Novel algorithms of greedy-type for probability density estimation as well as linear and nonlinear inverse problems

DISSERTATION
zur Erlangung des Grades eines Doktors
der Naturwissenschaften

vorgelegt von
Max Kontak, M. Sc.

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen
Siegen 2018

gedruckt auf alterungsbeständigem holz- und säurefreiem Papier

Gutachter: Prof. Dr. Volker Michel
Prof. Dr. Bernd Hofmann
Prof. Dr. Thomas Schuster

Tag der mündlichen Prüfung: 20. April 2018

Danksagung

Ich möchte diese Gelegenheit nutzen, um allen zu danken, die mich auf meinem bisherigen Weg unterstützt haben, insbesondere während der Promotion.

Zuerst danke ich meinem Betreuer Prof. Dr. Volker Michel für die Möglichkeit, bei ihm zu promovieren und für die Unterstützung während der letzten vier Jahre. Insbesondere bedanke ich mich für die Möglichkeit, meine Arbeit auf verschiedenen Tagungen in aller Welt vorzustellen.

Prof. Dr. Bernd Hofmann und Prof. Dr. Thomas Schuster danke ich für das Übernehmen der Zweitgutachten. Des Weiteren danke ich Prof. Dr. Hans-Peter Scheffler für die Bereitschaft, einerseits Mitglied der Promotionskommission zu sein und andererseits in der frühen Phase meiner Promotion als Ansprechpartner bei Stochastik-Fragen zu fungieren.

Dr. Simone Gramsch danke ich für die gute Zusammenarbeit in unserem gemeinsamen Forschungsprojekt und das Bereitstellen der entsprechenden Daten.

Mein Dank gebührt außerdem allen weiteren aktuellen und ehemaligen Mitgliedern der AG Geomathematik, insbesondere Sarah Leweke, Dr. Roger Telschow und Dr. Martin Gutting. Danke für die vielen Diskussionen und Denkanstöße, ohne die diese Arbeit in dieser Form nicht zustande gekommen wäre.

Zu guter Letzt danke ich meiner Familie, insbesondere meinen Eltern, meiner Oma und meiner Freundin. Ihr habt alle Euren Teil zu dieser Arbeit beigetragen. Herzlichen Dank dafür!

Zusammenfassung

Greedy-Algorithmen sind oft genutzte Methoden zur Generierung von sogenannten *sparsen* Approximationen. Funktionen auf diese Art zu approximieren ist aus verschiedenen Gründen vorteilhaft. Deshalb entwickeln wir Greedy-Algorithmen für zwei verschiedene Problemklassen, die Schätzung von Wahrscheinlichkeitsdichten einerseits und inverse Probleme andererseits.

Die Entwicklung eines Greedy-Algorithmus für die Dichteschätzung ist motiviert durch die Notwendigkeit, einen Simulationsalgorithmus für sogenannte Vliesstoffe zu implementieren, einem speziellen Typ technischer Textilien, die oft in industriellen Anwendungen verwendet werden. Wir werden solch einen Simulationsalgorithmus vorstellen, der eine Schätzung der Richtungsverteilung in einem Vliesstoff benötigt. Die Richtungen der Fäden in einem echten Vliesstoff können mit Computertomographen analysiert werden. Dies liefert Millionen von Datenpunkten. Benutzen wir die Wahrscheinlichkeitsdichte, die durch den neu entwickelten Greedy-Algorithmus generiert wird, so reduziert sich die Rechenzeit des Simulationsalgorithmus von 80 Tagen auf 150 Minuten um einen Faktor von 750 im Vergleich zur Verwendung von Kerndichteschätzern, einer Standardmethode für die Dichteschätzung.

Für inverse Probleme entwickeln wir zwei Verallgemeinerungen des Regularized Functional Matching Pursuit (RFMP)-Algorithmus, welcher ein Greedy-Algorithmus für lineare inverse Probleme ist. Für die erste Verallgemeinerung, die wir RWFMP nennen, legen wir verbesserte theoretische Ergebnisse im Vergleich zum RFMP vor. Außerdem kann durch den RWFMP die Rechenzeit des RFMP auf ein Zehntel reduziert werden, ohne viel Genauigkeit zu verlieren. Die zweite Verallgemeinerung ist ein RFMP für nichtlineare inverse Probleme. Wir wenden diesen Algorithmus auf das nichtlineare inverse Gravimetrieproblem an, welches sich mit der Bestimmung von Strukturen im Innern eines Planeten aus Gravitationsdaten befasst. Wir erhalten sehr gute numerische Resultate, betreffend sowohl die Genauigkeit und die *sparsity*, als auch die Interpretierbarkeit des Ergebnisses.

Abstract

Algorithms of greedy-type are a popular tool for sparse approximation. Sparse approximations of functions are beneficial for several reasons. Therefore, we will develop greedy algorithms for two classes of problems, probability density estimation and inverse problems.

The development of a greedy algorithm for density estimation was motivated by the desire to implement a simulation algorithm for so-called nonwovens, a particular type of technical textiles, which are widely used in industrial applications. We will propose such a simulation algorithm, which needs an estimation of the probability density of the fiber directions inside a nonwoven. Fortunately, these directions can be obtained from real nonwovens by a CT scan, which yields millions of data points. The incorporation of a probability density that is generated by the newly developed greedy algorithm reduces the computation time of the simulation algorithm from 80 days to 150 minutes by a factor of 750 in comparison to the use of a standard method for density estimation, namely kernel density estimators.

For inverse problems, we introduce two generalizations of the Regularized Functional Matching Pursuit (RFMP) algorithm, which is a greedy algorithm for linear inverse problems. For the first generalization, called RWFMP, an improved theoretical analysis is possible. Furthermore, using the RWFMP, it is possible to reduce the computation time of the RFMP by a factor of 10 without losing much of the accuracy. The second generalization is an RFMP for nonlinear inverse problems. We apply the algorithm to the nonlinear inverse gravimetric problem, which is concerned with the reconstruction of information about the interior of a planetary body from gravitational data. We obtain very good numerical results concerning the accuracy, the sparsity, and the interpretability of the results.

Contents

Chapter 1.

Introduction

In this work, we will derive greedy algorithms for the problem of density estimation in statistics and for linear and nonlinear inverse problems. Among other methods, which we will briefly discuss below, greedy algorithms are one possibility to generate so-called *sparse approximations* of more complicated structures. The sparse approximation of functions, signals, vectors, etc. has proved to be a very useful tool in, for example, compressive sensing, image compression and reconstruction, machine learning, medical imaging, signal processing, and statistics, in the last decades.

It turns out that the concept of sparsity and, in particular, greedy algorithms will also be helpful in the applications that we consider in this work. We will first describe these applications and will afterwards introduce the concept of sparsity and greedy algorithms in more detail.

The first problem that we are concerned with is the estimation of a probability density function on the sphere. The development of a greedy algorithm for this problem was motivated by a joint research project with the department Transport Processes of the Fraunhofer ITWM in Kaiserslautern. In recent years, at the Fraunhofer institute several models for the simulation of so-called nonwovens were developed (for an overview see, e. g., Klar et al. [100]). We will discuss these models in more detail and give more references in Section 5.4.2. Nonwovens are a particular type of technical textile, which consist of a large number of fibers that are laid down on a moving belt in the production process. As the name suggests, the filaments in nonwovens are not weaved or knitted, they mainly stick together due to adhesive forces. Due to the computational costs, most of the established models only yield a two-dimensional fiber laydown. Although nonwovens are very flat structures, which may be represented very well by a two-dimensional approximation, the fact that the fibers are lying over and underneath each other plays a significant role for the quality of the textile. Thus, it seems natural that the distribution of the directions of the fibers in the three-dimensional space are an important factor for the stability of the nonwoven and its industrial usability. Fortunately, the fiber

directions in real nonwovens can be obtained from a CT scan, whose high resolution gives us millions of data points. Consequently, we will propose a three-dimensional simulation algorithm for nonwoven fabrics, which is based on an estimation of the distribution of fiber directions from these data. We propose a greedy algorithm for the estimation of probability densities, which reduces the computation time of the simulation algorithm by a factor of around 750 compared to the use of kernel density estimators, which are a standard tool for density estimation.

The second type of problem that we deal with are linear and nonlinear inverse problems. In contrast to direct problems, where one computes the effect of a given cause, inverse problems are concerned with the determination of causes for given effects. Most inverse problems are ill-posed, that is,

- a solution does not exist,
- the solution is not unique, or
- the solution does not depend continuously on the data,

such that there is a need for a regularization. Several particular inverse problems are inverse gravimetric problems. Inverse gravimetry deals with the determination of the shape and the mass density of a body of mass, for example, the Earth, the Moon, or any other planetary body, from gravitational data. Whereas the determination of the mass density for a given shape is a linear inverse problem, the determination of the shape for a given mass density model is a nonlinear inverse problem. It is a well-known fact (see, e. g., Lauricella [107], Pizzetti [145, 146], and Weck [183]) that the solution to the linear inverse gravimetric problem is not unique due to the infinite-dimensional null space of the corresponding operator. Furthermore, this operator is compact and, thus, does not possess a continuous inverse. For both reasons, the linear inverse gravimetric problem is ill-posed. For the nonlinear problem, Isakov [88] has proved both a uniqueness and a stability result, which we will state in Chapter 7. Therefore, the nonlinear inverse gravimetric problem is not ill-posed as long as the existence of a solution is assumed. This is the reason why it is also interesting to consider the more difficult nonlinear problem instead of the seemingly easier linear one. Unfortunately, it turns out that the stability of the inversion is, although proved theoretically, not easy to deal with numerically. Thus, there is still a need for a regularization.

In recent years, the Geomathematics Group at the University of Siegen has developed a greedy algorithm for linear inverse problems, such as the linear inverse gravimetric problem, called the *Regularized Functional Matching Pursuit (RFMP)* as well as a variant called the *Regularized Orthogonal Functional Matching Pursuit (ROFMP)* (see, e. g., Fischer and Michel [46], Michel [123], and Michel and Telschow [128]).

In several publications, it was shown that the algorithms provide very good approximations of the solutions to the considered inverse problems concerning the accuracy and sparsity of the result, but also the interpretability of the result from the perspective of the application. In this work, we introduce two generalizations of the RFMP algorithm.

The first one is called the *Regularized Weak Functional Matching Pursuit (RWFMP)*. There, we apply a strategy to improve the iterative approximation procedure of the algorithm. On the one hand, this strategy guarantees the existence of the approximation in the next iteration from the theoretical perspective. On the other hand, we will show in a numerical example that this strategy leads to a reduction of the computation time of up to 90%. Furthermore, we can prove the convergence of the RWFMP in arbitrary infinite-dimensional Hilbert spaces, where only finite-dimensional spaces had been considered for the RFMP and the ROFMP. We will also provide an a-priori parameter choice rule for the RWFMP, which yields a convergent regularization.

The second generalization of the RFMP is an analogous algorithm for nonlinear inverse problems, which we call the RFMP_NL. After the derivation of the algorithm, we will apply it to the nonlinear inverse gravimetric problem, where, as stated above, there is a need for a regularization. The algorithm is based on an iterative minimization of a linearized Tikhonov functional, which makes it similar to the Levenberg-Marquardt method (see Levenberg [109] and Marquardt [115]) and the iteratively regularized Gauß-Newton method (see Bakushinsky [10]), and thus incorporates a regularization. We provide numerical results for two synthetic examples, where we contrive a solution and compute the corresponding gravitational data. Then, we use the RFMP_NL to obtain an approximation of the solution. A comparison of the proposed solution and the approximation shows that the algorithm provides very good approximations. For a qualitative comparison of the novel algorithm to existing algorithms for nonlinear inverse problems, see the considerations in Chapter 10.

Above, we have already mentioned that greedy algorithms yield sparse approximations. In the following, we provide a motivation for the term *sparsity* in this context for a finite-dimensional example. Assume that the element f, for which an approximation is sought, comes from some n-dimensional real Banach space $\mathcal{X}$, $n \in \mathbb{N}$. If $\mathcal{D} \subseteq \mathcal{X}$ is spanning the space $\mathcal{X}$ in the sense that every element $g \in \mathcal{X}$ has a (not necessarily unique) representation $g = \sum_{k=1}^{\infty} \alpha_k d_k$ for coefficients $\alpha_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, the set $\mathcal{D}$ is commonly called a *dictionary*. It is clear that one can always find such an expansion, where only n of the coefficients α_k are nonzero, by choosing n

linearly independent elements from $\mathcal{D}$. An expansion

$$f = \sum_{k=1}^{\infty} \alpha_k \, d_k \tag{1.1}$$

of the element f, which one wants to approximate, is called s-sparse if the number of nonzero coefficients is $s \leq n$, $s \in \mathbb{N}$ (cf. Foucart and Rauhut [50, Definition 2.1]). If we speak of a *sparse approximation* of the element f, we usually mean that there exists an expansion like in Eq. (1.1) that is s-sparse and s is much smaller than the dimension n of the underlying vector space.

Sparse approximations are advantageous for several reasons. First, many signals of the real world are sparse in some dictionary (cf. Foucart and Rauhut [50, Preface]) and often, one has too few measurements at hand to expand the signal in a suitable basis, for example, in orthonormal polynomials up to an appropriate order, which yields non-quadratic underdetermined systems of linear equations. In general, there exists no unique solution to such underdetermined systems, but by imposing a sparsity constraint (i. e., demanding that the signal is s-sparse in some prescribed dictionary), one can achieve a unique solution. In the infinite-dimensional situation, for example in inverse problems, a sparsity ansatz consequently acts like a regularization. Secondly, sparse approximations of, for example, images, allow for efficient compression algorithms. Whereas a full image consisting of n pixels, in principle, needs the memory for n coefficients, an s-sparse approximation of the image only needs the memory for s coefficients, which may be much more memory-efficient. This is, of course, true not only for images, but also for other signals, which need to be stored on a computer. Thirdly, a sparse representation of some element f is beneficial if the result should be used in further computations. Since all necessary operations need to be performed with only s instead of n basis elements, a large reduction of computation time is possible, depending on the ratio of s and n and the computational expensiveness of the operations. This effect will play a major role in parts of this work.

In the following, we will introduce the idea behind greedy algorithms. Prescribing a possibly overcomplete dictionary $\mathcal{D} \subseteq \mathcal{X}$, from which an approximation of the element $f \in \mathcal{X}$ should be constructed, greedy algorithms rely on the concept of local optimality. That is, starting with an initial approximation $f_0 \in \mathcal{X}$, one iteratively chooses coefficients $\alpha_{k+1} \in \mathbb{R}$ and elements $d_{k+1} \in \mathcal{D}$ for $k = 0, 1, 2, \ldots$, such that for given f_k the new approximation

$$f_{k+1} \coloneqq f_k + \alpha_{k+1} d_{k+1}$$

fulfills the optimization problem

$$L(f_k + \alpha_{k+1} d_{k+1}) \to \min! \qquad \text{s. t.} \qquad \alpha_{k+1} \in \mathbb{R},\ d_{k+1} \in \mathcal{D},$$

where L is a so-called loss function that fits to the problem to be solved. In approximation theory, greedy algorithms have been pushed forward mainly by Temlyakov [167–169]. Note that greedy algorithms are also called *matching pursuits* (Mallat and Zhang [112]) in signal processing and *projection pursuit* (Friedman and Stuetzle [60]) in statistics. They have been applied to inverse problems as the *Regularized Functional Matching Pursuit* algorithm in Fischer and Michel [46] and Michel [123]. For a more detailed discussion of greedy algorithms and the existing literature about these methods, see Chapter 3 of this work.

As already said, there are also other methods that yield sparse approximations, namely optimization methods and thresholding-based methods (for an overview see, e. g., Foucart and Rauhut [50, Chapter 3]). As the name suggests, the approximation $f^\star = \sum_{k=1}^{K} \alpha_k d_k$ that is generated by optimization methods is the solution of an optimization problem, in more detail, the problem

$$\|(\alpha_1, \ldots, \alpha_K)\|_1 \to \text{min!} \qquad \text{s. t.} \qquad L(f^\star) \leq \eta \tag{1.2}$$

for a fixed dictionary $\mathcal{D} = \{ d_1, \ldots, d_K \} \subseteq \mathcal{X}$, which needs to form a basis of the underlying vector space, and a parameter $\eta > 0$. In contrast, an example for thresholding methods are methods that are based on the idea to rewrite the equation $L(f^\star) = 0$ for the approximation $f^\star = \sum_{k=1}^{K} \alpha_k d_k$ as a fixed point equation and perform the corresponding fixed point iteration, where one only keeps the s largest coefficients α_k in every iteration. Here, the dictionary $\mathcal{D} = \{ d_1, \ldots, d_K \} \subseteq \mathcal{X}$ also needs to form a basis.

The latter is the main reason why we stick to greedy algorithms to obtain sparse approximations in this work. The great advantage of greedy algorithms is the fact that the dictionary does not need to form a basis. It turns out that, in the considered applications, it is beneficial to combine, for example, both global and localized function, if the signal that is to approximate consists of both global and local features. Furthermore, when using optimization methods, one needs to apply sophisticated optimization methods for the solution of the non-differentiable optimization problem in Eq. (1.2).

Nevertheless, we want to give a brief overview of the literature, in which both optimization methods and thresholding methods have been developed, analyzed, and successfully applied in various fields of research. For optimization methods, this includes the works by Berg and Friedlander [17], Candès and Tao [25], Chen et al. [28], Donoho and Elad [37], and Tibshirani [170]. Note that in statistics, these optimization methods are known as the LASSO *(least absolute shrinkage and selection operator)* estimator. For thresholding-based methods, we mention the works by Beck and Teboulle [16], Blumensath and Davies [22], Bredies and Lorenz [23], Daubechies et al. [31], and Donoho [36], who have applied thresholding algorithms in various

fields. Of course, this is only an incomplete selection from the vast amount of literature about these two categories of algorithms.

This work consists of four parts. In Part I, we will first recall several basic results in Chapter 2, which we will need in the course of the thesis. In Chapter 3, we will describe greedy algorithms in more detail and we will give references for the existing algorithms and their applications in different fields of research.

Part II will be dedicated to the derivation of a greedy algorithm for density estimation. We will first state some fundamental results of probability theory and statistics in Chapter 4. Then, we will develop a new greedy algorithm for the estimation of a probability density function in Chapter 5. There, we will apply the algorithm to a set of CT data of a nonwoven fabric. We will show that the greedy strategy leads to a large improvement of the computation time of a simple simulation algorithm.

Part III will deal with greedy algorithms for linear and nonlinear inverse problems. We will summarize the theory of inverse problems in Chapter 6. Since the nonlinear problem that we will deal with is the inverse gravimetric problem, Chapter 7 will be dedicated to inverse gravimetry. We will give some already existing results about the ill-posedness of the problem and compare the linear and nonlinear problems of inverse gravimetry. The new algorithms that we will develop are based on the Regularized Functional Matching Pursuit (RFMP), which we will derive in Chapter 8. In this chapter, we will also present some of the already established convergence results, describe properties of the algorithm, and the problems to which the RFMP and its variants have already been applied before. Then, a new algorithm for linear inverse problems, the *Regularized Weak Functional Matching Pursuit (RWFMP)* will be developed in Chapter 9. The approach, which is pursued, will enable us to prove not only the convergence of the algorithm in arbitrary Hilbert spaces, but we will also provide results about the rate of convergence. We will also show that there exists an a-priori parameter choice rule for the RWFMP such that we obtain a convergent regularization. Numerical tests will make clear that the considered *weak* approach can accelerate the iteration of the RFMP to reduce the computation by more than 90%. Finally, in Chapter 10, we will apply a similar strategy as in the derivation of the RFMP to obtain a greedy algorithm for nonlinear inverse problems. We will apply the algorithm to a synthetic example of the nonlinear inverse gravimetric problem and analyze the numerical results.

The final part of this work consists of a conclusion and an outlook.

Part I.

Basics

Chapter 2.

Notation and fundamentals of functional analysis

In this thesis, we denote by $\mathbb{N}$ the set of positive integers and define $\mathbb{N}_0 := \mathbb{N} \cup \{ 0 \}$. The set of real numbers is $\mathbb{R}$, whereas the set of complex numbers is $\mathbb{C}$. As usual, we denote by $\mathbb{R}^d$ the Euclidean vector space

$$\mathbb{R}^d := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \;\middle|\; x_1, \dots, x_d \in \mathbb{R} \right\},$$

and, for the particular case $d = 3$, the canonical basis is given by

$$\varepsilon^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \varepsilon^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \varepsilon^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

For vectors $x, y \in \mathbb{R}^d$, we define the *Euclidean inner product* by

$$x \cdot y = \sum_{n=1}^{d} x_n y_n,$$

and the *vector product* or *cross product* by

$$x \wedge y := \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

The *power set* of a set U is given by

$$\mathfrak{P}(U) := \{ M \subseteq U \}$$

and the *characteristic function* of $M \subseteq U$ is

$$\chi_M \colon U \to \{ 0, 1 \}, \qquad \chi_M(x) := \begin{cases} 1, & x \in M, \\ 0, & x \notin M. \end{cases}$$

Furthermore, we denote by $\#M$ the *cardinality* of the set M.

For open subsets $U \subseteq \mathbb{R}^d$, we denote by $\mathrm{C}(U)$ and $\mathrm{C}^{(k)}(U)$ the spaces of continuous and k-times continuously differentiable functions defined on U. Analogously, these spaces can be defined for functions on the closure of U such that all functions from $\mathrm{C}(\overline{U})$ ($\mathrm{C}^{(k)}(\overline{U})$) are continuous ($k$-times continuously differentiable) in U and the function (and its derivatives up to order k) can be continuously extended to the boundary ∂U.

For $f \in \mathrm{C}^{(1)}(U)$,

$$\nabla_x f(x) \coloneqq \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{pmatrix}$$

is the *gradient* of f, where the subscript stands for the variable of differentiation. The *Laplacian* of a function $f \in \mathrm{C}^{(2)}(U)$ is given by

$$\Delta_x f(x) \coloneqq \sum_{n=1}^{d} \frac{\partial^2 f}{\partial x_n^2}(x).$$

The Lebesgue integral of a function $f\colon \mathbb{R}^d \supseteq U \to \mathbb{R}$ is

$$\int_U f(x)\,\mathrm{d}x$$

such that $\mathrm{d}x$ stands for integration with respect to the Lebesgue measure. The Lebesgue measure itself will be denoted by λ such that for a measurable set U, $\lambda(U)$ denotes the volume of the set.

We stick to Freeden and Michel [56, Section 3.1.1] for the definition of a regular surface in $\mathbb{R}^3$.

Definition 2.1. A surface $\Sigma \subseteq \mathbb{R}^3$ is called *regular* if it satisfies

(a) the space $\mathbb{R}^3$ is divided into a bounded region Σ^{int} and an unbounded region Σ^{ext} such that

$$\Sigma^{\mathrm{ext}} = \mathbb{R}^3 \setminus \overline{\Sigma^{\mathrm{int}}}, \qquad \Sigma = \overline{\Sigma^{\mathrm{int}}} \cap \overline{\Sigma^{\mathrm{ext}}}, \qquad \Sigma^{\mathrm{int}} \cap \Sigma^{\mathrm{ext}} = \emptyset,$$

(b) Σ^{int} contains the origin,

(c) Σ is a closed and compact surface, which is free of double points,

(d) Σ has a continuously differentiable outer unit normal field ν.

The integral of a function $f\colon \Sigma \to \mathbb{R}$ over a regular surface is

$$\int_\Sigma f(x)\,\mathrm{d}\omega(x),$$

where ω is the surface measure on Σ.

Among the variety of existing integral theorems, we will only use Green's first identity in this work.

Theorem 2.2. Let $\Sigma \subseteq \mathbb{R}^3$ be a regular surface and let $f, g \in \mathrm{C}^{(2)}(\Sigma^{\mathrm{int}}) \cap \mathrm{C}^{(1)}(\overline{\Sigma^{\mathrm{int}}})$. Then,

$$\int_\Sigma f(x)\,\frac{\partial g}{\partial \nu}(x)\,\mathrm{d}\omega(x) = \int_{\Sigma^{\mathrm{int}}} f(x)\,\Delta_x g(x) + \nabla_x f(x)\cdot\nabla_x g(x)\,\mathrm{d}x,$$

where $\frac{\partial g}{\partial \nu}(x) = \nu(x)\cdot\nabla_x g(x)$ is the so-called *normal derivative* of g on Σ.

Note that for Green's identity to hold it is sufficient to demand only piecewise regularity of the outer normal field instead of condition (d) in Definition 2.1.

Since most of the problems that we will consider in this work are given on spherical domains, we will briefly discuss the geometry of the sphere in the following section.

2.1. Spherical geometry

This section is mainly based on Michel [122, Section 4.1]. We first define the unit sphere for arbitrary dimensions.

Definition 2.3 (see, e. g., Freeden and Gutting [54, Definition 6.1.5])**.** The unit sphere in $\mathbb{R}^d$, $d \in \mathbb{N}$, is denoted by

$$\mathbb{S}^{d-1} \coloneqq \left\{ \xi \in \mathbb{R}^d \;\middle|\; |\xi| = 1 \right\},$$

where $|\cdot|$ denotes the usual Euclidean norm in $\mathbb{R}^d$. Unit vectors will always be denoted by lowercase Greek letters.

The open ball with radius $R > 0$ in $\mathbb{R}^3$ is denoted by

$$\mathbb{B}_R \coloneqq \left\{ x \in \mathbb{R}^3 \;\middle|\; |x| < R \right\}.$$

The space $\mathbb{R}^3$ can also be parametrized by a spherical and a radial part, which leads to polar coordinates.

Definition 2.4 (see, e. g., Michel [122, Definition 4.3])**.** The parametrization of a point $x \in \mathbb{R}^3$ in polar coordinates is given by

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

where $r \in [0, \infty)$, $\varphi \in [0, 2\pi)$, and $t \in [-1, 1]$.

Analogously, every point $\xi \in \mathbb{S}^2$ can be described by

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

where $\varphi \in [0, 2\pi)$ and $t \in [-1, 1]$.

We define a local tripod on the sphere (i. e., an orthonormal system at every point $\xi \in \mathbb{S}^2$) by

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

$$\varepsilon^\varphi(\varphi) \coloneqq \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix},$$

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

Continuity and differentiability can also be transferred to the case of functions on spherical domains.

Definition 2.5. The spaces of continuous and k-times continuously differentiable functions on $\mathbb{S}^2$ are denoted by $\mathrm{C}(\mathbb{S}^2)$ and $\mathrm{C}^{(k)}(\mathbb{S}^2)$, respectively.

Remark. In the previous definition, continuity and differentiability of functions on the sphere have to be considered in terms of continuity and differentiability of functions on manifolds, cf. Lee [108, Definition 1.52].

For differentiable functions, it is possible to define certain differential operators on the sphere.

Theorem 2.6 (see, e. g., Michel [122, Theorems 4.5, 4.7, 4.8])**.** Using the *surface gradient* ∇^*,

$$\nabla^* := \varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} \frac{\partial}{\partial \varphi} + \varepsilon^t \sqrt{1-t^2} \frac{\partial}{\partial t},$$

the gradient ∇ in $\mathbb{R}^3$ can be decomposed as

$$\nabla = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r} \nabla^*.$$

The *surface curl gradient*

$$L^*_\xi F(\xi) := \xi \wedge \nabla^*_\xi F(\xi),$$

where $\xi \in \mathbb{S}^2$ and $F \in \mathrm{C}^{(1)}(\mathbb{S}^2)$, has the local coordinate representation

$$L^* = -\varepsilon^\varphi \sqrt{1-t^2} \frac{\partial}{\partial t} + \varepsilon^t \frac{1}{\sqrt{1-t^2}} \frac{\partial}{\partial \varphi}.$$

Furthermore, with the help of the *Beltrami operator*

$$\Delta^* := \frac{\partial}{\partial t}(1-t^2)\frac{\partial}{\partial t} + \frac{1}{1-t^2}\frac{\partial^2}{\partial \varphi^2},$$

the *Laplace operator* Δ can be decomposed as

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta^*.$$

The differentiation operators satisfy

$$\nabla^* \cdot \nabla^* = L^* \cdot L^* = \Delta^*$$

and for a function $F \in \mathrm{C}^{(1)}(\mathbb{S}^2)$ we have

$$(\xi\, F(\xi)) \cdot \left(\nabla^*_\xi F(\xi)\right) = (\xi\, F(\xi)) \cdot \left(L^*_\xi F(\xi)\right) = \left(\nabla^*_\xi F(\xi)\right) \cdot \left(L^*_\xi F(\xi)\right) = 0.$$

Integration on the sphere is defined as for regular surfaces such that the integral of a function $f\colon \mathbb{S}^2 \to \mathbb{R}$ is written as

$$\int_{\mathbb{S}^2} f(\xi)\, \mathrm{d}\omega(\xi),$$

where ω now is the surface measure on $\mathbb{S}^2$. We can decompose any integral over the space $\mathbb{R}^3$ into a radial and a spherical part such that

$$\int_{\mathbb{R}^3} f(x)\, \mathrm{d}x = \int_{\mathbb{S}^2} \int_0^\infty f(r\,\xi) r^2\, \mathrm{d}r\, \mathrm{d}\omega(\xi)$$

for $f\colon \mathbb{R}^3 \to \mathbb{R}$.

2.2. Fundamentals of functional analysis

In particular in the study of inverse problems, we will need function spaces and operators between them such that there is a need for the application of functional analysis. We will summarize the most important results in this section.

Definition 2.7. A real vector space $\mathcal{X}$ is a *normed space* if there exists a mapping called *norm* $\|\cdot\|_{\mathcal{X}} : \mathcal{X} \to [0, \infty)$ such that for all $f, g \in \mathcal{X}$ and $\lambda \in \mathbb{R}$, we have

(N1) $\|f\|_{\mathcal{X}} = 0$ implies $f = 0$ *(definiteness)*,

(N2) $\|\lambda f\|_{\mathcal{X}} = |\lambda| \, \|f\|_{\mathcal{X}}$ *(absolute homogeneity)*,

(N3) $\|f + g\|_{\mathcal{X}} \leq \|f\|_{\mathcal{X}} + \|g\|_{\mathcal{X}}$ *(triangle inequality)*.

A normed space $\mathcal{X}$ is called a *Banach space* if it is complete, that is, every Cauchy sequence is convergent.

A normed space $\mathcal{X}$ is called a *Pre-Hilbert space* if there exists an *inner product* $\langle \cdot, \cdot \rangle_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, that is, for all $f, g, h \in \mathcal{X}$ and $\lambda \in \mathbb{R}$, we have

(I1) $\langle f, f \rangle_{\mathcal{X}} = 0$ implies $f = 0$ *(definiteness)*,

(I2) $\langle f + \lambda g, h \rangle_{\mathcal{X}} = \langle f, h \rangle_{\mathcal{X}} + \lambda \langle g, h \rangle_{\mathcal{X}}$ *(linearity)*,

(I3) $\langle f, g \rangle_{\mathcal{X}} = \langle g, f \rangle_{\mathcal{X}}$ *(symmetry)*.

It is called a *Hilbert space* if it is a Banach space with the norm

$$\|f\|_{\mathcal{X}} := \sqrt{\langle f, f \rangle_{\mathcal{X}}}$$

that is induced by the inner product.

The closure of a subset $\mathcal{U} \subseteq \mathcal{X}$ of a Banach space $\mathcal{X}$ is denoted by $\overline{\mathcal{U}}$. If there is a norm $\|\cdot\|_*$ different from $\|\cdot\|_{\mathcal{X}}$, with respect to which the closure is applied, we denote it by $\overline{\mathcal{U}}^{\|\cdot\|_*}$. The *linear span* of a subset $\mathcal{U} \subseteq \mathcal{X}$ is defined as

$$\operatorname{span} \mathcal{U} := \left\{ \sum_{k=1}^{K} \alpha_k u_k \;\middle|\; K \in \mathbb{N}_0, \alpha_k \in \mathbb{R}, u_k \in \mathcal{U} \right\}.$$

Given a subset $\mathcal{U} \subseteq \mathcal{X}$ of a Hilbert space $\mathcal{X}$, the *orthogonal complement* $\mathcal{U}^\perp$ is given by

$$\mathcal{U}^\perp := \{ f \in \mathcal{X} \mid \langle f, u \rangle_{\mathcal{X}} = 0 \text{ for all } u \in \mathcal{U} \}.$$

In Hilbert spaces, we distinguish two types of convergence of sequences.

Definition 2.8. Let $(f_k)_{k\in\mathbb{N}} \subseteq \mathcal{X}$ be a sequence in a Hilbert space $\mathcal{X}$.

(a) The sequence is called *(strongly) convergent* to an element $f \in \mathcal{X}$ if

$$\lim_{k\to\infty} \|f - f_k\|_{\mathcal{X}} = 0$$

and

(b) it is called *weakly convergent* to an element $f \in \mathcal{X}$ if

$$\lim_{k\to\infty} \langle f_k, g\rangle_{\mathcal{X}} = \langle f, g\rangle_{\mathcal{X}}$$

for all $g \in \mathcal{X}$.

We abbreviate strong convergence by $f_k \to f$ $(k \to \infty)$ and weak convergence by $f_k \rightharpoonup f$ $(k \to \infty)$.

It is well-known that strong convergence implies weak convergence but not vice versa (as long as $\mathcal{X}$ is infinite-dimensional).

Mappings between Banach and Hilbert spaces are called *operators*. We will use the following convention in this thesis for Banach spaces $\mathcal{X}, \mathcal{Y}$: linear operators will always be denoted by $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$, and the image of an element $f \in \mathcal{X}$ is denoted by $\mathcal{T}f$ or $\mathcal{T}(f)$. An operator, which does not need to be linear (and, in general, is nonlinear), will be denoted by $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$, and the image of an element $f \in \mathcal{X}$ is $\mathcal{S}[f]$ such that the different parentheses indicate the linear or nonlinear dependence on the respective argument.

A particular linear operator is the identity operator

$$\mathcal{I}\colon \mathcal{X} \to \mathcal{X}, \qquad f \mapsto f.$$

For an operator $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ between Banach spaces, we denote by

$$\operatorname{ran} \mathcal{S} := \{\, \mathcal{S}[f] \mid f \in \mathcal{X} \,\}$$

the *range* or *image* of the operator and by

$$\operatorname{null} \mathcal{S} := \{\, f \in \mathcal{X} \mid \mathcal{S}[f] = 0 \,\}$$

the *null space* or *kernel* of the operator. The restriction of the operator to a subset $\mathcal{U} \subseteq \mathcal{X}$ is $\mathcal{S}|_{\mathcal{U}}$.

Of interest are several specific types of operators, which we will define in the following.

Definition 2.9. Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and let $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be an operator. We call this operator *continuous* if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $f_1, f_2 \in \mathcal{X}$

$$\|f_1 - f_2\|_{\mathcal{X}} \leq \delta \qquad \text{implies} \qquad \|\mathcal{S}[f_1] - \mathcal{S}[f_2]\|_{\mathcal{Y}} \leq \varepsilon.$$

It is well-known that for linear operators $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$, this is equivalent to the existence of a constant $C > 0$ such that

$$\|\mathcal{T}f_1 - \mathcal{T}f_2\|_{\mathcal{Y}} \leq C\,\|f_1 - f_2\|_{\mathcal{X}}$$

for all $f_1, f_2 \in \mathcal{X}$, which is called *boundedness*.

The set of *linear and continuous operators* is denoted by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$. It is a Banach space with the norm

$$\|\mathcal{T}\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \coloneqq \sup_{f \in \mathcal{X}\setminus\{0\}} \frac{\|\mathcal{T}f\|_{\mathcal{Y}}}{\|f\|_{\mathcal{X}}}, \qquad \mathcal{T} \in \mathcal{L}(\mathcal{X}, \mathcal{Y}).$$

As for functions on $\mathbb{R}^d$, there are several concepts of differentiability of operators. The first concept is the Gâteaux derivative, which is an analogue to directional derivatives.

Definition 2.10 (cf. Cea [26, Definition 1.1])**.** An operator $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ between Banach spaces $\mathcal{X}, \mathcal{Y}$ is called *Gâteaux differentiable* at $f \in \mathcal{X}$ in the direction $g \in \mathcal{X}$, if the Gâteaux derivative

$$\mathcal{S}'[f](g) \coloneqq \lim_{\varepsilon \searrow 0} \frac{\mathcal{S}[f + \varepsilon g] - \mathcal{S}[f]}{\varepsilon}$$

exists and the limit is an element of $\mathcal{Y}$.

The second concept is an analogue to the total differential in the finite-dimensional setting.

Definition 2.11 (cf. Cea [26, Definition 6.1])**.** An operator $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ between Banach spaces $\mathcal{X}, \mathcal{Y}$ is called *Fréchet differentiable* at $f \in \mathcal{X}$ if there exists a linear operator $\mathcal{T}_f\colon \mathcal{X} \to \mathcal{Y}$ such that

$$\lim_{\|g\|_{\mathcal{X}} \to 0} \frac{\|\mathcal{S}[f + g] - \mathcal{S}[f] - \mathcal{T}_f g\|_{\mathcal{Y}}}{\|g\|_{\mathcal{X}}}.$$

The operator $\mathcal{S}'[f] \coloneqq \mathcal{T}_f\colon \mathcal{X} \to \mathcal{Y}$ is called *Fréchet derivative* of $\mathcal{S}$ in $f \in \mathcal{X}$.

It is well-known that every Fréchet differentiable operator is also Gâteaux differentiable and that the Fréchet derivative $\mathcal{S}'[f]$ applied to some element $g \in \mathcal{X}$ is equal to the Gateaux derivative $\mathcal{S}'[f](g)$ in that direction. This is the reason, why we use the same symbol for both concepts.

There exists an analogue to Taylor's formula for differentiable operators, which is the basis for the linearization of operators.

Theorem 2.12 (cf. Cea [26, Proposition 2.2])**.** Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be twice Gâteaux differentiable in $f \in \mathcal{X}$ in the direction of $g \in \mathcal{X}$, that is,

$$\mathcal{S}''[f](g,g) := \lim_{\varepsilon \searrow 0} \frac{\mathcal{S}'[f+\varepsilon g](g) - \mathcal{S}'[f](g)}{\varepsilon}$$

exists in $\mathcal{Y}$.

Then there exists $\varepsilon \in [0,1]$ such that

$$\mathcal{S}[f+g] = \mathcal{S}[f] + \mathcal{S}'[f](g) + \frac{1}{2}\mathcal{S}''[f+\varepsilon g](g,g)$$

holds.

In other words, a linearization of the operator is given by

$$\mathcal{S}[f+g] \approx \mathcal{S}[f] + \mathcal{S}'[f](g).$$

Finally, we introduce the concept of the adjoint operator, since it plays an important role in the analysis of inverse problems.

Definition 2.13. Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a linear and bounded operator. The uniquely determined operator $\mathcal{T}^*\colon \mathcal{Y} \to \mathcal{X}$, which fulfills

$$\langle \mathcal{T}f, g\rangle_{\mathcal{Y}} = \langle f, \mathcal{T}^* g\rangle_{\mathcal{X}}$$

for all $f \in \mathcal{X}$ and $g \in \mathcal{Y}$, is called the *adjoint operator*.

If $\mathcal{T}\colon \mathcal{X} \to \mathcal{X}$ fulfills $\mathcal{T}^* = \mathcal{T}$, then the operator is called *self-adjoint*.

The most important Hilbert spaces for this work are function spaces, which we will define in the following section.

2.3. Function spaces

Both on compact subsets $U \subseteq \mathbb{R}^d$ and on $\mathbb{S}^2$, we have already defined spaces of continuous functions. It is well-known that, together with the supremum norm, these spaces are Banach spaces.

Theorem 2.14. Let $M \in \{ U, \mathbb{S}^2 \}$, where $U \subseteq \mathbb{R}^d$ is compact. Then

$$\|f\|_{\mathrm{C}(M)} \coloneqq \sup_{x \in M} |f(x)|, \qquad f \in \mathrm{C}(M),$$

is a norm on $\mathrm{C}(M)$ and with this norm $\mathrm{C}(M)$ is a Banach space.

Spaces of integrable functions can also be defined both on subsets of $\mathbb{R}^d$ and on the sphere $\mathbb{S}^2$.

Definition 2.15. Let $U \subseteq \mathbb{R}^d$ be measurable. For $1 \leq p < \infty$, the $\mathrm{L}^p(U)$-norm and the $\mathrm{L}^p(\mathbb{S}^2)$-norm of a measurable function $f \colon \mathbb{R}^d \to \mathbb{R}$ and $g \colon \mathbb{S}^2 \to \mathbb{R}$ is given by

$$\|f\|_{\mathrm{L}^p(U)} \coloneqq \left(\int_U |f(x)|^p \, \mathrm{d}x \right)^{1/p},$$

and

$$\|g\|_{\mathrm{L}^p(\mathbb{S}^2)} \coloneqq \left(\int_{\mathbb{S}^2} |g(\xi)|^p \, \mathrm{d}\omega(\xi) \right)^{1/p}.$$

The $\mathrm{L}^p(U)$-spaces and $\mathrm{L}^p(\mathbb{S}^2)$-spaces are consequently given as

$$\mathrm{L}^p(U) \coloneqq \left\{ f \colon U \to \mathbb{R} \;\middle|\; f \text{ is measurable and } \|f\|_{\mathrm{L}^p(U)} < \infty \right\}.$$
$$\mathrm{L}^p(\mathbb{S}^2) \coloneqq \left\{ g \colon \mathbb{S}^2 \to \mathbb{R} \;\middle|\; g \text{ is measurable and } \|g\|_{\mathrm{L}^p(\mathbb{S}^2)} < \infty \right\},$$

where we, as usual, formally identify functions with each other, which are equal almost everywhere.

Theorem 2.16. It is well-known that $\mathrm{L}^p(U)$ and $\mathrm{L}^p(\mathbb{S}^2)$ are Banach spaces for $1 \leq p < \infty$ and $\mathrm{L}^2(U)$ and $\mathrm{L}^2(\mathbb{S}^2)$ are Hilbert spaces with respect to the inner products

$$\langle f_1, f_2 \rangle_{\mathrm{L}^2(U)} \coloneqq \int_U f_1(x) f_2(x) \, \mathrm{d}x,$$
$$\langle g_1, g_2 \rangle_{\mathrm{L}^2(\mathbb{S}^2)} \coloneqq \int_{\mathbb{S}^2} g_1(\xi) g_2(\xi) \, \mathrm{d}\omega(\xi),$$

where $f_1, f_2 \in \mathrm{L}^2(U)$ and $g_1, g_2 \in \mathrm{L}^2(\mathbb{S}^2)$.

We note the following relation between the L^2-norm and the C-norm on both domains.

Theorem 2.17 (see, e. g., Michel [122, Theorem 4.10])**.** Let $f \in C(M)$ for $M \in \{ U, \mathbb{S}^2 \}$, where $U \subseteq \mathbb{R}^d$ is compact. Then,

$$\|f\|_{L^2(M)} \leq C \, \|f\|_{C(M)},$$

where $C = \sqrt{\lambda(U)}$ if $M = U$ and $C = \sqrt{4\pi}$ if $M = \mathbb{S}^2$.

A well-known result is that $\overline{C(M)}^{\|\cdot\|_{L^2(M)}} = L^2(M)$. The following section will deal with a certain orthonormal system of functions on the sphere called spherical harmonics.

2.4. Spherical harmonics

A well-known orthonormal system of functions on the sphere are the so-called spherical harmonics. We will introduce them in an axiomatic way here, and give closed formulas for the so-called *fully normalized spherical harmonics* afterwards. We begin with the definition of homogeneous and harmonic polynomials.

Definition 2.18 (see, e. g., Michel [122, Definition 5.2])**.** By $\mathrm{Hom}_n\left(\mathbb{R}^d\right)$, $n \in \mathbb{N}_0$, $d \in \mathbb{N}$, we denote the space of homogeneous polynomials of degree n on $\mathbb{R}^d$, that is, $P \in \mathrm{Hom}_n\left(\mathbb{R}^d\right)$ if and only if P is a polynomial of degree n and $P(\lambda\, x) = \lambda^n \, P(x)$ for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^d$.

If $D \subseteq \mathbb{R}^d$, then $\mathrm{Hom}_n(D)$, $n \in \mathbb{N}_0$, denotes the set of restrictions of homogeneous polynomials of degree n on $\mathbb{R}^d$ to the set D, that is,

$$\mathrm{Hom}_n(D) := \left\{ P|_D \;\middle|\; P \in \mathrm{Hom}_n\left(\mathbb{R}^d\right) \right\}.$$

Furthermore, we define for $n \in \mathbb{N}_0$ the spaces

$$\begin{aligned}
\mathrm{Harm}_n\left(\mathbb{R}^d\right) &:= \left\{ P \in \mathrm{Hom}_n\left(\mathbb{R}^d\right) \;\middle|\; \Delta P = 0 \right\}, \\
\mathrm{Harm}_{0\ldots n}\left(\mathbb{R}^d\right) &:= \bigoplus_{i=0}^{n} \mathrm{Harm}_i\left(\mathbb{R}^d\right), \\
\mathrm{Harm}_{0\ldots\infty}\left(\mathbb{R}^d\right) &:= \bigcup_{n=0}^{\infty} \mathrm{Harm}_{0\ldots n}\left(\mathbb{R}^d\right),
\end{aligned}$$

and for $D \subseteq \mathbb{R}^d$ we define

$$\begin{aligned}
\operatorname{Harm}_n(D) &:= \left\{ P|_D \;\middle|\; P \in \operatorname{Harm}_n\left(\mathbb{R}^d\right) \right\}, \\
\operatorname{Harm}_{0\ldots n}(D) &:= \left\{ P|_D \;\middle|\; P \in \operatorname{Harm}_{0\ldots n}\left(\mathbb{R}^d\right) \right\}, \\
\operatorname{Harm}_{0\ldots\infty}(D) &:= \left\{ P|_D \;\middle|\; P \in \operatorname{Harm}_{0\ldots\infty}\left(\mathbb{R}^d\right) \right\}.
\end{aligned}$$

If $d = 3$ and $D = \mathbb{S}^2$, we call the elements of $\operatorname{Harm}_n\left(\mathbb{S}^2\right)$, $n \in \mathbb{N}_0$, *(scalar) spherical harmonics.*

Theorem 2.19 (see, e. g., Michel [122, Theorems 5.3, 5.6])**.** The dimensions of the spaces introduced in Definition 2.18 for $d = 3$ are

$$\begin{aligned}
\dim \operatorname{Hom}_n\left(\mathbb{R}^3\right) &= \frac{(n+1)(n+2)}{2}, \\
\dim \operatorname{Harm}_n\left(\mathbb{R}^3\right) &= \dim \operatorname{Harm}_n\left(\mathbb{S}^2\right) = 2n+1.
\end{aligned}$$

We will always stick to the following convention.

Definition 2.20 (see, e. g., Michel [122, Definition 5.10])**.** By Y_n, we always denote a (scalar) spherical harmonic of degree $n \in \mathbb{N}_0$, that is, $Y_n \in \operatorname{Harm}_n\left(\mathbb{S}^2\right)$.

