vershynin. Signal Recovery from Incomplete and Inaccurate Measurements via Regularized Orthogonal Matching Pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 4:310–316, 2010.

[114] H. Nutz and K. Wolf. Time-Space Multiscale Analysis by Use of Tensor Product Wavelets and its Application to Hydrology and GRACE Data. *Studia Geophysica et Geodaetica*, 52:321–339, 2008.

[115] B.A. Olshausen and D.J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? *Vision Research*, 37:311–325, 1997.

[116] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. In *Asilomar Conference on Signals, Systems and Computers*, 1993.

[117] N.K. Pavlis, S.A. Holmes, S.C. Kenyon, and J.K. Factor. An Earth Gravitational Model to Degree 2160: EGM2008. presented at the 2008 General Assembly of the European Geosciences Union, Vienna, Austria, April 13-18, 2008.

[118] P. Pizzetti. Corpi Equivalenti Rispetto alla Attrazione Newtoniana Esterna. *Rendiconti dell'Accademia Nazionale dei Lincei*, 18:211–215, 1909.

[119] P. Pizzetti. Intorno alle Possibili Distribuzioni della Massa nell'Interno della Terra. *Annali di Matematica Pura ed Applicata*, 17:225–258, 1910.

[120] S. Qian and D. Chen. Signal Representation Using Adaptive Normalized Gaussian Functions. *Signal Processing*, 36:1–11, 1994.

[121] J.S. Resovsky and M.H. Ritzwoller. Regularization Uncertainty in Density Models Estimated from Normal Mode Data. *Geophysical Research Letters*, 26:2319–2322, 1999.

[122] J.S. Resovsky and J. Trampert. Reliable Mantle Density Error Bars: an Application of the Neighbourhood Algorithm to Normal-Mode and Surface Wave Data. *Geophysical Journal International*, 150:665–672, 2002.

[123] R. Reuter. *Integralformeln der Einheitssphäre und Harmonische Splinefunktionen*. PhD thesis, RWTH Aachen, 1982.

[124] A. Rieder. *Keine Probleme mit Inversen Problemen*. Vieweg, Braunschweig, Wiesbaden, 2003.

[125] M. Rodell, P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll. The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85:381–394, 2004.

[126] B. Romanowicz. Can we Resolve 3D Density Heterogeneity in the Lower Mantle?. *Geophysical Research Letters*, 28:1107–1110, 2001.

[127] F. Sansò, R. Barzaghi, and C.C. Tscherning. Choice of Norm for the Density Distribution of the Earth. *Geophysical Journal of the Royal Astronomical Society*, 87:123–141, 1986.

[128] T.I. Seidmann and C.R. Vogel. Well-Posedness and Convergence of Least-Square Estimation to Ill-Posed Problems. *Inverse Problems*, 5:227–238, 1989.

[129] N.A. Simmons, A.M. Forte, and S.P. Grand. Constraining Mantle Flow with Seismic and Geodynamic Data: A Joint Approach. *Earth and Planetary Science Letters*, 246:109–124, 2006.

[130] N.A. Simmons, A.M. Forte, and S.P. Grand. Joint Seismic, Geodynamic and Mineral Physical Constraints on Three-Dimensional Mantle Heterogeneity: Implications for the Relative Importance of Thermal versus Compositional Heterogeneity. *Geophysical Journal International*, 177:1284–1304, 2009.

[131] F.J. Simons. Slepian Functions and their Use in Signal Estimation and Spectral Analysis. In *Handbook of Geomathematics*, pages 891–923, 2010.

[132] F.J. Simons and F.A. Dahlen. Spherical Slepian Functions and the Polar Gap in Geodesy. *Geophysical Journal International*, 166:1039–1061, 2006.

[133] F.J. Simons, F.A. Dahlen, and M.A. Wieczorek. Spatiospectral Concentration on a Sphere. *SIAM Review*, 48:504–536, 2006.

[134] D. Slepian. Some Comments on Fourier-Analysis, Uncertainty and Modeling. *SIAM Review*, 25:379–393, 1983.

[135] M.F. Smith and G. Masters. Aspherical Structure Constraints from Free Oscillation Frequency and Attenuation Measurements. *Journal of Geophysical Research*, 94:1953–1976, 1989.