For a fixed degree $n \in \mathbb{N}_0$, $\left\{ Y_{n,j} \;\middle|\; j = 1, \ldots, 2n+1 \right\}$ will always stand for an orthonormal basis of $\operatorname{Harm}_n\left(\mathbb{S}^2\right)$.

The following result guarantees the orthogonality of spherical harmonics of different degrees.

Theorem 2.21 (see, e. g., Müller [134, Lemma 2])**.** For spherical harmonics of different degrees, that is, $Y_n \in \operatorname{Harm}_n\left(\mathbb{S}^2\right)$ and $Y_m \in \operatorname{Harm}_m\left(\mathbb{S}^2\right)$, where $n, m \in \mathbb{N}_0$, but $n \neq m$, we have

$$\langle Y_n, Y_m \rangle_{\mathrm{L}^2(\mathbb{S}^2)} = 0.$$

Corollary 2.22. From Definition 2.20 and Theorem 2.21, we obtain that

$$\left\{ Y_{n,j} \;\middle|\; n \in \mathbb{N}_0, j = 1, \ldots, 2n+1 \right\}$$

forms an orthonormal system in $\mathrm{L}^2(\mathbb{S}^2)$.

The following addition theorem is one of the most important and most useful result about spherical harmonics.

Theorem 2.23 (Addition theorem, Müller [134, Theorem 2])**.** For fixed degree $n \in \mathbb{N}_0$, let $\{ Y_{n,j} \mid j = 1, \dots, 2n+1 \}$ be an $L^2(\mathbb{S}^2)$-orthonormal system in the space $\mathrm{Harm}_n(\mathbb{S}^2)$. Then,

$$\sum_{j=1}^{2n+1} Y_{n,j}(\xi) Y_{n,j}(\eta) = \frac{2n+1}{4\pi} P_n(\xi \cdot \eta) \tag{2.1}$$

for all $\xi, \eta \in \mathbb{S}^2$, where P_n denotes the Legendre polynomial of degree n, that is, $\{ P_n \mid n \in \mathbb{N}_0 \}$ is a set of orthogonal polynomials in $L^2([-1,1])$ such that $P_n(1) = 1$ for all $n \in \mathbb{N}_0$.

Using the addition theorem, one can show that the right-hand side of Eq. (2.1) is a so-called *reproducing kernel*.

Definition 2.24 (Reproducing kernel, see, e. g., Michel [122, Definition 5.13])**.** Let $\mathcal{X}$ be a Hilbert space of functions on a domain $U \subseteq \mathbb{R}^d$. A function $k \colon U \times U \to \mathbb{R}$ is called a *reproducing kernel* if it satisfies

1. $k(x, \cdot) \in \mathcal{X}$ for all $x \in U$,
2. $\langle k(x, \cdot), f \rangle_{\mathcal{X}} = f(x)$ for all $f \in \mathcal{X}$ and $x \in U$.

If such a kernel exists, $\mathcal{X}$ is called a *reproducing kernel Hilbert space*.

It is well-known that the reproducing kernel is uniquely determined if it exists.

Theorem 2.25 (see, e. g., Freeden and Schreiner [58, Lemma 3.29])**.** The function

$$\mathbb{S}^2 \times \mathbb{S}^2 \ni (\xi, \eta) \mapsto \frac{2n+1}{4\pi} P_n(\xi \cdot \eta)$$

is the reproducing kernel of $(\mathrm{Harm}_n(\mathbb{S}^2), \langle \cdot, \cdot \rangle_{L^2(\mathbb{S}^2)})$.

Furthermore, the addition theorem can be used to prove the following inequalities.

Theorem 2.26 (e. g., Michel [122, Theorem 5.17])**.** If $Y_n \in \mathrm{Harm}_n(\mathbb{S}^2)$, $n \in \mathbb{N}_0$, then

$$\|Y_n\|_{C(\mathbb{S}^2)} \leq \sqrt{\frac{2n+1}{4\pi}} \|Y_n\|_{L^2(\mathbb{S}^2)}$$

and, in particular, we have

$$\|Y_{n,j}\|_{C(\mathbb{S}^2)} \leq \sqrt{\frac{2n+1}{4\pi}}.$$

We introduce the following kernel, which turns out to be an approximate identity for $\mathrm{C}(\mathbb{S}^2)$.

Definition 2.27 (Abel-Poisson kernel)**.** Let $Q_h \colon [-1,1] \to \mathbb{R}$, $h \in (0,1)$,

$$Q_h(t) \coloneqq \frac{1}{4\pi} \frac{1-h^2}{(1+h^2-2ht)^{3/2}}.$$

The kernel $\mathbb{S}^2 \times \mathbb{S}^2 \ni (\xi, \eta) \mapsto Q_h(\xi \cdot \eta)$ is called *Abel-Poisson kernel*.

For fixed $\xi \in \mathbb{S}^2$, we also define the function $Q_{h,\xi} \colon \mathbb{S}^2 \to \mathbb{R}$, $\eta \to Q_h(\xi \cdot \eta)$.

Note that one can show

$$Q_h(t) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} h^n P_n(t). \tag{2.2}$$

The Abel-Poisson kernel fulfills the following Poisson integral formula.

Theorem 2.28 (cf. Michel [122, Theorem 5.19])**.** Let $f \in \mathrm{C}(\mathbb{S}^2)$. Then,

$$\lim_{h \nearrow 1} \left\| \int_{\mathbb{S}^2} Q_h(\eta \cdot) f(\eta) \, \mathrm{d}\omega(\eta) - f \right\|_{\mathrm{C}(\mathbb{S}^2)} = 0,$$

which characterizes the kernel as an approximate identity since the arising integral is the spherical convolution of the kernel Q_h and the function f.

The Poisson integral formula can be used to show in several steps that the spherical harmonics form a complete orthonormal system of $\mathrm{L}^2(\mathbb{S}^2)$ such that every $f \in \mathrm{L}^2(\mathbb{S}^2)$ can be expanded in a Fourier series

$$f = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \left\langle f, Y_{n,j} \right\rangle_{\mathrm{L}^2(\mathbb{S}^2)} Y_{n,j},$$

where the convergence is in the sense of $\mathrm{L}^2(\mathbb{S}^2)$.

The arising coefficients are called *Fourier coefficients*. For these and for an analogue on the interval $[-1,1]$, we introduce the following abbreviations.

Definition 2.29. The *Fourier coefficient* of degree $n \in \mathbb{N}_0$ and order $j \in \{1, \dots, 2n+1\}$ of a function $f \in \mathrm{L}^2(\mathbb{S}^2)$ is defined as

$$f^\wedge(n,j) \coloneqq \left\langle f, Y_{n,j} \right\rangle_{\mathrm{L}^2(\mathbb{S}^2)},$$

and the *Legendre coefficient* of degree n of a function $g \in \mathrm{L}^2([-1,1])$ is given by

$$g^\wedge(n) \coloneqq 2\pi \langle g, P_n \rangle_{\mathrm{L}^2([-1,1])}.$$

Corollary 2.30 (for the second result, see Michel [122, Definition 3.15])**.** For $f \in \mathrm{L}^2(\mathbb{S}^2)$, we have

$$f = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} f^\wedge(n,j) Y_{n,j}$$

in the sense of $\mathrm{L}^2(\mathbb{S}^2)$ and for $g \in \mathrm{L}^2([-1,1])$, we have

$$g = \sum_{n=0}^{\infty} g^\wedge(n) \frac{2n+1}{4\pi} P_n$$

in the sense of $\mathrm{L}^2([-1,1])$.

Finally, we will give an example for a closed representation of spherical harmonics, the so-called *fully normalized spherical harmonics*.

Example 2.31 (cf. Michel [122, Section 5.2])**.** For $n \in \mathbb{N}_0$, $j = -n \ldots, n$, the fully normalized spherical harmonics are defined as

$$\begin{aligned} Y_{n,j}(\xi(\varphi,t)) \coloneqq \sqrt{\frac{(2n+1)(n-|j|)!(2-\delta_{j0})}{4\pi(n+|j|)!}}\, P_{n,|j|}(t) \\ \times \begin{cases} \sin(j\varphi), & j = 1,\ldots,n, \\ \cos(j\varphi), & j = -n,\ldots,0, \end{cases} \end{aligned} \tag{2.3}$$

where for $n \in \mathbb{N}_0$, $m = 0,\ldots,n$,

$$P_{n,m}(t) \coloneqq (1-t^2)^{m/2} \frac{\mathrm{d}^m}{\mathrm{d}t^m} P_n(t), \qquad t \in [-1,1],$$

are the associated Legendre functions. Note that we shifted the enumeration of the orders from $1,\ldots,2n+1$ to $-n,\ldots,n$ for the sake of easier readability of Eq. (2.3).

If we use spherical harmonics in numerical experiments in this thesis, these will always be fully normalized spherical harmonics.

The Abel-Poisson kernel belongs to a type of spherical functions called zonal functions, which we will discuss in the following section.

2.5. Zonal functions

The idea of zonal functions is that these functions only depend on the distance of the variable $\eta \in \mathbb{S}^2$ from some fixed point $\xi \in \mathbb{S}^2$. Since

$$|\xi - \eta|^2 = |\xi|^2 - 2\,\xi \cdot \eta + |\eta|^2 = 2 - 2\,\xi \cdot \eta,$$

this can be realized by a function, which only depends on the inner product $\xi \cdot \eta$. This results in the following definition.

Definition 2.32 (Michel [122, Definition 4.14])**.** Let $g\colon [-1,1] \to \mathbb{R}$ and $\xi \in \mathbb{S}^2$. A function $g_\xi\colon \mathbb{S}^2 \to \mathbb{R}$ of the form

$$\eta \mapsto g_\xi(\eta) \coloneqq g(\xi \cdot \eta)$$

is called ξ-*zonal function*. We also write $g(\xi\cdot) \coloneqq g_\xi$.

It turns out that both differentiation and integration of zonal functions can be done in terms of the function g on $[-1,1]$.

Theorem 2.33 (Michel [122, Theorem 4.15])**.** For $g \in \mathrm{C}^{(1)}([-1,1])$ and $\xi, \eta \in \mathbb{S}^2$, we have

$$\begin{aligned}\nabla^*_\xi g(\xi \cdot \eta) &= g'(\xi \cdot \eta)(\eta - (\xi \cdot \eta)\xi),\\ L^*_\xi g(\xi \cdot \eta) &= g'(\xi \cdot \eta)\xi \wedge \eta.\end{aligned}$$

Theorem 2.34 (Michel [122, Theorem 4.16])**.** Let $g\colon [-1,1] \to \mathbb{R}$ be integrable. Then

$$\int_{\mathbb{S}^2} g(\xi \cdot \eta)\,\mathrm{d}\omega(\eta) = 2\pi \int_{-1}^{1} g(t)\,\mathrm{d}t$$

for all $\xi \in \mathbb{S}^2$.

Corollary 2.35. A direct implication of Theorem 2.34 is that $g \in \mathrm{L}^p([-1,1])$, $1 \le p < \infty$, if and only if $g_\xi \in \mathrm{L}^p(\mathbb{S}^2)$.

Furthermore, the same is also true for $p = \infty$.

Another consequence of the theorem is the fact that the Fourier coefficients of a zonal function g_ξ relate to the Legendre coefficients of g as follows.

Theorem 2.36. Let $g_\xi = g(\xi\cdot)$ be a zonal function with $g_\xi \in \mathrm{L}^2(\mathbb{S}^2)$. Then

$$g_\xi{}^\wedge(n,j) = g^\wedge(n) Y_{n,j}(\xi).$$

Proof. On the one hand, we have

$$g(\xi\cdot) = \sum_{n=0}^{\infty} g^{\wedge}(n)\frac{2n+1}{4\pi}P_n(\xi\cdot) = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} g^{\wedge}(n)Y_{n,j}(\xi)Y_{n,j}$$

by virtue of the addition theorem. On the other hand,

$$g(\xi\cdot) = g_{\xi} = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} g_{\xi}{}^{\wedge}(n,j)Y_{n,j}.$$

Therefore, by comparison of coefficients,

$$g_{\xi}{}^{\wedge}(n,j) = g^{\wedge}(n)Y_{n,j}(\xi)$$

holds true. ■

Example 2.37 (Abel-Poisson kernel, Freeden et al. [53, Section 5.6.3])**.** We can employ the previous theorem to analyze the Abel-Poisson kernel. Due to the representation in Eq. (2.2), we have

$$Q_h{}^{\wedge}(n) = h^n$$

and Theorem 2.36 yields

$$Q_{h,\xi}{}^{\wedge}(n,j) = h^n\, Y_{n,j}(\xi).$$

An application of the Parseval identity and the addition theorem shows that the inner product of two Abel-Poisson kernels is again an Abel-Poisson kernel: Let $\xi,\eta\in\mathbb{S}^2$ and $h,s\in(0,1)$, then,

$$\begin{aligned}\left\langle Q_{h,\xi}, Q_{s,\eta}\right\rangle_{\mathrm{L}^2(\mathbb{S}^2)} &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} Q_{h,\xi}{}^{\wedge}(n,j)\, Q_{s,\eta}{}^{\wedge}(n,j)\\ &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} h^n\, Y_{n,j}(\xi)\, s^n\, Y_{n,j}(\eta)\\ &= \sum_{n=0}^{\infty}(h\,s)^n\, P_n(\xi\cdot\eta) = Q_{h\,s}(\xi\cdot\eta).\end{aligned}$$

Consequently, we have for the $\mathrm{L}^2(\mathbb{S}^2)$-norm of the Abel-Poisson kernel that

$$\begin{aligned}\left\|Q_{h,\xi}\right\|^2_{\mathrm{L}^2(\mathbb{S}^2)} &= \left\langle Q_{h,\xi}, Q_{h,\xi}\right\rangle_{\mathrm{L}^2(\mathbb{S}^2)} = Q_{h^2}(\xi\cdot\xi)\\ &= \frac{1}{4\pi}\frac{1-h^4}{(1+h^4-2\,h^2)^{3/2}} = \frac{1}{4\pi}\frac{(1+h^2)(1-h^2)}{(1-h^2)^3} = \frac{1}{4\pi}\frac{1+h^2}{(1-h^2)^2}\end{aligned}$$

holds.

Chapter 3.

Greedy algorithms and matching pursuits

Algorithms of greedy type have been developed in several subjects of research more or less independently in the last decades, namely in approximation theory (called *greedy algorithm*), signal processing (called *matching pursuit*), and statistics (called *projection pursuit regression*). We will present the application in statistics in more detail in Section 5.1. The signal processing approach will be shortly mentioned in this chapter. The main part of this chapter will be based on the results that were achieved in approximation theory, namely by V. N. Temlyakov (see, e. g., the book Temlyakov [169]).

In all of these areas, the term *greedy* is always used for iterative algorithms, which produce iterates that are locally optimal approximations to some function in a certain sense. In computer science and discrete mathematics, this term is also used for algorithms that are used to solve combinatorial problems like finding the shortest path problem (e. g., the Dijkstra algorithm) and the traveling salesman problem (e. g., the nearest neighbor algorithm). These algorithms also provide locally optimal approximations of the solution of the combinatorial problem. For an overview, see for example the book by Korte and Vygen [105].

The basic idea of greedy algorithms and matching pursuits is to generate a sequence $(f_k)_{k\in\mathbb{N}_0}$ of approximations of a solution f to some problem such that

$$f_{k+1} = f_k + \alpha_{k+1} d_{k+1},$$

and $(\alpha_{k+1}, d_{k+1}) \in \mathbb{R} \times \mathcal{D}$ are optimal in some sense. Here, $\mathcal{D}$ denotes a so-called *dictionary*, an arbitrary set from which the approximation can be constructed. The dictionary does not need to form a basis of the underlying space, instead it can be overcomplete to account for different structures in the solution.

In signal processing, Mallat and Zhang [112] derived the so-called *Matching Pursuit (MP)* algorithm for the approximation of signals in general Hilbert spaces. They proved convergence of the algorithm and presented several results, which yield a time-frequency decomposition of audio signals. It turns out that the convergence of the algorithm is relatively slow, since, in general, several dictionary elements

will be chosen more than once, which is of course not optimal. To account for this problem, the *Orthogonal Matching Pursuit (OMP)* was introduced by Pati et al. [141], which removes this disadvantage by an orthogonalization of the chosen dictionary elements. Another optimization of the OMP was given by Vincent and Bengio [179]. They included a so-called *pre-fitting* into the algorithm, which accounts for the fact that the chosen dictionary elements of the original OMP are no longer optimal after the orthogonalization such that this orthogonalization should have already been incorporated in the computation of the optimum. Note that both variants of the OMP are more expensive than the MP due to the orthogonalization steps, but the faster convergence of the methods justifies the higher effort.

Greedy algorithms in approximation theory possess a very similar history. The problem to be solved in approximation theory is always the following: we assume that a function $f \in \mathcal{X}$, where $\mathcal{X}$ is some separable Hilbert space of functions, should be approximated by elements from a dictionary $\mathcal{D} \subseteq \mathcal{X}$. That is, as described above, one wants to find sequences $(\alpha_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$, $(d_k)_{k\in\mathbb{N}} \subseteq \mathcal{D}$ such that

$$f \approx f_K := \sum_{k=1}^{K} \alpha_k d_k.$$

In DeVore and Temlyakov [32], the so-called *Pure Greedy Algorithm (PGA)* was introduced.

Algorithm 3.1 (Pure Greedy Algorithm, PGA)**.** Let $\mathcal{X}$ be a separable Hilbert space and $f \in \mathcal{X}$. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$ and assume that $\|d\|_{\mathcal{X}} = 1$ for all $d \in \mathcal{D}$. Generate a sequence $(f_k)_{k\in\mathbb{N}}$ by the following iteration:

1. Set $k := 0$ and set $f_0 := 0$.
2. Find a dictionary element d_{k+1} fulfilling
$$d_{k+1} = \operatorname*{argmax}_{d\in\mathcal{D}} |\langle f - f_k, d\rangle_{\mathcal{X}}|. \tag{3.1}$$
3. Set
$$\alpha_{k+1} := \langle f - f_k, d_{k+1}\rangle_{\mathcal{X}}.$$
4. Set $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$.
5. If a suitable stopping criterion is fulfilled: stop.
Otherwise: increase k by 1 and return to step 2 of the algorithm.

There is a geometrical interpretation of the maximization property in Eq. (3.1). On the one hand, the modulus of the inner product $|\langle f - f_k, d\rangle_{\mathcal{X}}|$ for $f, f_k \in \mathcal{X}$ and $d \in \mathcal{D}$ is a measure of how co-linear the dictionary element is to the approximation error $f - f_k$. It is maximal if the error is a multiple of the dictionary element and minimal if $f - f_k$ and d are perpendicular.

On the other hand, to obtain $\|f - f_{k+1}\|_{\mathcal{X}} = 0$, which would require the optimal choice for the dictionary element, one had to fulfill

$$\alpha_{k+1} d_{k+1} = f - f_k,$$

corresponding to the maximal value of the inner product above. Furthermore, it can be proved that $\|f - f_{k+1}\|_{\mathcal{X}} = \|f - f_k\|_{\mathcal{X}}$ (i. e., the approximation does not get better), if and only if the inner product above is 0, that is, minimal.

Consequently, the norm of the error improves most, if the inner product is maximal, which is the justification for Eq. (3.1). DeVore and Temlyakov [32] also proved the convergence of the PGA and a specific convergence rate.

Theorem 3.2 (Convergence of the PGA)**.** Let $(f_k)_{k\in\mathbb{N}_0}$ be the sequence of approximations generated by the PGA to approximate $f \in \mathcal{X}$ and let $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X}$.

Then $f_k \to f$ $(k \to \infty)$ and

$$\|f - f_k\|_{\mathcal{X}} \leq |f|_{\mathcal{A}_1(\mathcal{D})} k^{-1/6},$$

where $|f|_{\mathcal{A}_1(\mathcal{D})}$ is a certain norm of f (for more details, see DeVore and Temlyakov [32]).

In the same paper by DeVore and Temlyakov [32], an orthogonal algorithm called *Orthogonal Greedy Algorithm (OGA)*, where a similar strategy as in the OMP is applied, is introduced. The convergence and convergence rates are also proved for this algorithm. We do not go into more detail on the OGA here, since we will not use it in our work.

Later, Temlyakov [167] introduced so-called *weak* variants of the PGA and the OGA, called *Weak Greedy Algorithm (WGA)* and *Weak Orthogonal Greedy Algorithm (WOGA)*.

The motivation for these algorithm are difficulties with the maximization in Eq. (3.1). In fact, it is not clear if a maximizer exists if $\mathcal{X}$ is an infinite-dimensional Hilbert space such that also $\mathcal{D}$ has to consist of infinitely many elements to span $\mathcal{X}$. However, the supremum of the inner products in Eq. (3.1) always exists. The idea of the weak greedy algorithm consequently is to no longer search for the optimal dictionary element, but a dictionary element, which is near to the optimum in the specific way stated in the following algorithm.

Algorithm 3.3 (Weak Greedy Algorithm, WGA)**.** Let $\mathcal{X}$ be a separable Hilbert space and $f \in \mathcal{X}$. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$, a *weakness parameter* $\varrho \in (0,1]$, and assume that $\|d\|_{\mathcal{X}} = 1$ for all $d \in \mathcal{D}$. Generate a sequence $(f_k)_{k\in\mathbb{N}}$ by the following iteration:

1. Set $k := 0$ and set $f_0 := 0$.
2. Find a dictionary element d_{k+1} fulfilling
$$|\langle f - f_k, d_{k+1}\rangle_{\mathcal{X}}| \geq \varrho \sup_{d\in\mathcal{D}} |\langle f - f_k, d\rangle_{\mathcal{X}}|.$$
3. Set
$$\alpha_{k+1} := \langle f - f_k, d_{k+1}\rangle_{\mathcal{X}}.$$
4. Set $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$.
5. If a suitable stopping criterion is fulfilled: stop.
Otherwise: increase k by 1 and return to step 2 of the algorithm.

In analogy to the convergence results for the PGA, Temlyakov [167] also proves the following result for the WGA.

Theorem 3.4 (Convergence of the WGA)**.** Let $(f_k)_{k\in\mathbb{N}_0}$ be the sequence of approximations generated by the WGA to approximate $f \in \mathcal{X}$, let $\varrho \in (0,1]$, and let $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X}$.

Then $f_k \to f$ $(k \to \infty)$ and
$$\|f - f_k\|_{\mathcal{X}} \leq |f|_{\mathcal{A}_1(\mathcal{D})} (1 + k\varrho^2)^{-\varrho/(4+2\varrho)}. \tag{3.2}$$

Note that we stated a simplified version of the algorithm here. In the original paper, ϱ did not need to be constant throughout the iteration, instead a sequence $(\varrho_k)_{k\in\mathbb{N}_0}$ of weakness parameters can be used. As long as the sequence fulfills
$$\sum_{k=1}^{\infty} \frac{\varrho_k}{k} = \infty,$$
one can ensure convergence of the algorithm and the convergence rate in Eq. (3.2) still holds if the term $k\varrho^2$ is replaced by $\sum_{j=1}^{k} \varrho_j^2$.

One property of greedy algorithms that is important to notice, is that they are nonlinear. That is, the sequence $(\tilde{f}_k)_{k\in\mathbb{N}_0}$ of approximations of a sum $f + g$ of functions that is, for example, generated by the PGA, is not the sum of the sequences $(f_k)_{k\in\mathbb{N}_0}$, $(g_k)_{k\in\mathbb{N}_0}$ of approximations of f and g, respectively.

V. N. Temlyakov and several other authors developed multiple other versions of greedy approximation algorithms, which were all based on the general idea presented above. This includes, for example, the (Weak) Relaxed Greedy Algorithm (see DeVore and Temlyakov [32] and Temlyakov [167]), the Approximate Greedy Algorithm (see Gribonval and Nielsen [67]), and the Thresholding Greedy Algorithm (see Dilworth et al. [34]). The application of greedy algorithms to approximation in Banach spaces was studied in Temlyakov [168] using the Chebyshev Greedy Algorithm. Algorithms of greedy type were also applied in machine learning (see, e. g., Barron et al. [14]).

Finally, the idea of greedy algorithms and matching pursuits was applied to linear inverse problems by Fischer [45], Fischer and Michel [46], Gutting et al. [70], Michel [123], Michel and Orzlowski [126], Michel and Telschow [127, 128], and Telschow [166] yielding the Regularized Functional Matching Pursuit (RFMP) and the Regularized Orthogonal Functional Matching Pursuit (ROFMP) algorithms. The RFMP is a version of the MP and the PGA, whereas the ROFMP is an analogy to the OMP and the OGA. We will discuss the RFMP in more detail in Chapter 8, since in Chapters 9 and 10, we will derive new generalizations of these algorithms. In particular, the RWFMP in Chapter 9 will be an application of the WGA to linear inverse problems, whereas the RFMP_NL in Chapter 10 will be an MP/PGA-style algorithm for nonlinear inverse problems.

Part II.

A greedy algorithm for density estimation

Chapter 4.

Fundamentals of probability theory and statistics

In Chapter 5, we want to apply the idea of greedy algorithms to the problem of the estimation of probability density functions (PDFs). For this purpose, we will first summarize some basic results of probability theory in this Chapter 4. Then, we will review standard procedures for the estimation of probability density functions and the sampling from arbitrary probability distributions.

4.1. Basics of probability theory

We assume that the reader is familiar with the basics of measure and integration theory (see, e. g., Bauer [15]). For the basics of probability theory, we stick to the presentation in Pestman [144]. Note that we will mainly concentrate on $\mathbb{R}^d$-valued random variables for some $d \in \mathbb{N}$ in this section. In the end, we will shortly comment on how to transfer the presented concepts to $\mathbb{S}^{d-1}$-valued random variables, which are important for the application that we will present later on. We begin with the definition of a probability space.

Definition 4.1. Let Ω be a sample set, let $\mathfrak{A}$ be a σ-algebra of subsets of Ω, and let $\mathbb{P}$ be a probability measure on $\mathfrak{A}$, where

a) a *sample set* Ω is the model for a set of outcomes of a probability experiment, whose subsets are called *events*,

b) a *σ-algebra* $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ *of subsets of* Ω fulfills the three conditions

   i) $\Omega \in \mathfrak{A}$,
   ii) for all $A \in \mathfrak{A}$, we have $\Omega \setminus A \in \mathfrak{A}$, and
   iii) if $(A_k)_{k \in \mathbb{N}_0} \subseteq \mathfrak{A}$, then $\bigcup_{k \in \mathbb{N}_0} A_k \in \mathfrak{A}$,

c) a *probability measure* $\mathbb{P} \colon \mathfrak{A} \to [0, \infty)$ *on* $\mathfrak{A}$ is a measure that fulfills $\mathbb{P}(\Omega) = 1$.

Then the triplet $(\Omega, \mathfrak{A}, \mathbb{P})$ is called a *probability space*.

A particular σ-algebra on $\Omega := \mathbb{R}^d$ is the *collection* $\mathfrak{B}^d$ *of Borel sets*. It is defined as the smallest σ-algebra that contains all open subsets of $\mathbb{R}^d$, that is,

$$\mathfrak{B}^d := \bigcap_{\substack{\mathfrak{A} \text{ is a } \sigma\text{-algebra} \\ \mathfrak{O}(\mathbb{R}^d) \subseteq \mathfrak{A}}} \mathfrak{A},$$

where $\mathfrak{O}(\mathbb{R}^d)$ is the family of all open sets of $\mathbb{R}^d$. Any element of $\mathfrak{B}^d$ is called a *Borel set*, any measure on $\mathfrak{B}^d$ is called a *Borel measure*, and any measurable function $f\colon \mathbb{R}^d \to \mathbb{R}^p$ for $p \in \mathbb{N}$ is called a *Borel function*.

For $A \subseteq \Omega$ and an integrable function $f\colon \Omega \to \mathbb{R}$, by

$$\int_A f(x)\, \mathrm{d}\mathbb{P}(x)$$

we denote the integral of f over A with respect to the measure $\mathbb{P}$ as it is usually defined in integration theory.

Often, it is not possible to observe events (i. e., subsets of the sample space) directly in the real world. Therefore, one introduces the concept of random variables, which make it possible to describe also indirect outcomes of probability experiments.

Definition 4.2. Let $(\Omega, \mathfrak{A}, \mathbb{P})$ be a probability space and let $X\colon \Omega \to \mathbb{R}^d$ be a function. If we have $X^{-1}(A) \in \mathfrak{A}$ for every Borel set $A \subseteq \mathbb{R}^d$, we call X an $\mathbb{R}^d$*-valued random variable*.

We can obtain a Borel measure on $\mathbb{R}^d$ by

$$\mathfrak{B}^d \ni A \mapsto \mathbb{P}(X \in A) := \mathbb{P}_X(A) := \mathbb{P}\Big(X^{-1}(A)\Big),$$

which is called the *probability distribution of* X.

It is easy to see that for a random variable X, the triplet $\left(\mathbb{R}^d, \mathfrak{B}^d, \mathbb{P}_X\right)$ is a probability space. Due to the difficulty of observing the sample space Ω itself, $\left(\mathbb{R}^d, \mathfrak{B}^d, \mathbb{P}_X\right)$ is often the only space that is considered in probability theory. This is the reason, why we often omit the argument of X and also ignore the underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$.

From measure and integration theory, it is well known that there exist measures that cannot be rewritten as a Lebesgue integral (but of course as an integral with respect to the measure itself). In the following, we do not want to consider such measures, since we want to estimate probability density functions. In consequence, we introduce the notion of absolutely continuous measures.

Definition 4.3. Let $(\mathbb{R}^d, \mathfrak{B}^d, \mathbb{P})$ be a probability space. The probability measure $\mathbb{P}$ is called *absolutely continuous with respect to the Lebesgue measure* λ if for every set $A \in \mathfrak{B}^d$ we have $\lambda(A) = 0 \Rightarrow \mathbb{P}(A) = 0$. We write $\mathbb{P} \ll \lambda$.

By the Radon-Nikodym theorem (cf. Bauer [15, Theorem 17.10]), this is equivalent to the existence of a probability density function. For the considered case of a random variable, we can define the latter as follows.

Definition 4.4. Let X be an $\mathbb{R}^d$-valued random variable such that $\mathbb{P}_X \ll \lambda$ (we say that X has an *absolutely continuous probability distribution*). Then a function $f\colon \mathbb{R}^d \to [0, \infty)$ that fulfills $\mathbb{P}_X = f\,\lambda$ is called a *probability density function (PDF)* of X.

Note that the PDF is uniquely determined almost everywhere (i. e., two PDFs of the same distribution may only differ on a null set).

We denote the relation between X and f by $X \sim f$.

Presuming an absolutely continuous probability distribution of X with a PDF f, a simple application of measure and integration theory yields

$$\mathbb{P}(X \in A) = \int_A f(x)\,\mathrm{d}x$$

for every Borel set $A \in \mathfrak{B}^d$. Note that this implies

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \mathbb{P}\Big(X \in \mathbb{R}^d\Big) = \mathbb{P}_X\Big(\mathbb{R}^d\Big) = 1$$

from the fact that $\mathbb{P}_X$ is a probability measure.

If dealing with multiple random variables, two important concepts are the identical distribution and independence of these random variables.

Definition 4.5. Let $X_1, \ldots, X_N$ be $\mathbb{R}^d$-valued random variables with $X_n \sim f_n$ for probability density functions $f_n\colon \mathbb{R}^d \to [0, \infty)$ for all $n = 1, \ldots, N$. Then,

a) $X_1, \ldots, X_N$ are called *independent*, if we have for the $\left(\mathbb{R}^d\right)^n$-valued random variable $X = (X_1, \ldots, X_N)$ that

$$X \sim f,$$

where

$$f\colon \Big(\mathbb{R}^d\Big)^N \to [0, \infty), \quad f(x_1, \ldots, x_N) \coloneqq \prod_{n=1}^{N} f_n(x_n), \quad x_1, \ldots, x_N \in \mathbb{R}^d,$$

is the joint probability density function of $X_1, \ldots, X_N$.

b) $X_1, \ldots, X_N$ are called *identically distributed*, if $\mathbb{P}_{X_n} = \mathbb{P}_{X_m}$ holds. In other words, if we have $f_n = f_m$ almost everywhere for all $n, m = 1, \ldots, N$.

If both a) and b) are fulfilled, we call $X_1, \ldots, X_N$ *independent and identically distributed (i. i. d.)*.

Another important term in probability theory is the expectation, which will also be needed later on.

Definition 4.6. Let X be an $\mathbb{R}^d$-valued random variable and $g\colon \mathbb{R}^d \to \mathbb{R}$ a Borel function such that

$$\int_{\mathbb{R}^d} |g(x)| \, \mathrm{d}\mathbb{P}_X(x) < \infty.$$

Then

$$\mathbb{E}(g(X)) \coloneqq \int_{\mathbb{R}^d} g(x) \, \mathrm{d}\mathbb{P}_X(x)$$

is called the *expectation of* $g(X)$.

If $X \sim f$, we obtain

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x) f(x) \, \mathrm{d}x.$$

If $f \in \mathrm{L}^q(\mathbb{R}^d)$, a sufficient condition for the existence of $\mathbb{E}(g(X))$ is consequently that $g \in \mathrm{L}^p(\mathbb{R}^d)$ holds for $\frac{1}{p} + \frac{1}{q} = 1$, since then we have

$$\int_{\mathbb{R}^d} |g(x)| \, \mathrm{d}\mathbb{P}_X(x) = \int_{\mathbb{R}^d} |g(x)| \, f(x) \, \mathrm{d}x \leq \|g\|_{\mathrm{L}^p(\mathbb{R}^d)} \, \|f\|_{\mathrm{L}^q(\mathbb{R}^d)} < \infty \tag{4.1}$$

by Hölder's inequality.

An important result of probability theory that we will employ in the derivation of a greedy algorithm for the estimation of a PDF is the strong law of large numbers.

Theorem 4.7 (Strong law of large numbers, cf. Pestman [144, Theorem VII.2.14])**.** Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of i. i. d. $\mathbb{R}$-valued random variables such that $\mu = \mathbb{E}(X_1) < \infty$.

Then the sequence $(S_N)_{N \in \mathbb{N}}$ of means, that is,

$$S_N \coloneqq \frac{1}{N} \sum_{n=1}^{N} X_n$$

converges strongly (or almost surely) to μ, that is,

$$\mathbb{P}\left(\lim_{N\to\infty} S_N = \mu\right) = 1,$$

which means

$$\mathbb{P}\left(\left\{\, \omega \in \Omega \;\middle|\; \lim_{N\to\infty} S_N(\omega) = \mu \,\right\}\right) = 1.$$

Thus, in a certain sense, S_N is an approximation of μ.

The preceding concepts can easily be transferred to the case of $\mathbb{S}^{d-1}$-valued random variables. Analogously to the $\mathbb{R}^d$-valued case, the Borel algebra on $\mathbb{S}^{d-1}$ is defined as the smallest σ-algebra that contains all open subsets of $\mathbb{S}^{d-1}$. Note that, in this case, "open" has to be considered in terms of the subspace topology, that is, a subset of $\mathbb{S}^{d-1}$ is open if it is the intersection of an open set in $\mathbb{R}^d$ with $\mathbb{S}^{d-1}$. Using the so-defined Borel algebra, all definitions above can be transferred to the $\mathbb{S}^{d-1}$-valued case directly.

4.2. Kernel density estimators

An important problem in statistics is to find the probability distribution of a given sample of random variables $X_1, \dots, X_N \sim f$, where the PDF f is initially unknown. Thus, it is desirable to determine a so-called estimation $\hat{f}$ of the PDF f from the given sample of data. Standard techniques to solve this problem are, for example, histogram estimators and kernel density estimators (KDEs). As in the previous section, we will first present the classical $\mathbb{R}$-valued case and transfer the result to multiple dimensions and the sphere later. For this purpose, let $X_1, \dots, X_N \sim f$ be a set of i. i. d. $\mathbb{R}$-valued random variables and $x_1, \dots, x_N$ be the outcome of a single probability experiment (i. e., deterministic variables) such that x_n corresponds to the outcome of the random variable X_n for $n = 1, \dots, N$.

A simple method to obtain an estimate $\hat{f}$ is the so-called *empirical density function*, which is defined as

$$\hat{f}(x) := \frac{\#\{\, n \mid x_n = x \,\}}{N}$$

(cf. Pestman [144, Section VII.12]). Since $\{\, x_1, \dots, x_N \,\}$ is a null set, it is obvious that this is no reasonable PDF of an absolutely continuous probability distribution, because $\int_{\mathbb{R}} \hat{f}(x)\, \mathrm{d}x = 0$. The reason is that the empirical density function is a so-called discrete probability density, that is, it corresponds to a linear combination of Dirac

measures. To overcome this problem, one introduces methods that can be described as smoothed versions of the empirical density function.

Histogram estimators are based on the eponymous method to display data graphically. For real-valued random variables, the real line is divided into arbitrary equally sized subintervals and the resulting estimator is piecewise constant on these intervals, taking the value that corresponds to the fraction of all data points that fall into the specific interval. According to Härdle et al. [76, Chapter 3], the histogram estimator possesses several drawbacks, including the discontinuity of the estimated probability density function.

A continuous alternative are so-called *kernel density estimators*, which were first introduced by Parzen [140]. To introduce these estimators, we first have to define the notion of a kernel.

Definition 4.8 (cf. Pestman [144, Section VII.12])**.** A *kernel* $k\colon \mathbb{R} \to \mathbb{R}$ is a function that fulfills

a) $k(x) \geq 0$ and $k(-x) = k(x)$ for all $x \in \mathbb{R}$,
b) k is continuous and $\lim_{x\to\infty} k(x) = 0$,
c) $\int_{\mathbb{R}} k(x)\,\mathrm{d}x = 1$.

Having defined a kernel, we are now able to construct kernels of different so-called bandwidths $h > 0$ by

$$k_h(x) := \frac{1}{h}\, k\Big(\frac{x}{h}\Big).$$

We can prove the following results.

Theorem 4.9. The family of kernels $\{\, k_h \mid h > 0 \,\}$ fulfills:

a) $k_h(x) \geq 0$ for all $x \in \mathbb{R}$ and all $h > 0$,
b) $\int_{\mathbb{R}} k_h(x)\,\mathrm{d}x = 1$ for all $h > 0$,
c) for every constant $c > 0$, we have $\lim_{h\searrow 0} \int_{-c}^{c} k_h(x)\,\mathrm{d}x = 1$.

Proof. Part a) is obvious, since $k(x) \geq 0$ for all $x \in \mathbb{R}$ by Definition 4.8a).

Part b) can be seen by a change of variables $t = \frac{x}{h}$:

$$\int_{\mathbb{R}} k_h(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{h}\, k\Big(\frac{x}{h}\Big)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{h}\, k(t)\, h\;\mathrm{d}t = \int_{\mathbb{R}} k(t)\,\mathrm{d}t = 1,$$

due to Definition 4.8c).

Part c) can be proved as follows: we again use a change of variables, part b), and employ the symmetry of k to obtain

$$\int_{-c}^{c} k_h(x)\,\mathrm{d}x = 1 - 2\int_{c/h}^{\infty} k(t)\,\mathrm{d}t.$$

The latter integral has to converge to 0 for $h \searrow 0$, since $k \in \mathrm{L}^1(\mathbb{R})$. ■

It is remarkable that these properties also arise in approximation theory and Fourier analysis, where approximate identities are considered (cf. Grafakos [64, Section 1.2]). The linkage is the fact that the Dirac measure in measure and probability theory can be identified with the Dirac δ-distribution, which is studied in Fourier analysis. Actually, the idea of approximate identities is the approximation of the Dirac distribution by a sequence of smooth functions, which is similar to the idea behind kernel density estimators. Another fact that shows the similarity of KDEs and approximate identities is the following theorem.

Lemma 4.10 (cf. Pestman [144, Lemma VII.12.3])**.** Let $\{\, k_h \mid h > 0 \,\}$ be a family of kernels as defined above. If $x \in \mathbb{R}$ is a point of continuity of $g\colon \mathbb{R} \to \mathbb{R}$, then we have for the convolution

$$(k_h * g)(x) \coloneqq \int_{\mathbb{R}} k_h(x-t)\,g(t)\,\mathrm{d}t$$

that

$$\lim_{h \searrow 0} (k_h * g)(x) = g(x)$$

holds.

After these preparations, we can now define kernel density estimators on $\mathbb{R}$.

Definition 4.11. Let $X_1, \ldots, X_N \sim f$ be a set of $\mathbb{R}$-valued random variables. Then for a given bandwidth $h > 0$ the *kernel density estimator* $\hat{f}_h(X_1, \ldots, X_N)$ is defined as:

$$\hat{f}_h(X_1, \ldots, X_N)(x) \coloneqq \frac{1}{N}\sum_{n=1}^{N} k_h(x - X_n),$$

which strictly speaking is a random function itself (because it depends on the random variables $X_1, \ldots, X_N$).

For a specific outcome $x_1, \ldots, x_N \in \mathbb{R}$ of the probability experiment (e. g., a given data set), the estimator

$$\hat{f}_h(x_1, \ldots, x_N)(x) \coloneqq \frac{1}{N}\sum_{n=1}^{N} k_h(x - x_n)$$

is a deterministic function.

We will often omit the dependency on $X_1, \dots, X_N$ and $x_1, \dots, x_N$, respectively, if no confusion is likely to arise.

Note that $\hat{f}_h$ indeed is a PDF, since

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{f}_h(x)\,\mathrm{d}x &= \int_{-\infty}^{\infty} \frac{1}{N}\sum_{n=1}^{N} k_h(x - x_n)\,\mathrm{d}x = \frac{1}{N}\sum_{n=1}^{N}\int_{-\infty}^{\infty} k_h(x - x_n)\,\mathrm{d}x \\ &= \frac{1}{N}\sum_{n=1}^{N}\int_{-\infty}^{\infty} k_h(x)\,\mathrm{d}x = \frac{1}{N}\sum_{n=1}^{N} 1 = \frac{1}{N}\,N = 1,\end{aligned}$$

by the linearity of the integral, a change of variables, and the normalization of the kernels k_h.

We quote the following theorem, which is the main result about KDEs in statistics.

Theorem 4.12 (cf. Pestman [144, Theorem VII.12.5])**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}$-valued random variables such that $X_n \sim f$ for all $n \in \mathbb{N}$. Moreover, let $(h_N)_{N\in\mathbb{N}}$ be a sequence of bandwidths such that

$$\lim_{N\to\infty} h_N = 0, \qquad \lim_{N\to\infty} N\,h_N^2 = \infty.$$

Then we have

$$\hat{f}_{h_N}(X_1, \dots, X_N)(x) \to f(x)$$

for $N \to \infty$ in probability.

In this theorem, the convergence of a sequence $(Y_N)_{N\in\mathbb{N}}$ of random variables to a random variable Y for $N \to \infty$ *in probability* means that

$$\lim_{N\to\infty} \mathbb{P}(|Y_N - Y| > \varepsilon) = 0$$

for all $\varepsilon > 0$. Amongst other results from probability theory, the proof relies on Lemma 4.10, which again highlights the importance of the fact that the family $\{\, k_h \mid h > 0 \,\}$ of kernels also forms an approximate identity.

Theorem 4.12 can be interpreted as follows: if the bandwidth is chosen in a proper way in dependence of the number of data points N, then the kernel density estimator $\hat{f}_h(X_1, \dots, X_N)$ is a good approximation of f (in a certain sense) if the size of the sample N is large. The conditions on the sequence of bandwidths can be motivated from the fact that they yield asymptotic unbiasedness and an asymptotic vanishing

variance of the estimator, which are desirable properties in estimation theory. We do not want to go into further detail here. We refer to Pestman [144, Section VII.12] for more details. It remains to say that Theorem 4.12 only gives an asymptotic result. If a fixed number N of observations is given, it is not trivial to choose the bandwidth. For that reason, we will later on choose a bandwidth more or less arbitrarily. An overview of different methods for bandwidth selection in density estimation can be found in Jones et al. [93]. Note that this dilemma is similar to the choice of a regularization parameter in the solution of ill-posed inverse problems.

On the real line, another possibility for the estimation of a probability density is to take the empirical distribution function and use a finite difference approximation of the PDF (cf. Rosenblatt [154]), which is reasonable since analytically the PDF is the derivative of the distribution function. The drawback of this strategy is that it cannot be easily transferred to the multivariate or spherical case due to the non-existence of a distribution function. A distribution function doesn't exist in these cases because $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$ are not ordered sets. The same is true for the histogram estimator since one needs a suitable partition of the space at hand, where it would be desirable that it consists of subsets of equal area. This is of course possible in the multivariate case, but not that simple in the spherical case (for a viable solution see Górski et al. [63]). In contrast, kernel density estimators can be applied more easily in these cases such that KDEs are superior to both of the other methods.

Transferring KDEs to the sphere is pretty straight-forward and has first been done by Hall et al. [74]. Here, the most important difference is that in the definition of kernel density estimators the distance $x - x_n$ for $x, x_n \in \mathbb{R}$ is replaced by the inner product $\xi \cdot \xi_n$ for $\xi, \xi_n \in \mathbb{S}^{d-1}$. This is reasonable since the inner product of two unit vectors measures their distance. This is due to

$$|\xi - \xi_n|^2 = |\xi|^2 - 2\,\xi \cdot \xi_n + |\xi_n|^2 = 2 - 2\,\xi \cdot \xi_n.$$

Consequently, motivated by Definition 4.8 and Theorem 4.9, the definition of a kernel and the kernel density estimator on the sphere are as follows. Note that we also let no longer tend the bandwidth parameter to zero, but to one. This is done in preparation of the fact that we want to use the so-called Abel-Poisson kernel later, which exactly possesses this parametrization and fulfills the following definition.

Definition 4.13. A family $\{\, k_h \mid h \in (0,1) \,\}$, $k_h \colon [-1,1] \to \mathbb{R}$ is called a *family of spherical kernels* if

a) $k_h(t) \geq 0$ for all $t \in [-1,1]$,
b) $\int_{\mathbb{S}^{d-1}} k_h(\eta \cdot \xi)\, \mathrm{d}\omega(\xi) = 1$ for all $h \in (0,1)$ and $\eta \in \mathbb{S}^{d-1}$,

c) for every constant $c \in [-1, 1)$, we have

$$\lim_{h \nearrow 1} \int_{\{\, \xi \in \mathbb{S}^{d-1} \,\mid\, \eta \cdot \xi \geq c \,\}} k_h(\eta \cdot \xi) \, \mathrm{d}\omega(\xi) = 1$$

for all $\eta \in \mathbb{S}^{d-1}$.

Also in the spherical case, one can find a connection of these families of spherical kernels and approximate identities (see Freeden and Hesse [55, Section 2, especially Theorem 2.3(v)]).

Based on this definition of kernels, we can now introduce spherical kernel density estimators.

Definition 4.14 (cf. Hall et al. [74])**.** Given a set $\Xi_1, \ldots, \Xi_N \sim f$ of i. i. d. $\mathbb{S}^{d-1}$-valued random variables and a family of kernels $\{\, k_h \mid h \in (0,1) \,\}$, the associated kernel density estimator is defined as

$$\hat{f}_h(\Xi_1, \ldots, \Xi_N)(\xi) \coloneqq \frac{1}{N} \sum_{n=1}^{N} k_h(\Xi_n \cdot \xi).$$

Like in the real-valued case, for a fixed observation $\xi_1, \ldots, \xi_N$ of $\Xi_1, \ldots, \Xi_N$, we define

$$\hat{f}_h(\xi_1, \ldots, \xi_N)(\xi) \coloneqq \frac{1}{N} \sum_{n=1}^{N} k_h(\xi_n \cdot \xi).$$

Hall et al. [74] have proved results similar to Theorem 4.12 for spherical KDEs. Since this is not the focus of this work, we refer to this reference for further details.

4.3. Sampling from probability distributions on the real line and the sphere

An important task in computational statistics is the generation of non-uniformly distributed random variables, which is called *sampling* from a given distribution. There exists a vast amount of univariate and multivariate probability distributions that one can sample from, and for each distribution there might be several different methods (for an overview, see the book of Devroye [33]). Unfortunately, most methods are tailored specifically for a certain distribution and cannot be transferred to other probability distributions. Only some methods can be applied to almost arbitrary distributions. One of them is *rejection sampling*, which we will present in

this section, since it can also be used to sample from a PDF on the sphere. Since rejection sampling relies on the ability to sample from at least one other distribution, we first consider briefly the generation of uniform random variables on the real line and on the sphere. We will then discuss how the rejection sampling method can be used to transform uniform random variables to non-uniform random variables.

At first, it is to say that, in general, computers can only generate so-called pseudo-random numbers, that is, the generated numbers pass several statistical tests for randomness, although they are generated by a deterministic procedure (cf. Niederreiter [137, Section 7.2]). For the sake of simplicity, we will use the term *random* in the following, even if, technically, we should use the term *pseudo-random* instead.

4.3.1. Sampling from the uniform distribution

A very popular method for the generation of uniform random numbers on a real interval $[a, b)$ is the *linear congruential method* (cf. Niederreiter [137, Section 7.3]). Without going into too much detail, we pursue the idea to first generate a sequence of discrete uniformly distributed random variables from the set $\{ 0, \dots, M-1 \}$ for some $M \in \mathbb{N}$ and to rescale the result to the desired interval. The discrete random sequence $(y_n)_{n \in \mathbb{N}_0} \subseteq \{ 0, \dots, M-1 \}$ is obtained by the recursion

$$y_{n+1} \coloneqq (a\, y_n + c) \bmod M,$$

for $n \in \mathbb{N}_0$, starting with a so-called *seed value* $y_0 \in \{ 0, \dots, M-1 \}$ and prescribing the parameters a and c. The correct choice of these parameters is a crucial ingredient for the quality of random numbers generated by the linear congruential method. Since nowadays in every programming language there are random number generators available, which perform very well, we do not go into further detail here, but rely on these well-tested implementations.

Generating uniformly distributed $\mathbb{S}^{d-1}$-valued random variables can be done in different ways. We only present the following method for $\mathbb{S}^2$, which has been derived by Marsaglia [116] and is also used in the GNU Scientific Library (cf. Galassi et al. [62]), since we will use this library in the implementation. The method works as follows. First, sample two random variables Y_1, Y_2 from the uniform distribution on the interval $(-1, 1)$ (for example, by the linear congruential method mentioned above) until the pair (Y_1, Y_2) fulfills $S \coloneqq Y_1^2 + Y_2^2 < 1$. This actually yields random points that are uniformly distributed over the unit circle in $\mathbb{R}^2$. Consider the random vector $X = \left(2\, Y_1 \sqrt{1-S}, 2\, Y_2 \sqrt{1-S}, 1 - 2\, S\right)$. This random vector is then uniformly distributed on the unit sphere in $\mathbb{R}^3$.

4.3.2. Rejection sampling from arbitrary distributions

As common throughout this chapter, we first present the basic idea of *rejection sampling* (also called *acceptance-rejection method*) for the real-valued case. We base our considerations on the presentation in Devroye [33].

Let $f\colon \mathbb{R} \to [0,\infty)$ be a probability density function on $\mathbb{R}$ from which the generation of samples is desired. Rejection sampling in general needs the knowledge of another PDF g that dominates f such that $f \leq c\,g$ for some constant $c > 0$. Since it is the relevant case in our implementation, we restrict ourselves to the case that f is defined on an interval $[a,b]$ and g is the density corresponding to the uniform distribution on this interval, that is, we assume that f is (essentially) bounded. The following theorem provides the theoretical justification for the method of rejection sampling.

Theorem 4.15 (cf. Devroye [33, Theorem 3.1])**.** Let $X \sim f$ be an $[a,b]$-valued random variable and let U be a random variable that is uniformly distributed on $[0,1]$. Then, the $\mathbb{R}^2$-valued random variable $(X, U\,f(X))$ is uniformly distributed on the set

$$A_f = \{\, (x,u) \mid x \in [a,b], \quad 0 \leq u \leq f(x) \,\}.$$

On the other hand, if (X,U) is an $\mathbb{R}^2$-valued random variable uniformly distributed on A_f, then $X \sim f$.

In consequence, if we are able to obtain a sample from the random variable (X,U) that is uniformly distributed on A_f, the first component is a sample from the distribution on $[a,b]$ with PDF f. In practice, a sample (x,u) of (X,U) can be obtained by first generating a sample $(\tilde{x},\tilde{u})$ from a uniform distribution on

$$\tilde{A} = \{\, (y,t) \mid y \in [0,1], \quad 0 \leq t \leq C \,\}$$

for some upper bound $C > 0$ of f, which can be easily done by sampling from the uniform distribution on $[0,1]$ and $[0,C]$, respectively. If $\tilde{u} \leq f(\tilde{x})$, the sample is *accepted* and (x,u) is taken as a sample from the uniform distribution on A_f. If, on the other hand, $\tilde{u} > f(\tilde{x})$, the sample is *rejected* (hence the name of the method) and the generated pair is not taken into account for the samples from the uniform distribution on A_f.

Replacing the interval $[0,1]$ by the sphere $\mathbb{S}^2$, the generation of samples from a random variable $\Xi \sim f\colon \mathbb{S}^2 \to [0,\infty)$ is based on the generation of a uniform sample (ξ,u) on the set

$$A = \left\{\, (\xi,u) \;\middle|\; \xi \in \mathbb{S}^2, \quad 0 \leq u \leq f(x) \,\right\}.$$

Analogously, this can be done by generating a sample $(\tilde{\xi}, \tilde{u}) \in \mathbb{S}^2 \times [0, C]$ from the uniform distribution on

$$\tilde{A} = \{ (\eta, t) \mid \eta \in \mathbb{S}^2, \quad 0 \leq t \leq C \}$$

first and perform the same rejection strategy as above. Since methods for generating uniform distributions both on $\mathbb{S}^2$ and $[0, C]$ are well-known (see Section 4.3.1) and implemented in well-established software (like the already mentioned GNU Scientific Library), sampling from the uniform distribution on $\tilde{A}$ can easily be accomplished both in theory and in practice. Note that the efficiency of rejection sampling depends heavily on the quality of the upper bound C. If C is chosen too large, a lot of samples from $\tilde{A}$ are rejected. This is the reason why the next section deals with the determination of an upper bound in the case of a probability density that is given by a kernel density estimator.

4.3.3. Sampling from kernel density estimators

A problem that we are concerned with is to sample from a PDF $\hat{f}_h \colon \mathbb{S}^2 \to [0, \infty)$ that is given by a kernel density estimator using the rejection sampling method. The considerations in the previous section suggest that an upper bound C of the PDF is needed. Again, we first look at the one-dimensional case, since we can reuse the result for the spherical case.

Recall that the kernel density estimator for samples $x_1, \ldots, x_N$ of real-valued random variables $X_1, \ldots, X_N$ is given by

$$\hat{f}_h(x_1, \ldots, x_N)(x) = \frac{1}{N} \sum_{n=1}^{N} k_h(x - x_n), \qquad x \in \mathbb{R},$$

where $k_h \colon \mathbb{R} \to [0, \infty)$ is a kernel function. The easiest rigorous way to achieve an upper bound for $\hat{f}_h$ is using the triangle inequality to obtain

$$\begin{aligned} \left| \hat{f}_h(x) \right| &= \left| \frac{1}{N} \sum_{n=1}^{N} k_h(x - x_n) \right| \leq \frac{1}{N} \sum_{n=1}^{N} |k_h(x - x_n)| \\ &\leq \frac{1}{N} \sum_{n=1}^{N} \|k_h\|_{\mathrm{C}(\mathbb{R})} = \|k_h\|_{\mathrm{C}(\mathbb{R})} =: C_{\mathrm{tri}} \end{aligned} \tag{4.2}$$

for all $x \in \mathbb{R}$. It is easy to see that, in general, this upper bound is much too large, since we have $\hat{f}_h(x) = C_{\mathrm{tri}}$ for some $x \in \mathbb{R}$ if and only if $x_n = x$ for all $n = 1, \ldots, N$ if we choose a kernel that attains its maximum only at 0 and this is the case for most of the kernels that are typically used.

Therefore, we apply a different strategy for the determination of a bound for $\hat{f}_h$, which we could not find in the existing literature. We again restrict ourselves to the case of $[a,b]$-valued random variables. The idea is the following: the optimal upper bound of the kernel density estimator would be $\sup_{x\in[a,b]} \hat{f}_h(x)$, which cannot be easily determined due to the uncountability of $[a,b]$. What can be easily determined by evaluation of the KDE, is the term $\max_{t\in T_K} \hat{f}_h(t)$ for a given finite set of points $T_K := \{ t_1, \ldots, t_K \} \subseteq [a,b]$. Intuitively, one would expect that under suitable assumptions on the kernel function k_h and by using a sufficiently dense grid T_K, the maximum of the function on T_K is a good approximation of the supremum on $[a,b]$. We will quantify this intuition in the following for real-valued functions on $[a,b]$ and will afterwards apply the result to the KDE. Furthermore, we will consider the spherical case afterwards.

We denote by $\Theta_K := \max_{x\in[a,b]} \min_{t\in T_K} |x-t|$ the *nodal width* of the point set T_K (cf. the spherical setting in Freeden [51], Freeden et al. [53, Section 6.2], and Michel [122, Section 6.3]). The following theorem gives an answer to the risen question for Lipschitz continuous functions.

Theorem 4.16. Let $g\colon [a,b] \to \mathbb{R}$ be an arbitrary Lipschitz continuous function, that is,

$$|g(x) - g(y)| \leq L\,|x-y|$$

for all $x, y \in [a,b]$ and some constant $L > 0$. Let $T_K \subseteq [a,b]$ be a finite set of points and let Θ_K be its nodal width. Then,

$$\max_{x\in[a,b]} |g(x)| \leq \max_{t\in T_K} |g(t)| + L\,\Theta_K.$$

Proof. Let $x \in [a,b]$. Then by the definition of Θ_K there exists $t_x \in T_K$ such that

$$|x - t_x| \leq \Theta_K. \tag{4.3}$$

Thus,

$$\big||g(x)| - |g(t_x)|\big| \leq |g(x) - g(t_x)| \leq L\,|x - t_x| \leq L\,\Theta_K$$

due to the reversed triangle inequality, the Lipschitz-continuity of g and Eq. (4.3).

It follows that

$$|g(x)| \leq |g(t_x)| + L\,\Theta_K \leq \max_{t\in T_K} |g(t)| + L\,\Theta_K \tag{4.4}$$

and since $x \in [a,b]$ was chosen arbitrarily and the right-hand side of Eq. (4.4) does not depend on x, we obtain the desired result. ∎

We can apply this result to KDEs.

Corollary 4.17. Let $\hat{f}_h$ be the kernel density estimator defined by Definition 4.11 with respect to given observations $x_1, \ldots, x_N \in [a, b]$ of the $[a, b]$-valued random variables $X_1, \ldots, X_N \sim f$ and let the kernel k_h be Lipschitz continuous with Lipschitz constant $L > 0$. Furthermore, let $T_K \subseteq [a, b]$ be a finite set of points with nodal width Θ_K.

Then,

$$\max_{x \in [a,b]} \hat{f}_h(x) \leq \max_{t \in T_K} \hat{f}_h(t) + L\,\Theta_K.$$

Proof. It is well-known that the linear combination of Lipschitz continuous functions is again Lipschitz continuous and a Lipschitz constant of the linear combination is given by the linear combination of the Lipschitz constants (cf. Eriksson et al. [43, Section 12.5]).

Therefore, a Lipschitz constant $\hat{L}$ of $\hat{f}_h$ is given by

$$\hat{L} = \sum_{n=1}^{N} \frac{1}{N} L = N \frac{1}{N} L = L.$$

A combination of this result and the non-negativity of $\hat{f}_h$ with Theorem 4.16 proves the proposed inequality. ■

The following simple example shows that the upper bound that one obtains by the application of Corollary 4.17 is much smaller than the one obtained by applying the triangle inequality (see Eq. (4.2)).

Example 4.18. Let the kernel $k\colon \mathbb{R} \to [0, \infty)$ be given by

$$k(x) = \begin{cases} 1 + x, & -1 < x \leq 0, \\ 1 - x, & 0 < x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that k is actually a kernel as defined in Definition 4.8. For a visualization of this kernel, see Figure 4.1.

In this example, we assume that the observation $x_1 = 1, x_2 = 2, x_3 = 3$ of i. i. d. $[0, 4]$-valued random variables $X_1, X_2, X_3 \sim f$ is given and that we construct the

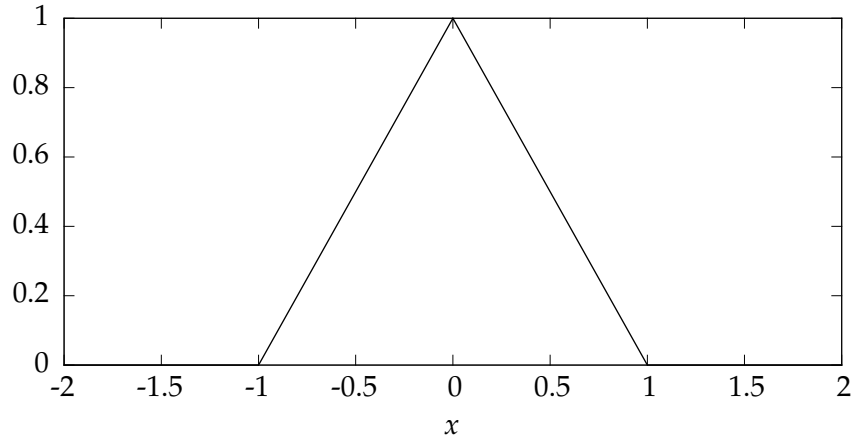


Figure 4.1.: Visualization of the kernel used in Example 4.18.

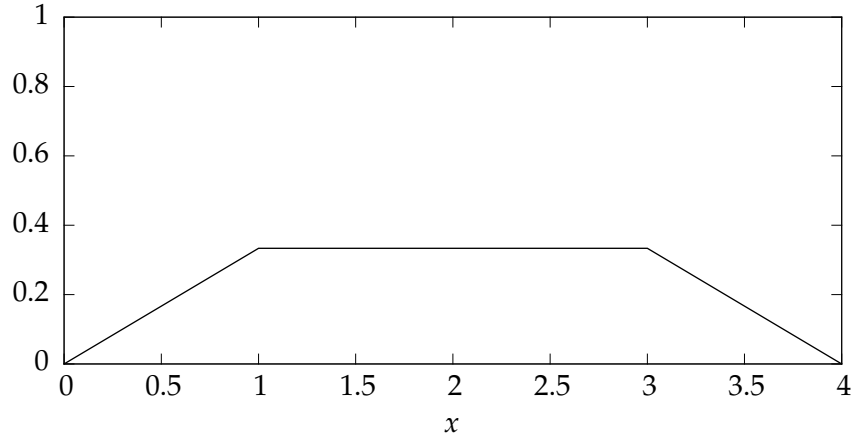


Figure 4.2.: Visualization of the kernel density estimator from Example 4.18.

kernel density estimator $\hat{f}_h(x_1, x_2, x_3)$ with $h = 1$. The KDE can be analytically computed to be

$$\hat{f}_1(x_1, x_2, x_3)(x) = \begin{cases} \frac{x}{3}, & 0 \leq x \leq 1, \\ \frac{1}{3}, & 1 < x < 3, \\ \frac{4-x}{3}, & 3 \leq x \leq 4 \end{cases}$$

and the optimal upper bound is thus $C_{\text{opt}} = \frac{1}{3}$. A visualization of the resulting KDE can be found in Figure 4.2.

It is easy to see that $\|k_1\|_{C(\mathbb{R})} = \|k\|_{C(\mathbb{R})} = 1$ such that by the triangle inequality, we obtain the upper bound

$$C_{\text{tri}} = 1,$$

which is much larger than C_{opt}.

To obtain an upper bound by the application of Corollary 4.17, we define an equally-spaced grid $T_{K+1} = \left\{ t_k = \frac{4k}{K} \;\middle|\; k = 0, \dots, K \right\}$ such that $\Theta_{K+1} = \frac{2}{K}$. If $K \geq 2$ there always exists a point $t \in T_{K+1}$, where $\hat{f}(t) = \frac{1}{3}$ such that

$$\max_{t \in T_{K+1}} \hat{f}_1(t) = \frac{1}{3}.$$

Since it is piecewise differentiable, a Lipschitz constant of the kernel k_1 is given by the maximum value of the piecewise defined derivative, that is, $L = 1$. Thus, we obtain by Corollary 4.17 that

$$\max_{x \in [0,4]} \hat{f}_1(x) \leq \frac{1}{3} + \frac{2}{K} =: C_{\text{eval},K},$$

depending on the number of points K. Since the second summand tends to zero for $K \to \infty$, we can obtain an upper bound arbitrarily close to C_{opt} by choosing K large enough.

For example, we have $C_{\text{eval},K} = \frac{1}{2}$ already for $K = 12$, which is a much better estimate for the upper bound than C_{tri}.

We return to the spherical case. Analogously to the real-valued case, we call a function $g \colon \mathbb{S}^{d-1} \to \mathbb{R}$ Lipschitz continuous with Lipschitz constant L, if

$$|g(\xi) - g(\eta)| \leq L \, |\xi - \eta|$$

for all $\xi, \eta \in \mathbb{S}^{d-1}$ (cf. Freeden and Schreiner [58, Section 2.4]). For a given finite point set $T_K \subseteq \mathbb{S}^{d-1}$, we define the nodal width similar to the real-valued case by $\Theta_K := \max_{\xi \in \mathbb{S}^{d-1}} \min_{\tau \in T_K} |\xi - \tau|$.

With these definitions, the proof of Theorem 4.16 can be directly transferred to the spherical case.

Theorem 4.19. Let $g \colon \mathbb{S}^{d-1} \to \mathbb{R}$ be an arbitrary Lipschitz continuous function with Lipschitz constant $L > 0$. Let $T_K \subseteq \mathbb{S}^{d-1}$ be a finite set of points and let Θ_K be its nodal width. Then,

$$\max_{\xi \in \mathbb{S}^{d-1}} |g(\xi)| \leq \max_{\tau \in T_K} |g(\tau)| + L \, \Theta_K.$$

To obtain a result like Corollary 4.17 for the spherical case, we have to determine the Lipschitz constant of a kernel function on the sphere. Let $k_h \colon [-1, 1] \to [0, \infty)$ be an element of a family of spherical kernels according to Definition 4.13 that is Lipschitz

continuous with Lipschitz constant L. Note that Lipschitz continuity is considered on $\mathbb{R}$ here. For fixed $\zeta \in \mathbb{S}^{d-1}$, we obtain

$$\begin{aligned} |k_h(\xi \cdot \zeta) - k_h(\eta \cdot \zeta)| &\leq L\,|\xi \cdot \zeta - \eta \cdot \zeta| \\ &= L\,|(\xi - \eta) \cdot \zeta| \leq L\,|\xi - \eta|\,|\zeta| = L\,|\xi - \eta| \end{aligned}$$

due to the Lipschitz continuity of k_h and the Cauchy-Schwarz inequality. This proves the following lemma.

Lemma 4.20. Let $k_h\colon [-1,1] \to [0,\infty)$ be an element of a family of spherical kernels that is Lipschitz continuous with Lipschitz constant L. Then for fixed $\zeta \in \mathbb{S}^{d-1}$, the spherical function $\xi \mapsto k_h(\xi \cdot \zeta)$ is also Lipschitz continuous with the identical Lipschitz constant L.

Finally, we obtain the following result in analogy to Corollary 4.17 for the KDE given by Definition 4.14.

Corollary 4.21. Let $\hat{f}_h$ be the KDE defined by Definition 4.14 with respect to given observations $\xi_1, \ldots, \xi_N \in \mathbb{S}^{d-1}$ of the spherical i. i. d. random variables $\Xi_1, \ldots, \Xi_N \sim f$. Let the kernel $k_h\colon [-1,1] \to [0,\infty)$ be Lipschitz continuous with Lipschitz constant $L > 0$. Moreover, let $T_K \subseteq \mathbb{S}^{d-1}$ be a finite set of points with nodal width Θ_K. Then,

$$\max_{\xi \in \mathbb{S}^{d-1}} \hat{f}_h(\xi) \leq \max_{\tau \in T_K} \hat{f}_h(\tau) + L\,\Theta_K.$$

The proof consists of the computation of a Lipschitz constant of $\hat{f}_h$. This can be done exactly like in the real-valued case.

In conclusion, we see that we can determine the maximum value of a kernel density estimator on the sphere to arbitrary precision by evaluating the KDE on a grid that is fine enough. For this purpose, we have to determine the Lipschitz constant of the real-valued kernel function k_h and the nodal width of the set of grid points, which may be done easily for a given kernel and point set.

Thus, for sampling from a kernel density estimator, the upper bound used in rejection sampling can be determined in the way presented above. Unfortunately, it may be the case that the number of evaluation points may be very high. However, this is actually only a minor problem when using rejection sampling, since the upper bound needs only to be determined once, whereas a worse upper bound (for example, determined by the triangle inequality) affects the performance of the sampling method in the generation of every single sample. In consequence, as long as the number of desired samples is high, the computational costs for determining the upper bound are worth it.

Chapter 5.

A greedy algorithm for the estimation of probability density functions

In this chapter, we will introduce a new greedy-type algorithm for the estimation of probability density functions (PDFs). We will apply the algorithm to a data set of fiber directions of a so-called nonwoven fabric. The output of the algorithm—a sparse approximation of the distribution of fiber directions—will be used in an elementary algorithm for the simulation of such technical textiles. Parts of this chapter have already been published in Gramsch et al. [66].

5.1. Greedy algorithms in statistics

Before introducing the newly developed greedy algorithm for the estimation of a probability density function, we want to give a literature review of the application of greedy algorithms in statistics and we want to demonstrate how the proposed algorithm will differ from established methods. Actually, in the retrospective view, it turns out that greedy algorithms have been developed in approximation theory, signal processing, and statistics, independently. The most widely known method of this type in statistics is *projection pursuit regression (PPR)*, which was developed by Friedman and Stuetzle [60]. As the name suggests, the method deals with the statistical problem of regression. This means that given observations $x_1, \dots, x_N \in \mathbb{R}$ and $y \in \mathbb{R}$ of real-valued random variables $X_1, \dots, X_N$ and Y, respectively, one is seeking a so-called *predictor function* $f\colon \mathbb{R}^N \to \mathbb{R}$ such that

$$y = f(x_1, \dots, x_N) + \varepsilon,$$

where ε is an error-term (cf. Sen and Srivastava [162, Section 1.3]). The most widely used technique for regression is linear regression, especially using the *ordinary least squares estimator*, which is no more than assuming that $f(x) = a \cdot x$ for $x, a \in \mathbb{R}^N$ and, given enough data points, solving the resulting over-determined system of linear equations by solving the normal equation (cf. Sen and Srivastava [162, Section 2.3]).

Note that this is very similar to the field of linear inverse problems and several problems arise in both fields, for example, overfitting.

In contrast to this standard procedure, PPR does not assume a linear predictor function f, but generates f by a method that is similar to greedy algorithms. In more detail, one generates an approximation f_M of f by a linear combination of some so-called real-valued ridge functions s_m such that

$$f_M(x) = \sum_{m=1}^{M} s_m(a_m \cdot x), \qquad x \in \mathbb{R}^N,$$

where $s_m \colon \mathbb{R} \to \mathbb{R}$, $a_m \in \mathbb{R}^N$. Given f_M, the functions s_{M+1} and the vectors a_{M+1} are determined such that the resulting approximation $f_{M+1} = f_M + s_{M+1}(a_{M+1}\cdot)$ is optimal in a certain sense. For the determination of the optimal choice of s_m and a_m Friedman and Stuetzle [60] use numerical optimization methods. For more details on PPR, consider the vast amount of literature about this method, for example, Friedman and Stuetzle [60], Huber [86], and Jones [92]. Important for us is the fact that the problem at hand, density estimation, is not a regression problem and thus, the PPR algorithm cannot be used for this purpose.

A greedy algorithm for density estimation called *projection pursuit density estimation (PPDE)* has been presented in Friedman et al. [61]. This method is different from our approach that will be presented in the following section. In PPDE, an approximation f_M of a given density function f is constructed by

$$f_M(x) = f_0(x) \prod_{m=1}^{M} s_m(a_m \cdot x),$$

where the s_m are again real-valued ridge functions. The greedy strategy in this case consists of finding s_{M+1} and a_{M+1} given f_M such that $f_{M+1} = f_M\, s_{M+1}(a_{M+1}\cdot)$ is again optimal in a certain sense (maximizing the so-called *Kullback-Leibler distance*) with the restriction that f_{M+1} should be a probability density function. As pointed out in Friedman et al. [61], in the case where f is unknown, which is always the case in real-world applications, again numerical optimization methods have to be used to compute the next iterates.

There will be no need for using sophisticated iterative numerical optimization methods in the iterations of the newly developed algorithm, which makes it superior to PPDE.

5.2. The greedy algorithm

In this section, we will develop a greedy algorithm for the estimation of a probability density function $f\colon \mathcal{M} \to [0,\infty)$, which is based on a sample $x_1,\dots,x_N \in \mathcal{M}$ of i. i. d. $\mathcal{M}$-valued random variables $X_1,\dots,X_N \sim f$. Here, $\mathcal{M}$ can be an arbitrary set that is equipped with a Borel measure, for example, $\mathbb{R}$, $[a,b]$ for $a,b \in \mathbb{R}$, $a < b$, or $\mathbb{S}^{d-1}$. Let furthermore $X \sim f$ be another $\mathcal{M}$-valued random variable such that $X, X_1,\dots,X_N$ are independent. In the following, we assume that $f \in \mathrm{L}^2(\mathcal{M})$ such that due to Eq. (4.1) we have $\mathbb{E}(g(X)) < \infty$ for every $g \in \mathrm{L}^2(\mathcal{M})$.

We will first motivate that in certain applications it might be advantageous to apply a greedy strategy to the problem of estimating a PDF. Assume that we have a large number N of samples. According to Section 4.2, given a family of kernels $\{\, k_h \mid h > 0 \,\}$, we can compute the kernel density estimator. The KDE is a sum of evaluations of the kernel on N points. Thus, for every evaluation of the KDE we have to evaluate the kernel N times. If we now want to generate new samples that are distributed approximately according to f we can sample from the KDE $\hat{f}_h$ instead, using the techniques presented in Sections 4.3.2 and 4.3.3. Unfortunately, when using rejection sampling, it is necessary to evaluate the KDE multiple times, depending on the quality of the upper bound and on the shape of the KDE itself. The more candidates for a sample are rejected, the worse the situation gets. Later, we will present computation times for the use of KDEs inside a rejection sampling algorithm and it turns out that in a realistic scenario this will lead to a computation time of days to weeks. Thus, determining a sparse estimation of the PDF is desirable, since, depending on the degree of sparsity, computation times can be drastically reduced when the sparse estimation is used inside a sampling method. Since greedy algorithms yield sparse representations of functions, it is straight-forward to apply a greedy strategy to the problem at hand. For details on the computation times when using KDEs and sparse PDFs in the application, see Table 5.1 in Section 5.4.5.

We do now try to apply the Pure Greedy Algorithm (PGA) (see Algorithm 3.1) to the problem of density estimation. For this purpose, let

$$\mathcal{D} \subseteq \left\{ g \in \mathrm{L}^2(\mathcal{M}) \;\middle|\; \|g\|_{\mathrm{L}^2(\mathcal{M})} = 1 \right\}$$

be a dictionary.

Recall that in the PGA, based on an initial approximation $f_0 \in \mathrm{L}^2(\mathcal{M})$, a sequence of approximations of the PDF f is generated by virtue of the recursion

$$f_{k+1} \coloneqq f_k + \alpha_{k+1}\, d_{k+1}, \qquad k \in \mathbb{N}_0,$$

where

$$d_{k+1} = \operatorname*{argmax}_{d \in \mathcal{D}} \left| \langle f - f_k, d \rangle_{\mathrm{L}^2(\mathcal{M})} \right|, \qquad \alpha_{k+1} = \langle f - f_k, d_{k+1} \rangle_{\mathrm{L}^2(\mathcal{M})}.$$

The difficulty in the application of the PGA to the problem at hand lies in the computation of the inner products in the determination of d_{k+1} and α_{k+1}. The problem is that the PDF f is typically unknown (otherwise one would not need to estimate it) such that the inner product $\langle f - f_k, d \rangle_{\mathrm{L}^2(\mathcal{M})}$ for $d \in \mathcal{D}$ cannot be computed. In more detail, we can use the bilinearity of the inner product to obtain

$$\langle f - f_k, d \rangle_{\mathrm{L}^2(\mathcal{M})} = \langle f, d \rangle_{\mathrm{L}^2(\mathcal{M})} - \langle f_k, d \rangle_{\mathrm{L}^2(\mathcal{M})}. \tag{5.1}$$

Since $f_k = f_0 + \sum_{j=1}^{k} \alpha_j d_j$, the latter inner product can be rewritten as

$$\langle f_k, d \rangle_{\mathrm{L}^2(\mathcal{M})} = \langle f_0, d \rangle_{\mathrm{L}^2(\mathcal{M})} + \sum_{j=1}^{k} \alpha_j \langle d_j, d \rangle_{\mathrm{L}^2(\mathcal{M})}.$$

Thus, if the inner products $\langle f_0, d \rangle_{\mathrm{L}^2(\mathcal{M})}$ is known as well as the inner products $\langle d, d' \rangle_{\mathrm{L}^2(\mathcal{M})}$ for arbitrary elements $d, d' \in \mathcal{D}$ from the dictionary, then we can compute the third inner product in Eq. (5.1) easily. The problem to compute the second inner product in this equation remains. Since we do not have any values of the function f or the evaluation of some operator applied to f at hand, methods like the RFMP (cf. Chapter 8) cannot be applied directly. Instead, we have only given data points that are distributed according to the PDF that is to be approximated.

The key to overcome this difficulty is the strong law of large numbers (see Theorem 4.7). We observe that for $d \in \mathcal{D}$, we have

$$\langle f, d \rangle_{\mathrm{L}^2(\mathcal{M})} = \int_{\mathcal{M}} d(x)\, f(x)\, \mathrm{d}x = \mathbb{E}(d(X))$$

such that, given the observations $x_1, \ldots, x_N$, we can approximate the inner product by

$$\langle f, d \rangle_{\mathrm{L}^2(\mathcal{M})} = \mathbb{E}(d(X)) \approx \frac{1}{N} \sum_{n=1}^{N} d(x_n). \tag{5.2}$$

Inserting this approximation into the PGA yields the following algorithm.

Algorithm 5.1 (Greedy algorithm for the estimation of PDFs)**.** Let $f \in \mathrm{L}^2(\mathcal{M})$ be an unknown probability density function, let $X_1, \ldots, X_N \sim f$ be $\mathcal{M}$-valued random variables, and let $x_1, \ldots, x_N \in \mathcal{M}$ be the corresponding observations. Let a dictionary $\mathcal{D} \subseteq \{\, g \in \mathrm{L}^2(\mathcal{M}) \mid \|g\|_{\mathrm{L}^2(\mathcal{M})} = 1 \,\}$ be given.

Generate a sequence $(f_k)_{k \in \mathbb{N}_0}$ of approximations of f iteratively according to the following scheme:

1. Set $k := 0$ and choose an initial approximation $f_0 \in \mathrm{L}^2(\mathscr{M})$.
2. Find a dictionary element $d_{k+1} \in \mathscr{D}$ fulfilling the maximization property

$$d_{k+1} = \operatorname*{argmax}_{d \in \mathscr{D}} \left| \frac{1}{N} \sum_{n=1}^{N} d(x_n) - \langle f_k, d \rangle_{\mathrm{L}^2(\mathscr{M})} \right|. \tag{5.3}$$

3. Set

$$\alpha_{k+1} := \frac{1}{N} \sum_{n=1}^{N} d_{k+1}(x_n) - \langle f_k, d_{k+1} \rangle_{\mathrm{L}^2(\mathscr{M})}. \tag{5.4}$$

4. Set $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$.
5. If a suitable stopping criterion is fulfilled: stop.
   Otherwise: increase k by 1 and return to step 2 of the algorithm.

We will note multiple times in this thesis that in theory we would expect a condition on the dictionary like $\operatorname{span} \mathscr{D} = \mathrm{L}^2(\mathscr{M})$. However, from the computational point of view, this condition can never be fulfilled, since $\mathrm{L}^2(\mathscr{M})$ is infinite-dimensional and thus, also $\mathscr{D}$ would have to be comprised of infinitely many elements and this cannot be realized on computers. For a finite dictionary, the maximization in Eq. (5.3) can be carried out by evaluating the term that is to be maximized for every dictionary element and by choosing the dictionary element with the maximal value. Note also that the maximizer in Eq. (5.3) does not need to be unique. In this case, we choose one arbitrary maximizer among all available maximizers.

Note furthermore that it is still a challenge to find a suitable stopping criterion for Algorithm 5.1. A first idea is to stop the algorithm if the error does no longer exceed a prescribed bound. In the kth iteration of the algorithm, we can compute the current error as

$$\|f - f_k\|^2_{\mathrm{L}^2(\mathscr{M})} = \|f\|^2_{\mathrm{L}^2(\mathscr{M})} - 2\langle f, f_k \rangle_{\mathrm{L}^2(\mathscr{M})} + \|f_k\|^2_{\mathrm{L}^2(\mathscr{M})}.$$

The latter term can be computed exactly, since f_k is a linear combination of f_0 and of dictionary elements. The penultimate term could be approximated by the strong law of large numbers like it was done in Eq. (5.2). Unfortunately, the first term on the right-hand side of the equation cannot be approximated in any way, since the PDF f is typically unknown. Hence, it is not possible to use the error as a stopping criterion. Another idea would be to use the absolute value $|\alpha_k|$ of the coefficients as a stopping criterion. In Figure 5.8 in Section 5.4.4, we will see that it cannot be recommended to do so. Due to the stochastic nature of the problem, these

coefficients are also equipped with a certain amount of noise. Stopping the iteration as soon as the absolute value of a coefficients α_k falls below a certain threshold may consequently stop the iteration too early. In conclusion, it is very difficult to find a stopping criterion for the iteration in Algorithm 5.1. In our numerical experiments, we will simply stop after a certain fixed number of iterations, since there is currently no better alternative.

We return to the spherical case, that is, $\mathcal{M} = \mathbb{S}^{d-1}$, and remember Definition 4.14 of a kernel density estimator in Section 4.2 for a given observed data set $\xi_1, \dots, \xi_N \in \mathbb{S}^{d-1}$. An obvious choice for a dictionary, which we will also use in the upcoming application, is to put all the kernels that arise in the kernel density estimator into the dictionary, that is,

$$\mathcal{D} := \{\, \xi \mapsto k_h(\xi_n \cdot \xi) \mid n = 1, \dots, N \,\}, \tag{5.5}$$

for a given kernel $k_h \colon [-1, 1] \to [0, \infty)$. The output of Algorithm 5.1 will then be a linear combination of the kernels that also arise in the KDE. Different to the KDE, the kernels are not combined with a weight of $\frac{1}{N}$ each, but instead, more or less arbitrary weights α are assigned to the kernels centered at the data points. Depending on the stopping index, several of the kernels will not be chosen for the approximation of the PDF at all. In the following, we will call a function

$$\xi \mapsto \sum_{n=1}^{N} \beta_n \, k_h(\xi_n \cdot \xi),$$

which might be generated by Algorithm 5.1, and where most of the coefficients $(\beta_n)_{n=1,\dots,N}$ are zero, a *sparse kernel density estimator*, although technically it is no longer a kernel density estimator in the strict sense.

The motivation for applying greedy algorithms to estimate a PDF was that KDEs may lead to a very high computational effort when they are used inside a sampling method. Of course, after introducing the greedy algorithm, the question arises if this disadvantage of KDEs can be overcome by the application of Algorithm 5.1. If the dictionary is given as in Eq. (5.5), in a naive implementation of the algorithm the kernel k_h has to be evaluated N^2 times (namely, in every inner product $x_r \cdot x_s$, $r, s = 1, \dots, N$) in Eq. (5.3). Performing K iterations, this leads to $K N^2$ evaluations of the kernel. Due to the symmetry of the inner product, this can be easily reduced to $(N + 1) N/2$ evaluations, yielding $K N (N + 1)/2$ evaluations in total. But still, at first sight, we have to evaluate the kernel very often to obtain a sparse estimation of the PDF such that there seems to be no real advantage over the use of KDE.

However, looking at the algorithm more thoroughly, there is much space for optimization. The empirical expectation value

$$\frac{1}{N}\sum_{n=1}^{N} d(x_n), \tag{5.6}$$

which is present in Eq. (5.3), does not depend on the iteration index k. Thus, it can be computed once before starting the iteration and stored to be used in every step of the iteration. Consequently, the computational effort for K iterations of the greedy algorithm in total reduces from KN^2 evaluations of the kernel to $N(N+1)/2$ evaluations. Note that the values computed in Eq. (5.6) can also be used in Eq. (5.4) such that this generates no additional computational effort.

Furthermore, it is clear that the greedy algorithm has to be applied only once per data set to obtain a sparse estimation of the PDF. The pairs $(\alpha_k, d_k) \in \mathbb{R} \times \mathcal{D}$ can be stored and reused every time the estimation is needed, for example, when using a sampling method. On the other hand, when using a KDE in the sampling method, the kernel has to be evaluated multiple times for the generation of one single sample. This clearly yields a higher computational effort if many samples are needed. The fact that the greedy algorithm needs only to be applied once makes it possible to execute this algorithm on a powerful computer system to obtain a list of chosen coefficients and kernels, and subsequently use a slower computer, for example, an ordinary desktop computer, for sampling.

In principle, the dictionary can consist of many more types of functions other then kernels. For example, orthogonal polynomials or wavelet basis functions could be used. On the real line also a spline basis could be used (on the sphere this coincides with kernels). Since, in general, these functions are not non-negative, it cannot be guaranteed that the approximation, which is generated by the greedy algorithm, is a PDF, since it may attains negative values. Of course, including only non-negative dictionary functions does not guarantee a non-negative approximation since the algorithm can choose negative coefficients. Nevertheless, this is one of the reasons why we restrict ourselves to kernels in the dictionary. Furthermore, one could add kernels with different concentration parameters h to the dictionary, which would eliminate the necessity for a parameter choice method. However, we expect that this could lead to problems with overfitting if kernels are included in the dictionary whose concentration parameter is too close to 0 (in the real-valued case) or 1 (in the spherical case) are present in the dictionary. Thus, we would only shift the burden of the choice of one bandwidth to the choice of an appropriate bound for all of the bandwidths of the kernels that are included in the dictionary. This is why we also do not pursue this approach. Moreover, by including other functions than kernels, we would lose the comparability to kernel density estimators, which is also a desirable

property since these are a very well-established method in statistics. Furthermore, we would ignore that kernels are also approximate identities, which also affects the interpretability of the resulting estimator. Finally, dictionaries consisting of kernels have been successfully used in other greedy algorithms, for example, when the *Regularized Functional Matching Pursuit* and the *Regularized Orthogonal Functional Matching Pursuit* has been applied to spherical inverse problems in Berkel et al. [18], Fischer [45], Fischer and Michel [46–48], Michel [123], Michel and Telschow [127, 128], and Telschow [166], as well as in the *kernel matching pursuit* in Vincent and Bengio [179] in the field of machine learning.

5.3. A synthetic example

In this section, we apply Algorithm 5.1 to synthetic data, in particular, points on the sphere $\mathbb{S}^2$. For the Pure Greedy Algorithm, it is well known that it converges. We derived Algorithm 5.1 by using an approximation by the strong law of large numbers in Eq. (5.2) instead of the exact value that is used in the PGA. Thus, the convergence results for the PGA cannot be applied to the presented algorithm. This is the reason why we will empirically study the convergence of the algorithm using this synthetic example.

As a kernel on the sphere, which we put into our dictionary, we choose the Abel-Poisson kernel, which has already been introduced in Definition 2.27, and which is given by

$$Q_h(t) := \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} h^n P_n(t) = \frac{1}{4\pi} \frac{1-h^2}{(1+h^2-2ht)^{3/2}}. \tag{5.7}$$

We choose the Abel-Poisson kernel for several reasons. First, from the computational point of view, the Abel-Poisson kernel possesses several advantages over other kernels on the sphere. In Theorem 2.36, we have already seen that every zonal function on the sphere has a representation as a Legendre series, as long as it is in $\mathrm{L}^2(\mathbb{S}^2)$. Since we assumed in Definition 4.13 that kernels on the sphere are zonal functions, this is also true in this application. However, many kernels on the sphere (e. g., the *Gauß-Weierstraß* kernel or the *Shannon kernel*) *only* possess a representation as a Legendre series, but there is no closed representation. In consequence, when using such kernels in practice, one always has to truncate the Legendre series and probably has a high number of summands. This may lead to a very high computational effort due to the high number of kernel evaluations that is needed. The Abel-Poisson kernel, however, possesses the closed representation in Eq. (5.7), which makes it much cheaper to evaluate in the implementation.

Secondly, we have already mentioned the similarity between families of kernels for KDEs and approximate identities in Section 4.2 due to the similarity of the Dirac measure to the Dirac δ-distribution. The family of Abel-Poisson kernels indeed is a Dirac family in $\mathrm{L}^2(\mathbb{S}^2)$ (cf. Freeden et al. [53, Chapter 8], Freeden and Schreiner [57], Freeden and Schreiner [58, Chapter 7], Freeden and Windheuser [59], and Michel [122, Chapter 7]) and thus it is also an approximate identity. This is the actual reason why we will use the Abel-Poisson kernel in the following. As already mentioned in Section 4.2, we do not deal with the problem of bandwidth selection for the Abel-Poisson kernel here. We choose $h = 0.9$ for the Abel-Poisson kernels that we put into the dictionary, since the results seemed to be reasonably good.

To create a synthetic example to test the convergence of the greedy algorithm, we choose the PDF

$$f(\xi) := Q_{0.6}(\xi \cdot \varepsilon^3), \qquad \xi \in \mathbb{S}^2,$$

such that the Abel-Poisson kernel is concentrated around the North Pole. From this PDF, we sampled a data set $\xi_1, \ldots, \xi_N$ of the i. i. d. random variables $X_1, \ldots, X_N \sim f$ by applying the rejection sampling method, which was described in Section 4.3.2. Here, we chose to sample $N = 10^6$ points on the sphere. The density function and a reduced data set with 2000 points can be found in Figure 5.1. This selection was performed by choosing the 2000 points at random to overcome the difficulty that the data set may be sorted in some way such that simply taking the first 2000 data points would not create a representative figure.

Starting with $f_0 = 0$, we performed 10 000 iterations of the greedy algorithm to obtain an approximation of the original PDF. In this synthetic case, it is possible to compute the relative $\mathrm{L}^2(\mathbb{S}^2)$-error $\|f - f_k\|_{\mathrm{L}^2(\mathbb{S}^2)} / \|f\|_{\mathrm{L}^2(\mathbb{S}^2)}$ explicitly, since

$$\langle Q_h(\eta \cdot), Q_{\bar{h}}(\zeta \cdot) \rangle_{\mathrm{L}^2(\mathbb{S}^2)} = Q_{h\bar{h}}(\eta \cdot \zeta)$$

(see Example 2.37). A semi-logarithmic plot of this relative error in the conducted 10 000 iterations can be found in Figure 5.2.

For comparison with the upcoming real-world example, where we cannot compute the approximation error, we also provide a plot of the absolute value of coefficients $|\alpha_k|$ during the iteration in Figure 5.3.

In both figures, we observe a convergent behavior of the algorithm in these first 10 000 iterations. It is natural that the error does not tend to zero if the number of data points is fixed. This is due to the approximation in Eq. (5.2). Multiple performed numerical experiments show that the number of data points influences the convergence of the algorithm. The lower the number of data points is, the

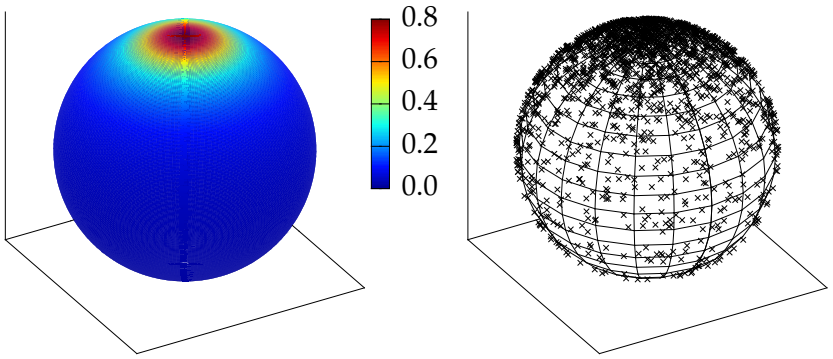


Figure 5.1.: Left: PDF from the synthetic example, an Abel-Poisson kernel with parameter $h = 0.6$, centered at the North Pole.
Right: 2000 of in total 10^6 data points sampled from this PDF using rejection sampling.