[136] J.-L. Starck, E.J. Candès, and D.L. Donoho. The Curvelet Transform for Image Denoising. *IEEE Transactions on Image Processing*, 10:131–141, 2002.

[137] J.-L. Starck, M. Elad, and D.L. Donoho. Image Decomposition: Separation of Texture from Piecewise Smooth Content. In *SPIE Meeting*, 2003.

[138] G. Szegö. *Orthogonal Polynomials*. AMS Colloquium Publications, Volume 23, Providence, Rhode Island, 1939.

[139] B.D. Tapley, S. Bettadpur, J.C. Ries, P.F. Thompson, and M.M. Watkins. GRACE Measurements of Mass Variability in the Earth System. *Science*, 305:503–505, 2004.

[140] V.N. Temlyakov. Greedy Algorithms and m-Term Approximation. *Journal of Approximation Theory*, 98:117–145, 1999.

[141] V.N. Temlyakov. Greedy Algorithms with Regard to Multivariate Systems with Special Structure. *Constructive Approximation*, 16:399–425, 1999.

[142] V.N. Temlyakov. Nonlinear Methods of Approximation. *Foundations of Computational Mathematics*, 3:33–107, 2003.

[143] M. Thalhammer, Y. Ricard, R. Rummel, and K.H. Ilk. Application of Space-Borne Gravimetry to Research on the Interior of the Earth. Technical report, ESA study – CIGAR 4, 1996.

[144] J.A. Tropp and A.C. Gilbert. Signal Recovery from Random Measurements via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53:4655–4666, 2007.

[145] C.C. Tscherning. Density-Gravity Covariance Functions Produced by Overlapping Rectangular Blocks of Constant Density. *Geophysical Journal International*, 105:771–776, 1991.

[146] C.C. Tscherning. Isotropic Reproducing Kernels for the Inner of a Sphere or Spherical Shell and their Use as Density Covariance Functions. *Mathematical Geology*, 28:161–168, 1996.

[147] C.C. Tscherning and G. Strykowski. Quasi-Harmonic Inversion of Gravity Field Data. In *Model Optimization in Exploration Geophysics 2, Proceedings of the 5th International Mathematical Geophysics Seminar held at the Free University of Berlin*, 1987.

[148] M. Tücks. *Navier-Splines und ihre Anwendung in der Deformationsanalyse*. PhD thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, Shaker, Aachen, 1996.

[149] R. Venkataramani and Y. Bresler. Sub-Nyquist Sampling of Multiband Signals: Perfect Reconstruction and Bounds on Aliasing Error. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3:1633–1636, 1998.

[150] P. Vincent and Y. Bengio. Kernel Matching Pursuit. *Machine Learning*, 48:169–191, 2002.

[151] A. Voigt and J. Wloka. *Hilberträume und elliptische Differentialoperatoren.* Bibliographisches Institut, Mannheim, Wien, Zürich, 1975.

[152] N. Weck. Zwei Inverse Probleme in der Potentialtheorie. *Mitteilungen aus dem Institut für Theoretische Geodäsie in Bonn*, 4:27–36, 1972.

[153] J. Weidmann. *Lineare Operatoren in Hilberträumen.* Mathematische Leitfäden, B.G. Teubner, Stuttgart, 1976.

[154] R. Widmer-Schnidrig. Application of Regionalized Multiplet Stripping to Retrieval of Aspherical Structure Constraints. *Geophysical Journal International*, 148:201–213, 2002.

[155] B. Wohlberg. Noise Sensitivity of Sparse Signal Representations: Reconstruction Error Bounds for the Inverse Problem. *IEEE Transactions on Signal Processing*, 51:3053–3060, 2003.

[156] H. Xiao, V. Rokhlin, and N. Yarvin. Prolate Spheroidal Wave Functions, Quadrature, and Interpolation. *Inverse Problems*, 17:805–838, 2001.

[157] K. Yosida. *Functional Analysis.* Springer, Berlin, Heidelberg, New York, 1980.

[158] E. Zeidler. *Nonlinear Functional Analysis and its Applications.* Springer, Berlin, Heidelberg, New York, 1985.

Index