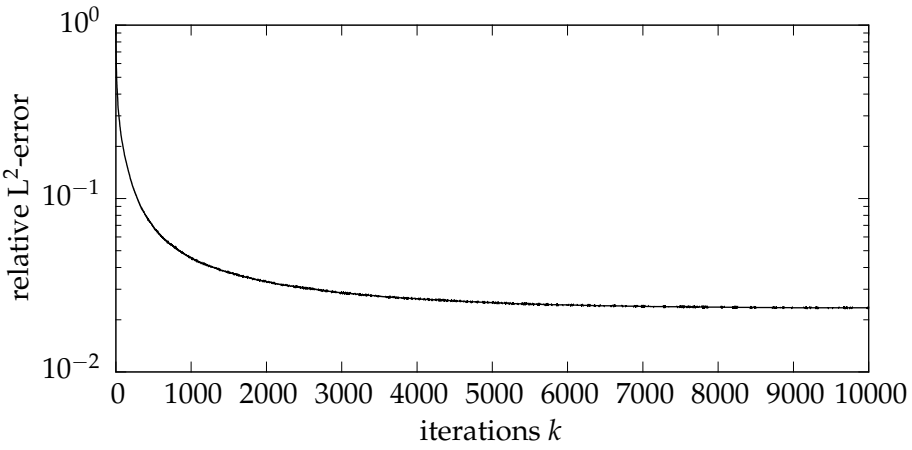


Figure 5.2.: Semi-logarithmic plot of the relative $L^2(\mathbb{S}^2)$-error $\|f - f_k\|_{L^2(\mathbb{S}^2)} / \|f\|_{L^2(\mathbb{S}^2)}$ in the first 10 000 iterations of the greedy algorithm, when it is applied to the synthetic example.

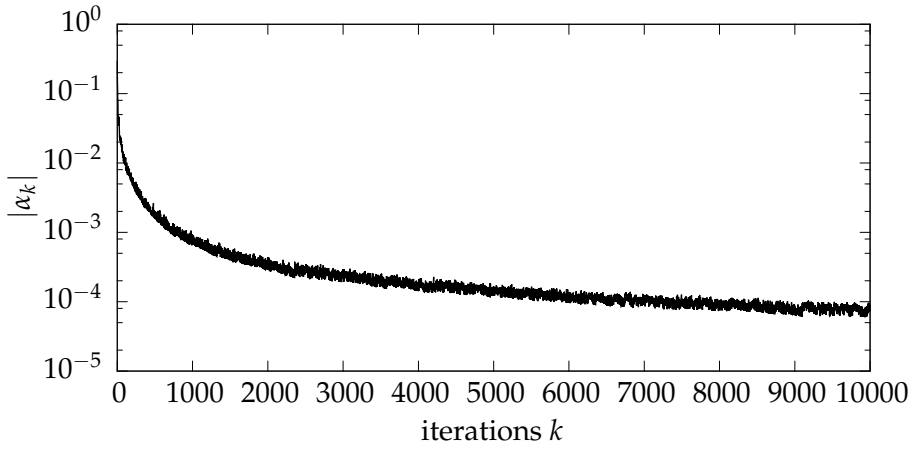


Figure 5.3.: Semi-logarithmic plot of the absolute values of the chosen coefficients in the iterations of the greedy algorithm, when it is applied to the synthetic data set.

earlier the error stops to tend to zero, but instead remains static or even rises again. This is plausible, since, having reached a certain level of accuracy, the approximation can only be improved with additional information on the function to be approximated. In this case, *more information* means *more data points*. The dependence of the convergence on the number of data points is a possible subject of future research.

5.4. An application: analysis and simulation of nonwoven fabrics

Finally, we will conduct the analysis and simulation of so-called nonwoven fabrics, using the greedy algorithm for the estimation of probability density functions, which was presented in Section 5.2.

For this purpose, we will first describe the characteristics of nonwoven fabrics and their production process as well as the need for its simulation. Then, we will discuss existing simulation algorithms and why there is a need for a new simulation algorithm. The simulation algorithm, which we present, will be based on a data set generated by a CT scanner. The data set represents fiber directions inside a real sample of a nonwoven fabric. After we have described this data set, we will present the novel simulation algorithm, where there is a need to estimate the PDF of the fiber directions. Thus, we apply both the greedy algorithm and the simulation algorithm to the given data set. Note again that parts of the following considerations have already been published in Gramsch et al. [66].

5.4.1. Nonwoven fabrics

A *nonwoven* (or *nonwoven fabric*) is defined as "...a sheet of fibres, continuous filaments, or chopped yarns of any nature or origin, that have been formed into a web by any means, and bonded together by any means, with the exception of weaving or knitting" (cited in Mao and Russel [113]) according to EDANA, the *European Disposables and Nonwovens Association*. Nonwovens are widely used in products of daily life, since they can be absorbent, antistatic, breathable, conductive or nonconductive, elastic, flame resistant, impermeable or permeable, smooth, and stiff to name but a few. These very diverse properties can be achieved by combining the right raw materials with specific production processes.

This versatility also has a drawback. Since the market of nonwovens demands more and more customized products, the development cycles have to be shortened. Thus, the simulation of nonwoven production processes is a mathematical key technology for the engineers to modify the processes according to customer-specific needs.

Therefore, we discuss some already existing simulation algorithms in the following section.

Mainly, there are three different nonwoven production processes: dry-lay processes, wet-lay processes, and extrusion processes. In the following, we concentrate on a typical extrusion process, the so-called spunbond process. A sketch of this process can be found in Figure 5.4. In the spunbond production process, a polymer melt is extruded through spinnerets. The evolving fibers are cooled and stretched by air and are swirled around by turbulent air streams until they are laid down on a perforated conveyor belt. By suction, they are fixed to the belt and form a random web. Finally, several post-processing steps like bonding and finishing are implemented to produce a nonwoven fabric. More details on the production process can, for example, be found in Albrecht et al. [5].

5.4.2. Existing simulation algorithms

Several papers have already dealt with the mathematical modeling of nonwoven production processes. For example, a first principles physics model has been developed in Klar et al. [100] and Wegener et al. [184] and this model is also used in the simulation software FIDYST (fiber dynamics simulation tool, see Gramsch et al. [65]). In principle, the model would have to deal with a two-way coupling problem of the aerodynamic forces and the fiber dynamics, which is not solvable to industrial scales due to the required resolution of the mesh (cf. Marheineke and

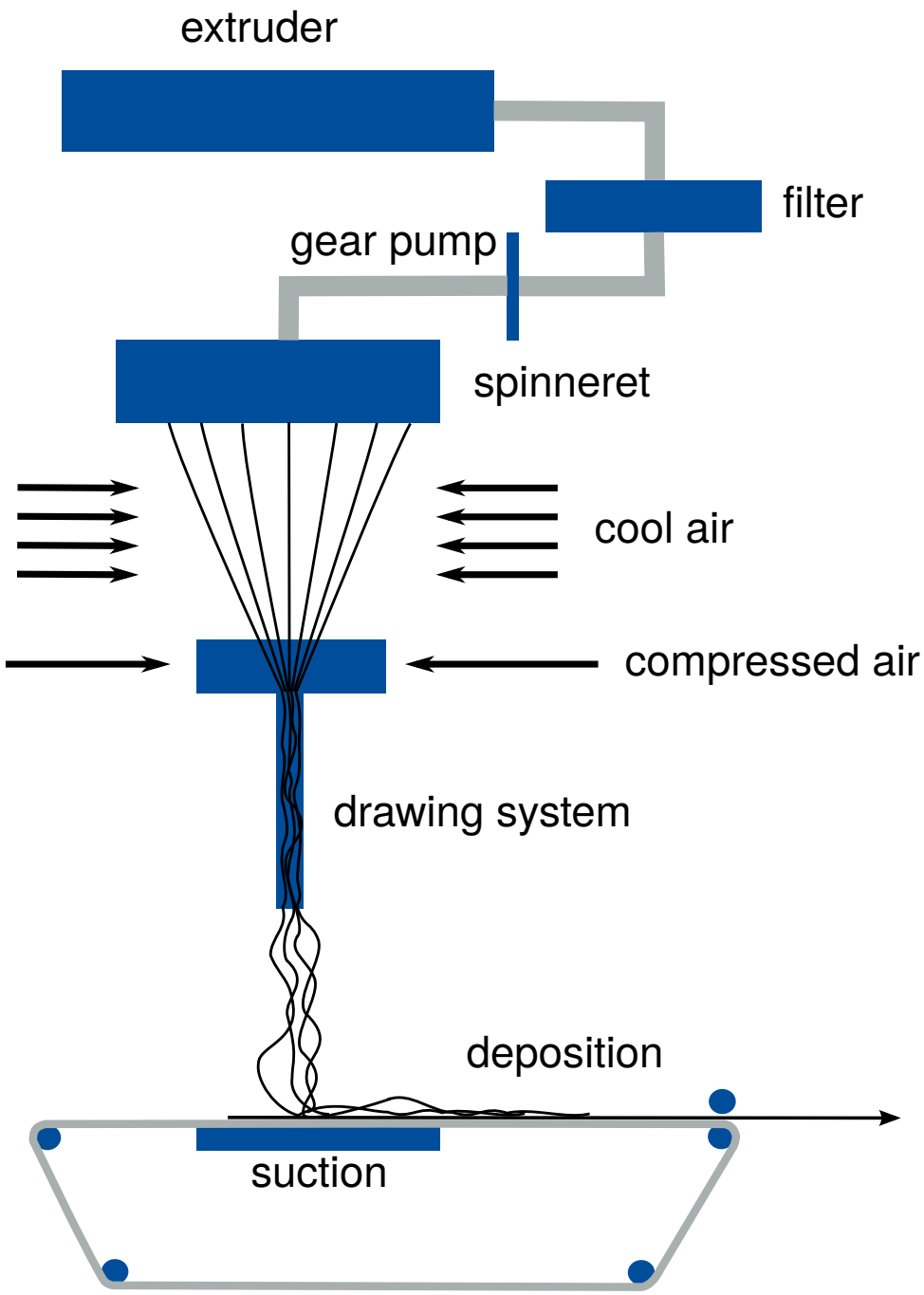


Figure 5.4.: Sketch of the spunbond process.
(Figure provided by S. Gramsch, Fraunhofer ITWM, Kaiserslautern)

Wegener [114]). Therefore, the model only incorporates a one-way coupling by the application of an air drag model. Up to now, interactions of the filaments with itself or with other filaments, which have already been laid down on the conveyor belt, are neglected. Consequently, the simulated laydown of the filaments on the belt is only two-dimensional.

A two-dimensional simulated fabric represents typical nonwovens, which are very flat, in an excellent way. However, several properties of nonwovens originate in the fact that the fibers are lying on top of each other. Thus, a three-dimensional simulation of the fiber laydown is desirable.

It seems to be unrealistic to simulate a real nonwoven, which consists of thousands of fibers, in reasonable time due to the high computational effort that is needed with the current first principles model. Hence, a different approach was developed in Klar et al. [101]. There, the three-dimensional laydown is described by a surrogate model based on a stochastic differential equation. The parameters for this model are estimated by combining two-dimensional data obtained by a simulation with the first principles model mentioned above and three-dimensional data obtained by CT scans (cf. Grothaus et al. [68]). In contrast, we will focus on a surrogate model that uses these three-dimensional data from CT scans, only. Therefore, in the following, we will briefly describe the used CT data set.

5.4.3. Description of the CT data

The industrial company Oerlikon Neumag has provided real samples of nonwovens that were produced by a spunbond process. The samples are made from polypropylene fibers, which have a diameter of $1.2 \times 10^{-5}\,\mathrm{m}$ and a density of $0.9\,\mathrm{g\,m^{-3}}$. The area density of the sample is $12.8\,\mathrm{g\,m^{-2}}$. The samples were analyzed by the department *Image Processing* of the Fraunhofer ITWM with its 3D-microtomography scanner, which delivers a real-valued third-order tensor with so-called gray values. These gray values are mapped to local fiber orientations at each voxel using an eigenvalue analysis of the Hessian matrix of the second partial derivatives of the gray values. For more details, the reader is referred to Redenbach et al. [148]. The microtomography scanner possesses a resolution of 1 µm to 70 µm while the sample size is $1\,\mathrm{mm}^3$ to $10\,\mathrm{mm}^3$. For the data set that we use, this high resolution leads to $N = 9\,600\,558$ points on $\mathbb{S}^2$, which will be the basis for our considerations. A reduced version of the data set with 2000 points can be found in Figure 5.5. Again, the reduction has been performed by a random selection for the reason that we have already given, when the synthetic data set was presented. To get an impression of the distribution of fiber directions on the sphere, we have also computed the KDE for the data set, where the kernel is the Abel-Poisson kernel with parameter $h = 0.9$.

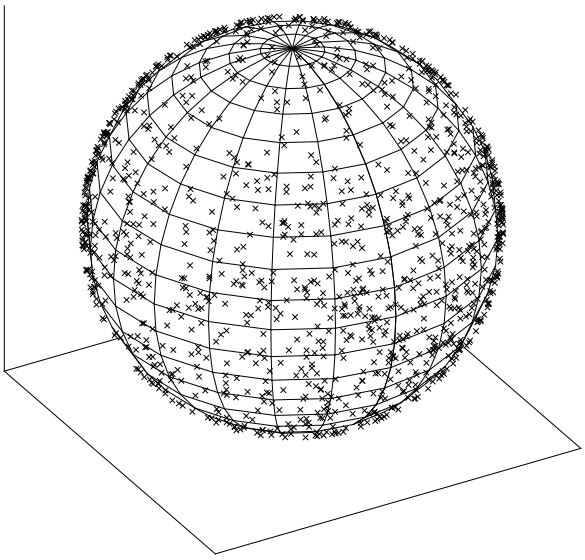


Figure 5.5.: 2000 out of 9 600 558 points from the CT data set of directions in a real nonwoven fabric.

A plot of the KDE as a function on the sphere can be found in Figure 5.6. The figure shows that there exists a dominant direction of the fibers, which corresponds to the direction in which the belt moves in the production process.

5.4.4. Application of the greedy algorithm to the CT data

We applied the greedy algorithm (Algorithm 5.1) to the CT data set that has been described in the previous section. We chose to perform 10 000 iterations of the algorithm and the result is a sparse kernel density estimator, which is depicted in Figure 5.7. Since, in contrast to the synthetic example in Section 5.3, the error cannot be explicitly calculated, in Figure 5.8 the absolute values of the chosen coefficients $|\alpha_k|$ are plotted to get an impression of the size of corrections that the greedy algorithm performs throughout the 10 000 iterations. This indicator for the size of corrections looks very similar to the plot of the coefficients of the synthetic example in Figure 5.3 in Section 5.3.

We observe that the KDE in Figure 5.6 and the sparse KDE in Figure 5.7 show a clear qualitative similarity. Quantitatively, one recognizes certain differences both in the value of the estimator (see the color bars that range up to 0.18 in contrast to 0.22) and in the structure of the estimator (the structures seem to be rougher in the result of the greedy approximation). However, this is no contradiction, since both the KDE and

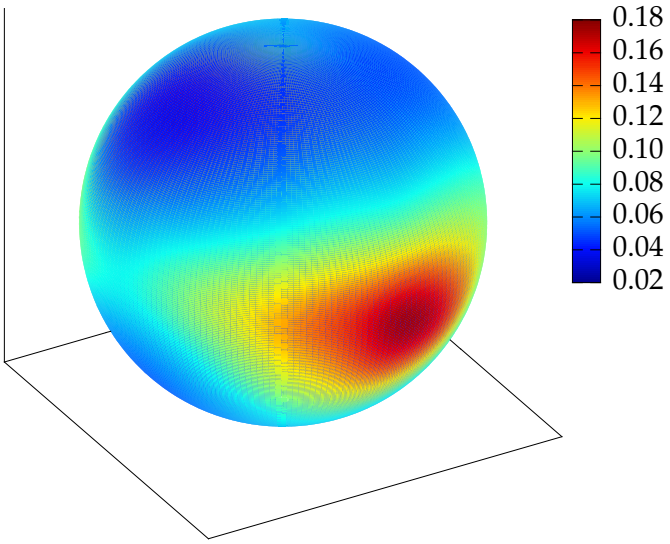


Figure 5.6.: Plot of the KDE obtained from the CT data set where the Abel-Poisson kernel with parameter $h = 0.9$ was used.

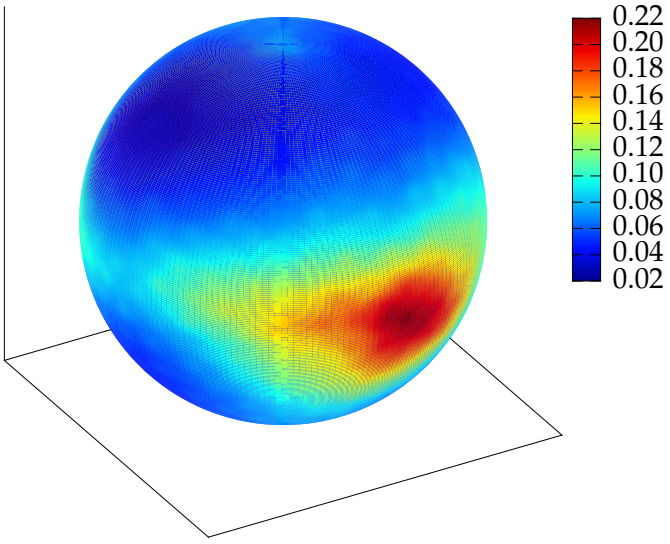


Figure 5.7.: Sparse kernel density estimation of the PDF of a real CT data set after 10 000 iterations of the greedy algorithm.

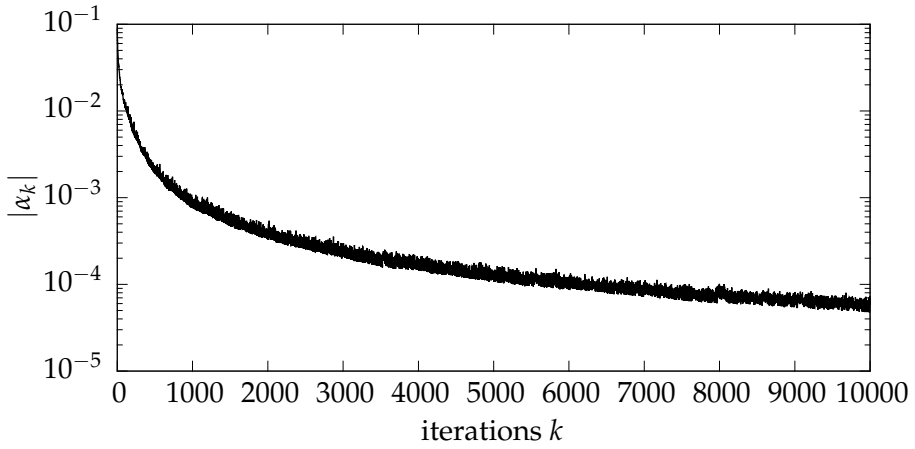


Figure 5.8.: Absolute value of the chosen coefficients in the iterations of the greedy algorithm, when it is applied to the real CT data set.

the sparse KDE are estimates of the same probability density function, but the greedy approximation is *not* an approximation of the KDE. Thus, depending on the choice of the kernels, the dictionary, and the corresponding parameters, both estimators can be different although they estimate the same PDF.

5.4.5. A simple simulation algorithm and numerical results

To conclude this chapter, we will present a simple simulation algorithm for nonwoven fabrics, which yields a three-dimensional simulated fiber. As already indicated, we want to use an estimation $\hat{f}$ of the PDF f, generated by a KDE or the greedy algorithm. The concept of the algorithm is to start at an arbitrary point in $\mathbb{R}^3$ and to choose the next point of a fiber by sampling a direction from the estimated PDF and moving into this direction by a given step width. By construction, the resulting fiber approximately has the same distribution of fiber directions as the real nonwoven from which the data set was generated. The resulting algorithm looks as follows.

Algorithm 5.2. Let an estimate $\hat{f}$ of the PDF f as well as a discretization parameter $s \in (0, \infty)$ be given. Generate a discretization $(Z_j)_{j=0,1,2,\ldots} \subseteq \mathbb{R}^3$ of a fiber by the following iteration:

1. Set $j := 0$ and choose an initial point $Z_0 \in \mathbb{R}^3$.
2. Sample a direction $\eta_{j+1} \in \mathbb{S}^2$ from the estimated PDF $\hat{f}$.
3. Set $Z_{j+1} := Z_j + s\,\eta_{j+1}$.

4. If the desired number of points is reached: stop.
   Otherwise: increase j by 1 and return to step 2.

Note that this algorithm looks similar to the stochastic time discrete approximation of stochastic differential equations, in particular the Euler-Maruyama method, where Gaußian pseudo-random numbers are used in a similar way if the stochastic differential equation involves Brownian motion (cf. Kloeden and Platen [102, Chapter 9]). However, it is not trivial if there is an interpretation of our algorithm as an Euler-Maruyama scheme for a certain stochastic differential equation and even if it was such a scheme, it is completely unclear what such a stochastic equation or a stochastic process solving this equation would look like. We further want to remark that we do not consider the movement of the conveyor belt explicitly in this simulation algorithm like the methods in Klar et al. [100] do. Instead, the belt movement is implicitly contained in the CT data set since, of course, in the production of the real nonwoven sample the belt moved with a certain velocity. This could already be seen in Figure 5.6 and Figure 5.7, where we observed a dominant direction of the fiber directions.

To perform step 2 of Algorithm 5.2, we use the method of rejection sampling that was presented in Section 4.3.2. A discretized fiber that has been obtained by the application of the simulation algorithm using the estimated PDF generated by the greedy algorithm is shown in Figure 5.9.

We have already mentioned two computational aspects of sampling with the use of kernel density estimators. First, the motivation for the development of the greedy algorithm in Section 5.2 was that using a KDE in rejection sampling leads to a very high number of evaluations of the kernel. This is in particular the case for the given data set, where the number of samples is $N = 9\,600\,558$. Secondly, we discussed the importance of a preferably low upper bound of the estimate of the PDF for rejection sampling in Section 4.3.3.

In consequence, we performed a numerical study and the results can be found in Table 5.1. In this table, we compare the use of the KDE and the sparse KDE as well as the application of two different upper bounds for the PDF (using the triangle inequality and evaluations on a grid, cf. Section 4.3.3). The following raw data are listed in the table: the number of samples that have been generated, the number of evaluations of the estimated PDF that were needed to generate these samples, and the CPU time that was consumed. Note that there is a difference in the number of samples between the four presented scenarios due to the high computational effort that would be needed for the more inefficient methods. For that reason, to achieve a better comparison, we add the average number of evaluations of the estimated PDF per sample, the average CPU time per sample, and an extrapolation of the time that

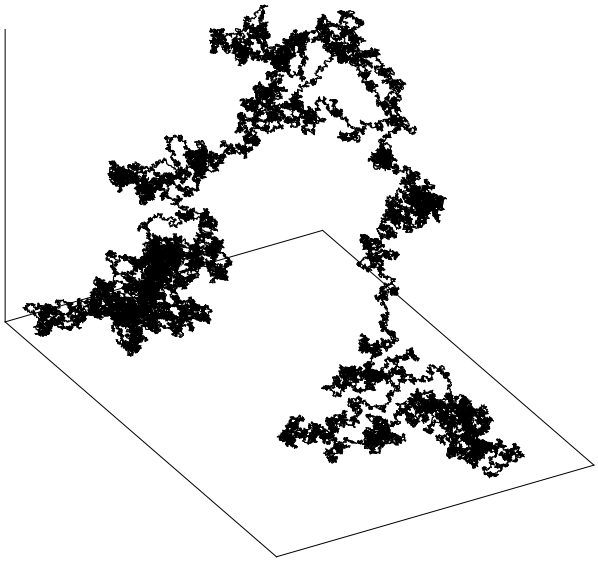


Figure 5.9.: Fiber, consisting of 100 000 points, simulated by Algorithm 5.2 in conjunction with a greedy approximation of the PDF.

Table 5.1.: Computation times for different estimates of the PDF and different upper bounds for the estimated PDF. For the comparison "CPU time/nonwoven", it has been assumed that a simulated nonwoven consists of 100 fibers with 100 000 line segments each. The presented data show an enormous saving of computation time when using the newly developed greedy algorithm.

	KDE		Greedy approximation	
upper bound	triangle ineq.	evaluation	triangle ineq.	evaluation
samples	500	10 000	10 000	100 000
evaluations	91 356	23 639	1 881 041	303 081
CPU time	26 690 s	6902.6 s	566.51 s	91.517 s
eval./sample	183	2.36	188	3.03
CPU time/sample	5.33×10^{1} s	6.90×10^{-1} s	5.66×10^{-2} s	9.15×10^{-4} s
CPU time/nonwoven	148 283 h	1917 h	157 h	2.50 h

would be needed to simulate a nonwoven fabric with 100 fibers with 100 000 line segments each, to the table.

One can see that the use of the greedy approximation makes Algorithm 5.2 more efficient by multiple orders of magnitude. In the more efficient case of determining an upper bound by the evaluation on a fine grid, the simulation time for a nonwoven with 100 fibers with 100 000 line segments each, is reduced from 1917 h to 2.5 h, thus from nearly 80 days to 150 minutes, a factor of around 750.

In conclusion, applying a greedy strategy to the problem of estimating a probability density function, makes it possible to apply a simple simulation algorithm for nonwoven fabrics much more efficiently than by using the standard approach of kernel density estimators.

Part III.

Greedy algorithms for inverse problems

Chapter 6.

Theory of inverse problems

According to Keller [96], two problems are called *inverses* of each other "if the formulation of each involves all or part of the solution of the other". Although this is, consequently, a symmetric term, one of the two problems is commonly called the *direct problem* and the other is called the corresponding *inverse problem*. The reason for this is that the direct problem often is an established well-understood problem of applied mathematics, which is—more or less—easy to solve. In contrast, the inverse problem is often newer and, for example, has arisen in the development of new measuring equipment in the applications. Moreover, the inverse problem frequently is more difficult to solve. Characterizations of the facts, *why* inverse problems are a lot more difficult to solve, will be given in this chapter.

A different characterization of the *direct* and the *inverse problem* is based on physics (cf. Richter [152, Preface]). Here, it is assumed that there exists a set $\mathcal{X}$ of *causes* and a set $\mathcal{Y}$ of *effects*, which are connected by a mapping

$$\mathcal{S}\colon \mathcal{X} \to \mathcal{Y},$$

modeling a physical law. In this context, the direct problem is

given the cause $f \in \mathcal{X}$, compute the effect $\mathcal{S}[f] \in \mathcal{Y}$,

whereas the inverse problem is

given the effect $g \in \mathcal{Y}$, find a cause $f \in \mathcal{X}$ such that $\mathcal{S}[f] = g$.

This variant can easily be transferred to a functional analytic formulation, where $\mathcal{X}, \mathcal{Y}$ are function spaces and $\mathcal{S}$ is an operator between these spaces. Note that we stick to the convention that we have already introduced in Section 2.2, where linear operators are denoted by $\mathcal{T}$ and general operators (including nonlinear ones) are denoted by $\mathcal{S}$. Additionally, regular or no parentheses are used for linear operators and square brackets are used for nonlinear operators.

In this work, we will restrict ourselves to Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. The study of inverse problems in more general spaces, in particular Banach spaces, is also an

established approach in the analysis of inverse problems (cf. Schuster et al. [160]). The restriction to Hilbert spaces leads to the following notion of an inverse problem.

Problem 6.1 (General inverse problem)**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and let the operator $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be continuous. Find $f^* \in \mathcal{X}$ such that

$$\mathcal{S}[f^*] = g \tag{6.1}$$

for given data $g \in \mathcal{Y}$.

Note that we require at least some kind of stability of the direct problem, which is characterized by the continuity of the operator $\mathcal{S}$. In consequence, we exclude, for example, chaotic systems, where the effects are very sensitive to small changes in the causes, from our considerations.

An important term in this context is the well-posedness and ill-posedness of inverse problems. It goes back to Hadamard [73] and can be formulated as follows.

Definition 6.2 (Well-posedness (Hadamard), cf. Engl et al. [42, Chapter 2])**.** Problem 6.1 is said to be *well-posed* in the sense of Hadamard if the following three conditions are fulfilled:

(W1) Eq. (6.1) possesses a solution $f^* \in \mathcal{X}$ for all data $g \in \mathcal{Y}$,
(W2) the solution of Eq. (6.1) is unique for all data $g \in \mathcal{Y}$,
(W3) the solution $f^* \in \mathcal{X}$ of Eq. (6.1) depends continuously on the data $g \in \mathcal{Y}$.

If one of the conditions (W1)–(W3) is violated, Problem 6.1 is called *ill-posed*.

In terms of the continuous operator $\mathcal{S}$, this can be rephrased as follows.

(W1′) $\operatorname{ran} \mathcal{S} = \mathcal{Y}$ (*surjectivity*),
(W2′) $\mathcal{S}[f] = \mathcal{S}[f']$ implies $f = f'$ (*injectivity*),
(W3′) $\mathcal{S}^{-1}$ is continuous (*stability of inversion*).

Note that, in general, all measured data are equipped with noise. Thus, apart from the non-existence and the non-uniqueness of the solution, the instability of the inversion is a crucial difficulty. Due to the discontinuity of the inverse operator, small changes in the data (like the measurement noise) lead to large changes in the obtained solution, which is not desirable from the physical point of view.

The following section deals with the topic of ill-posedness for *linear* operators, whereas we discuss *nonlinear* inverse problems in Section 6.2.

6.1. Linear inverse problems

This section is based on Engl et al. [42, Chapter 2]. Here, $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ will always denote a linear and bounded operator between Hilbert spaces $\mathcal{X}, \mathcal{Y}$. Consequently, the problem to solve is the following one.

Problem 6.3 (Linear inverse problem)**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ a linear and bounded operator. Find $f^* \in \mathcal{X}$ such that

$$\mathcal{T} f^* = g \tag{6.2}$$

for given data $g \in \mathcal{Y}$.

We remark that for linear operators boundedness and continuity are equivalent. Furthermore, in the characterization of well-posedness, (W2′) can be replaced by

(W2″) $\operatorname{null} \mathcal{T} = \{ 0 \}$,

since injectivity of a linear operator is equivalent to having a trivial null space.

Introducing the notion of a *generalized inverse*, one can reduce (W1′)–(W3′) to the stability of the inversion. This will be discussed in the subsequent section.

6.1.1. The Moore-Penrose generalized inverse

The motivation for the reduction to a generalized solution is that, having noisy data at hand, one can only determine some $f^\times \in \mathcal{X}$, which is a solution to Eq. (6.2) in some approximate way, anyway. Additionally, if $f^\times$ is not unique, one may impose additional conditions on the solution, which yields a unique generalized solution $f^+ \in \mathcal{X}$. In more detail, we will define the terms of a *least squares* and *best approximate solution* in the following definition.

Definition 6.4 (cf. Engl et al. [42, Definition 2.1])**.** Let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a bounded linear operator. We call $f^\times \in \mathcal{X}$ a *least squares solution* of Problem 6.3 if

$$\left\| g - \mathcal{T} f^\times \right\|_{\mathcal{Y}} = \inf_{h \in \mathcal{X}} \| g - \mathcal{T} h \|_{\mathcal{Y}}.$$

We call $f^+ \in \mathcal{X}$ a *best approximate solution* if it is a least squares solution and

$$\left\| f^+ \right\|_{\mathcal{X}} = \inf \{ \, \| h \|_{\mathcal{X}} \mid h \text{ is a least squares solution} \, \},$$

that is, f^+ is a least squares solution whose norm is minimal among all least squares solutions.

The term of a best approximate solution is related to the so-called *Moore-Penrose generalized inverse*, which we will define in the following.

Definition 6.5 (cf. Engl et al. [42, Definition 2.2])**.** Let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a bounded linear operator. Define

$$\tilde{\mathcal{T}} := \mathcal{T}|_{(\operatorname{null}\mathcal{T})^\perp}\colon (\operatorname{null}\mathcal{T})^\perp \to \operatorname{ran}\mathcal{T}.$$

Then, the *Moore-Penrose generalized inverse* $\mathcal{T}^+$ is the unique linear extension of $\tilde{\mathcal{T}}^{-1}$ to the domain

$$\operatorname{dom}\mathcal{T}^+ := \operatorname{ran}\mathcal{T} \oplus (\operatorname{ran}\mathcal{T})^\perp$$

such that

$$\operatorname{null}\mathcal{T}^+ = (\operatorname{ran}\mathcal{T})^\perp.$$

In the following, we will cite some important features of the Moore-Penrose generalized inverse.

Theorem 6.6 (cf. Engl et al. [42, Proposition 2.3])**.** Let $\mathcal{P}_{\operatorname{null}\mathcal{T}}$ and $\mathcal{P}_{\overline{\operatorname{ran}\mathcal{T}}}$ be orthogonal projections to the respective sets. Then we have $\operatorname{ran}\mathcal{T}^+ = (\operatorname{null}\mathcal{T})^\perp$ and the so-called *Moore-Penrose equations*

$$\begin{aligned}
\mathcal{T}\mathcal{T}^+\mathcal{T} &= \mathcal{T}\\
\mathcal{T}^+\mathcal{T}\mathcal{T}^+ &= \mathcal{T}^+\\
\mathcal{T}^+\mathcal{T} &= \mathcal{I}_{\mathcal{X}} - \mathcal{P}_{\operatorname{null}\mathcal{T}}\\
\mathcal{T}\mathcal{T}^+ &= \mathcal{P}_{\overline{\operatorname{ran}\mathcal{T}}}|_{\operatorname{dom}\mathcal{T}^+}
\end{aligned}$$

hold. Furthermore, $\mathcal{T}^+$ is uniquely characterized by these equations.

Theorem 6.7 (cf. Engl et al. [42, Proposition 2.4])**.** The Moore-Penrose generalized inverse $\mathcal{T}^+$ has a closed graph and $\mathcal{T}^+$ is bounded if and only if $\overline{\operatorname{ran}\mathcal{T}} = \operatorname{ran}\mathcal{T}$.

The relation between $\mathcal{T}^+$ and the best approximate solution is given by the following theorem.

Theorem 6.8 (cf. Engl et al. [42, Theorem 2.5])**.** If the data fulfills $g \in \operatorname{dom}\mathcal{T}^+$, the best approximate solution f^+ of Problem 6.3 is uniquely determined and is given by

$$f^+ = \mathcal{T}^+ g$$

such that the set of all least squares solutions is $f^+ + \operatorname{null}\mathcal{T}$.

The preceding theorems are the reason for a different characterization of ill-posedness, which was given by Nashed (cf. Rieder [153, Section 2.1]). It is based on the observation that after introducing a generalized inverse, the only difficulty arises if $\mathcal{T}^+$ is not continuous.

Definition 6.9 (Nashed [135])**.** Problem 6.3 is said to be *well-posed* in the sense of Nashed if $\overline{\operatorname{ran}\mathcal{T}} = \operatorname{ran}\mathcal{T}$.

It is called *ill-posed* in the sense of Nashed if $\overline{\operatorname{ran}\mathcal{T}} \neq \operatorname{ran}\mathcal{T}$.

Analogously to the finite-dimensional case of ordinary least squares regression (cf. the considerations in Section 5.1), least squares solutions can also be characterized by the normal equation.

Theorem 6.10 (Normal equation)**.** Let $g \in \operatorname{dom}\mathcal{T}^+$. The element $f^\times \in \mathcal{X}$ is a least squares solution of Problem 6.3 if and only if

$$\mathcal{T}^*\mathcal{T}f^\times = \mathcal{T}^*g$$

holds, which is called the *normal equation*.

Thus, $f^+ = \mathcal{T}^+g$ is the solution of the normal equation that has the minimal norm. In other words,

$$\mathcal{T}^+ = (\mathcal{T}^*\mathcal{T})^+\mathcal{T}^*$$

holds.

6.1.2. Compact operators and their spectral analysis

Compact linear operators play an important role in the field of inverse problems, since it turns out that Problem 6.3 is *always* ill-posed if the operator $\mathcal{T}$ is compact and the space $\mathcal{X}$ is not finite-dimensional. Additionally, under certain conditions, integral operators are compact operators.

We first give the definition of compact operators.

Definition 6.11 (cf. Kirsch [99, Definition A.31])**.** An operator $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ is called *compact* if every bounded set $S \subseteq \mathcal{X}$ is mapped to a relatively compact set $\mathcal{T}(S) \subseteq \mathcal{Y}$, that is, $\overline{\mathcal{T}(S)}$ is compact.

Next, we define the spectrum of an arbitrary operator.

Definition 6.12 (cf. Kirsch [99, Definition A.49])**.** Let $\mathcal{T}\colon \mathscr{X} \to \mathscr{X}$ be a linear operator. The *spectrum* of $\mathcal{T}$ is defined as the set

$$\sigma(\mathcal{T}) \coloneqq \{\, \lambda \in \mathbb{C} \mid \mathcal{T} - \lambda\mathcal{I} \text{ has no bounded inverse on } \mathscr{X} \,\}\,.$$

Every $\lambda \in \sigma(\mathcal{T})$ is called an *eigenvalue* of $\mathcal{T}$ as long as $\mathcal{T} - \lambda\mathcal{I}$ is not injective. For every eigenvalue λ, all elements $f \in \operatorname{null}(\mathcal{T} - \lambda\mathcal{I}) \setminus \{\, 0 \,\}$ are called *eigenvectors* of $\mathcal{T}$.

The basis for the spectral analysis of compact operators is the following theorem for self-adjoint compact operators.

Theorem 6.13 (cf. Kirsch [99, Theorem A.51])**.** Let $\mathcal{T}\colon \mathscr{X} \to \mathscr{X}$ be compact and self-adjoint (i. e. $\mathcal{T}^* = \mathcal{T}$) and $\mathcal{T} \neq 0$. Then

(a) Every $\lambda \in \sigma(\mathcal{T})$ is an eigenvalue or $\lambda = 0$.

(b) Every $\lambda \in \sigma(\mathcal{T})$ is real.

(c) The operator $\mathcal{T}$ has at least one and up to a countable number of eigenvalues. The only possible accumulation point of $\sigma(\mathcal{T})$ is 0.

(d) For every eigenvalue $\lambda \neq 0$ there exist only finitely many linearly independent eigenvectors. Eigenvectors, which correspond to different eigenvalues, are orthogonal.

(e) If we order the eigenvalues such that

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots$$

and denote the projection onto the eigenspaces corresponding to the eigenvalue λ_j by $\mathcal{P}_j\colon \mathscr{X} \to \operatorname{null}(\mathcal{T} - \lambda_j\mathcal{I})$, then

$$\mathcal{T} = \sum_{j=1}^{J} \lambda_j \mathcal{P}_j$$

if there are only finitely many eigenvalues $\lambda_1, \ldots, \lambda_J$, and

$$\mathcal{T} = \sum_{j=1}^{\infty} \lambda_j \mathcal{P}_j$$

if there is an infinite sequence of eigenvalues, where the convergence is with respect to the operator norm. In the following, we will use the index set $\mathbb{J}$ to account for both cases, that is $\mathbb{J} = \{\, 1, \ldots, J \,\}$ and $\mathbb{J} = \mathbb{N}$, respectively.

This leads to the so-called *singular value decomposition (SVD)* of a (possibly not self-adjoint) compact operator.

Definition 6.14 (cf. Kirsch [99, Definition A.52])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces, let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a compact linear operator, and let $\mathcal{T}^*\colon \mathcal{Y} \to \mathcal{X}$ be its adjoint operator. Denote by λ_j, $j \in \mathbb{J}$, the eigenvalues of the self-adjoint operator $\mathcal{T}^*\mathcal{T}\colon \mathcal{X} \to \mathcal{X}$. Then the numbers

$$\sigma_j = \sqrt{\lambda_j}, \qquad j \in \mathbb{J},$$

are called *singular values* of $\mathcal{T}$.

Theorem 6.15 (cf. Kirsch [99, Theorem A.53])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces, let the operator $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be compact and linear, and let $\mathcal{T}^*\colon \mathcal{Y} \to \mathcal{X}$ be its adjoint operator. Let $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots > 0$ be the ordered sequence of positive singular values of $\mathcal{T}$. Then there exist orthonormal systems $(f_j)_{j\in\mathbb{J}} \subseteq \mathcal{X}$ and $(g_j)_{j\in\mathbb{J}} \subseteq \mathcal{Y}$ such that

$$\mathcal{T} f_j = \sigma_j g_j, \qquad \mathcal{T}^* g_j = \sigma_j f_j$$

for all j.

The system $(\sigma_j, f_j, g_j)_{j\in\mathbb{J}}$ is called a *singular system* for $\mathcal{T}$. Every element $f \in \mathcal{X}$ can be represented in terms of the singular system as

$$f = f_0 + \sum_{j\in\mathbb{J}} \langle f, f_j \rangle_{\mathcal{X}} f_j$$

for some $f_0 \in \operatorname{null}(\mathcal{T})$ and

$$\mathcal{T} f = \sum_{j\in\mathbb{J}} \sigma_j \langle f, f_j \rangle_{\mathcal{X}} g_j,$$

which we both call a *singular value decomposition (SVD)*.

The SVD enables us to prove the following existence theorem for a best approximate solution f^+ of Problem 6.3, which also allows for a series representation for f^+.

Theorem 6.16 (cf. Engl et al. [42, Theorem 2.8])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces, let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a compact linear operator, and let $(\sigma_j, f_j, g_j)_{j\in\mathbb{J}}$ be its singular system.

1. We have $g \in \operatorname{dom} \mathcal{T}^+$ if and only if

   $$\sum_{j\in\mathbb{J}} \frac{1}{\sigma_j^2} \left| \langle g, g_j \rangle_{\mathcal{Y}} \right|^2 < \infty,$$

   which is called the *Picard criterion*.

2. If $g \in \operatorname{dom} \mathcal{T}^+$, then

$$f^+ = \mathcal{T}^+ g = \sum_{j \in \mathbb{J}} \frac{1}{\sigma_j} \langle g, g_j \rangle_{\mathcal{Y}} f_j \tag{6.3}$$

is a series representation of the Moore-Penrose inverse.

The central result, which shows why compact operators and ill-posed inverse problems have such a strong connection, is the following theorem.

Theorem 6.17 (cf. Rieder [153, Satz 2.2.8(e)])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces, let $\mathcal{T} \colon \mathcal{X} \to \mathcal{Y}$ be a compact linear operator. If $\mathcal{T}$ is invertible and $\mathcal{X}$ is not finite-dimensional, then $\mathcal{T}^{-1}$ is not continuous.

The most important type of compact operators, which will also arise in Chapter 7, are Fredholm integral operators. We cite the following result, which is based on the theory of Hilbert-Schmidt operators.

Theorem 6.18 (cf. Heuser [81, Chapter 87] and Rieder [153, Satz 2.2.7])**.** Let $\emptyset \neq S \subseteq \mathbb{R}^n$ be compact. If $k \in \mathrm{L}^2(S \times S)$, then the Fredholm integral operator $\mathcal{T} \colon \mathrm{L}^2(S) \to \mathrm{L}^2(S)$,

$$\mathcal{T} f = \int_S k(x, \cdot)\, f(x)\, \mathrm{d}x,$$

is compact.

The next section will deal with the regularization of ill-posed inverse problems.

6.1.3. Regularization methods

Let Problem 6.3 be the problem to be solved, that is, find $f^* \in \mathcal{X}$ such that

$$\mathcal{T} f^* = g,$$

where $\mathcal{T} \colon \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator between Hilbert spaces $\mathcal{X}, \mathcal{Y}$ and $g \in \mathcal{Y}$ is the given exact data. We assume that g is not available, but only a noisy version $g^\delta \in \mathcal{Y}$ such that $\left\| g - g^\delta \right\|_{\mathcal{Y}} \leq \delta$, where $\delta > 0$ is called the *noise level*. Furthermore, in this section we always assume an ill-posed problem, that is, $\overline{\operatorname{ran} \mathcal{T}} \neq \operatorname{ran} \mathcal{T}$. As already pointed out by Engl et al. [42, Section 3.1], due to the unboundedness of the Moore-Penrose generalized inverse, the term $\mathcal{T}^+ g^\delta$ is not a good approximation of $\mathcal{T}^+ g$. The basic idea of regularization is to replace this bad approximation by a better one, which we call f_λ^δ. It should depend continuously

on the noisy data g^δ and on a so-called regularization parameter λ such that the determination of f_λ^δ is a well-posed problem. In a way, one consequently replaces an ill-posed problem by a well-posed one, which is an approximation of the original problem in a certain manner. It is then desirable that f_λ^δ tends to f^+ for $\delta \searrow 0$ if the regularization parameter is chosen in a suitable way. We can formulate all of this in the language of operators in the following way.

Definition 6.19 (cf. Engl et al. [42, Definition 3.1])**.** Let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a bounded linear operator between Hilbert spaces $\mathcal{X}, \mathcal{Y}$. Let $\Lambda = (0, \lambda_0)$ for some $\lambda_0 > 0$ be a set of regularization parameters such that for every $\lambda \in \Lambda$, a continuous (but not necessarily linear) operator $\mathcal{R}_\lambda\colon \mathcal{Y} \to \mathcal{X}$ is given.

The family $\{ \mathcal{R}_\lambda \mid \lambda \in \Lambda \}$ is called a *regularization* of $\mathcal{T}^+$ if for every $g \in \operatorname{dom} \mathcal{T}^+$ there exists a so-called *parameter choice rule* $\lambda\colon \mathbb{R}^+ \times \mathcal{Y} \to \Lambda$ such that

$$\limsup_{\delta \searrow 0} \left\{ \left\| \mathcal{R}_{\lambda(\delta, g^\delta)}[g^\delta] - \mathcal{T}^+ g \right\|_{\mathcal{X}} \;\middle|\; g^\delta \in \mathcal{Y}, \left\| g^\delta - g \right\|_{\mathcal{Y}} \leq \delta \right\} = 0$$

holds. Furthermore, we require that

$$\limsup_{\delta \searrow 0} \left\{ \lambda(\delta, g^\delta) \;\middle|\; g^\delta \in \mathcal{Y}, \left\| g^\delta - g \right\|_{\mathcal{Y}} \leq \delta \right\} = 0.$$

For a fixed $g \in \operatorname{dom} \mathcal{T}^+$, the pair $(\mathcal{R}_\lambda, \lambda)$ is called a *convergent regularization method* for Problem 6.3 if the two limit relations above hold for this particular choice of g.

If the parameter choice rule does only depend on δ, but not on g^δ, then it is called an *a-priori parameter choice rule*. Otherwise, it is called an *a-posteriori parameter choice rule*.

If $(\mathcal{R}_\lambda, \delta)$ is a convergent regularization method, it can be shown that

$$\lim_{\delta \searrow 0} \mathcal{R}_{\lambda(\delta, g)}[g] = \mathcal{T}^+ g$$

(cf. Engl et al. [42, Remark 3.5]) such that regularizations are pointwise approximations of the Moore-Penrose generalized inverse. If we define the regularized solution $f_{\lambda(\delta,g)}^\delta := \mathcal{R}_{\lambda(\delta,g)}[g]$, we have, consequently,

$$\lim_{\delta \searrow 0} f_{\lambda(\delta,g)}^\delta = f^+$$

such that the regularized solution converges to the best approximate solution.

We do not want to go deeper into the theory of regularization methods at this point. For an overview over subjects like *order optimality* and general results about parameter choice rules, we refer to the monographs on that matter, for example, Engl et al. [42], Kirsch [99], and Rieder [153].

In the following section, we will discuss Tikhonov regularization, which is one of the most popular regularization methods, in more detail.

6.1.4. Tikhonov regularization

We will base this section on Kirsch [99, Section 2.2].

Definition 6.20 (Tikhonov functional)**.** Let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a linear and bounded operator, let $g \in \mathcal{Y}$, and $\lambda > 0$. Then $\mathcal{A}_{\lambda,g}\colon \mathcal{X} \to \mathbb{R}$,

$$\mathcal{A}_{\lambda,g}[f] \coloneqq \|g - \mathcal{T}f\|_{\mathcal{Y}}^2 + \lambda \|f\|_{\mathcal{X}}^2$$

is called the *Tikhonov functional*.

The idea of Tikhonov regularization is based on the fact that the least squares functional is identical to the Tikhonov functional for $\lambda = 0$. Since f^+ is a minimizer of the least squares functional, one hopes that the minimizers of the Tikhonov functional converge to f^+ for $\lambda \searrow 0$ (and it will be shown that this is indeed the case). At first, we provide a theorem, which shows that the minimizer of the Tikhonov functional is uniquely determined and fulfills a regularized normal equation.

Theorem 6.21 (cf. Kirsch [99, Theorem 2.11])**.** Let $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ be a linear and bounded operator between Hilbert spaces, let $g \in \mathcal{Y}$, and $\lambda > 0$. Then the Tikhonov functional $\mathcal{A}_{\lambda,g}$ has a unique minimizer $f_\lambda \in \mathcal{X}$, which is the unique solution of

$$(\mathcal{T}^*\mathcal{T} + \lambda\mathcal{I}) f_\lambda = \mathcal{T}^* g,$$

which is called the *regularized normal equation*. Consequently,

$$f_\lambda = (\mathcal{T}^*\mathcal{T} + \lambda\mathcal{I})^{-1}\mathcal{T}^* g \tag{6.4}$$

holds.

The formulation in Eq. (6.4) can be used to define a family of regularization operators $\{\, \mathcal{R}_\lambda \mid \lambda > 0 \,\}$ of $\mathcal{T}^+$ by

$$\mathcal{R}_\lambda \coloneqq (\mathcal{T}^*\mathcal{T} + \lambda\mathcal{I})^{-1}\mathcal{T}^*.$$

In the following theorem, we will state that there is a parameter choice method such that $(\mathcal{R}_\lambda, \lambda)$ forms a convergent regularization method if $\mathcal{T}$ is compact. A heuristic justification for this result can be obtained by choosing a singular system $(\sigma_j, f_j, g_j)_{j\in\mathbb{J}}$ of the operator $\mathcal{T}$. Then we have

$$\mathcal{R}_\lambda g = \sum_{j\in\mathbb{J}} \frac{\sigma_j}{\sigma_j^2 + \lambda} \langle g, g_j \rangle_{\mathcal{Y}} f_j$$

for $g \in \operatorname{dom} \mathcal{T}^+$. We observe that we have

$$\frac{\sigma_j}{\sigma_j^2 + \lambda} \to \frac{1}{\sigma_j}$$

for $\lambda \searrow 0$, which is exactly the term arising in the series representation of $\mathcal{T}^+$ in Eq. (6.3). Additionally, the theorem provides a convergence rate of Tikhonov regularization.

Theorem 6.22 (cf. Kirsch [99, Theorem 2.12])**.** Let $\mathcal{T} : \mathcal{X} \to \mathcal{Y}$ be a linear compact operator between Hilbert spaces and let $g \in \mathcal{Y}$. Then

(a) The family $\{ \mathcal{R}_\lambda \mid \lambda > 0 \}$ is a regularization of $\mathcal{T}^+$. Every a-priori parameter choice method $\lambda : (0, \infty) \to (0, \infty)$, which fulfills

$$\lambda(\delta) \searrow 0, \qquad \frac{\delta^2}{\lambda(\delta)} \searrow 0 \qquad (\delta \to 0)$$

is feasible.

(b) If we define

$$f_\lambda^\delta := \mathcal{R}_\lambda g^\delta = (\mathcal{T}^* \mathcal{T} + \lambda \mathcal{I})^{-1} \mathcal{T}^* g^\delta$$

for noisy data $g^\delta \in \mathcal{Y}$ with $\left\| g - g^\delta \right\|_{\mathcal{Y}} \leq \delta$, if we assume that $f^+ = \mathcal{T}^* \mathcal{T} h \in \operatorname{ran}(\mathcal{T}^* \mathcal{T})$ with $\|h\|_{\mathcal{X}} \leq \tau$, and if we define the parameter choice $\lambda(\delta) = m \, (\delta / \tau)^{2/3}$ for constants $m, \tau > 0$, then

$$\left\| f_{\lambda(\delta)}^\delta - f^+ \right\|_{\mathcal{X}} \leq \left(\frac{1}{2\sqrt{m}} + m \right) \tau^{1/3} \delta^{2/3}$$

holds.

It can be proved that this order of convergence is optimal and cannot be improved (cf. Kirsch [99, Theorem 2.13]).

6.2. Nonlinear inverse problems

In contrast to linear inverse problems, the theory for nonlinear inverse problems is much more difficult. In particular, there is no general regularization method, which works for a larger class of nonlinear inverse problems, or as Kirsch [99] states it, "every nonlinear inverse problem has its own characteristic features that should be used for a successful solution". Nevertheless, we will state some of the few general theoretical results that exist about nonlinear problems. We start by defining a nonlinear inverse problem.

Problem 6.23 (Nonlinear inverse problem)**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and let $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be a Fréchet-differentiable and possibly nonlinear operator. For given data $g \in \mathcal{Y}$, find $f^* \in \mathcal{X}$ such that

$$\mathcal{S}[f^*] = g.$$

For simplicity, we assume that a unique solution exists. We introduce a more sophisticated definition of well-posedness and ill-posedness for nonlinear inverse problems.

Definition 6.24 (cf. Rieder [153, Definition 7.1.1])**.** We call Problem 6.23 *locally well-posed in* $f^+ \in \mathcal{X}$ if there exists $\varrho > 0$ such that $f_k \to f^+$ for all sequences $(f_k)_{k\in\mathbb{N}} \subseteq \mathcal{X}$ with $\|f_k - f^+\|_{\mathcal{X}} < \varrho$ and $\mathcal{S}[f_k] \to \mathcal{S}[f^+]$ for $k \to \infty$.

Otherwise, we call the inverse problem *locally ill-posed*.

Note that this characterization is a proper generalization of the instability of the inverse operator for linear inverse problems, since a linear operator is locally well-posed either for *all* or for *no* $f^+ \in \mathcal{X}$. A similar result to Theorem 6.17 for nonlinear inverse problems is the following.

Theorem 6.25 (cf. Rieder [153, Satz 7.3.4])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces such that $\mathcal{X}$ is infinite-dimensional and separable. Let $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be continuous, compact, and weakly sequentially closed, that is, if $f_k \rightharpoonup f$ and $\mathcal{S}[f_k] \rightharpoonup g$ for $k \to \infty$ implies $\mathcal{S}[f] = g$.

Then, Problem 6.23 is locally ill-posed everywhere in $\mathcal{X}$.

Note that we define *compact* for nonlinear operators in the same way as we did for linear operators in Definition 6.11. Moreover, note that nonlinear compact operators are not continuous in general. This is why we need continuity as an additional condition. Finally, the result of the previous theorem is also true if $\mathcal{S}$ is not defined on the whole space $\mathcal{X}$, but only on some domain $\operatorname{dom}\mathcal{S} \subseteq \mathcal{X}$. In this case, the problem is locally ill-posed in all interior points of $\operatorname{dom}\mathcal{S}$.

In Section 2.2, we have already considered derivatives of nonlinear operators and their linearization. Most numerical methods for nonlinear inverse problems are based on such a linearization of the operator. Thus, it makes sense to consider the relation of ill-posedness and well-posedness of the original nonlinear and the linearized problem. For this purpose, we first define the linearized problem.

Problem 6.26. The linearized problem consists of finding $f \in \mathcal{X}$ such that

$$\mathcal{S}'[f^+](f) = g,$$

for given $f^+ \in \mathcal{X}$, $g \in \mathcal{Y}$.

The study of the relationship between the ill-posedness of Problem 6.23 and Problem 6.26 was carried out by Hofmann and Scherzer [82, 83]. We cite the versions as given by Rieder [153].

Theorem 6.27 (cf. Rieder [153, Satz 7.3.5])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and let $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ be a Fréchet differentiable operator, whose derivative fulfills a local Lipschitz condition in $f^+ \in \mathcal{X}$, that is, there exists $L, \varrho > 0$ such that

$$\left\|\mathcal{S}'[f] - \mathcal{S}'[f^+]\right\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \leq L \left\|f - f^+\right\|_{\mathcal{X}}$$

for all $f \in \mathcal{X}$ with $\|f - f^+\|_{\mathcal{X}} < \varrho$.

If Problem 6.23 is locally ill-posed in $f^+ \in \mathcal{X}$, then also Problem 6.26 is ill-posed.

It is important to notice that the converse is not generally true. The equivalence of the well-posedness of both problems can be proved under an additional assumption.

Theorem 6.28 (cf. Rieder [153, Satz 7.3.7])**.** Let the Fréchet differentiable operator $\mathcal{S}\colon \mathcal{X} \to \mathcal{Y}$ fulfill the so-called *tangential cone condition*, that is, there exists $\varrho > 0$ and $0 < \eta < 1$ such that

$$\left\|\mathcal{S}[f] - \mathcal{S}\left[\bar{f}\right] - \mathcal{S}'[\bar{f}](f - \bar{f})\right\|_{\mathcal{Y}} \leq \eta \left\|\mathcal{S}[f] - \mathcal{S}\left[\bar{f}\right]\right\|_{\mathcal{Y}}$$

for $f, \bar{f} \in \mathcal{X}$ with $\|f - f^+\|_{\mathcal{X}}, \left\|\bar{f} - f^+\right\|_{\mathcal{X}} < \varrho$.

Then, Problem 6.23 is locally well-posed (ill-posed) in f^+ if and only if Problem 6.26 is well-posed (ill-posed).

There exist a variety of numerical methods for the solution of nonlinear ill-posed inverse problems including nonlinear Tikhonov regularization and iterative methods. We will mention several of them in Section 10.2 after we have introduced our own new algorithm for nonlinear inverse problems.

Chapter 7.

Inverse gravimetry

Gravimetry is the method of the determination of the gravitational field of the Earth or other planetary bodies, whereas *inverse gravimetry* is concerned with the analysis of the Earth's interior based on these measurements of the gravitational field. Models of the gravitational field itself, determined by gravimetry, are useful in several applications, for example, the computation of satellite orbits (see Savet [158]). First and foremost, these models are used in inverse gravimetry, which yields interesting information about the interior of the Earth used in the fields of geodesy and geophysics as well as in industrial applications. For example, models of the Earth's interior can be used in the analysis of seismic waves (see, e. g., Komatitsch and Tromp [103]), but also for oil and gas exploration (see Pawlowski [143]) and deep geothermal systems (see Blick et al. [21]). Also information about surface densities can be obtained by models of the gravitational field, which are interesting for the analysis of mass transport and climate research (see, e. g., Velicogna and Wahr [178]). The most widely used model of the Earth's gravitational field currently is the *Earth Gravitational Model 2008 (EGM2008)* [142]. It is a spherical harmonic model, which is complete up to degree and order 2159. An established model of the Earth's interior is the *Preliminary Reference Earth Model (PREM)* (see Dziewonski and Anderson [40]), which includes radial models of the mass density, velocities of seismic waves, and several other parameters. Nowadays, the gravitational field of the Earth is measured by satellite missions like CHAMP (see Reigber et al. [151]), GRACE (see Tapley et al. [165]), GOCE (see Drinkwater et al. [38]), as well as the upcoming GRACE follow-on mission (see Flechtner et al. [49]), which will allow for even more precise models both of the gravitational field and the interior of the Earth in the future. Satellite missions to the Moon (GRAIL, see Zuber et al. [187]) and Jupiter (Juno, see Matousek [117]) also allow for the study of the interior of these celestial bodies.

Since inverse gravimetry has a variety of applications, the theoretical analysis of the gravitational potential and the corresponding inverse problems is of great importance. In the following sections, we will therefore first discuss the Newtonian potential and several variants of the inverse gravimetric problem.

7.1. Newton's gravitational potential

The basis of modeling the gravitational field is Newton's law of gravitation, which was first described qualitatively in Newton [136, Proposition LXXV]. It states that the gravitational potential U, measured at a point $y \in \mathbb{R}^3 \setminus \{ x \}$, of a point mass M, which is located at a point $x \in \mathbb{R}^3$, is given by

$$U(y) = \frac{G\,M}{|x-y|}, \qquad y \in \mathbb{R}^3 \setminus \{ x \}$$

(cf., e. g., Kellogg [97, Section I.2]), where

$$G := 6.674\,08 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$$

(see Mohr et al. [133]) is the gravitational constant. However, to simplify the notation, we assume that the units of all physical quantities are chosen such that $G = 1$.

Assuming a continuously distributed mass density inside a given body, one obtains the following definition of the gravitational potential.

Definition 7.1 (cf. Kellogg [97, Section III.3])**.** Let a body fill the region $\mathcal{E} \subseteq \mathbb{R}^3$ with the density function $\varrho\colon \mathcal{E} \to \mathbb{R}$. Then, the function $U_{\mathcal{E},\varrho}\colon \mathbb{R}^3 \setminus \overline{\mathcal{E}} \to \mathbb{R}$,

$$U_{\mathcal{E},\varrho}(y) = \int_{\mathcal{E}} \frac{\varrho(x)}{|x-y|}\,\mathrm{d}x,$$

is called *(Newton's) gravitational potential*.

From now on, we assume that $\mathcal{E}$ is a bounded open domain in $\mathbb{R}^3$ with a piecewise smooth boundary and that the mass density function $\varrho\colon \mathcal{E} \to \mathbb{R}$ is measurable and bounded.

Note that, in theory, we allow a negative mass density, which is not reasonable from the geophysical perspective. This is the case, because later on we will also discuss inverse problems, where ϱ plays the role of a mass anomaly, which can attain negative values.

It is well-known that the Newtonian potential is continuously differentiable in the whole space, harmonic outside of the region $\mathcal{E}$, and regular at infinity.

Theorem 7.2 (cf. Mikhlin [130, Theorem 11.6.1])**.** Under the assumptions on $\mathcal{E}$ and ϱ stated above, the Newtonian potential $U_{\mathcal{E},\varrho}$ defined in Definition 7.1 is continuous and continuously differentiable in $\mathbb{R}^3$.

Theorem 7.3 (cf. Mikhlin [130, Theorem 11.6.2])**.** Under the assumptions on $\mathcal{E}$ and ϱ stated above, the Newtonian potential fulfills $U_{\mathcal{E},\varrho} \in \mathrm{C}^{(2)}(\mathbb{R}^3 \setminus \overline{\mathcal{E}})$ and

$$\Delta_y U_{\mathcal{E},\varrho}(y) = 0$$

for all $y \in \mathbb{R}^3 \setminus \overline{\mathcal{E}}$.

Theorem 7.4 (cf. Freeden and Gerhards [52, Introduction])**.** Under the assumptions stated above, the Newtonian potential fulfills

$$\left|U_{\mathcal{E},\varrho}(x)\right| = \mathcal{O}\Big(|x|^{-1}\Big), \qquad \left|\nabla U_{\mathcal{E},\varrho}(x)\right| = \mathcal{O}\Big(|x|^{-2}\Big), \qquad |x| \to \infty,$$

which is what we call *regular at infinity*.

Often, in the functional analytic approach, it is not known whether ϱ is bounded. Instead, one assumes that $\varrho \in \mathrm{L}^2(\mathcal{E})$, which does not imply (essential) boundedness of ϱ. Using the theory of Fredholm integral equations, the following result can be proved.

Theorem 7.5. Let $\varrho \in \mathrm{L}^2(\mathcal{E})$ and a regular surface S be given such that $\overline{\mathcal{E}} \subseteq S^{\mathrm{int}}$. Then we have $U_{\mathcal{E},\varrho}|_S \in \mathrm{L}^2(S)$.

Furthermore, for fixed $\mathcal{E}$, the operator $\mathcal{T}_{\mathcal{E}} \colon \mathrm{L}^2(\mathcal{E}) \to \mathrm{L}^2(S)$, $\varrho \mapsto U_{\mathcal{E},\varrho}|_S$ is a compact linear operator.

Proof. Since $\overline{\mathcal{E}} \subseteq S^{\mathrm{int}}$, there exists $d > 0$ such that $|x - y| > d$ for all $x \in \mathcal{E}$, $y \in S$. The kernel $k \colon \mathcal{E} \times S \to \mathbb{R}$, $k(x,y) = 1/|x-y|$ fulfills $k \in \mathrm{L}^2(\mathcal{E} \times S)$, since

$$\begin{aligned}
\int_{\mathcal{E}} \int_S |k(x,y)|^2 \,\mathrm{d}\omega(y)\,\mathrm{d}x &= \int_{\mathcal{E}} \int_S \frac{1}{|x-y|^2} \,\mathrm{d}\omega(y)\,\mathrm{d}x \\
&\leq \int_{\mathcal{E}} \int_S \frac{1}{d^2} \,\mathrm{d}\omega(y)\,\mathrm{d}x \\
&= \frac{\omega(S)\,\lambda(\mathcal{E})}{d^2} < \infty.
\end{aligned}$$

The fact that $U_{\mathcal{E},\varrho}|_S \in \mathrm{L}^2(S)$ and the compactness of the operator follow from the theory of Fredholm integral equations (see Yosida [186, Example 1 in Section VII.3, Example 2 in Section X.2], and Theorem 6.18). ∎

Furthermore, the Poisson equation is fulfilled inside the domain $\mathcal{E}$.

Theorem 7.6 (cf. Mikhlin [130, Theorem 11.6.3])**.** Under the assumptions on $\mathcal{E}$ and ϱ stated above and the additional requirement that ϱ is Lipschitz continuous on $\mathcal{E}$, the Newtonian potential U fulfills

$$-\Delta_y U_{\mathcal{E},\varrho}(y) = 4\,\pi\,\varrho(y),$$

for all $y \in \mathcal{E}$.

An important result of potential theory states that harmonic functions are arbitrarily often continuously differentiable and all its derivatives are also harmonic (cf. Freeden and Gerhards [52, Corollary 3.8]). Thus, also the Newtonian potential has this property outside the body $\mathcal{E}$.

7.2. Inverse gravimetric problems

Inverse gravimetry is concerned with the gathering of information about the shape $\mathcal{E}$ and the mass density ϱ from the gravitational potential U of a body. This information will mostly be values of U, ∇U, or $\nabla \otimes \nabla U$ (the Hessian) on a regular surface S such that $\overline{\mathcal{E}} \subseteq S^{\text{int}}$.

Indeed, concerning the inverse problem, the following theorem shows that, at least for convex domains with analytic boundaries, there is no difference between measuring the potential itself and its gradient on the regular surface S.

Theorem 7.7 (cf. Isakov [88, Lemma 2.1.1])**.** Let $S \subseteq \mathbb{R}^3$ be an analytic regular surface such that S^{int} is convex. If $\Sigma_1, \Sigma_2 \subseteq S^{\text{int}}$ are regular surfaces, $\varrho_j \colon \Sigma_j^{\text{int}} \to (0,\infty)$ for $j = 1, 2$ are positive mass density functions, and

$$\left|\nabla U_{\Sigma_1^{\text{int}},\varrho_1}(y)\right| = \left|\nabla U_{\Sigma_2^{\text{int}},\varrho_2}(y)\right|$$

for all $y \in S$, then

$$U_{\Sigma_1^{\text{int}},\varrho_1}(y) = U_{\Sigma_2^{\text{int}},\varrho_2}(y)$$

for all $y \in S^{\text{ext}}$.

Therefore, we restrict to the case of a measured potential and we formulate the following general inverse gravimetric problem.

Problem 7.8 (General inverse gravimetric problem)**.** Let $S \subseteq \mathbb{R}^3$ be a regular surface and $g \in \mathrm{L}^2(S)$ be a given function. Find $\mathcal{E} \subseteq S^{\text{int}}$ and $\varrho \in \mathrm{L}^2(\mathcal{E})$ such that

$$U_{\mathcal{E},\varrho}|_S = g.$$

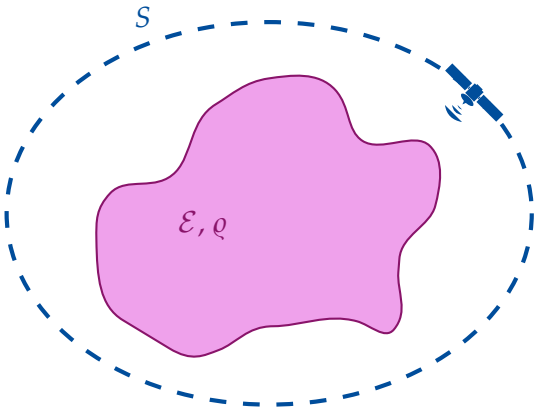


Figure 7.1.: Sketch of the general inverse gravimetric problem: The body $\mathcal{E}$ is filled with mass of the density ϱ (purple). The gravitational potential is measured (for example by a satellite) on a regular surface S (blue).

A sketch of the situation can be found in Figure 7.1.

Determining both $\mathcal{E}$ and ϱ uniquely from gravitational data fails already in a very elementary problem setting incorporating constant densities and simple geometries, as the following example shows.

Example 7.9. It is well-known that the gravitational potential of a ball $\mathbb{B}_R$ with radius $R > 0$ with the constant mass density $\varrho > 0$ is given by

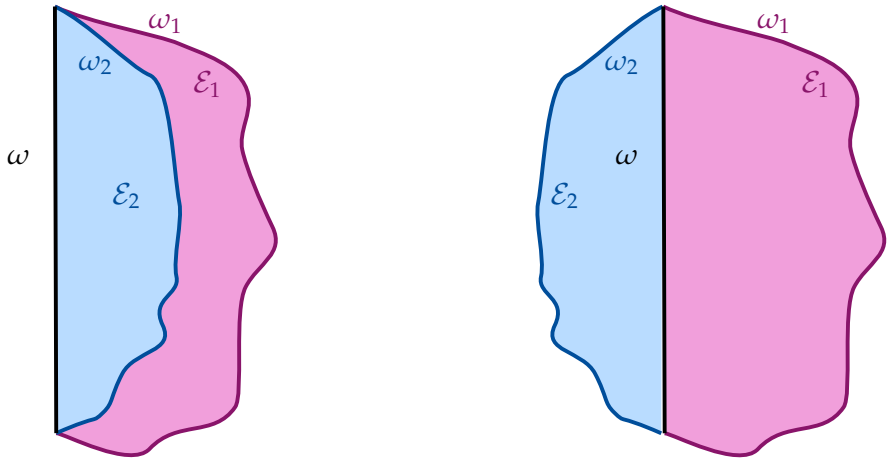
$$U_{\mathbb{B}_R,\varrho}(y) = \frac{4\,\pi\,R^3}{3}\,\frac{\varrho}{|y|} \tag{7.1}$$

for $y \in \mathbb{R}^3$ with $|y| > R$.

Thus, the gravitational potentials of two balls $\mathbb{B}_{R_1}$ and $\mathbb{B}_{R_2}$ are equal in all points $y \in \mathbb{R}^3$ with $|y| > \max\{ R_1, R_2 \}$ as long as the corresponding mass densities fulfill $R_1^3\,\varrho_1 = R_2^3\,\varrho_2$, say for example, $R_1 = 2$, $R_2 = 1$, $\varrho_1 = 1$, $\varrho_2 = 8$.

As a consequence of this severe non-uniqueness of the general inverse gravimetric problem, in most of the literature either the domain $\mathcal{E}$ or the mass density function ϱ is assumed to be known and the inverse problem is to find the respective other unknown. This is why we will deal with these two problems in the following sections.

However, we mention the following result, which was stated in Isakov [88]. It shows that under several technical assumptions both the domain and the mass density can be recovered uniquely from gravitational data.



(a) The sets intersect.

(b) The sets are disjoint.

Figure 7.2.: Sketch of the two possible situations in Theorem 7.10. On the one hand, the sets $\mathcal{E}_1, \mathcal{E}_2$ are both bounded by the graph of the function ω (black). On the other hand they are bounded by the graphs of the functions ω_1 (purple) and ω_2 (blue), respectively.

Theorem 7.10 (cf. Isakov [88, Theorem 3.2.1])**.** Suppose that there exist functions $\omega, \omega_1, \omega_2 \colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathcal{E}_j = \left\{ x \in \mathbb{R}^3 \;\middle|\; \omega(x_2, x_3) < x_1 < \omega_j(x_2, x_3) \text{ or } \omega_j(x_2, x_3) < x_1 < \omega(x_2, x_3) \right\},$$

for $j = 1, 2$. Thus, we assume that $\mathcal{E}_1$ and $\mathcal{E}_2$ are bounded on the one hand by the graph of the function ω and on the other hand by the graph of ω_1 and ω_2, respectively, that is, the boundaries of both sets have a common part (the so-called *contact domain*). The situation is depicted in Figure 7.2. Furthermore, we assume that there exists a regular surface S such that $\mathcal{E}_1, \mathcal{E}_2 \subseteq S^{\text{int}}$.

Additionally, we require that the density functions $\varrho_j \colon S^{\text{int}} \to \mathbb{R}$ are continuous and do not depend on the first component x_1.

Then the equality

$$U_{\mathcal{E}_1, \varrho_1}(y) = U_{\mathcal{E}_2, \varrho_2}(y) \qquad \text{for all } y \in S^{\text{ext}} \tag{7.2}$$

implies that $\mathcal{E}_1 = \mathcal{E}_2$ and $\varrho_1 = \varrho_2$ on $\mathcal{E}_1$.

Note that, in order to be well-defined in every $x \in \mathcal{E}$ for all possible sets $\mathcal{E} \subseteq \mathbb{R}^3$, we have to define $\varrho(x)$ for all $x \in S^{\text{int}}$, where all feasible sets $\mathcal{E}$ fulfill $\overline{\mathcal{E}} \subseteq S^{\text{int}}$. Note, furthermore, that Eq. (7.2) is implied by

$$U_{\mathcal{E}_1, \varrho_1}(y) = U_{\mathcal{E}_2, \varrho_2}(y) \qquad \text{for all } y \in S \tag{7.3}$$

if the functions ϱ_j are bounded, since then S is a regular surface, the potentials $U_{\mathcal{E}_j,\varrho_j}$ are continuous on S, and the gravitational potential on S^{ext} is the solution of the exterior Dirichlet problem, which is known to be unique. Consequently, Eq. (7.2) can be replaced by Eq. (7.3) in this case.

In the following sections, we will consider the important special cases of the linear and nonlinear inverse gravimetric problems, where either $\mathcal{E}$ or ϱ are assumed to be known.

7.2.1. Linear inverse gravimetric problem

Since the solution to Problem 7.8 is difficult to obtain, we first restrict to the so-called *linear inverse gravimetric problem*, where only the mass density is assumed to be unknown, whereas the shape of the Earth $\mathcal{E}$ is assumed to be known. This results in the following problem.

Problem 7.11 (Linear inverse gravimetric problem)**.** Let $S \subseteq \mathbb{R}^3$ be a regular surface, let $\mathcal{E}$ be a bounded open domain, and let a function $g \in \mathrm{L}^2(S)$ be given.

Find $\varrho \in \mathrm{L}^2(\mathcal{E})$ such that

$$U_{\mathcal{E},\varrho}|_S = g.$$

The operator that maps $\varrho \in \mathrm{L}^2(\mathcal{E})$ to $U_{\mathcal{E},\varrho}|_S$ for fixed $\mathcal{E}$ is denoted by $\mathcal{T}_\mathcal{E} \colon \mathrm{L}^2(\mathcal{E}) \to \mathrm{L}^2(S)$ and the operator equation

$$\mathcal{T}_\mathcal{E}(\varrho) = g \tag{7.4}$$

is called the *linear inverse gravimetric problem*.

Remark. The operator $\mathcal{T}_\mathcal{E}$ is indeed linear, since for $\varrho_1, \varrho_2 \in \mathrm{L}^2(\mathcal{E})$ and $\lambda \in \mathbb{R}$ we have

$$\begin{aligned}\mathcal{T}_\mathcal{E}(\varrho_1 + \lambda\varrho_2) &= \int_\mathcal{E} \frac{\varrho_1(x) + \lambda\varrho_2(x)}{|x - \cdot|}\,\mathrm{d}x \\ &= \int_\mathcal{E} \frac{\varrho_1(x)}{|x - \cdot|}\,\mathrm{d}x + \lambda \int_\mathcal{E} \frac{\varrho_2(x)}{|x - \cdot|}\,\mathrm{d}x \\ &= \mathcal{T}_\mathcal{E}(\varrho_1) + \lambda\mathcal{T}_\mathcal{E}(\varrho_2)\end{aligned}$$

due to the linearity of the Lebesgue integral.

Furthermore, in Theorem 7.5 we have already shown that the operator is bounded and compact.

Below, we will discuss the ill-posedness of the inverse problem in Eq. (7.4). We begin with a well-known result about the uniqueness of the solution. It was given in the following form by Weck [183].

Theorem 7.12 (cf. Weck [183, Lemma 1])**.** The null space of the operator $\mathcal{T}_{\mathcal{E}}$ is given by

$$\operatorname{null} \mathcal{T}_{\mathcal{E}} = \left\{ \Delta f \mid f \in \mathrm{H}_0^2(\mathcal{E}) \right\},$$

where $\mathrm{H}_0^2(\mathcal{E})$ is the completion of the space of arbitrary often differentiable functions with compact support in $\mathcal{E}$ with respect to the well-known $\mathrm{H}^2(\mathcal{E})$-Sobolev norm. Furthermore, the orthogonal complement of the null space is given by

$$(\operatorname{null} \mathcal{T}_{\mathcal{E}})^{\perp} = \operatorname{null} \Delta := \left\{ f \in \mathrm{C}^{(\infty)}(\mathcal{E}) \;\middle|\; \Delta f = 0 \right\}, \tag{7.5}$$

that is, it consists of all harmonic functions.

Note that Weyl's Lemma (cf. Freeden and Gerhards [52, Section 4.1.2]) states that every harmonic distribution can be represented by a function, which leads to the use of the space $\mathrm{C}^{(\infty)}(\mathcal{E})$ in Eq. (7.5). Note furthermore that the result in Eq. (7.5) has been proved before, see Lauricella [107] and Pizzetti [145, 146].

Concerning the existence and stability, we can refer to the theory of compact operators, which has already been summarized in Section 6.1.2. Since $\mathcal{T}_{\mathcal{E}}$ is compact, there exists a singular system $(\sigma_j, f_j, g_j)_{j\in\mathbb{N}}$ and in terms of this singular system a necessary and sufficient condition for the existence of a solution of Eq. (7.4) for given data g is the Picard condition

$$\sum_{j=1}^{\infty} \frac{1}{\sigma_j^2} \left| \langle g, g_j \rangle_{\mathrm{L}^2(S)} \right|^2 < \infty$$

as already stated in Theorem 6.16. Furthermore, we have already proved in Theorem 6.17 that, as long as the domain is infinite-dimensional, every linear compact operator has an unbounded inverse such that the inverse problem is unstable.

For the special case of $\mathcal{E} = \mathbb{B}_1$, which is relevant in geoscientific applications due to the nearly spherical structure of the Earth, we have consequently that

$$\operatorname{null} \mathcal{T}_{\mathcal{E}} = \operatorname{Anharm}(\mathbb{B}_1)$$

(cf. Michel [118, Theorem 2.2.3]), where

$$\operatorname{Anharm}(\mathbb{B}_1) := \left(\overline{\operatorname{Harm}_{0\ldots\infty}(\mathbb{B}_1)}^{\mathrm{L}^2(\mathbb{B}_1)} \right)^{\perp}$$

is the set of anharmonic functions on $\mathbb{B}_1$. Thus,

$$\mathrm{L}^2(\mathbb{B}_1) = \overline{\mathrm{Harm}_{0\ldots\infty}(\mathbb{B}_1)}^{\mathrm{L}^2(\mathbb{B}_1)} \oplus \mathrm{Anharm}(\mathbb{B}_1).$$

Similar results for L^p and Sobolev spaces have been given by Sansò [156]. Furthermore, Ballani et al. [13] and Michel [118, Section 2.3] derived bases for the spaces of anharmonic polynomials on $\mathbb{B}_1$.

Additionally, Michel and Fokas [124] showed the following existence result including a closed formula for the solution of Eq. (7.4).

Theorem 7.13 (cf. Michel and Fokas [124, Corollary 4.1])**.** Let $V\colon \overline{\mathbb{R}^3 \setminus \mathbb{B}_1} \to \mathbb{R}$ be an arbitrary function satisfying

- $V|_{\mathbb{S}^2} \in \mathrm{L}^2(\mathbb{S}^2)$,
- $\sum_{n=0}^{\infty} \sum_{j=-n}^{n} \langle V|_{\mathbb{S}^2}, Y_{n,j} \rangle^2_{\mathrm{L}^2(\mathbb{S}^2)} n^3 < \infty$,
- $\Delta V = 0$ in $\mathbb{R}^3 \setminus \mathbb{B}_1$ and V is regular at infinity.

Then the unique solution $\varrho \in \mathrm{C}^{(2)}(\mathbb{B}_1)$ with $V = U_{\mathbb{B}_1,\varrho}$ and $\Delta\varrho = 0$ in $\mathbb{B}_1$ is given by

$$\varrho(x) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi}(2n+3)|x|^n \sum_{j=-n}^{n} \langle V|_{\mathbb{S}^2}, Y_{n,j} \rangle_{\mathrm{L}^2(\mathbb{S}^2)} Y_{n,j}\left(\frac{x}{|x|}\right)$$

provided that the series converges with respect to $\mathrm{L}^2(\mathbb{B}_1)$.

Unfortunately, the harmonicity condition for ϱ lacks a physical interpretation (cf. Michel and Fokas [124]), since the maximum principle for harmonic functions (cf. Helms [77, Theorem 1.5.9]) states that the density ϱ would have to attain its maximum at the Earth's surface. This is not reasonable from the perspective of application.

Several further conditions on the density ϱ, which yield a unique solution, are discussed in Michel and Fokas [124], but all of them share the disadvantage of having little physical interpretation.

Not only the non-uniqueness, but also the instability of the inverse problem has an impact on its usability in real-world applications. Since measurements are always equipped with a certain amount of noise, induced by measurement errors, regularization strategies have to be applied. A vast amount of literature about the numerical solution of Problem 7.11 with the use of different regularization methods exists. The methods, which were applied, include point mass models (Stromeyer and Ballani [164]), using basis functions with local support (Sansò et al. [157]), a truncated singular value decomposition (Tscherning and Strykowski [172]), and a Tikhonov regularization approach (Weck [183]), as well as methods using spherical splines

(Fengler et al. [44] and Michel and Wolf [129]) and spherical wavelets (Michel [118–121] and Michel and Fokas [124]). For an extensive overview, see Section 9 of Michel and Fokas [124]. Finally, we mention the work of Fischer [45] and Fischer and Michel [46], where the Regularized Functional Matching Pursuit was applied to the linear inverse gravimetric problem. For the nonlinear inverse gravimetric problem, which will be the topic of the following section, we will present a similar algorithm in Chapter 10.

7.2.2. Nonlinear inverse gravimetric problem

In the previous section, Problem 7.11 was derived from Problem 7.8 by assuming that the shape of the Earth $\mathcal{E}$ is known and the mass density function ϱ is unknown. In the following, we will look at the problem the other way around: we will assume that a model for the mass density function is available (for example, we could use PREM [40]) and that we want to determine the shape of the Earth. Furthermore, we will see that the resulting nonlinear integral operator equation can also be used to determine boundary layers inside the Earth.

Problem 7.14 (Nonlinear inverse gravimetric problem)**.** Let $S \subseteq \mathbb{R}^3$ be a regular surface, let $\varrho \in \mathrm{L}^2(S^{\mathrm{int}})$ be a mass density model, and let a function $g \in \mathrm{L}^2(S)$ be given. Find $\mathcal{E}$ such that $\overline{\mathcal{E}} \subseteq S^{\mathrm{int}}$ and

$$U_{\mathcal{E},\varrho}|_S = g.$$

The operator that maps $\mathcal{E} \subseteq S^{\mathrm{int}}$ to $U_{\mathcal{E},\varrho}|_S$ for fixed ϱ is denoted by $\tilde{\mathcal{S}}_\varrho$ and the operator equation

$$\tilde{\mathcal{S}}_\varrho[\mathcal{E}] = g$$

is called the *nonlinear inverse gravimetric problem*.

Since $\tilde{\mathcal{S}}_\varrho$ should formally be defined on a space of subsets of S^{int}, which is difficult to handle, we restrict to the case of domains that are star-shaped with respect to the origin. This is also motivated by the following example for non-uniqueness for Problem 7.14 with domains that are not star-shaped. It is based on Example 7.9, which we have already given before.

Example 7.15 (cf. Isakov [88, Section 2.2])**.** From Eq. (7.1) we can conclude that for a constant density $\varrho > 0$ and radii $R_1 > R_2 > 0$, we have

$$U_{\mathbb{B}_{R_1} \setminus \mathbb{B}_{R_2},\varrho}(y) = \frac{4\pi}{2} \left(R_1^3 - R_2^3\right) \frac{\varrho}{|y|}.$$

Consequently, the gravitational potential of two spherical shells $\mathbb{B}_{R_1} \setminus \mathbb{B}_{R_2}$ and $\mathbb{B}_{R_3} \setminus \mathbb{B}_{R_4}$ with identical constant mass density $\varrho > 0$ is equal if and only if $R_1^3 - R_2^3 = R_3^3 - R_4^3$.

This simple example already shows that without any restrictions on the shape of $\mathcal{E}$ there is a severe non-uniqueness associated to the nonlinear inverse gravimetric problem.

The restriction to star-shaped domains leads to the following problem.

Problem 7.16 (Nonlinear inverse gravimetric problem, star-shaped)**.** Let $S \subseteq \mathbb{R}^3$ be a regular surface, let $\varrho \in \mathrm{L}^2(S^{\mathrm{int}})$ be a mass density model, and let a function $g \in \mathrm{L}^2(S)$ be given.

Find a function $\sigma\colon \mathbb{S}^2 \to (0,\infty)$ such that $\mathcal{E} = \Sigma^{\mathrm{int}}$, where the regular surface $\Sigma \subseteq S^{\mathrm{int}}$ is given by

$$\Sigma := \left\{ r\xi \in \mathbb{R}^3 \;\middle|\; \xi \in \mathbb{S}^2, r = \sigma(\xi) \right\} \tag{7.6}$$

and

$$U_{\Sigma^{\mathrm{int}},\varrho}|_S = g.$$

The operator that maps the function σ to $U_{\Sigma^{\mathrm{int}},\varrho}|_S$ for fixed ϱ is denoted by $\mathcal{S}_\varrho$ and the operator equation

$$\mathcal{S}_\varrho[\sigma] = g$$

is called the *nonlinear inverse gravimetric problem (with a star-shaped domain)*.

Note that, using polar coordinates, the nonlinear integral operator has the expression

$$\begin{aligned} \mathcal{S}_\varrho[\sigma]\,(y) &= U_{\Sigma^{\mathrm{int}},\varrho}(y) \\ &= \int_{\Sigma^{\mathrm{int}}} \frac{\varrho(x)}{|x-y|}\,\mathrm{d}x \\ &= \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2\,\mathrm{d}r\,\mathrm{d}\omega(\xi). \end{aligned} \tag{7.7}$$

A sketch of the setting for Problem 7.16 can be found in Figure 7.3.

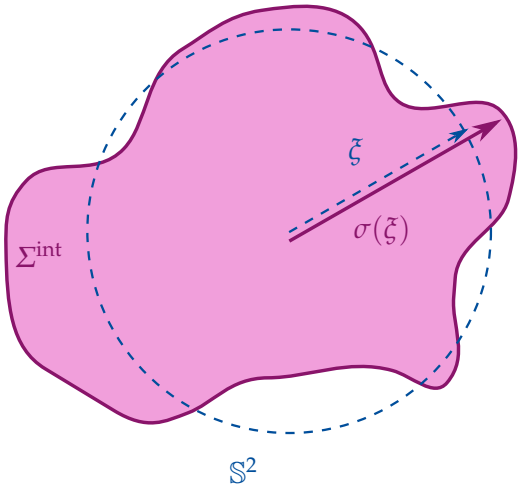


Figure 7.3.: Geometrical representation of the situation in Problem 7.16: the boundary Σ of the star-shaped body of mass Σ^{int} is parametrized by a function σ on $\mathbb{S}^2$.

Remark. In general, this operator is indeed nonlinear, since for $\sigma, \tau \colon \mathbb{S}^2 \to (0, \infty)$, we have

$$\begin{aligned}\mathcal{S}_\varrho[\sigma+\tau](y) &= \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)+\tau(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi) \\ &= \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r + \int_{\sigma(\xi)}^{\sigma(\xi)+\tau(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi),\end{aligned}$$

which is only equal to $\mathcal{S}_\varrho[\sigma](y) + \mathcal{S}_\varrho[\tau](y)$ if $\sigma(\xi) = 0$ for all $\xi \in \mathbb{S}^2$.

We will discuss the well-/ill-posedness of the nonlinear inverse gravimetric problem in the following paragraphs.

It is clear that a necessary condition for the existence of a solution to Problem 7.16 is the harmonicity of the potential, that is, g needs to be a restriction of some harmonic function to S. In more detail, Weck [183] mentions that g needs to be analytic. Isakov [88, Chapter 5] deals with a more detailed discussion of the topic of existence. In Isakov [90], the same author states that "it is not possible to obtain (even local) existence results", even after providing a "special local existence theorem" in Isakov [88], which shows that the topic of existence of a solution is a very difficult problem. We refer to the works by Isakov for further details, since we want to concentrate on the issues of uniqueness and stability of a solution in this section. These properties differ most between the linear and the nonlinear inverse gravimetric problem.

Isakov [88] proves uniqueness results for Problem 7.16 both in the case of a constant and a variable mass density. Below, we will state both theorems and present the proof for a constant mass density.

First, we begin with an important lemma about harmonic functions. It is actually the main idea for the proof of uniqueness of the nonlinear inverse gravimetric problem and was first presented by Novikov [138] for the analogous two-dimensional problem using the logarithmic potential.

Lemma 7.17 (cf. the proof of Isakov [88, Theorem 2.2.1])**.** Suppose $A \subseteq \mathbb{R}^3$ is an open domain. If $f\colon A \to \mathbb{R}$ is harmonic, then also the function

$$x = r\xi \mapsto r\,\frac{\partial f}{\partial r}(r\xi),$$

where $r \geq 0$ and $\xi \in \mathbb{S}^2$ are the polar coordinates, is harmonic.

Proof. Let $f\colon A \to \mathbb{R}$ be harmonic, that is $\Delta_x f(x) = 0$ for all $x \in A$. Since $\frac{\partial f}{\partial r}$ is the directional derivative in the direction of the vector $x/|x|$ and $r = |x|$, we obtain for arbitrary $x = r\xi \in A$ that

$$r\,\frac{\partial f}{\partial r}(r\xi) = |x|\,\frac{x}{|x|} \cdot \nabla_x f(x) = x \cdot \nabla_x f(x)$$

for $x \neq 0$. For $x = 0$, we obtain the same result due to the continuity of both sides of the equation. Thus,

$$\begin{aligned}
\Delta_{r\xi}\left(r\,\frac{\partial f}{\partial r}\right)(r\xi) &= \Delta_x\left(x \cdot \nabla_x f(x)\right) \\
&= \sum_{k=1}^{3} \Delta_x\left(x_k\,\frac{\partial f}{\partial x_k}(x)\right) \\
&= \sum_{k=1}^{3}\left(\Delta_x x_k\,\frac{\partial f}{\partial x_k}(x) + 2\,\nabla_x x_k \cdot \nabla_x \frac{\partial f}{\partial x_k}(x) + x_k\,\Delta_x \frac{\partial f}{\partial x_k}(x)\right) \\
&= 2\sum_{k=1}^{3}\sum_{j=1}^{3} \frac{\partial x_k}{\partial x_j}\,\frac{\partial^2 f}{\partial x_j x_k}(x) \qquad (7.8) \\
&= 2\sum_{k=1}^{3}\sum_{j=1}^{3} \delta_{jk}\,\frac{\partial^2 f}{\partial x_j x_k}(x) \\
&= 2\sum_{k=1}^{3} \frac{\partial^2 f}{\partial x_j^2}(x) \\
&= 2\,\Delta_x f(x) = 0,
\end{aligned}$$

where Eq. (7.8) holds because

$$\Delta_x x_k = \sum_{j=1}^{3} \frac{\partial^2 x_k}{\partial x_j^2} = 0 \qquad \text{and} \qquad \Delta_x \frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \Delta_x f(x) = 0$$

by Schwarz's theorem. ■

The use of Green's identity (see Theorem 2.2) for particular functions yields the following result.

Lemma 7.18. Let $A \subseteq \mathbb{R}^3$ be an open domain with a sufficiently smooth boundary and let $f \in \mathrm{C}^{(2)}(A)$. Then,

$$\int_A x \cdot \nabla f(x) + 3 f(x) \, \mathrm{d}x = \int_{\partial A} f(x) \, x \cdot \nu(x) \, \mathrm{d}\omega(x),$$

where $\nu \colon \partial A \to \mathbb{R}^3$ is the outer unit normal field of A.

Proof. Let $g \colon A \to \mathbb{R}$, $g(x) := |x|^2/2$. Then,

$$\nabla_x g(x) = x \qquad \text{and} \qquad \Delta_x g(x) = \operatorname{div}_x x = 3.$$

Therefore, we have

$$\begin{aligned} \int_{\partial A} f(x) \, x \cdot \nu(x) \, \mathrm{d}\omega(x) &= \int_{\partial A} f(x) \, \nabla_x g(x) \cdot \nu(x) \, \mathrm{d}\omega(x) \\ &= \int_{\partial A} f(x) \, \frac{\partial g}{\partial \nu}(x) \, \mathrm{d}\omega(x) \\ &= \int_A f(x) \, \Delta_x g(x) + \nabla f(x) \cdot \nabla g(x) \, \mathrm{d}x \qquad (7.9) \\ &= \int_A 3 f(x) + x \cdot \nabla_x f(x) \, \mathrm{d}x \end{aligned}$$

by using Green's identity in Eq. (7.9), which completes the proof. ■

These lemmas will be used in the proof of the following theorem. We carry out the proof for two reasons: first, the theorem itself highlights the usefulness of the nonlinear inverse gravimetric problem due to the uniqueness of a solution in contrast to the linear problem. Secondly, we try to give an improved and expanded presentation of the proof such that it is easier to understand for the reader.

Theorem 7.19 (cf. Isakov [88, Theorem 2.2.1])**.** Let $S \subseteq \mathbb{R}^3$ be a regular surface. Suppose that $\Sigma_1, \Sigma_2 \subseteq S^{\mathrm{int}}$ are regular surfaces such that $\Sigma_1^{\mathrm{int}}, \Sigma_2^{\mathrm{int}}$ are star-shaped.

If

$$U_{\Sigma_1^{\mathrm{int}},1}|_{S^{\mathrm{ext}}} = U_{\Sigma_2^{\mathrm{int}},1}|_{S^{\mathrm{ext}}}, \qquad (7.10)$$

then $\Sigma_1 = \Sigma_2$.

Proof. We rewrite Eq. (7.10) as

$$\begin{aligned}
0 = U_{\Sigma_1^{\mathrm{int}},1}(y) - U_{\Sigma_2^{\mathrm{int}},1}(y) &= \int_{\Sigma_1^{\mathrm{int}}} \frac{1}{|x-y|}\,\mathrm{d}x - \int_{\Sigma_2^{\mathrm{int}}} \frac{1}{|x-y|}\,\mathrm{d}x \\
&= \int_{S^{\mathrm{int}}} \frac{1}{|x-y|}\chi_{\Sigma_1^{\mathrm{int}}}(x)\,\mathrm{d}x - \int_{S^{\mathrm{int}}} \frac{1}{|x-y|}\chi_{\Sigma_2^{\mathrm{int}}}(x)\,\mathrm{d}x \\
&= \int_{S^{\mathrm{int}}} \frac{1}{|x-y|}\Big(\chi_{\Sigma_1^{\mathrm{int}}}(x) - \chi_{\Sigma_2^{\mathrm{int}}}(x)\Big)\,\mathrm{d}x \\
&= U_{S^{\mathrm{int}},\chi_{\Sigma_1^{\mathrm{int}}}-\chi_{\Sigma_2^{\mathrm{int}}}}(y)
\end{aligned}$$

for all $y \in S^{\mathrm{ext}}$. Thus, for $\varrho\colon S^{\mathrm{int}} \to \mathbb{R}$, $\varrho(x) \coloneqq \chi_{\Sigma_1^{\mathrm{int}}}(x) - \chi_{\Sigma_2^{\mathrm{int}}}(x)$, we have $\varrho \in \operatorname{null} \mathcal{T}_{S^{\mathrm{int}}}$ such that we can apply Theorem 7.12 to obtain

$$\begin{aligned}
0 &= \int_{S^{\mathrm{int}}} f(x)\,\varrho(x)\,\mathrm{d}x \\
&= \int_{S^{\mathrm{int}}} f(x)\,\Big(\chi_{\Sigma_1^{\mathrm{int}}}(x) - \chi_{\Sigma_2^{\mathrm{int}}}(x)\Big)\,\mathrm{d}x \\
&= \int_{\Sigma_1^{\mathrm{int}}} f(x)\,\mathrm{d}x - \int_{\Sigma_2^{\mathrm{int}}} f(x)\,\mathrm{d}x
\end{aligned}$$

for every harmonic function $f\colon S^{\mathrm{int}} \to \mathbb{R}$.

Therefore, by Lemma 7.17 we obtain

$$\int_{\Sigma_1^{\mathrm{int}}} x \cdot \nabla f(x) + 3\,f(x)\,\mathrm{d}x = \int_{\Sigma_2^{\mathrm{int}}} x \cdot \nabla f(x) + 3\,f(x)\,\mathrm{d}x \tag{7.11}$$

for every harmonic function f, which yields

$$\int_{\Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}} x \cdot \nabla f(x) + 3\,f(x)\,\mathrm{d}x = \int_{\Sigma_2^{\mathrm{int}} \setminus \Sigma_1^{\mathrm{int}}} x \cdot \nabla f(x) + 3\,f(x)\,\mathrm{d}x, \tag{7.12}$$

observing that $\Sigma_1^{\mathrm{int}} \cap \Sigma_2^{\mathrm{int}}$ is part of the domain of integration on both sides of Eq. (7.11).

We now define the boundary sets

$$\begin{aligned}
\Sigma_{1\mathrm{e}} &\coloneqq \Sigma_1 \setminus \Sigma_2^{\mathrm{int}}, & \Sigma_{2\mathrm{e}} &\coloneqq \Sigma_2 \setminus \Sigma_1^{\mathrm{int}}, \\
\Sigma_{1\mathrm{i}} &\coloneqq \Sigma_2 \cap \overline{\Sigma_1^{\mathrm{int}}}, & \Sigma_{2\mathrm{i}} &\coloneqq \Sigma_1 \cap \overline{\Sigma_2^{\mathrm{int}}}
\end{aligned}$$

such that $\partial\big(\Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}\big) = \Sigma_{1\mathrm{e}} \cup \Sigma_{1\mathrm{i}}$ and $\partial\big(\Sigma_2^{\mathrm{int}} \setminus \Sigma_1^{\mathrm{int}}\big) = \Sigma_{2\mathrm{e}} \cup \Sigma_{2\mathrm{i}}$ (see Figure 7.4 for a sketch of the situation).

Then we obtain from Eq. (7.12) by using Lemma 7.18 that for every harmonic function f

$$I(f) \coloneqq \int_{\Sigma_{1\mathrm{e}} \cup \Sigma_{1\mathrm{i}}} f(x)\,x \cdot \nu(x)\,\mathrm{d}\omega(x) - \int_{\Sigma_{2\mathrm{e}} \cup \Sigma_{2\mathrm{i}}} f(x)\,x \cdot \nu(x)\,\mathrm{d}\omega(x) = 0, \tag{7.13}$$

where the outer normal fields ν on $\Sigma_{1\mathrm{e}} \cup \Sigma_{1\mathrm{i}}$ and $\Sigma_{2\mathrm{e}} \cup \Sigma_{2\mathrm{i}}$ are oriented with respect to the sets $\Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}$ and $\Sigma_2^{\mathrm{int}} \setminus \Sigma_1^{\mathrm{int}}$, respectively.

Suppose $\Sigma_1^{\mathrm{int}} \not\subseteq \Sigma_2^{\mathrm{int}}$. An application of Lemma 1.7.4 from Isakov [88] yields that there exists a sequence of harmonic functions $(f_k)_{k\in\mathbb{N}}$, $f_k \colon \Sigma_1^{\mathrm{int}} \cup \Sigma_2^{\mathrm{int}} \to \mathbb{R}$ such that $0 \leq f_k(x) \leq 1$ for all $x \in \Sigma_1^{\mathrm{int}} \cup \Sigma_2^{\mathrm{int}}$ and $f_k \to 1$ a. e. on $\Sigma_{1\mathrm{e}}$ and $f_k \to 0$ almost everywhere on $\Sigma_{2\mathrm{e}}$.

On the one hand, we obtain from Eq. (7.13) that

$$
\begin{aligned}
0 &= I(f_k) \\
&= \int_{\Sigma_{1\mathrm{e}}\cup\Sigma_{1\mathrm{i}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) - \int_{\Sigma_{2\mathrm{e}}\cup\Sigma_{2\mathrm{i}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) \\
&= \int_{\Sigma_{1\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) + \int_{\Sigma_{1\mathrm{i}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) \\
&\qquad - \int_{\Sigma_{2\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) - \int_{\Sigma_{2\mathrm{i}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) && (7.14) \\
&\geq \int_{\Sigma_{1\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) + \int_{\Sigma_{1\mathrm{i}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) \\
&\quad - \int_{\Sigma_{2\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) && (7.15) \\
&\geq \int_{\Sigma_{1\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) + \int_{\Sigma_{1\mathrm{i}}} x\cdot\nu(x)\,\mathrm{d}\omega(x) \\
&\quad - \int_{\Sigma_{2\mathrm{e}}} f_k(x)\, x\cdot\nu(x)\,\mathrm{d}\omega(x) && (7.16) \\
&\overset{k\to\infty}{\longrightarrow} \int_{\Sigma_{1\mathrm{e}}} x\cdot\nu(x)\,\mathrm{d}\omega(x) + \int_{\Sigma_{1\mathrm{i}}} x\cdot\nu(x)\,\mathrm{d}\omega(x) && (7.17) \\
&= \int_{\Sigma_{1\mathrm{e}}\cup\Sigma_{1\mathrm{i}}} x\cdot\nu(x)\,\mathrm{d}\omega(x)
\end{aligned}
$$

where Eq. (7.14) holds since $\Sigma_{1\mathrm{e}} \cap \Sigma_{1\mathrm{i}}$ and $\Sigma_{2\mathrm{e}} \cap \Sigma_{2\mathrm{i}}$ have measure zero. Eq. (7.15) is true since $f_k(x) \geq 0$ and $x \cdot \nu(x) \leq 0$ for $x \in \Sigma_{2\mathrm{i}}$, whereas Eq. (7.16) holds because $f_k(x) \leq 1$ and $x \cdot \nu(x) \leq 0$ for all $x \in \Sigma_{1\mathrm{i}}$. Note that we used the star-shapedness of $\Sigma_1^{\mathrm{int}} \cap \Sigma_2^{\mathrm{int}}$ with respect to the same point here to obtain that the outer normal $\nu(x)$ and the vector x need to point into opposite half space. Finally, Eq. (7.17) is fulfilled due to the dominated convergence theorem (the integrands are dominated by the integrable function $x \mapsto |x|$).

On the other hand, Lemma 7.18 for $f \equiv 1$ on $A = \Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}$ gives

$$
\int_{\Sigma_{1\mathrm{e}}\cup\Sigma_{1\mathrm{i}}} x\cdot\nu(x)\,\mathrm{d}x = 3\,\lambda\big(\Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}\big) \geq 0.
$$

Consequently,

$$
\lambda\big(\Sigma_1^{\mathrm{int}} \setminus \Sigma_2^{\mathrm{int}}\big) = 0
$$

such that $\Sigma_1^{\text{int}} \subseteq \Sigma_2^{\text{int}}$ since the sets are star-shaped. This is a contradiction to the assumption that $\Sigma_1^{\text{int}} \not\subseteq \Sigma_2^{\text{int}}$.

Similarly, one shows that $\Sigma_2^{\text{int}} \subseteq \Sigma_1^{\text{int}}$ and thus $\Sigma_1^{\text{int}} = \Sigma_2^{\text{int}}$ and consequently, $\Sigma_1 = \Sigma_2$. ■

A generalization of this theorem, whose proof uses the same technique as the proof of the previous theorem, is the following.

Theorem 7.20 (cf. Isakov [88, Theorem 3.1.1])**.** Suppose the assumptions of Theorem 7.19 are fulfilled, and additionally a mass density model $\varrho\colon S^{\text{int}} \to \mathbb{R}$ is given, for which

$$\varrho \in \mathrm{C}(S^{\text{int}} \setminus \{0\}), \quad \frac{\partial \varrho}{\partial r} \in \mathrm{C}(S^{\text{int}} \setminus \{0\}), \text{ and } \quad \varrho \frac{\partial}{\partial r}\left(r^3 \varrho\right) \geq 0 \text{ on } S^{\text{int}} \setminus \{0\}$$

hold.

If

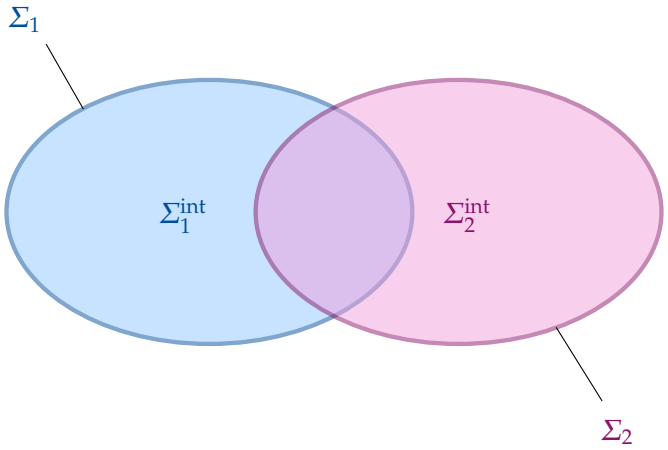
$$U_{\Sigma_1^{\text{int}},\varrho}|_{S^{\text{ext}}} = U_{\Sigma_2^{\text{int}},\varrho}|_{S^{\text{ext}}},$$

then $\Sigma_1 = \Sigma_2$.

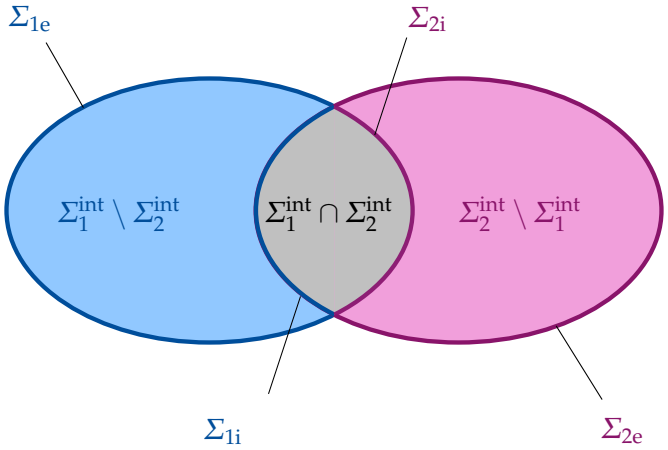
For several other uniqueness results with different assumptions on the shape of the domain and on the mass density model, we refer to Isakov [88, Chapter 3] and Isakov [89, Section 4.1].

In addition to the provided uniqueness results, Isakov [88] also gives an insight into the stability of the nonlinear inverse gravimetric problem. It turns out that the inversion of the operator $\mathcal{S}_\varrho$ is indeed stable (in theory), but that difficulties arise from the numerical perspective in this case. We will discuss this topic further after stating the theorem. In Theorem 7.7, we have already stated that it does not matter if the potential itself or its gradient is given as the right-hand side of the inverse problem. The original theorem deals with the problem, where the modulus of the gradient of the potential is given on a certain subset of a sphere with arbitrary center and radius. We will only present a simplified version such that the main idea of the result gets clear.

Theorem 7.21 (cf. Isakov [88, Theorem 3.6.1])**.** Let $S = \mathbb{S}^2$ and let two regular surfaces $\Sigma_1, \Sigma_2 \subseteq S^{\text{int}}$ be given, which are parametrized by functions $\sigma_1, \sigma_2\colon \mathbb{S}^2 \to (0,1)$ such that $\Sigma_1^{\text{int}}, \Sigma_2^{\text{int}}$ are star-shaped. Additionally, it is required that there exists a constant $h > 0$ such that $\sigma_1(\xi), \sigma_2(\xi) \in (h, 1-h)$ for all $\xi \in \mathbb{S}^2$ and that $\sigma_1, \sigma_2 \in \mathrm{C}^{(2)}(\mathbb{S}^2)$.



(a) The sets Σ_1^{int} and Σ_2^{int} are bounded by the regular surfaces Σ_1 and Σ_2, respectively.



(b) The sets $\Sigma_{1e}, \Sigma_{1i}, \Sigma_{2e}, \Sigma_{2i}$ are defined such that $\Sigma_1 = \Sigma_{1e} \cup \Sigma_{2i}$, $\Sigma_2 = \Sigma_{2e} \cup \Sigma_{1i}$. The boundaries of the difference sets are given by $\partial\left(\Sigma_1^{\text{int}} \setminus \Sigma_2^{\text{int}}\right) = \Sigma_{1e} \cup \Sigma_{1i}$ and $\partial\left(\Sigma_2^{\text{int}} \setminus \Sigma_1^{\text{int}}\right) = \Sigma_{2e} \cup \Sigma_{2i}$.

Figure 7.4.: Sketch of the sets arising in the proof of Theorem 7.19. Note that in this sketch, we choose to depict the special case of convex sets, which are star-shaped with respect to every point inside the sets itself. The origin, which is the point with respect to which both sets need to be star-shaped to fulfill the preconditions of the proof, can therefore be any of the points in the intersection.

Then there is a constant $C > 0$ such that if

$$\left|\left|\nabla U_{\Sigma_1^{\text{int}},1}(y)\right| - \left|\nabla U_{\Sigma_2^{\text{int}},1}(y)\right|\right| < \varepsilon, \qquad \text{for all } y \in S,$$

then,

$$|\sigma_1(\xi) - \sigma_2(\xi)| < C\,|\log \varepsilon|^{-1/C}, \qquad \text{for all } \xi \in \mathbb{S}^2. \tag{7.18}$$

Clearly, this is a stability estimate for the nonlinear inverse gravimetric problem, since it shows that the solution depends continuously on the data.

Unfortunately, the stability of the inverse problem can be described as being *weak* (cf. Isakov [89, Section 1.1]) due to the logarithmic nature of the estimate, which leads to "numerical difficulties" (Isakov [89]). The reason for these difficulties is the fact that the right-hand side of Eq. (7.18) changes dramatically near $\varepsilon = 0$. This can be quantified by the derivative of the function $f_C(\varepsilon) \coloneqq C\,|\log \varepsilon|^{-1/C}$, which is given by

$$\frac{\partial f_C}{\partial \varepsilon}(\varepsilon) = \frac{(-\log \varepsilon)^{-(1+C)/C}}{\varepsilon} = \left(\frac{\varepsilon^{-C/(1+C)}}{-\log \varepsilon}\right)^{(1+C)/C}$$

if $0 < \varepsilon < 1$. Since both the numerator and the denominator of the fraction on the right-hand side tend to infinity for $\varepsilon \to 0$, we obtain by L'Hospital's rule that

$$\begin{aligned}\lim_{\varepsilon \searrow 0} \frac{\varepsilon^{-C/(1+C)}}{-\log \varepsilon} &= \lim_{\varepsilon \searrow 0} \frac{\frac{C}{1+C}\varepsilon^{-C/(1+C)-1}}{\varepsilon^{-1}} \\ &= \lim_{\varepsilon \searrow 0} \frac{C}{1+C}\varepsilon^{-C/(1+C)} = \infty\end{aligned}$$

where the last equality is true since the power $-C/(1+C)$ is negative. Consequently,

$$\lim_{\varepsilon \searrow 0} \frac{\partial f_C}{\partial \varepsilon}(\varepsilon) = \infty.$$

This shows that if one restricts the image space of the operator in such a way that it is surjective (and thus bijective by the uniqueness theorem), the inverse problem is no longer ill-posed, but still ill-conditioned from the numerical perspective (cf. Isakov [90, Section 2]). In consequence, a regularization of the inverse problem is unavoidable to obtain a stable solution. In consequence, in Section 10.3 we will apply the newly developed Regularized Functional Matching Pursuit algorithm for nonlinear inverse problems to the nonlinear inverse gravimetric problem.

In conclusion, by addressing the nonlinear inverse gravimetric problem instead of the linear inverse gravimetric problem, the problem gets more difficult because of the nonlinearity. On the other hand, the solution of the nonlinear problem is both unique and stable (at least in theory), which is advantageous and beneficial for the numerical solution of the inverse problem.

Note that one can also formulate the nonlinear inverse gravimetric problem as an inverse source problem for a partial differential equation, namely the Poisson equation. This was, for example, done by Hettlich and Rundell [78]. Other inverse source problems of this type are related to the heat equation (see, e. g., Hettlich and Rundell [79]) and the Helmholtz equation (see, e. g., Elschner and Yamamoto [41] and Hettlich and Rundell [80]).

7.2.3. Gâteaux and Fréchet differentiability of the nonlinear operator

Many algorithms for nonlinear inverse problems use either the Fréchet or the Gâteaux derivative of the involved operator. In this section, we will compute the Gateaux derivative of $\mathcal{S}_\varrho$, which was defined in Problem 7.16.

To compute the Gâteaux derivative, we will use the following special case of Leibniz' rule for differentiation of integrals (cf. Holmes [84, Theorem 6.2]).

Lemma 7.22. Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be sufficiently smooth. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^{g(t)} f(x)\,\mathrm{d}x = f(g(t))\,g'(t).$$

The Gâteaux derivative $\mathcal{S}'_\varrho[\sigma](\tau)\colon \mathrm{C}(\mathbb{S}^2) \to \mathrm{C}(S)$ at $\sigma \in \mathrm{C}(\mathbb{S}^2)$ in the direction $\tau \in \mathrm{C}(\mathbb{S}^2)$ can now be obtained by an application of the preceding lemma to the expression in Eq. (7.7):

$$\begin{aligned}
\mathcal{S}'_\varrho[\sigma](\tau)(y) &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathcal{S}_\varrho[\sigma + \varepsilon\tau]\,(y)\bigg|_{\varepsilon=0} \\
&\sim \int_{\mathbb{S}^2} \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \int_0^{\sigma(\xi)+\varepsilon\tau(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|} r^2\,\mathrm{d}r\bigg|_{\varepsilon=0} \mathrm{d}\omega(\xi) \\
&= \int_{\mathbb{S}^2} \frac{\varrho((\sigma(\xi) + \varepsilon\tau(\xi))\,\xi)}{|(\sigma(\xi) + \varepsilon\tau(\xi))\,\xi - y|} (\sigma(\xi) + \varepsilon\tau(\xi))^2\,\tau(\xi)\bigg|_{\varepsilon=0} \mathrm{d}\omega(\xi) \\
&= \int_{\mathbb{S}^2} \frac{\varrho(\sigma(\xi)\,\xi)}{|\sigma(\xi)\,\xi - y|} (\sigma(\xi))^2\,\tau(\xi)\,\mathrm{d}\omega(\xi) \qquad (7.19)
\end{aligned}$$

for all $y \in S$. The symbol $\sim$ in the second line should indicate that we assumed that the interchanging of differentiation and integration is possible. Of course, this

would have to be proved. Since we prove in the following that the term in Eq. (7.19) is not only the Gâteaux, but also the Fréchet derivative under certain assumptions, we can omit this proof, since the Fréchet derivative is always also the Gâteaux derivative.

Theorem 7.23. Let $\varrho \in \mathrm{C}^{(1)}(S^{\mathrm{int}})$. Then the Fréchet derivative of the operator $\mathcal{S}_\varrho \colon \mathrm{C}(\mathbb{S}^2) \to \mathrm{C}(S)$ is given as

$$\mathcal{S}'_\varrho[\sigma](\tau)(y) = \int_{\mathbb{S}^2} \frac{\varrho(\sigma(\xi)\,\xi)}{|\sigma(\xi)\,\xi - y|} (\sigma(\xi))^2\, \tau(\xi)\, \mathrm{d}\omega(\xi)$$

for all $\sigma, \tau \in \mathrm{C}(\mathbb{S}^2)$ and $y \in S$, assuming there exists $C > 0$ such that $|\sigma(\xi)\xi - y| > C$ for all $\xi \in \mathbb{S}^2$ and $y \in S$ (i. e., $\Sigma \subseteq S^{\mathrm{int}}$).

Proof. For the sake of brevity, we define

$$k(x, y) := \frac{\varrho(x)}{|x - y|} |x|^2$$

such that

$$\mathcal{S}_\varrho[\sigma]\,(y) = \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi).$$

Note that

$$\nabla_x k(x, y) = \frac{|x|^2}{|x - y|} \nabla_x \varrho(x) + 2 \frac{\varrho(x)}{|x - y|} x - \varrho(x) |x|^2 \frac{x - y}{|x - y|^3}$$

and

$$|\nabla_x k(x, y)| \leq \frac{R^2_{S^{\mathrm{int}}}}{C} \|\varrho\|_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)} + 2 \frac{\|\varrho\|_{\mathrm{C}(S^{\mathrm{int}}_*)}}{C} R_{S^{\mathrm{int}}} + \|\varrho\|_{\mathrm{C}(S^{\mathrm{int}}_*)} R^2_{S^{\mathrm{int}}} \frac{1}{C^2} < \infty \tag{7.20}$$

for all $x \in S^{\mathrm{int}}_*$ and $y \in S$, where

$$S^{\mathrm{int}}_* := \left\{ x \in S^{\mathrm{int}} \;\middle|\; |x - y| > C \text{ for all } y \in S \right\}$$

and $R_{S^{\mathrm{int}}} := \max_{y \in S} |y|$. Consequently,

$$\|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)} < \infty$$

and since the last term in Eq. (7.20) does not depend on y, we even have

$$\sup_{y \in S} \|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)} < \infty. \tag{7.21}$$

Let $y \in S$ be fixed. Then, if we consider the term from the definition of the Fréchet derivative, we obtain for sufficiently small $\tau \in \mathrm{C}(\mathbb{S}^2)$ that

$$\begin{aligned}
&\left| \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) - \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) \right. \\
&\quad \left. - \int_{\mathbb{S}^2} k(\sigma(\xi)\xi, y)\, \tau(\xi)\, \mathrm{d}\omega(\xi) \right| && (7.22) \\
&= \left| \int_{\mathbb{S}^2} \int_{\sigma(\xi)}^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) - \int_{\mathbb{S}^2} k(\sigma(\xi)\xi, y)\, \tau(\xi)\, \mathrm{d}\omega(\xi) \right| \\
&= \left| \int_{\mathbb{S}^2} \int_{\sigma(\xi)}^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r - k(\sigma(\xi)\xi, y)\, \tau(\xi)\, \mathrm{d}\omega(\xi) \right| \\
&\leq \int_{\mathbb{S}^2} \left| \int_{\sigma(\xi)}^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r - k(\sigma(\xi)\xi, y)\, \tau(\xi) \right| \mathrm{d}\omega(\xi) && (7.23) \\
&= \int_{\mathbb{S}^2} \left| \int_0^{\tau(\xi)} k((\sigma(\xi)+r)\xi, y)\, \mathrm{d}r - k(\sigma(\xi)\xi, y)\, \tau(\xi) \right| \mathrm{d}\omega(\xi) && (7.24) \\
&= \int_{\mathbb{S}^2} |k((\sigma(\xi)+\bar{r})\xi, y)\tau(\xi) - k(\sigma(\xi)\xi, y)\, \tau(\xi)|\, \mathrm{d}\omega(\xi) && (7.25) \\
&= \int_{\mathbb{S}^2} |k((\sigma(\xi)+\bar{r})\xi, y) - k(\sigma(\xi)\xi, y)||\tau(\xi)|\, \mathrm{d}\omega(\xi) \\
&= \int_{\mathbb{S}^2} \left| \bar{r} \left[\frac{\partial}{\partial r} k((\sigma(\xi)+r)\xi, y) \right]_{r=\hat{r}} \right| |\tau(\xi)|\, \mathrm{d}\omega(\xi) && (7.26) \\
&\leq \|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S_*^{\mathrm{int}})} \|\tau\|^2_{\mathrm{L}^2(\mathbb{S}^2)} && (7.27) \\
&\leq 4\pi \|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S_*^{\mathrm{int}})} \|\tau\|^2_{\mathrm{C}(\mathbb{S}^2)}, && (7.28)
\end{aligned}$$

where we used the triangle inequality in Eq. (7.23), a substitution in Eq. (7.24), the existence of $\bar{r} \in [0, \tau(\xi)]$ such that the equality holds due to the intermediate value theorem for integrals in Eq. (7.25), and the existence of $\hat{r} \in [0, \bar{r}]$, such that the identity holds due to the intermediate value theorem for differentiation in Eq. (7.26). Furthermore, we employed $\bar{r} \leq |\tau(\xi)|$ and Eq. (7.20) in Eq. (7.27) and Theorem 2.17 in Eq. (7.28). Note that we have to require τ to be sufficiently small such that $|(\sigma(\xi) + \tau(\xi))\xi - y| > C$ holds, implying a finite $\mathrm{C}^{(1)}(S_*^{\mathrm{int}})$-norm of k.

It follows that

$$\begin{aligned}
&\sup_{y \in S} \left| \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) - \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) \right. \\
&\qquad \left. - \int_{\mathbb{S}^2} k(\sigma(\xi)\xi, y)\, \tau(\xi)\, \mathrm{d}\omega(\xi) \right| \\
&\leq 4\pi \sup_{y \in S} \|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S^{\mathrm{int}})} \|\tau\|^2_{\mathrm{C}(\mathbb{S}^2)},
\end{aligned}$$

which proves the assertion since the latter term tends to 0 even if it is divided by $\|\tau\|_{\mathrm{C}(\mathbb{S}^2)}$ and the term in Eq. (7.22) is exactly the term arising in the definition of the Fréchet derivative. ■

Often, one would like to apply Hilbert space techniques in the analysis and numerical solution of inverse problems. Up to now, we have only considered the operator $\mathcal{S}_\varrho$ as an operator $\mathrm{C}(\mathbb{S}^2) \to \mathrm{C}(S)$. To consider it as an operator $\mathrm{L}^2(\mathbb{S}^2) \to \mathrm{L}^2(S)$, we have to ensure the existence of the integrals in the definition of the operator. Furthermore, we can prove that the image is an $\mathrm{L}^2(S)$-function.

Theorem 7.24. Let $\varrho \in \mathrm{L}^\infty(S^{\mathrm{int}})$ and let $\sigma \in \mathrm{L}^2(\mathbb{S}^2)$ such that $|\sigma(\xi)\xi - y| > C$ for almost all $\xi \in \mathbb{S}^2$, almost all $y \in S$, and some constant $C > 0$. Then, we have for almost all $y \in S$ that

$$\int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|} r^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi) < \infty.$$

Furthermore, $\mathcal{S}_\varrho[\sigma] \in \mathrm{L}^2(S)$.

Proof. Let k, S_*^{int}, and $R_{S^{\mathrm{int}}}$ be defined as in the proof of the previous theorem. Then, we have

$$|k(x,y)| = \frac{|\varrho(x)|}{|x-y|}|x|^2 \leq \frac{\|\varrho\|_{\mathrm{L}^\infty(S^{\mathrm{int}})}}{C} R_{S^{\mathrm{int}}}^2 < \infty \tag{7.29}$$

for almost all $(x,y) \in S_*^{\mathrm{int}} \times S$ such that

$$\|k(\cdot,y)\|_{\mathrm{L}^\infty(S_*^{\mathrm{int}})} < \infty.$$

Thus,

$$\begin{aligned}
\left| \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y) \,\mathrm{d}r \,\mathrm{d}\omega(\xi) \right| &\leq \int_{\mathbb{S}^2} \|k(\cdot,y)\|_{\mathrm{L}^\infty(S_*^{\mathrm{int}})} \left| \int_0^{\sigma(\xi)} \mathrm{d}r \right| \mathrm{d}\omega(\xi) \\
&\leq \|k(\cdot,y)\|_{\mathrm{L}^\infty(S_*^{\mathrm{int}})} \|\sigma\|_{\mathrm{L}^1(\mathbb{S}^2)} \\
&\leq \sqrt{4\pi} \|k(\cdot,y)\|_{\mathrm{L}^\infty(S_*^{\mathrm{int}})} \|\sigma\|_{\mathrm{L}^2(\mathbb{S}^2)} < \infty,
\end{aligned}$$

for almost all $y \in S$, which proves the first assertion.

Since the penultimate term in Eq. (7.29) does not depend on y, we observe that

$$\begin{aligned}
\int_S \left(\|k(\cdot,y)\|_{\mathrm{L}^\infty(S_*^{\mathrm{int}})} \right)^2 \mathrm{d}\omega(y) &\leq \int_S \frac{\|\varrho\|_{\mathrm{L}^\infty(S^{\mathrm{int}})}^2}{C^2} R_{S^{\mathrm{int}}}^4 \,\mathrm{d}\omega(y) \\
&= \omega(S) \frac{\|\varrho\|_{\mathrm{L}^\infty(S^{\mathrm{int}})}^2}{C^2} R_{S^{\mathrm{int}}}^4 < \infty
\end{aligned}$$

such that $\mathcal{S}_\varrho[\sigma] \in \mathrm{L}^2(S)$. ■

Using the same technique as in Theorem 7.23, we can prove the Fréchet differentiability of the operator $\mathcal{S}_\varrho \colon \mathrm{L}^2(\mathbb{S}^2) \to \mathrm{L}^2(S)$.

Theorem 7.25. Let $\varrho \in \mathrm{C}^{(1)}(S^{\mathrm{int}})$. Then, the Fréchet derivative of the operator $\mathcal{S}_\varrho \colon \mathrm{L}^2(\mathbb{S}^2) \to \mathrm{L}^2(S)$ is given as

$$\mathcal{S}'_\varrho[\sigma](\tau)(y) = \int_{\mathbb{S}^2} \frac{\varrho(\sigma(\xi)\,\xi)}{|\sigma(\xi)\,\xi - y|} (\sigma(\xi))^2\, \tau(\xi)\, \mathrm{d}\omega(\xi)$$

for all $\sigma, \tau \in \mathrm{L}^2(\mathbb{S}^2)$ and $y \in S$, assuming there exists $C > 0$ such that $|\sigma(\xi)\xi - y| > C$ for almost all $\xi \in \mathbb{S}^2$ and $y \in S$.

Proof. Define k, S^{int}_*, and $R_{S^{\mathrm{int}}}$ as in the proof of Theorem 7.23. From Eq. (7.21), we obtain that

$$\int_S \|k(\cdot, y)\|^2_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)}\, \mathrm{d}\omega(y) \leq \omega(S) \sup_{y \in S} \|k(\cdot, y)\|^2_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)} < \infty. \tag{7.30}$$

The inequalities in Eqs. (7.22)–(7.27) are still true such that

$$\begin{aligned} &\left| \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)+\tau(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) - \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y)\, \mathrm{d}r\, \mathrm{d}\omega(\xi) \right. \\ &\qquad \left. - \int_{\mathbb{S}^2} k(\sigma(\xi)\xi, y)\, \tau(\xi)\, \mathrm{d}\omega(\xi) \right| \\ &\leq \|k(\cdot, y)\|_{\mathrm{C}^{(1)}(S^{\mathrm{int}}_*)} \|\tau\|^2_{\mathrm{L}^2(\mathbb{S}^2)} \end{aligned}$$

for almost all $y \in S$. An application of Eq. (7.30) proves the assertion. ■

7.2.4. The determination of boundary layers and topographies

We will briefly discuss two problems that are related to the nonlinear inverse gravimetric problem, namely the determination of boundary layers and the determination of topographies (i. e., the boundary of a body of mass deviates from a prescribed model). It will be shown that, under certain assumptions, both problems can be reduced to a modified version of Problem 7.16.

The first problem to solve is the determination of boundary layers inside a body of mass.

Problem 7.26. Let $S \subseteq \mathbb{R}^3$ be a regular surface and let $\Sigma_1 \subseteq S^{\mathrm{int}}$ be a given star-shaped regular surface such that Σ_1^{int} is a body of mass (e. g., the Earth). Assume that there exists a boundary layer inside Σ_1^{int} that is described by a regular surface $\Sigma_2 \subseteq \Sigma_1^{\mathrm{int}}$ such that Σ_2^{int} is star-shaped and a mass density model inside Σ_2^{int} is given

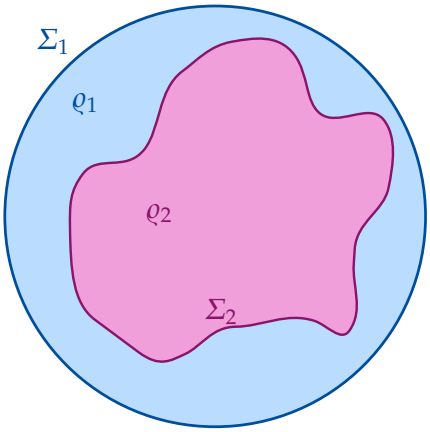


Figure 7.5.: Sketch of the situation in Problem 7.26: There is a boundary layer Σ_2 (purple) inside a body of mass, whose boundary is given by Σ_1 (blue). The mass density in the purple region is given by ϱ_2, whereas the mass density in the blue region is given by ϱ_1.

by a function $\varrho_2 \colon S^{\text{int}} \to \mathbb{R}$ and a mass density model inside $\Sigma_1^{\text{int}} \setminus \overline{\Sigma_2^{\text{int}}}$ is given by a function $\varrho_1 \colon S^{\text{int}} \to \mathbb{R}$ (see also the sketch in Figure 7.5).

Given a function $g \in \mathrm{L}^2(S)$, find $\sigma_2 \colon \mathbb{S}^2 \to (0, \infty)$ that parametrizes Σ_2 like in Eq. (7.6) such that

$$U_{\Sigma_1^{\text{int}}, \varrho_1 \chi_{\Sigma_1^{\text{int}} \setminus \overline{\Sigma_2^{\text{int}}}} + \varrho_2 \chi_{\Sigma_2^{\text{int}}}} |_S = g.$$

To obtain a formulation equivalent to Problem 7.16, we take a look at the (almost everywhere) pointwise formulation of the problem. For almost every $y \in S$, we have

$$\begin{aligned} g(y) &= U_{\Sigma_1^{\text{int}}, \varrho_1 \chi_{\Sigma_1^{\text{int}} \setminus \overline{\Sigma_2^{\text{int}}}} + \varrho_2 \chi_{\Sigma_2^{\text{int}}}}(y) \\ &= \int_{\Sigma_1^{\text{int}}} \frac{\varrho_1(x) \chi_{\Sigma_1^{\text{int}} \setminus \overline{\Sigma_2^{\text{int}}}}(x) + \varrho_2(x) \chi_{\Sigma_2^{\text{int}}}(x)}{|x - y|} \, \mathrm{d}x \\ &= \int_{\Sigma_1^{\text{int}}} \frac{\varrho_1(x) \left(\chi_{\Sigma_1^{\text{int}}}(x) - \chi_{\overline{\Sigma_2^{\text{int}}}}(x) \right) + \varrho_2(x) \chi_{\Sigma_2^{\text{int}}}(x)}{|x - y|} \, \mathrm{d}x \qquad (7.31) \\ &= \int_{\Sigma_1^{\text{int}}} \frac{\varrho_1(x)}{|x - y|} \, \mathrm{d}x + \int_{\Sigma_2^{\text{int}}} \frac{\varrho_2(x) - \varrho_1(x)}{|x - y|} \, \mathrm{d}x, \end{aligned}$$

where in Eq. (7.31) we used $\chi_{A \setminus B} = \chi_A - \chi_B$ if $B \subseteq A$. From these considerations, we obtain that by introducing the density difference function

$$\varrho_{\text{diff}} \colon S^{\text{int}} \to \mathbb{R}, \qquad \varrho_{\text{diff}}(x) \coloneqq \varrho_2(x) - \varrho_1(x) \qquad \text{for } x \in S^{\text{int}},$$

and a modified right-hand side

$$\tilde{g} \in \mathrm{L}^2(S), \qquad \tilde{g}(y) := g(y) - \int_{\Sigma_1^{\mathrm{int}}} \frac{\varrho_1(x)}{|x-y|}\,\mathrm{d}x \qquad \text{for } y \in S,$$

Problem 7.26 is equivalent to the problem of finding σ_2 such that

$$U_{\Sigma_2^{\mathrm{int}},\varrho_{\mathrm{diff}}}|_S = \tilde{g},$$

which is exactly Problem 7.16 for a particular choice of the body and the mass density.

The nonlinear inverse gravimetric problem has indeed often been used in the literature to determine boundary layers inside the Earth, in particular, the so-called Mohorovičić discontinuity, which is the boundary between the crust and the mantle (see Clauser [29, Section 1.5]). The corresponding publications are mostly written from the geoscientific point of view, for example, see Aitken [2], Chappell and Kusznir [27], Guimera et al. [69], Hsieh and Yen [85], and Reguzzoni and Sampietro [149, 150].

The second problem that we want to discuss is the following one, where we assume that it is known that the body of mass is already well approximated by a ball of radius $R > 0$. The topography of the surface can thus be described by the deviation from the sphere with radius R, that is, by a function $\tilde{\sigma}\colon \mathbb{S}^2 \to \mathbb{R}$, $\tilde{\sigma}(\xi) = \sigma(\xi) - R$ for $\xi \in \mathbb{S}^2$, where σ is the parametrization always used above.

Problem 7.27. Let $S \subseteq \mathbb{R}^3$ be a regular surface, let $\varrho \in \mathrm{L}^2(S)$ be a mass density model, and let a function $g \in \mathrm{L}^2(S^{\mathrm{int}})$ be given. Assume that the boundary of the unknown body of mass is given by a regular surface $\tilde{\Sigma} \subseteq S^{\mathrm{int}}$, which is parametrized by a function $\tilde{\sigma}\colon \mathbb{S}^2 \to \mathbb{R}$ such that

$$\tilde{\Sigma} := \left\{ r\xi \in \mathbb{R}^3 \;\middle|\; \xi \in \mathbb{S}^2, r = R + \tilde{\sigma}(\xi) \right\}$$

for some radius $R > 0$.

Find $\tilde{\sigma}$ such that

$$U_{\tilde{\Sigma}^{\mathrm{int}},\varrho}|_S = g.$$

First, it is obvious that the preceding problem is equivalent to Problem 7.16 if the relation $\sigma(\xi) = R + \tilde{\sigma}(\xi)$ for $\xi \in \mathbb{S}^2$ is used. Using polar coordinates this can be further simplified to obtain

$$g(y) = \int_{\tilde{\Sigma}^{\mathrm{int}}} \frac{\varrho(x)}{|x-y|}\,\mathrm{d}x$$

$$
\begin{aligned}
&= \int_{\mathbb{S}^2} \int_0^{R+\tilde{\sigma}(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi) \\
&= \int_{\mathbb{S}^2} \int_0^{R} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r + \int_R^{R+\tilde{\sigma}(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|}\, r^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi) \\
&= \int_{\mathbb{B}_R} \frac{\varrho(x)}{|x-y|} \,\mathrm{d}x + \int_{\mathbb{S}^2} \int_0^{\tilde{\sigma}(\xi)} \frac{\varrho((R+r)\xi)}{|(R+r)\xi - y|} (R+r)^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi)
\end{aligned}
$$

for $y \in S$. Thus, by introducing the modified right-hand side

$$
\tilde{g} \in \mathrm{L}^2(S), \qquad \tilde{g}(y) := g(y) - \int_{\mathbb{B}_R} \frac{\varrho(x)}{|x-y|} \,\mathrm{d}x \qquad \text{for } y \in S,
$$

Problem 7.27 is equivalent to the nonlinear integral equation

$$
\int_{\mathbb{S}^2} \int_0^{\tilde{\sigma}(\xi)} \frac{\varrho((R+r)\xi)}{|(R+r)\xi - y|} (R+r)^2 \,\mathrm{d}r \,\mathrm{d}\omega(\xi) = \tilde{g}(y) \tag{7.32}
$$

for all $y \in S$. This is not exactly a reduction of the problem to Problem 7.16, but the only difference in the integral operators in Eq. (7.7) and in Eq. (7.32) are the terms r (in Eq. (7.7)) and $R + r$ (in Eq. (7.32)). Thus, if an implementation of the operator in Eq. (7.7) is already available, it is easy to modify it to represent the operator in Eq. (7.32).

7.2.5. Euclidean formulation of the nonlinear inverse gravimetric problem

In this section, we will give a short overview over a different formulation of the inverse gravimetric problem, which we call *Euclidean formulation* in contrast to the *spherical formulation* given in Problem 7.16. We call the latter a spherical formulation, since the boundary of the star-shaped body of mass is given by a function on the sphere. In the Euclidean setting, one assumes that the unknown body $\mathcal{E}$ can be described in Euclidean coordinates by

$$
\mathcal{E} = \left\{ (x,y,z) \in \mathbb{R}^3 \;\middle|\; -H \le z \le \sigma(x,y) - H, \quad x,y \in \Omega \right\},
$$

where $H > 0$ is a constant representing a *depth*, $\Omega \subseteq \mathbb{R}^2$ is a bounded domain, and the function σ fulfills

$$
\begin{aligned}
0 < \sigma(x,y) < H \qquad & \text{for all } (x,y) \in \Omega \text{ and} \\
\sigma(x,y) = 0 \qquad & \text{for all } (x,y) \in \partial\Omega
\end{aligned}
$$

such that $\mathcal{E}$ can be described as a body of matter that is bounded in the (x,y)-directions by a given domain Ω. In the z-direction it is bounded by a plane that is parallel to the x-y-plane with the depth $-H$ on the bottom, and the graph of the function $-H + \sigma$ at the top. For a sketch of the situation, see Figure 7.6.

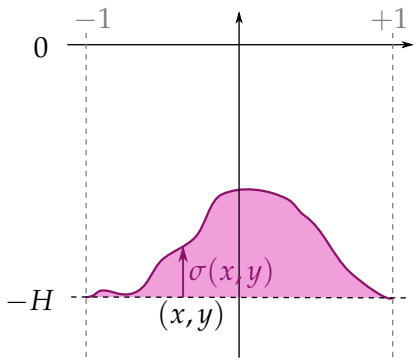


Figure 7.6.: Two-dimensional slice of the situation in the Euclidean formulation for $\Omega = [-1,1]^2$: the function σ describes the deviation of the body of mass from the plane at the depth $-H$.

The corresponding inverse problem is based on the assumption that the gradient of the gravitational potential is measured at height 0 in the set Ω, which for a constant density $\varrho = 1$ results in the nonlinear integral equation

$$\int_\Omega \frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + (\sigma(x',y') - H)^2}} \,\mathrm{d}x'\,\mathrm{d}y' = f(x,y), \tag{7.33}$$

where $f\colon \Omega \to \mathbb{R}$ is related to the derivative of the gravitational potential in the z-direction (cf. Ang et al. [7]). Due to the included derivative in this formulation, the integral in the z-direction can be eliminated by using the fundamental theorem of calculus, such that Eq. (7.33) contains only a two-dimensional integral. In Ang et al. [7], a uniqueness result for the inverse problem in Eq. (7.33) is proved.

This problem, also in a lower-dimensional version using the logarithmic potential, has been used as a prototype for a nonlinear integral equation extensively both in mathematics and the applications, for example by Airapetyan et al. [1], Akimova et al. [3], Akimova and Vasin [4], Ang et al. [7–9], Bakushinsky and Kokurin [11], Bakushinsky et al. [12], Bertete-Aguirre et al. [20], Diniz-Ehrhardt et al. [35], Haber and Oldenburg [72], Imomnazarov et al. [87], Korolev [104], Loukianov [111], Misici and Zirilli [131], Ramm and Smirnova [147], Reguzzoni and Sampietro [149], Richter [152], Seidman and Vogel [161], Shubha et al. [163], Tikhonov and Glasko [171], Vasin [173–175], Vasin and Perestoronina [176], Vasin and Skorik [177], and Wang and Yuan [182].

We mentioned this problem in this context here, because it is often called *the* inverse gravimetric problem in the literature mentioned above and we want to make clear what the connection is between both formulations of the nonlinear inverse gravimetric problem.

Chapter 8.

Regularized Functional Matching Pursuit (RFMP)

In Berkel et al. [18], Fischer [45], Fischer and Michel [46–48], Michel [123], Michel and Telschow [127, 128], and Telschow [166] several variants of a novel algorithm for the regularization of linear inverse problems, the *Regularized Functional Matching Pursuit (RFMP)*, have been developed.

In this chapter, we will present the basic form of this algorithm as it is described in Michel [123] and give previously achieved theoretical results concerning the convergence of the algorithm. This will be the basis for the considerations in the following chapters, where we develop new modifications of this algorithm.

8.1. Problem setting

The algorithm deals with the solution of Problem 6.3, where $\mathcal{Y} = \mathbb{R}^\ell$ for $\ell \in \mathbb{N}$, corresponding to the most common setting in practical applications, where only a finite amount of (possibly noisy) measurements is available. We do not assume that all measurements are of the same physical quantity and thereby related to the same specific operator. Instead, the operator $\mathcal{T}: \mathcal{X} \to \mathbb{R}^\ell$ can be comprised of distinct functionals $\mathcal{T}_1, \ldots, \mathcal{T}_\ell: \mathcal{X} \to \mathbb{R}$ such that

$$\mathcal{T}(f) = \begin{pmatrix} \mathcal{T}_1(f) \\ \vdots \\ \mathcal{T}_\ell(f) \end{pmatrix} \in \mathbb{R}^\ell, \qquad f \in \mathcal{X}.$$

It is therefore possible to perform a joint inversion of different data types with this algorithm, which has been a challenge in the field of applied inverse problems and, in particular, in geophysics since the 1970s (see, for example, Haber and Oldenburg [71], Vozoff and Jupp [180], and Yokota et al. [185]).

Note that if $\mathcal{T}_k$ is a bounded functional for every $k = 1, \ldots, \ell$, then $\mathcal{T}$ is a linear and bounded operator.

In the following section we will first derive the so-called *Functional Matching Pursuit (FMP)*, followed by the *Regularized Functional Matching Pursuit (RFMP)*, which incorporates a regularization term, which is needed due to the possibility of ill-posedness of Problem 6.3. Although the FMP is a particular case of the RFMP, we introduce both algorithms separately, since similarities of the FMP to the greedy algorithm from Chapter 3 can be drawn. For both algorithms, we prescribe a dictionary $\mathcal{D} \subseteq \mathcal{X}$ and iteratively generate a sequence

$$\left(f_K = f_0 + \sum_{k=1}^{K} \alpha_k d_k \right)_{K \in \mathbb{N}_0}$$

of approximations to a solution $f \in \mathcal{X}$, where $\alpha_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$ for all $k \in \mathbb{N}_0$. An initial approximation is given by $f_0 \in \mathcal{X}$.

8.2. Functional Matching Pursuit

In Fischer [45] and Fischer and Michel [46], the FMP was derived by explicitly carrying out the iterative minimization of the least squares functional

$$\mathcal{A}_{0,g}[f_k + \alpha d] = \|g - \mathcal{T}(f_k + \alpha d)\|_{\mathbb{R}^\ell}^2$$

for $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$, given $f_k \in \mathcal{X}$ and data $g \in \mathcal{Y}$. The same strategy will be used in Chapter 10 to obtain a related algorithm for nonlinear inverse problems, the RFMP_NL.

Here, we present a different approach based on the results of Chapter 3, which yields the same algorithm. The idea is to apply Algorithm 3.1 (the Pure Greedy Algorithm, PGA) in the range $\mathcal{Y}$ of the operator $\mathcal{T}$.

The function to approximate by the PGA is consequently the data $g \in \mathcal{Y}$. The dictionary for the PGA needs to be normalized and the elements need to be from the image space, so we choose $\{ \mathcal{T}d / \|\mathcal{T}d\|_{\mathcal{Y}} \mid d \in \mathcal{D} \}$ as the dictionary, such that we need to assume $d \notin \operatorname{null} \mathcal{T}$.

If we then apply the PGA, we obtain a sequence $(g_K)_K \subseteq \mathcal{Y}$ of approximations of g, which have the form

$$g_k = \mathcal{T}f_0 + \sum_{k=1}^{K} \tilde{\alpha}_k \frac{\mathcal{T}d_k}{\|\mathcal{T}d_k\|_Y},$$

where the dictionary elements $d_k \in \mathcal{D}$ fulfill

$$d_k = \operatorname*{argmax}_{d \in \mathcal{D}} \left| \left\langle g - g_k, \frac{\mathcal{T}d}{\|\mathcal{T}d\|_{\mathcal{Y}}} \right\rangle_{\mathcal{Y}} \right|$$

and the coefficients $\tilde{\alpha}_k \in \mathbb{R}$ are given by

$$\tilde{\alpha}_k = \left\langle g - g_k, \frac{\mathcal{T} d_k}{\|\mathcal{T} d_k\|_{\mathcal{Y}}} \right\rangle_{\mathcal{Y}}$$

for $k = 1, \ldots, K$. If we include the normalization of $\mathcal{T} d_k$ in the coefficient such that $\alpha_k := \frac{\tilde{\alpha}_k}{\|\mathcal{T} d_k\|_{\mathcal{Y}}}$, this yields the following algorithm, which is identical to the FMP in Fischer [45], Fischer and Michel [46], and Michel [123].

Algorithm 8.1 (Functional Matching Pursuit, FMP)**.** Let $\mathcal{T}$ and g be given as in Problem 6.3. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X} \setminus \{0\}$ and an initial approximation $f_0 \in \mathcal{X}$.

1. Set $k := 0$, define the residual $r_0 := g - \mathcal{T} f_0$ and choose a stopping criterion.
2. Find

$$d_{k+1} := \operatorname*{argmax}_{d \in \mathcal{D}} \left| \frac{\langle r_k, \mathcal{T} d \rangle_{\mathcal{Y}}}{\|\mathcal{T} d\|_{\mathcal{Y}}} \right| \tag{8.1}$$

   and set

$$\alpha_{k+1} := \frac{\langle r_k, \mathcal{T} d_{k+1} \rangle_{\mathcal{Y}}}{\|\mathcal{T} d_{k+1}\|_{\mathcal{Y}}^2},$$

   as well as $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$ and $r_{k+1} := g - \mathcal{T}(f_{k+1}) = r_k - \alpha_{k+1} \mathcal{T}(d_{k+1})$.
3. If the stopping criterion is fulfilled, then f_{k+1} is the output. Otherwise, increase k by 1 and return to step 2.

As a direct consequence of Theorem 3.2, we obtain the following theorem.

Theorem 8.2 (Convergence of the residuals)**.** Let $\mathcal{T}$ and g be given as in Problem 6.3 and choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$ such that

$$\overline{\operatorname{span} \mathcal{T}(\mathcal{D})} = \mathbb{R}^\ell \qquad \text{and} \qquad \operatorname{null} \mathcal{T} \cap \mathcal{D} = \emptyset.$$

Then the sequence $(r_k)_{k \in \mathbb{N}_0}$ of residuals converges and

$$\lim_{k \to \infty} r_k = 0.$$

The convergence of the sequence of iterates $(f_k)_{k \in \mathbb{N}_0}$ cannot be proved using the results from Chapter 3. Nevertheless, a proof for this fact is given in Fischer [45], this proof has been extended to the regularized setting in Michel [123] and it has been improved further by Michel and Orzlowski [126]. Since the result for the FMP is just a particular case of the result for the RFMP, we refer to the following section.

8.3. Regularized Functional Matching Pursuit

Analogously to the FMP, the RFMP was derived in Fischer [45] and Fischer and Michel [46] by iteratively minimizing the Tikhonov functional

$$\mathcal{A}_{\lambda,g}[f_k + \alpha d] = \|g - \mathcal{T}(f_k + \alpha d)\|_{\mathbb{R}^\ell}^2 + \lambda \|f_k + \alpha d\|_{\mathcal{X}}^2.$$

for $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$, given $f_k \in \mathcal{X}$ and data $g \in \mathcal{Y}$. Here, $\lambda > 0$ is a regularization parameter. Since we will use the same technique for the RFMP_NL in Chapter 10, we again omit the derivation of the algorithm here.

Algorithm 8.3 (Regularized Functional Matching Pursuit, RFMP)**.** Let $\mathcal{T}$ and g be given as in Problem 6.3. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X} \setminus \{0\}$, an initial approximation $f_0 \in \mathcal{X}$ and a regularization parameter $\lambda > 0$.

1. Set $k := 0$, define the residual $r_0 := g - \mathcal{T}f_0$ and choose a stopping criterion.
2. Find

$$d_{k+1} := \operatorname*{argmax}_{d \in \mathcal{D}} \frac{\left(\langle r_k, \mathcal{T}d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{T}d\|_{\mathcal{Y}}^2 + \lambda \|d\|_{\mathcal{X}}^2} \tag{8.2}$$

   and set

$$\alpha_{k+1} := \frac{\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1}\rangle_{\mathcal{X}}}{\|\mathcal{T}d_{k+1}\|_{\mathcal{Y}}^2 + \lambda \|d_{k+1}\|_{\mathcal{X}}^2},$$

   as well as $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$ and $r_{k+1} := g - \mathcal{T}f_{k+1} = r_k - \alpha_{k+1}\mathcal{T}d_{k+1}$.
3. If the stopping criterion is fulfilled, then f_{k+1} is the output. Otherwise, increase k by 1 and return to step 2.

For the Tikhonov functional of the iterates, the following property could be derived in Fischer [45] and Michel [123].

Theorem 8.4. The sequence

$$\left(\|r_k\|_{\mathbb{R}^\ell}^2 + \lambda \|f_k\|_{\mathcal{X}}^2\right)_{k \in \mathbb{N}_0} \tag{8.3}$$

is monotonically decreasing and convergent.

As already mentioned in the previous section, in Michel [123] and Michel and Orzlowski [126], a convergence result for the sequence $(f_k)_{k \in \mathbb{N}_0}$ of approximation was given. We state this result in the following.

Theorem 8.5. Let the regularization parameter fulfill $\lambda \geq 0$ and let the dictionary $\mathcal{D} \subseteq \mathcal{X}$ fulfill the following properties:

1. "semi-frame-condition": There exists a constant $c > 0$ and an integer $M \in \mathbb{N}$ such that, for all expansions $H = \sum_{k=1}^{\infty} \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k are not necessarily pairwise distinct but $\left|\left\{ j \in \mathbb{N} \;\middle|\; d_j = d_k \right\}\right| \leq M$ for each $k \in \mathbb{N}$, the following inequality is valid:

$$c\|H\|_{\mathcal{X}} \leq \sum_{k=1}^{\infty} \beta_k^2.$$

2. $C_1 := \inf_{d \in \mathcal{D}} \left(\|\mathcal{T}d\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{X}} \right)$.

If the sequence $(f_k)_{k \in \mathbb{N}_0}$ is produced by the RFMP and no dictionary element is chosen more than M times, then $(f_k)_{k \in \mathbb{N}_0}$ converges in $\mathcal{X}$ to $f_\infty := f_0 + \sum_{k=1}^{\infty} \alpha_k d_k \in \mathcal{X}$. Moreover, if $\overline{\operatorname{span} \mathcal{D}} = \mathcal{X}$ and $C_2 := \sup_{d \in \mathcal{D}} \|d\|_{\mathcal{X}} < \infty$, then f_∞ solves the Tikhonov-regularized normal equation

$$(\mathcal{T}^*\mathcal{T} + \lambda \mathcal{I}) f_\infty = \mathcal{T}^* g,$$

see Section 6.1.4.

We omit the proof of this result here, since we will present it in a more general setting in Chapter 9. Note that we included the case $\lambda = 0$ here, such that this theorem also covers the convergence of the FMP algorithm, which was described in Section 8.2.

Combining Theorem 8.4 and Theorem 8.5 with Theorem 6.21, we obtain that the sequence in Eq. (8.3) converges monotonically decreasing to the minimum value of the Tikhonov functional.

8.4. Properties of the algorithm and applications

In this section, we want to briefly summarize several properties of the RFMP and the inverse problems to which they were applied to in the past.

A big advantage of the RFMP is its ability to construct an approximation of the solution that is adapted to the structure of the signal. This is due to the fact that the dictionary may be any set of functions, in particular, it does not need to be a basis. Many other methods for linear inverse problems, when implemented on a computer, need to expand the solution in a basis, which may be a restriction if the

solution consists of several different features. For example, on the sphere, one could use global basis functions like spherical harmonics in a Tikhonov regularization to obtain an approximation of the solution. If one tries to approximate highly localized features with this basis, one will see that the solution will not be sparse, but instead has a lot of non-zero coefficients in the expansion. On the other hand, one could use a spline method using radial basis functions, which will yield the same problems if one tries to capture the global trend of the solution. The RFMP makes it possible to—in a sense—interpolate between both extremes.

Furthermore, after choosing a basis, one needs to solve a (possibly large and ill-conditioned) system of linear equations in many other methods. For example, for a spline method like in Berkel and Michel [19], a system of linear equations with a dense matrix has to be solved. Also, when applying a Tikhonov regularization, the regularized normal equation in general includes a dense matrix. In the RFMP, there is no need for the solution of a linear system of equations. In a way, the RFMP is an iterative optimization algorithm for the Tikhonov functional, which avoids the explicit solution of these linear systems. Several other methods (Tikhonov regularization, Landweber iteration) need the knowledge of the adjoint operator $\mathcal{T}^*$, which is not necessary in the RFMP. The only thing that needs to be provided for the algorithm is an implementation of the operator $\mathcal{T}$ and the desired dictionary.

From the algorithmic point of view, the RFMP can be implemented very efficiently. Since the approximations f_k are only linear combinations of dictionary elements, most of the terms that arise in the algorithm can be computed in a preprocessing step. For instance, if one computes the inner products $\langle d_i, d_j \rangle_{\mathcal{X}}$ and $\langle \mathcal{T} d_i, \mathcal{T} d_j \rangle_{\mathbb{R}^\ell}$ for $d_i, d_j \in \mathcal{D}$ in advance, one can speed up the iteration by several orders of magnitude. Since these inner products only depend on the operator and the dictionary, one can reuse them if the algorithm is applied to different data sets belonging to the same inverse problem using the same dictionary.

One could certainly argue that the RFMP is an iterative method instead of a direct one, like for example Tikhonov regularization (in its pure form), and that using an iterative method for a linear inverse problem is inefficient. However, on the one hand, this is also true for the well-established Landweber iteration method. On the other hand, the solution of linear systems with a large and possibly ill-conditioned matrix can only efficiently and accurately been done with iterative solvers such that this argument is not feasible.

Theorem 8.5 shows that the approximations generated by the RFMP converge to the solution of the Tikhonov regularized normal equation, which is also the minimizer of the Tikhonov functional. For that reason one could also simply use a steepest descent

method or comparable optimization methods for the determination of the minimizer. However, in this case, one is again restricted to a representation of the minimizer in a certain basis, that is, in general this will not lead to a sparse representation of the function. Using the RFMP, it is possible to put very diverse types of functions into the dictionary, which makes a sparser representation possible. Note that it is not necessary to know in which basis the solution has a sparse representation (otherwise one could indeed use a steepest descent method or the like with this basis). Instead, one can put all the functions that come to one's mind into the dictionary and the algorithm will choose those functions, which fit best to the solution. This does not only provide a sparse solution from the mathematical perspective, but it also makes the solution more interpretable from the practitioner's point of view. For example, a geophysicist may be able to distinguish global from local features in the provided approximate solution, which often is of particular interest.

The RFMP (and a variant called the ROFMP) has already been applied to various inverse problems since its development. This includes inverse gravimetry (Fischer and Michel [46–48]), the approximation of functions on the sphere such as gravitational field modelling (Michel and Telschow [127]), downward continuation (Michel and Telschow [128] and Telschow [166]), and a joint inversion of gravimetry and normal mode data (Fischer [45]). As already mentioned in the conclusions of Michel and Orzlowski [125], work in progress is an application to medical imaging data (Leweke [110]). In all of these applications, the RFMP proved to provide very good results concerning not only the sparsity and the interpretability of the result, but also concerning the accuracy of the approximate solutions. Additionally, it turned out that the algorithm, in particular the orthogonalized variant ROFMP, can also deal with very scattered data points, which is where "classical methods" are stretched to their limits due to the arising ill-conditioned matrices.

Chapter 9.

Regularized Weak Functional Matching Pursuit (RWFMP)

In this chapter, we deal with a generalization of the RFMP algorithm, which has already been presented in Chapter 8.

Consequently, the problem to solve is a linear inverse problem

$$\mathcal{T} f = g \tag{9.1}$$

as defined in Problem 6.3. So far, the theoretical analysis of the RFMP was restricted to a finite-dimensional data space $\mathcal{Y}$. This is a reasonable assumption in practical cases where the data are, for example, samples of an observable. However, there are also applications where the right-hand side is given as a function (e. g., derived as a model of some data). Furthermore, from a theoretical point of view, the range of a linear operator is closed if it is finite-dimensional (cf. Engl et al. [42], Chapter 2.2). In consequence, the problems which were handled by the RFMP are actually well-posed in the sense of Nashed (but probably ill-conditioned).

Here, we will present the *Regularized Weak Functional Matching Pursuit (RWFMP)*, which can handle data from arbitrary (possibly infinite-dimensional) Hilbert spaces.

When considering infinite-dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, another difficulty arises in the analysis of the algorithm. Unfortunately, it is not clear if there exists a minimizing dictionary element in Eqs. (8.1) and (8.2). The novel RWFMP algorithm instead adds a dictionary element to the approximation that is near to the optimum in a certain sense such that this difficulty is surmounted. This is exactly the same motivation that was given for the Weak Greedy Algorithm (WGA) in Chapter 3.

In the following, we will first introduce the non-regularized version of the algorithm, called WFMP, which is a greedy algorithm for *well-posed* linear inverse problems. For this algorithm, we will prove the convergence to a solution, as well as convergence rates. Subsequently, we will derive the regularized variant of the algorithm

for *ill-posed* inverse problems, the RWFMP, and transfer the convergence results to this case. It turns out that, using the proved convergence rate, we can also prove that there exists an a-priori parameter choice rule for the RWFMP such that we obtain a convergent regularization method (in the sense as presented in Section 6.1.3).

9.1. Weak Functional Matching Pursuit (WFMP)

In this section, we will present the Weak Functional Matching Pursuit (WFMP), which we obtain by applying the idea of the Weak Greedy Algorithm (WGA) (see Algorithm 3.3) to the Functional Matching Pursuit (see Section 8.2). We will derive the convergence of the algorithm in the weak and the strong sense both in the range $\mathcal{Y}$ and the domain $\mathcal{X}$ of the operator $\mathcal{T}$. As for the FMP, it will be shown that, for given data $g \in \mathcal{Y}$, the algorithm converges to a solution f^+ of the normal equation

$$\mathcal{T}^*\mathcal{T}f^+ = \mathcal{T}^*g.$$

In this section, we will assume the well-posedness of the inverse problem in Eq. (9.1) in the sense of Nashed, that is, $\overline{\operatorname{ran}\mathcal{T}} = \operatorname{ran}\mathcal{T}$. We will drop this constraint in Section 9.2 when a regularization is applied to the ill-posed inverse problem.

9.1.1. The algorithm

Remember that the Functional Matching Pursuit (FMP) as presented in Section 8.2 is based on the following concept: let a dictionary $\mathcal{D} \subseteq \mathcal{X}$ be given. Beginning with an initial approximation $f_0 \in \mathcal{X}$, we iteratively define

$$f_{k+1} := f_k + \alpha_{k+1}d_{k+1},$$

where $d_{k+1} \in \mathcal{D}$ is chosen such that

$$\left|\frac{\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}}}{\|\mathcal{T}d_{k+1}\|_{\mathcal{Y}}}\right| = \max_{d\in\mathcal{D}}\left|\frac{\langle r_k, \mathcal{T}d\rangle_{\mathcal{Y}}}{\|\mathcal{T}d\|_{\mathcal{Y}}}\right|, \tag{9.2}$$

and

$$\alpha_{k+1} := \frac{\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}}}{\|\mathcal{T}d_{k+1}\|_Y^2},$$

where $r_k \coloneqq g - \mathcal{T} f_k$ denotes the residual in step k and $\mathcal{Y} = \mathbb{R}^\ell$ for $\ell \in \mathbb{N}$.

The FMP possesses several drawbacks both in theory and in practice.

First, the range of the operator is assumed to be finite-dimensional. On the one hand, this is no problem in practice, since an infinite-dimensional range is not realizable on computers and there is always only a finite amount of measured data. On the other hand, this is not the usual setting in the theoretical analysis of inverse problems (see Chapter 6), since often the infinite-dimensionality of the range causes the ill-posedness of the problem and makes it therefore more interesting. Also, as already mentioned in the introduction to this chapter, one might be interested in data that are given by a model function.

Secondly, similar to the considerations for the PGA in Chapter 3, in Eq. (9.2) one assumes that the maximum exists. It is trivial that this holds true if $\#\mathcal{D} < \infty$, which is the only case that can be realized on a computer in practice. However, if $\#\mathcal{D} = \infty$, which is necessary in theory to span the infinite-dimensional Hilbert space $\mathcal{X}$, it is not clear that the supremum is attained and the maximum exists. Nevertheless, if the dictionary is finite but very large, it may also be computationally expensive to find the maximum in practice, even if one can be sure that it exists.

Both drawbacks are fixed by applying the basic concept of the Weak Greedy Algorithm (WGA, see Algorithm 3.3) to the FMP, which yields the following algorithm. In contrast to the FMP, we use a normalization of the dictionary elements in the codomain here, that is, we require that $\|\mathcal{T} d\|_{\mathcal{Y}} = 1$ for all $d \in \mathcal{D}$. This implies that $\operatorname{null} \mathcal{T} \cap \mathcal{D} = \emptyset$. In the theoretical analysis of the algorithm, it will turn out that this is no restriction at all.

Algorithm 9.1 (Weak Functional Matching Pursuit, WFMP)**.** Let $\mathcal{X}, \mathcal{Y}, \mathcal{T}$ be given as in Problem 6.3. Furthermore, let data $g \in \mathcal{Y}$, a weakness parameter $\varrho \in (0, 1]$, and the initial approximation $f_0 = 0 \in \mathcal{X}$ be given. Choose a dictionary $\mathcal{D} \subseteq \{\, d \in \mathcal{X} \mid \|\mathcal{T} d\|_{\mathcal{Y}} = 1 \,\} \subseteq \mathcal{X}$.

1. Set $k \coloneqq 0$, define the residual $r_0 \coloneqq g - \mathcal{T} f_0 = g$ and choose a stopping criterion.
2. Find an element $d_{k+1} \in \mathcal{D}$ which fulfills

$$|\langle r_k, \mathcal{T} d_{k+1} \rangle_{\mathcal{Y}}| \geq \varrho \sup_{d \in \mathcal{D}} |\langle r_k, \mathcal{T} d \rangle_{\mathcal{Y}}|. \tag{9.3}$$

   Set

$$\alpha_{k+1} \coloneqq \langle r_k, \mathcal{T} d_{k+1} \rangle_{\mathcal{Y}}, \tag{9.4}$$

   as well as $f_{k+1} \coloneqq f_k + \alpha_{k+1} d_{k+1}$ and $r_{k+1} \coloneqq g - \mathcal{T} f_{k+1} = r_k - \alpha_{k+1} \mathcal{T} d_{k+1}$.

3. If the stopping criterion is fulfilled, then f_{k+1} is the output. Otherwise, increase k by 1 and return to step 2.

Remark. If $\varrho = 1$, then the WFMP is equivalent to the FMP (up to the normalization of the dictionary). Thus, all the following results also apply to the FMP, even with an infinite-dimensional range space $\mathcal{Y}$. Additionally, if $\varrho < 1$, the existence of d_{k+1} in Eq. (9.3) is guaranteed.

In the following sections, we will prove the convergence of the WFMP both in the data space as well as in the domain of the operator $\mathcal{T}$. This is done in several steps: first, we prove weak convergence of the residuals in the data space. Secondly, strong convergence in the data space is shown. Finally, we prove the convergence of the iteration also in the domain.

9.1.2. Weak convergence of the residuals

To prove that the residuals converge to zero in the weak sense, we first prove convergence of the norm of residuals. Note that this section is similar to the proofs for the FMP in Fischer [45] and Fischer and Michel [46].

Lemma 9.2. Let $(r_k)_{k\in\mathbb{N}_0}$ be the sequence of residuals arising in Algorithm 9.1. Then the following holds true:

(a) The sequence $(\|r_k\|_{\mathcal{Y}})_{k\in\mathbb{N}_0}$ is monotonically decreasing.

(b) The sequence $(\|r_k\|_{\mathcal{Y}})_{k\in\mathbb{N}_0}$ converges.

Proof. To prove part (a), we observe that

$$\begin{aligned}
\|r_{k+1}\|_{\mathcal{Y}}^2 &= \|r_k - \alpha_{k+1}\mathcal{T}d_{k+1}\|_{\mathcal{Y}}^2 \\
&= \|r_k\|_{\mathcal{Y}}^2 + \alpha_{k+1}^2 - 2\alpha_{k+1}\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}} \\
&= \|r_k\|_{\mathcal{Y}}^2 + \langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}}^2 - 2\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}}^2 \\
&= \|r_k\|_{\mathcal{Y}}^2 - \langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}}^2 \qquad (9.5) \\
&\leq \|r_k\|_{\mathcal{Y}}^2.
\end{aligned}$$

Part (b) follows from the fact that the sequence is bounded from below by 0. ■

Note that we do not know yet that the limit of the sequence of norms of residuals is 0, we only know that it exists. To prove weak convergence of the residuals to 0, we need the following lemma.

Lemma 9.3. The sequence

$$(\alpha_{k+1})_{k\in\mathbb{N}_0} = (\langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}})_{k\in\mathbb{N}_0}$$

is square-summable.

Proof. The following equations hold by expressing $\|r_0\|_{\mathcal{Y}}$ as a telescoping sum for an arbitrary $m \in \mathbb{N}$ and using the relation in Eq. (9.5):

$$\begin{aligned}\|r_0\|_{\mathcal{Y}}^2 &= \sum_{k=0}^{m-1} \Big(\|r_k\|_{\mathcal{Y}}^2 - \|r_{k+1}\|_{\mathcal{Y}}^2 \Big) + \|r_m\|_{\mathcal{Y}}^2 \\ &= \sum_{k=0}^{m-1} \langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}}^2 + \|r_m\|_{\mathcal{Y}}^2.\end{aligned}$$

Thus, by Lemma 9.2, we obtain

$$\begin{aligned}\sum_{k=0}^{\infty} \alpha_{k+1}^2 = \sum_{k=0}^{\infty} \langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}}^2 &= \lim_{m\to\infty} \sum_{k=0}^{m-1} \langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}}^2 \\ &= \|r_0\|_{\mathcal{Y}}^2 - \lim_{m\to\infty} \|r_m\|_{\mathcal{Y}}^2 < \infty.\end{aligned}$$

Thus, the sequence is square-summable. ■

Lemma 9.3 gives rise to the two following additional results.

Corollary 9.4. Since every square-summable sequence converges to zero, we have

$$\lim_{k\to\infty} \alpha_{k+1} = \lim_{k\to\infty} \langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}} = 0.$$

Corollary 9.5. We have

$$\lim_{k\to\infty} \langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}} = 0$$

for all $d \in \mathcal{D}$.

Proof. Because of Eq. (9.3), we obtain for every $d \in \mathcal{D}$ that

$$0 \leq |\langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}}| \leq \frac{1}{\varrho} |\langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}}|.$$

Since $\varrho \in (0, 1]$ is fixed and the right-hand side tends to 0 for $k \to \infty$ by the preceding corollary, this proves our claim. ■

Finally, we obtain weak convergence of the WFMP in the following theorems. For the sake of readability, we define $\mathcal{V} := \operatorname{span}\{\, \mathcal{T} d \mid d \in \mathcal{D} \,\} \subseteq \mathcal{Y}$.

Theorem 9.6. We have weak convergence of $(r_k)_{k\in\mathbb{N}_0}$ to zero in the space $\overline{\mathcal{V}}$, that is,

$$\lim_{k\to\infty} \langle r_k, \overline{v}\rangle_{\mathcal{Y}} = 0$$

for all $\overline{v} \in \overline{\mathcal{V}}$.

Proof. By Corollary 9.5, we have

$$\lim_{k\to\infty} \langle r_k, v\rangle_{\mathcal{Y}} = 0$$

for all $v \in \mathcal{V}$ due to the bilinearity of the inner product. Since $(r_k)_{k\in\mathbb{N}_0}$ is a bounded sequence, we obtain

$$\lim_{k\to\infty} \langle r_k, \overline{v}\rangle_{\mathcal{Y}} = 0$$

for all $\overline{v} \in \overline{\mathcal{V}}$, too. ■

Theorem 9.7. Let the given data fulfill $g \in \overline{\mathcal{V}}$. Then $r_k \rightharpoonup 0$ in $\mathcal{Y}$ for $k \to \infty$.

Proof. Let $z \in \mathcal{Y}$ be arbitrary. Since $\overline{\mathcal{V}}$ is a closed subspace of $\mathcal{Y}$, we obtain the decomposition $\mathcal{Y} = \overline{\mathcal{V}} \oplus \overline{\mathcal{V}}^{\perp}$. Thus, there exist uniquely defined $z_{\parallel} \in \overline{\mathcal{V}}$, $z_{\perp} \in \overline{\mathcal{V}}^{\perp}$ such that $z = z_{\parallel} + z_{\perp}$. It follows that

$$\lim_{k\to\infty} \langle r_k, z\rangle_{\mathcal{Y}} = \lim_{k\to\infty} \Big(\langle r_k, z_{\parallel}\rangle_{\mathcal{Y}} + \langle r_k, z_{\perp}\rangle_{\mathcal{Y}} \Big).$$

Since $r_k = y - \mathcal{T} f_k \in \overline{\mathcal{V}}$ and $z_{\perp} \perp \overline{\mathcal{V}}$, the latter term vanishes and

$$\lim_{k\to\infty} \langle r_k, z\rangle_{\mathcal{Y}} = \lim_{k\to\infty} \langle r_k, z_{\parallel}\rangle_{\mathcal{Y}} = 0$$

by Theorem 9.6, since $z_{\parallel} \in \overline{\mathcal{V}}$. ■

Note that in the case $g \in \overline{\mathcal{V}}$, we have proved so far that the sequence $(\|r_k\|_{\mathcal{Y}})_{k\in\mathbb{N}_0}$ is convergent and that $r_k \rightharpoonup 0$ ($k \to \infty$) in $\mathcal{Y}$. Unfortunately, we cannot conclude convergence $r_k \to 0$ in $\mathcal{Y}$ in the strong sense from these facts, since $\mathcal{Y}$ may be infinite-dimensional. This is different in the considerations in Fischer [45] and Fischer and Michel [46], where it was assumed that $\dim \mathcal{Y} = \ell \in \mathbb{N}$. The next section is dedicated to the proof of strong convergence of the residuals, which requires a more complicated technique.

9.1.3. Strong convergence of the residuals

The following proofs are based on the technique introduced in Jones [92] for *projection pursuit regression* (see also Section 5.1), a variant of the WGA in statistics.

Lemma 9.8. Let $(a_k)_{k\in\mathbb{N}_0}$ be a square-summable sequence, where $a_k \geq 0$ for all $k \in \mathbb{N}_0$. Then the identity

$$\liminf_{k\to\infty}\left(a_k \sum_{j=1}^{k} a_j\right) = 0.$$

holds.

Proof. Let $\varepsilon > 0$. Since $(a_k)_{k\in\mathbb{N}_0}$ is square-summable, there exists $K \in \mathbb{N}_0$ such that

$$\sum_{j=K}^{\infty} a_j^2 < \frac{\varepsilon}{2}. \tag{9.6}$$

Let $k > K$ be arbitrary. Since $\lim_{j\to\infty} a_j = 0$, there exists $\tilde{p} = \tilde{p}(k) > k$ for which

$$a_{\tilde{p}} \sum_{j=1}^{k} a_j < \frac{\varepsilon}{2}. \tag{9.7}$$

Finally, let $p = p(k) \in \mathbb{N}_0$ be an index with $k < p \leq \tilde{p}$, which fulfills

$$a_p \leq a_j \tag{9.8}$$

for all $j \in \{k+1, \ldots, \tilde{p}\}$.

Then,

$$\begin{aligned}
\inf_{m\geq k}\left(a_m \sum_{j=1}^{m} a_j\right) &\leq a_p \sum_{j=1}^{p} a_j = a_p \sum_{j=1}^{k} a_j + a_p \sum_{j=k+1}^{p} a_j \\
&\leq a_{\tilde{p}} \sum_{j=1}^{k} a_j + \sum_{j=k+1}^{p} a_j^2 \\
&\leq a_{\tilde{p}} \sum_{j=1}^{k} a_j + \sum_{j=K}^{\infty} a_j^2 \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

where the first inequality is due to the properties of the infimum, the second inequality follows from Eq. (9.8) and the non-negativity of the sequence, and the last inequality is true due to Eq. (9.6) and Eq. (9.7). ■

Corollary 9.9. We have

$$\liminf_{k\to\infty} |\langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}}| \sum_{j=1}^{k} \left|\langle r_j, \mathcal{T} d_{j+1}\rangle_{\mathcal{Y}}\right| = 0.$$

Proof. This follows from Lemma 9.3 in conjunction with Lemma 9.8. ■

The preceding corollary is a crucial ingredient in the proof of the following theorem, where it is shown that the sequence of residuals converges strongly in $\mathcal{Y}$.

Theorem 9.10. The sequence $(r_k)_{k\in\mathbb{N}_0}$ is a Cauchy sequence in $\mathcal{Y}$ and thus convergent.

Proof. This proof is an extended version of a similar proof in Jones [92, Section 2].

Assume that the sequence is not a Cauchy sequence in $\mathcal{Y}$. Then we have

$$\exists \varepsilon > 0 \colon \forall K \in \mathbb{N}_0 \colon \exists m, k \geq K \colon \|r_m - r_k\|_{\mathcal{Y}} > \varepsilon. \tag{9.9}$$

Let $\gamma > 0$ be an arbitrary constant.

From Lemma 9.2(b), we obtain the existence of $R := \lim_{j\to\infty} \|r_j\|_{\mathcal{Y}}$. Thus, there exists $K \in \mathbb{N}_0$ such that $\|r_K\|_{\mathcal{Y}}^2 < R^2 + \gamma$. Since $\left(\|r_j\|_{\mathcal{Y}}\right)_{j\in\mathbb{N}_0}$ is monotonically decreasing due to Lemma 9.2(a) and by Eq. (9.9), we obtain that there exist $m, k \geq K$ which fulfill

$$\begin{aligned} \|r_m - r_k\|_{\mathcal{Y}} &> \varepsilon, \\ \|r_m\|_{\mathcal{Y}}^2 &< R^2 + \gamma, \\ \|r_k\|_{\mathcal{Y}}^2 &< R^2 + \gamma. \end{aligned} \tag{9.10}$$

Furthermore, by Corollary 9.9 there exists $p > \max\{m, k\}$ such that

$$\left|\langle r_p, \mathcal{T} d_{p+1}\rangle_{\mathcal{Y}}\right| \sum_{j=1}^{p} \left|\langle r_j, \mathcal{T} d_{j+1}\rangle_{\mathcal{Y}}\right| < \gamma. \tag{9.11}$$

Since

$$\varepsilon < \|r_m - r_k\|_{\mathcal{Y}} \leq \|r_m - r_p\|_{\mathcal{Y}} + \|r_p - r_k\|_{\mathcal{Y}},$$

we have $\|r_m - r_p\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$ or $\|r_p - r_k\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$.

Without loss of generality, let $\left\|r_m - r_p\right\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$. We obtain

$$\begin{aligned}
\left\|r_m - r_p\right\|_{\mathcal{Y}}^2 &= \|r_m\|_{\mathcal{Y}}^2 + \left\|r_p\right\|_{\mathcal{Y}}^2 - 2\langle r_m, r_p\rangle_{\mathcal{Y}} \\
&= \|r_m\|_{\mathcal{Y}}^2 + \left\|r_p\right\|_{\mathcal{Y}}^2 - 2\left\langle r_p + \sum_{j=m}^{p-1} \alpha_{j+1}\mathcal{T}d_{j+1}, r_p \right\rangle_{\mathcal{Y}} \\
&\leq \|r_m\|_{\mathcal{Y}}^2 - \left\|r_p\right\|_{\mathcal{Y}}^2 + 2\sum_{j=m}^{p-1} \left|\alpha_{j+1}\right| \left|\left\langle \mathcal{T}d_{j+1}, r_p\right\rangle_{\mathcal{Y}}\right| \\
&= \|r_m\|_{\mathcal{Y}}^2 - \left\|r_p\right\|_{\mathcal{Y}}^2 + 2\sum_{j=m}^{p-1} \left|\left\langle r_j, \mathcal{T}d_{j+1}\right\rangle_{\mathcal{Y}}\right| \underbrace{\left|\left\langle r_p, \mathcal{T}d_{j+1}\right\rangle_{\mathcal{Y}}\right|}_{\leq \frac{1}{\varrho}\left|\left\langle r_p, \mathcal{T}d_{p+1}\right\rangle_{\mathcal{Y}}\right|} \\
&\leq \underbrace{\|r_m\|_{\mathcal{Y}}^2}_{\leq R^2+\gamma} \underbrace{-\left\|r_p\right\|_{\mathcal{Y}}^2}_{\leq -R^2} + \frac{2}{\varrho} \underbrace{\left|\left\langle r_p, \mathcal{T}d_{p+1}\right\rangle_{\mathcal{Y}}\right| \sum_{j=m}^{p-1} \left|\left\langle r_j, \mathcal{T}d_{j+1}\right\rangle_{\mathcal{Y}}\right|}_{<\gamma} \\
&\leq R^2 + \gamma - R^2 + \frac{2}{\varrho}\gamma = \left(1 + \frac{2}{\varrho}\right)\gamma,
\end{aligned}$$

where Eq. (9.4), Eq. (9.3), Eq. (9.10), Lemma 9.2(a), and Eq. (9.11) have been used in this order.

Since $\gamma > 0$ was chosen arbitrarily, one can choose γ small enough such that $\left(1 + \frac{2}{\varrho}\right)\gamma < \frac{\varepsilon^2}{4}$, which yields a contradiction to $\left\|r_m - r_p\right\|_{\mathcal{Y}} > \frac{\varepsilon}{2}$. ■

Since $\mathcal{Y}$ is a Hilbert space, the sequence $(r_k)_{k\in\mathbb{N}_0}$ converges in $\mathcal{Y}$ in the strong sense. In the following, we will prove several properties of the limit of this sequence.

Theorem 9.11. For $r_\infty \coloneqq \lim_{k\to\infty} r_k$, we have

$$r_\infty \perp \overline{\mathcal{V}}.$$

Proof. Let $\overline{v} \in \overline{\mathcal{V}}$. Since the inner product is a continuous function of its arguments due to the Cauchy-Schwarz inequality, we may interchange the limit and the inner product to obtain

$$\langle r_\infty, \overline{v}\rangle_{\mathcal{Y}} = \left\langle \lim_{k\to\infty} r_k, \overline{v} \right\rangle_{\mathcal{Y}} = \lim_{k\to\infty} \langle r_k, \overline{v}\rangle_{\mathcal{Y}} = 0$$

by Theorem 9.6. ■

Corollary 9.12. If $g \in \overline{\mathcal{V}}$, we have $r_\infty = 0$ due to the fact that the weak limit is 0 according to Theorem 9.7, the strong limit has been proved to exist in Theorem 9.10, and the latter has to coincide with the weak limit.

In the next section, we will prove convergence of the algorithm also in the domain $\mathcal{X}$ of the operator $\mathcal{T}$.

9.1.4. Convergence in the domain

So far, we have only considered convergence of the WFMP in the image space $\mathcal{Y}$ of the operator $\mathcal{T}$. To achieve convergence also in the domain $\mathcal{X}$ of the operator, we can adopt the proof of the analogous statement for the FMP, which has first been stated in Fischer [45]. We will adhere to the improved version of the proof that has recently been given in Michel and Orzlowski [126]. Due to the normalization of the dictionary, we can omit the second condition that is required in the latter reference.

Theorem 9.13. Let the assumptions of Algorithm 9.1 be fulfilled. Furthermore, let the dictionary $\mathcal{D} \subseteq \mathcal{X}$ fulfill the following condition:

(SFC) The *semi-frame condition*:

There exists a constant $c > 0$ and an integer $M \in \mathbb{N}$ such that for all expansions $H = \sum_{k=1}^{\infty} \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k are not necessarily pairwise distinct, but $\# \left\{ j \in \mathbb{N} \;\middle|\; d_j = d_k \right\} \leq M$ for all $k \in \mathbb{N}$, the inequality

$$c\|H\|_{\mathcal{X}}^2 \leq \sum_{k=1}^{\infty} \beta_k^2$$

is valid.

If the sequence $(f_k)_{k\in\mathbb{N}_0}$ is produced by the WFMP and no dictionary element is chosen more than M times, then $(f_k)_{k\in\mathbb{N}_0}$ converges in $\mathcal{X}$ to $f_\infty := \sum_{k=1}^{\infty} \alpha_k d_k$.

Proof. From Lemma 9.3, we obtain that $(\alpha_k)_{k\in\mathbb{N}}$ is square-summable. The latter and the semi-frame condition (SFC) give rise to the fact that the series $\sum_{k=1}^{\infty} \alpha_k d_k$ converges in the strong sense in $\mathcal{X}$ and hence, $f_\infty \in \mathcal{X}$ as defined above exists. It is also clear that

$$f_\infty = \lim_{K\to\infty} f_K$$

holds in the sense of $\mathcal{X}$, since

$$\lim_{K\to\infty} \|f_\infty - f_K\|_{\mathcal{X}}^2 = \lim_{K\to\infty} \left\| \sum_{k=K+1}^{\infty} \alpha_k d_k \right\|_{\mathcal{X}}^2 \leq \frac{1}{c} \lim_{K\to\infty} \sum_{k=K+1}^{\infty} \alpha_k^2 = 0,$$

due to the square-summability of the sequence of coefficients (see Lemma 9.3). ■

Corollary 9.14. Now that we know that $(f_k)_{k\in\mathbb{N}_0}$ converges, it follows from Theorem 9.11 that

$$\mathcal{T} f_\infty = \mathcal{P}_{\overline{\mathcal{V}}} g,$$

where $\mathcal{P}_{\overline{\mathcal{V}}}$ denotes the orthogonal projection in $\mathcal{Y}$ onto $\overline{\mathcal{V}}$. In other words, $\mathcal{T} f_\infty$ is the best approximation of g in $\overline{\mathcal{V}}$.

After showing that the sequence of iterates $(f_k)_{k\in\mathbb{N}_0}$ converges, we will prove that the limit satisfies the associated normal equation.

Theorem 9.15. If the assumptions of Theorem 9.13 are fulfilled and additionally $\overline{\operatorname{span}\mathcal{D}} = (\operatorname{null}\mathcal{T})^\perp$ holds, then the limit f_∞ satisfies the normal equation

$$\mathcal{T}^*\mathcal{T} f_\infty = \mathcal{T}^* g. \tag{9.12}$$

Proof. Since the sequences $(r_k)_{k\in\mathbb{N}_0} = (g - \mathcal{T} f_k)_{k\in\mathbb{N}_0}$ and $(f_k)_{k\in\mathbb{N}_0}$ both converge and $\mathcal{T}$ is continuous, we have $r_\infty = g - \mathcal{T} f_\infty$.

By Theorem 9.11, we have $r_\infty \perp \overline{\mathcal{V}}$. Hence,

$$g - \mathcal{T} f_\infty \perp \overline{\operatorname{span}\{\, \mathcal{T} d \mid d \in \mathcal{D} \,\}}$$

such that for all $d \in \mathcal{D}$ we have

$$0 = \langle g - \mathcal{T} f_\infty, \mathcal{T} d\rangle_{\mathcal{Y}} = \langle \mathcal{T}^*(g - \mathcal{T} f_\infty), d\rangle_{\mathcal{X}}.$$

Furthermore, for all $d \in \operatorname{null}\mathcal{T}$ we also obtain

$$\langle \mathcal{T}^*(g - \mathcal{T} f_\infty), d\rangle_{\mathcal{X}} = \langle g - \mathcal{T} f_\infty, \mathcal{T} d\rangle_{\mathcal{Y}} = \langle g - \mathcal{T} f_\infty, 0\rangle_{\mathcal{Y}} = 0.$$

Since $\operatorname{span}\mathcal{D} \oplus \operatorname{null}\mathcal{T}$ is dense in $\mathcal{X}$ and the inner product is a non-degenerate bilinear form, we obtain

$$\mathcal{T}^*\mathcal{T} f_\infty = \mathcal{T}^* g$$

as desired. ■

Remark. We make the following three observations:

(a) We have $\overline{\mathcal{V}} = \overline{\operatorname{ran}\mathcal{T}} = \operatorname{ran}\mathcal{T}$ due to the condition $\overline{\operatorname{span}\mathcal{D}} = (\operatorname{null}\mathcal{T})^{\perp}$ and the well-posedness of the inverse problem.

(b) Thus, as remarked earlier, $r_\infty = 0$ if $g \in \operatorname{ran}\mathcal{T}$. From Theorem 9.13, we obtain that the limit $f_\infty \in \mathcal{X}$ exists and in the preceding proof, we have already employed that $r_\infty = g - \mathcal{T}f_\infty$ due to the continuity of $\mathcal{T}$. Thus, f_∞ is a solution of the inverse problem Eq. (9.1), since $\mathcal{T}f_\infty = g$.

(c) It is well-known that the solution of the normal equation Eq. (9.12) is also a least squares solution of the inverse problem (see Theorem 6.10). That is in our case

$$\|g - \mathcal{T}f_\infty\|_{\mathcal{Y}} = \min_{f\in\mathcal{X}} \|g - \mathcal{T}f\|_{\mathcal{Y}}.$$

This means that the WFMP (as well as the FMP) has an interpretation as a minimization algorithm for the optimization problem

$$\|g - \mathcal{T}f\|_{\mathcal{Y}} \to \text{min!} \quad \text{subject to} \quad f \in \mathcal{X}.$$

In fact, the FMP was originally motivated as an iterative minimization of exactly the functional $f \mapsto \|g - \mathcal{T}f\|_{\mathcal{Y}}^2$ and so things have come full circle.

9.1.5. Convergence rates

In analogy to the convergence result for the WGA in Temlyakov [167, Theorem 5.1], see also Chapter 3, we can prove a convergence rate of the WFMP in the data space in the following.

We first state the following lemma, which is an analogy to DeVore and Temlyakov [32, Lemma 3.4] and Temlyakov [167, Lemma 3.1].

Lemma 9.16. Let $\varrho \in (0, 1]$ and let $(a_k)_{k\in\mathbb{N}_0}$ be a sequence of non-negative numbers, which satisfies

$$a_0 \le 1, \qquad a_{k+1} \le a_k\big(1 - \varrho^2 a_k\big) \text{ for all } k \ge 0.$$

Then,

$$a_k \le \frac{1}{1 + k\varrho^2} \text{ for all } k \in \mathbb{N}_0. \tag{9.13}$$

Proof. We proof the claim by induction on k. For $k = 0$ the statement is obvious.

Assuming that the inequality in Eq. (9.13) is true for some $k \in \mathbb{N}_0$, we have to show that

$$a_{k+1} \leq \frac{1}{1+(k+1)\varrho^2}.$$

The latter is clearly true, if $a_{k+1} = 0$. If $a_{k+1} > 0$ then also $a_k > 0$ and we obtain

$$\begin{aligned} a_{k+1} &\leq a_k(1-\varrho^2 a_k) \leq a_k \frac{1}{1+\varrho^2 a_k} = \frac{1}{\frac{1}{a_k}+\varrho^2} \\ &\leq \frac{1}{(1+k\varrho^2)+\varrho^2} = \frac{1}{1+(k+1)\varrho^2} \end{aligned}$$

since $(1-x) \leq (1+x)^{-1}$ for $x \geq 0$. ■

Next, we define the following norm on $\overline{\mathcal{V}}$.

Definition 9.17. Given a linear and bounded operator $\mathcal{T}\colon \mathcal{X} \to \mathcal{Y}$ and a dictionary $\mathcal{D}$, for $z \in \overline{\mathcal{V}}$ we define

$$|z|_{\mathcal{T}\mathcal{D}} := \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; z = \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T} \bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\}, \tag{9.14}$$

where the limit of the second series is considered in the sense of $\mathcal{Y}$.

Furthermore, we define the set

$$\widehat{\overline{\mathcal{V}}} := \left\{ z \in \overline{\mathcal{V}} \;\middle|\; |z|_{\mathcal{T}\mathcal{D}} < \infty \right\}.$$

In general, an element $z \in \overline{\mathcal{V}}$ may be represented as a linear combination of images of dictionary elements in several different ways, since the dictionary itself and its image do not need to form a basis of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Thus, the quantity $|z|_{\mathcal{T}\mathcal{D}}$ can be read as a measure of how sparse the element z can be expressed as a linear combination of images of dictionary elements.

However, note that $|z|_{\mathcal{T}\mathcal{D}}$ does not need to be finite. Even if $\mathcal{T}\mathcal{D}$ was an orthonormal basis in $\mathcal{Y}$, it is not natural that the Fourier coefficients of some element z are absolutely summable.

Lemma 9.18. Let $z \in \overline{\mathcal{V}}$. Then we have

$$\|z\|_{\mathcal{Y}} \leq |z|_{\mathcal{T}\mathcal{D}}.$$

Proof. If $z \in \overline{\mathcal{V}} \setminus \widehat{\overline{\mathcal{V}}}$, then the inequality is clear, since the right-hand side is infinite.

If $z \in \widehat{\overline{\mathcal{V}}}$, then for all $\varepsilon > 0$ there exist $\bar{\beta}_k \in \mathbb{R}$ and $\bar{d}_k \in \mathcal{D}$ such that $z = \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T} \bar{d}_k$ and $\sum_{k=1}^{\infty} |\bar{\beta}_k| \leq |z|_{\mathcal{T}\mathcal{D}} + \varepsilon$. Thus,

$$\|z\|_{\mathcal{Y}} = \left\| \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T} \bar{d}_k \right\|_{\mathcal{Y}} \leq \sum_{k=1}^{\infty} |\bar{\beta}_k| \|\mathcal{T} \bar{d}_k\|_{\mathcal{Y}} = \sum_{k=1}^{\infty} |\bar{\beta}_k| \leq |z|_{\mathcal{T}\mathcal{D}} + \varepsilon.$$

Since ε was chosen arbitrarily, this proves the inequality. ■

Theorem 9.19. The mapping

$$z \mapsto |z|_{\mathcal{T}\mathcal{D}}$$

is a norm on $\widehat{\overline{\mathcal{V}}}$.

Proof. We show the axioms for a norm one by one.

Definiteness On the one hand, it is obvious that $|0|_{\mathcal{T}\mathcal{D}} = 0$, since the zero sequence is a feasible choice in Eq. (9.14).

Assume on the other hand that $|z|_{\mathcal{T}\mathcal{D}} = 0$ for $z \in \widehat{\overline{\mathcal{V}}}$. Then by Lemma 9.18, we obtain $0 \leq \|z\|_{\mathcal{Y}} \leq |z|_{\mathcal{T}\mathcal{D}} = 0$, thus $z = 0$.

Absolute homogeneity We have to prove that $|\lambda z|_{\mathcal{T}\mathcal{D}} = |\lambda| |z|_{\mathcal{T}\mathcal{D}}$ for $z \in \widehat{\overline{\mathcal{V}}}$ and $\lambda \in \mathbb{R}$.

For $\lambda = 0$, the absolute homogeneity follows from the definiteness of the norm.

Let $\lambda \neq 0$ and $z = \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T} \bar{d}_k \in \widehat{\overline{\mathcal{V}}}$. Thus, we have $\lambda z = \sum_{k=1}^{\infty} (\lambda \bar{\beta}_k) \mathcal{T} \bar{d}_k$ such that $|\lambda z|_{\mathcal{T}\mathcal{D}} \leq |\lambda| |z|_{\mathcal{T}\mathcal{D}}$ is guaranteed.

It remains to show that there are no $\hat{\beta}_k$, $\hat{d}_k$ such that $\lambda z = \sum_{k=1}^{\infty} \hat{\beta}_k \mathcal{T} \hat{d}_k$ and $\sum_{k=1}^{\infty} |\hat{\beta}_k| < |\lambda| |z|_{\mathcal{T}\mathcal{D}}$. Assume the contrary, then consequently $z = \sum_{k=1}^{\infty} \frac{\hat{\beta}_k}{\lambda} \mathcal{T} \hat{d}_k$ and $\sum_{k=1}^{\infty} \left| \frac{\hat{\beta}_k}{\lambda} \right| = \frac{1}{|\lambda|} \sum_{k=1}^{\infty} |\hat{\beta}_k| < |z|_{\mathcal{T}\mathcal{D}}$, which is a contradiction to the definition of $|z|_{\mathcal{T}\mathcal{D}}$.

Triangle inequality Let $z, w \in \widehat{\overline{\mathcal{V}}}$ and $\varepsilon > 0$. Due to the definition of the norm, there exist $\bar{\beta}_k, \hat{\beta}_k \in \mathbb{R}$ and $\bar{d}_k, \hat{d}_k \in \mathcal{D}$ such that

$$z = \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T} \bar{d}_k, \qquad \sum_{k=1}^{\infty} |\bar{\beta}_k| \leq |z|_{\mathcal{T}\mathcal{D}} + \frac{\varepsilon}{2},$$

$$w = \sum_{k=1}^{\infty} \hat{\beta}_k \mathcal{T} \hat{d}_k, \qquad \sum_{k=1}^{\infty} |\hat{\beta}_k| \leq |w|_{\mathcal{T}\mathcal{D}} + \frac{\varepsilon}{2}.$$

Set

$$\gamma_k := \begin{cases} \bar{\beta}_{k/2}, & \text{for even } k, \\ \hat{\beta}_{(k+1)/2}, & \text{else,} \end{cases} \qquad f_k := \begin{cases} \bar{d}_{k/2} & \text{for even } k, \\ \hat{d}_{(k+1)/2} & \text{else.} \end{cases}$$

Note that there is no problem with rearranging the series, since the series is absolutely convergent (see the proof of Lemma 9.18) and, thus, unconditionally convergent, since $\overline{\mathcal{V}}$ is complete with respect to $\|\cdot\|_{\mathcal{Y}}$ (cf. Kadets and Kadets [94, Chapter 1, §3]).

Consequently, $z + w = \sum_{k=1}^{\infty} \gamma_k \mathcal{T} f_k$ and

$$\sum_{k=1}^{\infty} |\gamma_k| \leq \sum_{k=1}^{\infty} |\beta_k| + \sum_{k=1}^{\infty} |\hat{\beta}_k| \leq |z|_{\mathcal{T}\mathcal{D}} + |w|_{\mathcal{T}\mathcal{D}} + \varepsilon.$$

Thus,

$$|z + w|_{\mathcal{T}\mathcal{D}} \leq \sum_{k=1}^{\infty} |\gamma_k| \leq |z|_{\mathcal{T}\mathcal{D}} + |w|_{\mathcal{T}\mathcal{D}},$$

by the definition of the norm and the fact that ε was chosen arbitrarily.

■

Theorem 9.20. Let $(r_k)_{k \in \mathbb{N}_0}$ be the sequence of residuals generated by the WFMP and let $g \in \overline{\mathcal{V}}$. Then

$$\|r_k\|_{\mathcal{Y}} \leq |g|_{\mathcal{T}\mathcal{D}} \left(1 + k\varrho^2\right)^{-\varrho/(4+2\varrho)}. \tag{9.15}$$

Proof. If $g \in \overline{\mathcal{V}} \setminus \widehat{\overline{\mathcal{V}}}$, then the inequality is clear, since the right-hand side is infinite. If furthermore $|g|_{\mathcal{T}\mathcal{D}} = 0$ then both sides of the inequality are zero, since $\|r_k\|_{\mathcal{Y}} = \|g\|_{\mathcal{Y}} = 0$ for all $k \in \mathbb{N}_0$, because for all coefficients we have $\alpha_k = 0$.

Let $g \in \widehat{\overline{\mathcal{V}}}$ and $|g|_{\mathcal{T}\mathcal{D}} > 0$. For $k \in \mathbb{N}_0$, define the sequence

$$b_k := |g|_{\mathcal{T}\mathcal{D}} + \sum_{j=1}^{k} |\alpha_j|.$$

Then, we obtain

$$|r_k|_{\mathcal{T}\mathcal{D}} = |g - \mathcal{T} f_k|_{\mathcal{T}\mathcal{D}} = \left| g - \sum_{j=1}^{k} \alpha_j \mathcal{T} d_j \right|_{\mathcal{T}\mathcal{D}}$$

$$\leq |g|_{\mathcal{T}\mathcal{D}} + \sum_{j=1}^{k} |\alpha_j \mathcal{T} d_j|_{\mathcal{T}\mathcal{D}} \leq |g|_{\mathcal{T}\mathcal{D}} + \sum_{j=1}^{k} |\alpha_j| = b_k$$

by using the triangle inequality, the absolute homogeneity of the norm, and the fact that $|\mathcal{T} d_j|_{\mathcal{T}\mathcal{D}} \leq 1$.

For arbitrary $\varepsilon > 0$, there exist $\bar{\beta}_j \in \mathbb{R}$ and $\bar{d}_j \in \mathcal{D}$ such that $r_k = \sum_{j=1}^{\infty} \bar{\beta}_j \mathcal{T} \bar{d}_j$ and $\sum_{j=1}^{\infty} |\bar{\beta}_j| \leq b_k + \varepsilon$. Thus,

$$\begin{aligned} \|r_k\|_{\mathcal{Y}}^2 = |\langle r_k, r_k \rangle_{\mathcal{Y}}| &= \left| \left\langle r_k, \sum_{j=1}^{\infty} \bar{\beta}_j \mathcal{T} \bar{d}_j \right\rangle_{\mathcal{Y}} \right| \\ &\leq \sum_{j=1}^{\infty} |\bar{\beta}_j| \left| \langle r_k, \mathcal{T} \bar{d}_j \rangle_{\mathcal{Y}} \right| \leq (b_k + \varepsilon) \sup_{d \in \mathcal{D}} |\langle r_k, Td \rangle_{\mathcal{Y}}| \end{aligned}$$

and since ε was arbitrary this yields

$$\sup_{d \in \mathcal{D}} |\langle r_k, Td \rangle_{\mathcal{Y}}| \geq \frac{\|r_k\|_{\mathcal{Y}}^2}{b_k}$$

and

$$|\alpha_{k+1}| \geq \frac{\varrho \|r_k\|_{\mathcal{Y}}^2}{b_k}$$

by the definition of α_{k+1} in Eq. (9.4).

Since $\|r_{k+1}\|_{\mathcal{Y}}^2 = \|r_k\|_{\mathcal{Y}}^2 - |\alpha_{k+1}|^2$ (see Eq. (9.5)), we obtain on the one hand that

$$\|r_{k+1}\|_{\mathcal{Y}}^2 \leq \|r_k\|_{\mathcal{Y}}^2 - \frac{\varrho^2 \|r_k\|_{\mathcal{Y}}^4}{b_k^2} = \|r_k\|_{\mathcal{Y}}^2 \left(1 - \frac{\varrho^2 \|r_k\|_{\mathcal{Y}}^2}{b_k^2} \right) \tag{9.16}$$

and, on the other hand,

$$\|r_{k+1}\|_{\mathcal{Y}}^2 \leq \|r_k\|_{\mathcal{Y}}^2 - |\alpha_{k+1}| \frac{\varrho \|r_k\|_{\mathcal{Y}}^2}{b_k} = \|r_k\|_{\mathcal{Y}}^2 \left(1 - \frac{\varrho |\alpha_{k+1}|}{b_k} \right). \tag{9.17}$$

Since $(b_k)_{n \in \mathbb{N}_0}$ is monotonically increasing, Eq. (9.16) gives

$$\frac{\|r_{k+1}\|_{\mathcal{Y}}^2}{b_{k+1}^2} \leq \frac{\|r_k\|_{\mathcal{Y}}^2}{b_k^2} \left(1 - \frac{\varrho^2 \|r_k\|_{\mathcal{Y}}^2}{b_k^2} \right).$$

The application of Lemma 9.16 with $a_k := \frac{\|r_k\|_{\mathcal{Y}}^2}{b_k^2}$ (note that $a_0 = \frac{\|r_0\|_{\mathcal{Y}}^2}{b_0^2} = \frac{\|g\|_{\mathcal{Y}}^2}{b_0^2} \leq \frac{|g|_{\mathcal{T}\mathcal{D}}}{|g|_{\mathcal{T}\mathcal{D}}} = 1$) yields

$$\frac{\|r_k\|_{\mathcal{Y}}^2}{b_k^2} \leq \left(1 + k\varrho^2 \right)^{-1}. \tag{9.18}$$

From the inequality in Eq. (9.17) and the fact that $b_{k+1} = b_k\left(1 + \frac{|\alpha_{k+1}|}{b_k}\right)$ combined with the generalized Bernoulli inequality

$$(1+x)^\alpha \leq 1 + \alpha x, \qquad 0 \leq \alpha \leq 1, \qquad x \geq 0,$$

(cf. Mitrinovic [132, Section 2.4]) we obtain

$$\begin{aligned}
\|r_{k+1}\|_{\mathcal{Y}}^2 b_{k+1}^\varrho &= \|r_{k+1}\|_{\mathcal{Y}}^2 b_k^\varrho \left(1 + \frac{|\alpha_{k+1}|}{b_k}\right)^\varrho \\
&\leq \|r_{k+1}\|_{\mathcal{Y}}^2 b_k^\varrho \left(1 + \varrho\frac{|\alpha_{k+1}|}{b_k}\right) \\
&\leq \|r_k\|_{\mathcal{Y}}^2 \left(1 - \varrho\frac{|\alpha_{k+1}|}{b_k}\right) b_k^\varrho \left(1 + \varrho\frac{|\alpha_{k+1}|}{b_k}\right) \\
&\leq \|r_k\|_{\mathcal{Y}}^2 \left(1 - \varrho^2\frac{|\alpha_{k+1}|^2}{b_k^2}\right) b_k^\varrho \\
&\leq \|r_k\|_{\mathcal{Y}}^2 b_k^\varrho.
\end{aligned}$$

Thus, by induction, the following sequence of inequalities holds true for all $k \in \mathbb{N}_0$

$$\begin{aligned}
\|r_{k+1}\|_{\mathcal{Y}}^2 b_{k+1}^\varrho &\leq \|r_k\|_{\mathcal{Y}}^2 b_k^\varrho \leq \cdots \\
&\leq \|r_1\|_{\mathcal{Y}}^2 b_1^\varrho \leq \|r_0\|_{\mathcal{Y}}^2 b_0^\varrho = \|g\|_{\mathcal{Y}}^2 |g|_{\mathcal{T}\mathcal{D}}^\varrho \leq |g|_{\mathcal{T}\mathcal{D}}^{2+\varrho},
\end{aligned}$$

where the last inequality is true due to Lemma 9.18.

Thus, using Eq. (9.18), we obtain

$$\|r_k\|_{\mathcal{Y}}^{4+2\varrho} = \|r_k\|_{\mathcal{Y}}^{2\varrho} \|r_k\|_{\mathcal{Y}}^4 \leq \|r_k\|_{\mathcal{Y}}^{2\varrho} b_k^{-2\varrho} |g|_{\mathcal{T}\mathcal{D}}^{4+2\varrho} \leq |g|_{\mathcal{T}\mathcal{D}}^{4+2\varrho} \left(1 + k\varrho^2\right)^{-\varrho},$$

which implies Eq. (9.15). ■

As already mentioned, the constant $|g|_{\mathcal{T}\mathcal{D}}$ in the rate of convergence is connected to how good the dictionary matches the data, which is plausible. Also, it has already been laid out that the constant does not need to be finite such that the inequality Eq. (9.15) is meaningless if it is infinite. Nevertheless, the proof of convergence of the algorithm did not need the finiteness of $|g|_{\mathcal{T}\mathcal{D}}$ such that only the rate of convergence depends on that property.

Corollary 9.21. If it is not known whether $g \in \overline{\mathcal{V}}$, one still obtains

$$\|r_k - r_\infty\|_{\mathcal{Y}} = \left\|\mathcal{P}_{\overline{\mathcal{V}}} g - \mathcal{T} f_k\right\|_{\mathcal{Y}} \leq \left|\mathcal{P}_{\overline{\mathcal{V}}} g\right|_{\mathcal{T}\mathcal{D}} \left(1 + k\varrho^2\right)^{-\varrho/(4+2\varrho)}.$$

Proof. First, notice that

$$r_k - r_\infty = g - \mathcal{T}f_k - (g - \mathcal{T}f_\infty) = \mathcal{T}f_\infty - \mathcal{T}f_k = \mathcal{P}_{\overline{\mathcal{V}}}g - \mathcal{T}f_k$$

holds, where the last equality is true due to Corollary 9.14. Thus, for arbitrary $d \in \mathcal{D}$, we have

$$\langle r_k - r_\infty, \mathcal{T}d \rangle_{\mathcal{Y}} = \langle \mathcal{P}_{\overline{\mathcal{V}}}g - \mathcal{T}f_k, \mathcal{T}d \rangle_{\mathcal{Y}}.$$

Application of Theorem 9.11 yields $\langle r_\infty, \mathcal{T}d \rangle_{\mathcal{Y}} = 0$ such that

$$\langle g - \mathcal{T}f_k, \mathcal{T}d \rangle_{\mathcal{Y}} = \langle \mathcal{P}_{\overline{\mathcal{V}}}g - \mathcal{T}f_k, \mathcal{T}d \rangle_{\mathcal{Y}}. \tag{9.19}$$

We will now prove that the iterates $(f_k)_{k\in\mathbb{N}_0}$, which are provided by the WFMP when it is applied to the inverse problem

$$\mathcal{T}f = g \tag{9.20}$$

may also be generated by the WFMP, when it is applied to the inverse problem

$$\mathcal{T}f = \mathcal{P}_{\overline{\mathcal{V}}}g. \tag{9.21}$$

For this purpose, consider the characterization of d_{k+1} in Eq. (9.3) for both inverse problems. It turns out that due to Eq. (9.19) every choice of d_{k+1} for the inverse problem Eq. (9.20) is also a valid choice for the inverse problem Eq. (9.21) and vice versa. Furthermore, the definition of α_{k+1} in Eq. (9.4) yields the same result in both settings if the same element d_{k+1} from the dictionary is chosen.

Thus, the sequence $(f_k)_{k\in\mathbb{N}_0}$ could also be generated by the WFMP, when it is applied to Eq. (9.21). Since $\mathcal{P}_{\overline{\mathcal{V}}}g \in \overline{V}$, we can apply Theorem 9.20 and obtain the desired result. ■

Remark. Note that, in general, the iterates of the WFMP for the inverse problems Eq. (9.20) and Eq. (9.21) are not identical due to the non-uniqueness of the choice of d_{k+1}. Nevertheless, the proof shows that the convergence rate is the same for both problems.

In consequence, for the FMP, that is, the case $\varrho = 1$, we obtain the following convergence rate.

Corollary 9.22. For the sequence $(r_k)_{k\in\mathbb{N}_0}$ generated by the FMP and its limit r_∞, we have

$$\|r_k - r_\infty\|_{\mathcal{Y}} = \left\|\mathcal{P}_{\overline{\mathcal{V}}}g - \mathcal{T}f_k\right\|_{\mathcal{Y}} \leq \left|\mathcal{P}_{\overline{\mathcal{V}}}g\right|_{\mathcal{T}\mathcal{D}}(1+k)^{-1/6}. \tag{9.22}$$

In Fischer [45] and Fischer and Michel [46], the convergence rate

$$\|r_k\|_{\mathcal{Y}} \leq \|g\|_{\mathcal{Y}} \left(1 - I(\mathcal{D})^2\right)^{k/2} \tag{9.23}$$

is proved, where $\mathcal{Y} = \mathbb{R}^\ell$ and

$$I(\mathcal{D}) = \inf_{v \in \mathbb{S}^{\ell-1}} \sup_{d \in \mathcal{D}} |\langle v, \mathcal{T}d \rangle_{\mathbb{R}^\ell}|.$$

If we compare Eq. (9.23) to Eq. (9.22) for $g \in \overline{\mathcal{V}}$, that is,

$$\|r_k\|_{\mathcal{Y}} \leq |g|_{\mathcal{T}\mathcal{D}} (1+k)^{-1/6}, \tag{9.24}$$

where

$$|g|_{\mathcal{T}\mathcal{D}} = \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; g = \sum_{k=1}^{\infty} \bar{\beta}_k \mathcal{T}\bar{d}_k, \bar{\beta}_k \in \mathbb{R}, \bar{d}_k \in \mathcal{D} \right\},$$

we can make a few observations. Concerning the dependence of the convergence rate on the iteration index k, both inequalities are qualitatively different. Eq. (9.23) is exponential in k, whereas Eq. (9.24) is rational in k such that the first convergence rate seems to be preferable at the first sight. On the other hand, the basis of the exponential term in Eq. (9.23) depends on the dictionary and can be arbitrarily close to 1 (note that it has been shown in Fischer [45] and Fischer and Michel [46], that it is not equal to 1). Furthermore, the term $I(\mathcal{D})$ is difficult to compute for a given operator $\mathcal{T}$ and a dictionary $\mathcal{D}$. Although this is also true for the term $|g|_{\mathcal{T}\mathcal{D}}$ in Eq. (9.24), in the latter inequality the dependence on $\mathcal{D}$ is at least only contained in the constant, whereas the asymptotic convergence rate depends on the iteration index only. From these considerations we conclude that none of the two inequalities should be preferred over the other one, since both of them have their advantages and disadvantages.

A similarity of both inequalities is that the convergence rate gets better if the dictionary is better adapted to the data, which is characterized by a large value of $I(\mathcal{D})$ and a small value of $|g|_{\mathcal{T}\mathcal{D}}$. Fischer [45, Section 3.3] has already concluded from Eq. (9.23) that one should consequently choose a dictionary, which is adapted to the data, at least if some information about the signal structure is known. The newly proved convergence rate in Eq. (9.24) thus supports this argument, which has already been given when the FMP was developed. The topic of the PhD thesis by Schneider [159] (work in progress) in the Geomathematics Group at the University of Siegen will be exactly the problem of finding an optimal dictionary for classes of inverse problems, which relates to the concept of *dictionary learning* that is often applied in machine learning.

In this section we only considered inverse problems that are well-posed in the sense of Nashed. In the following section, we will deduce a regularized version of the WFMP, which can be applied to ill-posed inverse problems.

9.2. The Regularized Weak Functional Matching Pursuit (RWFMP)

As already mentioned in Chapter 6, many inverse problems are ill-posed, a term which we paraphrase according to Nashed as $\overline{\operatorname{ran} \mathcal{T}} \neq \operatorname{ran} \mathcal{T}$. In consequence, the WFMP cannot be applied to the inverse problem, since the convergence could only be proved for well-posed inverse problems in the previous section. Thus, a regularization technique has to be applied.

9.2.1. The algorithm

As already mentioned in Section 8.3, Fischer [45], Fischer and Michel [46], and Michel [123] realized a regularization of the inverse problem by adding a penalty term, which is equivalent to the application of a Tikhonov regularization, yielding the so-called Regularized Functional Matching Pursuit (RFMP) algorithm. As for the FMP, we apply the strategy of the WGA to the RFMP to obtain the following algorithm.

Algorithm 9.23 (Regularized Weak Functional Matching Pursuit, RWFMP)**.** Let $\mathcal{X}, \mathcal{Y}, \mathcal{T}$ be given as in Problem 6.3. Furthermore, let data $g \in \mathcal{Y}$, a weakness parameter $\varrho \in (0, 1]$, a regularization parameter $\lambda > 0$, and the initial approximation $f_0 = 0 \in \mathcal{X}$ be given. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$, whose elements $d \in \mathcal{D}$ satisfy $\|\mathcal{T}d\|_{\mathcal{Y}}^2 + \lambda \|d\|_{\mathcal{X}}^2 = 1$.

1. Set $k := 0$, define the residual $r_0 := g - \mathcal{T}f_0 = g$ and choose a stopping criterion.
2. Find an element $d_{k+1} \in \mathcal{D}$, which fulfills

$$|\langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1}\rangle_{\mathcal{X}}| \geq \varrho \sup_{d \in \mathcal{D}} |\langle r_k, \mathcal{T}d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}|. \tag{9.25}$$

   Set

$$\alpha_{k+1} := \langle r_k, \mathcal{T}d_{k+1}\rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1}\rangle_{\mathcal{X}}, \tag{9.26}$$

   as well as $f_{k+1} := f_k + \alpha_{k+1} d_{k+1}$ and $r_{k+1} := g - \mathcal{T}f_{k+1} = r_k - \alpha_{k+1}\mathcal{T}d_{k+1}$.
3. If the stopping criterion is fulfilled, then f_{k+1} is the output. Otherwise, increase k by 1 and return to step 2.

Similarly to the non-regularized case, the RWFMP algorithm coincides with the original RFMP if $\varrho = 1$, up to the normalization of the dictionary.

It now comes into play that we were able to show convergence of the WFMP for arbitrary (especially infinite-dimensional) Hilbert spaces $\mathcal{Y}$. The following strategy could not be pursued in the previous setting, where $\mathcal{Y} = \mathbb{R}^\ell$, or at least $\dim \mathcal{Y} < \infty$, was required. We will give an interpretation of the RWFMP as WFMP for a modified well-posed inverse problem

$$\tilde{\mathcal{T}}_\lambda f = \tilde{g}, \tag{9.27}$$

where $\tilde{\mathcal{T}}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$. Let us first equip the space $\mathcal{Y} \times \mathcal{X}$ with an inner product to obtain a Hilbert space.

Lemma 9.24. Let $(\mathcal{X}, \langle \cdot, \cdot \rangle_{\mathcal{X}})$ and $(\mathcal{Y}, \langle \cdot, \cdot \rangle_{\mathcal{Y}})$ be Hilbert spaces. Then

$$\left\langle \begin{pmatrix} g_1 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_2 \\ f_2 \end{pmatrix} \right\rangle_{\mathcal{Y} \times \mathcal{X}} := \langle g_1, g_2 \rangle_{\mathcal{Y}} + \langle f_1, f_2 \rangle_{\mathcal{X}}, \qquad \begin{pmatrix} g_1 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_2 \\ f_2 \end{pmatrix} \in \mathcal{Y} \times \mathcal{X},$$

defines an inner product on $\mathcal{Y} \times \mathcal{X}$ and this space is complete with respect to the given inner product.

Furthermore, the associated norm is given by

$$\left\| \begin{pmatrix} g \\ f \end{pmatrix} \right\|_{\mathcal{Y} \times \mathcal{X}} := \sqrt{\|g\|_{\mathcal{Y}}^2 + \|f\|_{\mathcal{X}}^2}, \qquad \begin{pmatrix} g \\ f \end{pmatrix} \in \mathcal{Y} \times \mathcal{X}.$$

Using this topology on the space $\mathcal{Y} \times \mathcal{X}$, we obtain the following lemma.

Lemma 9.25. When using the same data, initial approximation, and weakness parameter, the RWFMP in Algorithm 9.23 produces iterates $(f_k)_{k \in \mathbb{N}_0}$, which are also valid iterates generated by the WFMP in Algorithm 9.1 if the latter is applied to the inverse problem

$$\tilde{\mathcal{T}}_\lambda f = \tilde{g},$$

where $\tilde{\mathcal{T}}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$, $\tilde{\mathcal{T}}_\lambda f := \begin{pmatrix} \mathcal{T} f \\ \sqrt{\lambda} f \end{pmatrix}$, and $\tilde{g} := \begin{pmatrix} g \\ 0 \end{pmatrix} \in \mathcal{Y} \times \mathcal{X}$.

Proof. Since all of the input parameters are the same, it remains to show that one obtains Eq. (9.25) and Eq. (9.26) if one inserts $\tilde{\mathcal{T}}_\lambda$ and $\tilde{y}$ into Eq. (9.3) and Eq. (9.4), respectively.

Fortunately, this can easily be seen, since for $d \in \mathcal{D}$, we have

$$\begin{aligned}\langle \tilde{r}_k, \tilde{\mathcal{T}}_\lambda d\rangle_{\mathcal{Y}\times\mathcal{X}} &= \langle \tilde{g} - \tilde{\mathcal{T}}_\lambda f_k, \tilde{\mathcal{T}}_\lambda d\rangle_{\mathcal{Y}\times\mathcal{X}} \\ &= \left\langle \begin{pmatrix} g - \mathcal{T} f_k \\ 0 - \sqrt{\lambda} f_k \end{pmatrix}, \begin{pmatrix} \mathcal{T} d \\ \sqrt{\lambda} d \end{pmatrix} \right\rangle_{\mathcal{Y}\times\mathcal{X}} \\ &= \langle g - \mathcal{T} f_k, \mathcal{T} d\rangle_{\mathcal{Y}} + \left\langle -\sqrt{\lambda} f_k, \sqrt{\lambda} d \right\rangle_{\mathcal{X}} \\ &= \langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}\end{aligned}$$

and

$$\begin{aligned}\|\tilde{\mathcal{T}}_\lambda d\|^2_{\mathcal{Y}\times\mathcal{X}} &= \left\| \begin{pmatrix} \mathcal{T} d \\ \sqrt{\lambda} d \end{pmatrix} \right\|^2_{\mathcal{Y}\times\mathcal{X}} \\ &= \|\mathcal{T} d\|^2_{\mathcal{Y}} + \left\| \sqrt{\lambda} d \right\|^2_{\mathcal{X}} \\ &= \|\mathcal{T} d\|^2_{\mathcal{Y}} + \lambda \|d\|^2_{\mathcal{X}}.\end{aligned}$$

Note that in accordance to the other definitions, we used the notation $\tilde{r}_k \coloneqq \tilde{g} - \tilde{\mathcal{T}}_\lambda f_k$ and $r_k \coloneqq g - \mathcal{T} f_k$ in the considerations above. ■

Note that the remark that was stated after Corollary 9.21 is also true in this case. We cannot expect that the RWFMP for the original problem and the WFMP for the modified problem produce identical iterates. The important result is that the iterates of one algorithm fulfill the selection criterion of the other algorithm and could thus be chosen there.

In Section 9.1, we assumed well-posedness of the inverse problem. Since the idea of a regularization is to substitute an ill-posed problem by a related well-posed problem, it is well-known that the minimization of the Tikhonov functional is well-posed. This can also been characterized in terms of the modified operator $\tilde{\mathcal{T}}_\lambda$.

Lemma 9.26. For the operator

$$\tilde{\mathcal{T}}_\lambda \colon \mathcal{X} \to \mathcal{Y} \times \mathcal{X}, \qquad \tilde{\mathcal{T}}_\lambda f = \begin{pmatrix} \mathcal{T} f \\ \sqrt{\lambda} f \end{pmatrix},$$

we have $\overline{\operatorname{ran} \tilde{\mathcal{T}}_\lambda} = \operatorname{ran} \tilde{\mathcal{T}}_\lambda$ such that the inverse problem in Eq. (9.27) is well-posed in the sense of Nashed.

Proof. Neglecting a permutation of the components, for $\lambda > 0$ the set $\operatorname{ran} \tilde{\mathcal{T}}_\lambda$ is the graph of the operator $\lambda^{-1/2}\mathcal{T}$. The assertion follows, since every continuous operator has a closed graph (cf. Rudin [155], Proposition 2.14). ■

9.2.2. Convergence results

By the application of Lemma 9.25, the following results about the convergence of the RWFMP are direct consequences of the corresponding results for the WFMP. In the following, we always assume that the sequence $(f_k)_{k\in\mathbb{N}_0}$ is generated by the RWFMP.

Lemma 9.27. The sequence $\left(\|r_k\|_{\mathcal{Y}}^2 + \lambda\|f_k\|_{\mathcal{X}}^2\right)_{k\in\mathbb{N}_0}$ is monotonically decreasing and convergent.

Proof. Since

$$\|\tilde{r}_k\|_{\mathcal{Y}\times\mathcal{X}}^2 = \|r_k\|_{\mathcal{Y}}^2 + \lambda\|f_k\|_{\mathcal{X}}^2,$$

this is a consequence of Lemma 9.2. ■

An immediate consequence of Theorem 9.13 is the following result.

Theorem 9.28. Under the condition (SFC) of Theorem 9.13, the sequence $(f_k)_{k\in\mathbb{N}_0}$ that is generated by the RWFMP converges to $f_\infty = \sum_{k=1}^{\infty} \alpha_k d_k \in \mathcal{X}$.

Theorem 9.29. If $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X}$, the limit f_∞ fulfills the Tikhonov-regularized normal equation

$$(\mathcal{T}^*\mathcal{T} + \lambda\mathcal{I})f_\infty = \mathcal{T}^* g.$$

Proof. First, observe that $\tilde{\mathcal{T}}_\lambda : \mathcal{X} \to \mathcal{Y}\times\mathcal{X}$ is injective due to the identity operator in the second component (and since $\lambda > 0$). Hence, $\operatorname{null}\tilde{\mathcal{T}}_\lambda = \{0\}$ and $(\operatorname{null}\tilde{\mathcal{T}}_\lambda)^\perp = \mathcal{X}$. Thus, since $\overline{\operatorname{span}\mathcal{D}} = \mathcal{X} = (\operatorname{null}\tilde{\mathcal{T}}_\lambda)^\perp$, from Theorem 9.15 we obtain that f_∞ fulfills

$$\tilde{\mathcal{T}}_\lambda^*\tilde{\mathcal{T}}_\lambda f_\infty = \tilde{\mathcal{T}}_\lambda^*\tilde{g}. \tag{9.28}$$

The adjoint operator of $\tilde{\mathcal{T}}_\lambda$ is given by

$$\tilde{\mathcal{T}}_\lambda^* : \mathcal{Y}\times\mathcal{X} \to \mathcal{X}, \quad \tilde{\mathcal{T}}_\lambda^*\tilde{z} = \mathcal{T}^* z + \sqrt{\lambda}w, \quad \text{where } \tilde{z} = \begin{pmatrix} z \\ w \end{pmatrix} \in \mathcal{Y}\times\mathcal{X},$$

since for $\tilde{z} = \begin{pmatrix} z \\ w \end{pmatrix} \in \mathcal{Y}\times\mathcal{X}$ and $f \in \mathcal{X}$, we obtain

$$\begin{aligned}\langle \tilde{z}, \tilde{\mathcal{T}}_\lambda f\rangle_{\mathcal{Y}\times\mathcal{X}} &= \left\langle \begin{pmatrix} z \\ w \end{pmatrix}, \begin{pmatrix} \mathcal{T}f \\ \sqrt{\lambda}f \end{pmatrix} \right\rangle_{\mathcal{Y}\times\mathcal{X}} = \langle z, \mathcal{T}f\rangle_{\mathcal{Y}} + \left\langle w, \sqrt{\lambda}f \right\rangle_{\mathcal{X}} \\ &= \langle \mathcal{T}^* z, f\rangle_{\mathcal{X}} + \left\langle \sqrt{\lambda}w, f \right\rangle_{\mathcal{X}} = \left\langle \mathcal{T}^* z + \sqrt{\lambda}w, f \right\rangle_{\mathcal{X}}.\end{aligned}$$

Finally, elaborating both sides of Eq. (9.28) as

$$\tilde{\mathcal{T}}_\lambda^* \tilde{\mathcal{T}}_\lambda f_\infty = \tilde{\mathcal{T}}_\lambda^* \begin{pmatrix} \mathcal{T} f_\infty \\ \sqrt{\lambda} f_\infty \end{pmatrix} = \mathcal{T}^* \mathcal{T} f_\infty + \lambda f_\infty = (\mathcal{T}^* \mathcal{T} + \lambda \,\mathrm{id}) f_\infty,$$

and

$$\tilde{\mathcal{T}}_\lambda^* \tilde{g} = \tilde{\mathcal{T}}_\lambda^* \begin{pmatrix} g \\ 0 \end{pmatrix} = \mathcal{T}^* g,$$

we have proved the assertion. ■

Remark. From the theory of Tikhonov regularization, it is well-known that a solution of the regularized normal equation is also the unique minimizer of the Tikhonov functional

$$f \mapsto \|g - \mathcal{T} f\|_{\mathcal{Y}}^2 + \lambda \|f\|_{\mathcal{X}}^2 \tag{9.29}$$

(see Section 6.1.4). Thus, by Theorem 9.29, the limit f_∞ of the RWFMP iteration is the unique minimizer of Eq. (9.29). The RWFMP, consequently, has an interpretation of an iterative minimization algorithm for the Tikhonov functional. Actually, the RFMP was derived as such an algorithm in Fischer [45] such that this interpretation again fits well.

From Theorem 9.20 and Corollary 9.21, we can also derive a convergence rate of the algorithm, measured in the Tikhonov functional. The following lemma states that the constant $|\tilde{\mathcal{T}}_\lambda f_\infty|_{\tilde{\mathcal{T}}_\lambda \mathcal{D}}$, which arises consequently, can be rewritten to represent how sparse the solution f_∞ can be expressed as a linear combination of dictionary elements.

Lemma 9.30. Let $f \in \mathcal{X}$, then

$$|\tilde{\mathcal{T}}_\lambda f|_{\tilde{\mathcal{T}}_\lambda \mathcal{D}} = |f|_{\mathcal{D}} := \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; f = \sum_{k=1}^{\infty} \bar{\beta}_k \bar{d}_k,\ \bar{\beta}_k \in \mathbb{R},\ \bar{d}_k \in \mathcal{D} \right\}.$$

Proof. We have

$$\begin{aligned} |\tilde{\mathcal{T}}_\lambda f|_{\tilde{\mathcal{T}}_\lambda \mathcal{D}} &= \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; \tilde{\mathcal{T}}_\lambda f = \sum_{k=1}^{\infty} \bar{\beta}_k \tilde{\mathcal{T}}_\lambda \bar{d}_k, \bar{\beta}_k \in \mathbb{R}, \bar{d}_k \in \mathcal{D} \right\} \\ &= \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; \begin{pmatrix} \mathcal{T} f \\ \sqrt{\lambda} f \end{pmatrix} = \sum_{k=1}^{\infty} \bar{\beta}_k \begin{pmatrix} \mathcal{T} \bar{d}_k \\ \sqrt{\lambda} \bar{d}_k \end{pmatrix}, \bar{\beta}_k \in \mathbb{R}, \bar{d}_k \in \mathcal{D} \right\} \\ &= \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; \begin{pmatrix} \mathcal{T} f \\ f \end{pmatrix} = \sum_{k=1}^{\infty} \bar{\beta}_k \begin{pmatrix} \mathcal{T} \bar{d}_k \\ \bar{d}_k \end{pmatrix}, \bar{\beta}_k \in \mathbb{R}, \bar{d}_k \in \mathcal{D} \right\} \end{aligned}$$

$$= \inf \left\{ \sum_{k=1}^{\infty} |\bar{\beta}_k| \;\middle|\; f = \sum_{k=1}^{\infty} \bar{\beta}_k \bar{d}_k, \bar{\beta}_k \in \mathbb{R}, \bar{d}_k \in \mathcal{D} \right\} = |f|_{\mathcal{D}},$$

where the second to last equality is true since the equation in the second component implies the equation in the first component due to the linearity and continuity of the operator $\mathcal{T}$. ■

Consequently, we obtain the following convergence rate result.

Theorem 9.31. For the sequence $(f_k)_{k\in\mathbb{N}_0}$ generated by the RWFMP, we have

$$\|r_k - r_\infty\|_{\mathcal{Y}}^2 + \lambda \|f_k - f_\infty\|_{\mathcal{X}}^2 \leq |f_\infty|_{\mathcal{D}}^2 \left(1 + k\varrho^2\right)^{-\varrho/(2+\varrho)}.$$

For the RFMP, we obtain the following convergence rate by setting $\varrho = 1$.

Corollary 9.32. For the sequences $(r_k)_{k\in\mathbb{N}_0}$ and $(f_k)_{k\in\mathbb{N}_0}$ generated by the RFMP and their limits r_∞ and f_∞, we have

$$\|r_k - r_\infty\|_{\mathcal{Y}}^2 + \lambda \|f_k - f_\infty\|_{\mathcal{X}}^2 \leq |f_\infty|_{\mathcal{D}}^2 (1 + k)^{-1/3}.$$

Note that the right-hand side does not seem to depend on the operator $\mathcal{T}$ and the regularization parameter λ. However, this is not true, since f_∞ is the minimizer of the Tikhonov functional corresponding to $\mathcal{T}$ and λ such that there is an implicit dependence on both of them.

In Section 9.1 we presented a different convergence rate for the FMP that was given by Fischer [45] and Fischer and Michel [46]. This result could not be transferred to the RFMP. Thus, the previous theorem and the corollary are the first convergence rates that were ever proved for the regularized versions of the greedy algorithms for linear inverse problems.

Moreover, it is now possible to estimate the approximation error $\|f_k - f_\infty\|_{\mathcal{X}}$, which could not be done before.

It turns out that the results of this section enable us to prove that the RWFMP is a convergent regularization method as defined in Section 6.1.3.

9.2.3. The RWFMP as a convergent regularization method

This section is dedicated to the analysis of the RWFMP as an iterative regularization algorithm. We will present an a-priori parameter choice for the Tikhonov regularization parameter λ and the number of iterations K after which the RWFMP is stopped in dependence on the noise level δ of the data. For optimal choices of λ and K, we obtain a convergence rate of $\delta^{2/3}$ for $\delta \to 0$, which is the optimal rate for Tikhonov regularization, on which the RWFMP is based.

For this purpose, by $(f^\delta_{\lambda,k})_{k\in\mathbb{N}_0}$ we denote the sequence of iterates of the RWFMP, when the algorithm is applied to the inverse problem $\mathcal{T}f = g^\delta$ using the regularization parameter $\lambda > 0$. Here, $g^\delta \in \mathcal{Y}$ denotes noisy data, which fulfill $\|g^\delta - g\|_{\mathcal{Y}} \leq \delta$.

Let $f^\delta_{\lambda,\infty}$ be the minimizer of the Tikhonov functional and $f^+ = \mathcal{T}^+ g$ be the best approximate solution of $\mathcal{T}f = g$ as already defined in Section 6.1. We will again state a version of part (b) of Theorem 6.22, since we will apply it in our analysis.

Theorem 9.33. If $f^+ = \mathcal{T}^*\mathcal{T}h$ for some $h \in \mathcal{X}$ with $\|h\|_{\mathcal{X}} \leq \tau$ and if $\lambda(\delta) = m_1 (\delta/\tau)^{2/3}$ for constants $m_1, \tau > 0$, then

$$\left\| f^\delta_{\lambda(\delta),\infty} - f^+ \right\|_{\mathcal{X}} \leq \left(\frac{1}{2\sqrt{m_1}} + m_1 \right) \tau^{1/3}\, \delta^{2/3} = C_1\, \delta^{2/3}$$

for all $\delta > 0$, where $C_1 := \left(\frac{1}{2\sqrt{m_1}} + m_1\right) \tau^{1/3}$.

Furthermore, from Theorem 9.31, we immediately obtain the following corollary.

Corollary 9.34. Let $(f^\delta_{\lambda,k})_{k\in\mathbb{N}_0}$ be the sequence of iterates of the RWFMP with weakness parameter $\varrho \in (0,1]$, when the algorithm is applied to the inverse problem $\mathcal{T}f = g^\delta$ using the regularization parameter $\lambda > 0$. Furthermore, assume $\left| f^\delta_{\lambda,\infty} \right|_{\mathcal{D}} < \infty$. Then, for all $k \in \mathbb{N}_0$, we have

$$\left\| f^\delta_{\lambda,k} - f^\delta_{\lambda,\infty} \right\|_{\mathcal{X}} \leq \left| f^\delta_{\lambda,\infty} \right|_{\mathcal{D}} \lambda^{-1/2}\, \varrho^{-\varrho/(2+\varrho)}\, k^{-\varrho/(4+2\varrho)} = C_2(\lambda)\, \lambda^{-1/2}\, k^{-\varrho/(4+2\varrho)},$$

where $C_2(\lambda) := \left| f^\delta_{\lambda,\infty} \right|_{\mathcal{D}} \varrho^{-\varrho/(2+\varrho)}$.

The preceding theorem and the corollary enable us to prove the following result.

Theorem 9.35. Let $(f^\delta_{\lambda,k})_{k\in\mathbb{N}_0}$ be the sequence of iterations of the RWFMP with weakness parameter $\varrho \in (0,1]$, when the algorithm is applied to the inverse problem $\mathcal{T}f = g^\delta$ using the regularization parameter $\lambda > 0$.

Furthermore, we assume that there exists $C_3 > 0$ such that $C_2(\lambda) \leq C_3$ for all $\lambda > 0$.

Additionally, let $f^+ = \mathcal{T}^*\mathcal{T}h$ for some $h \in \mathcal{X}$ with $\|h\|_{\mathcal{X}} \leq \tau$ and let $\lambda(\delta) = m_1 (\delta/\tau)^{2/3}$ as well as $K(\delta) = m_2\, \delta^{-(4+2\varrho)/\varrho}$ for some constants $m_1, m_2 > 0$. Then there exists $C > 0$ such that

$$\left\| f^\delta_{\lambda(\delta),K(\delta)} - f^+ \right\|_{\mathcal{X}} \leq C\, \delta^{2/3}$$

for all $\delta > 0$.

Proof. Let $\delta > 0$ be chosen arbitrarily. Since all conditions of Theorem 9.33 are fulfilled, we obtain

$$\left\| f^\delta_{\lambda(\delta),\infty} - f^+ \right\|_{\mathcal{X}} \leq C_1\, \delta^{2/3},$$

where $f^\delta_{\lambda(\delta),\infty}$ is the minimizer of the Tikhonov functional with regularization parameter $\lambda(\delta)$.

Corollary 9.34 and the assumption $C_2(\lambda) \leq C_3$ yield

$$\begin{aligned}\left\| f^\delta_{\lambda(\delta),K(\delta)} - f^\delta_{\lambda(\delta),\infty} \right\|_{\mathcal{X}} &\leq C_2(\lambda(\delta))\, (\lambda(\delta))^{-1/2}\, (K(\delta))^{-\varrho/(4+2\varrho)} \\ &\leq C_3 \cdot (\lambda(\delta))^{-1/2}\, (K(\delta))^{-\varrho/(4+2\varrho)}.\end{aligned}$$

The insertion of the definition of $\lambda(\delta)$ and $K(\delta)$ gives

$$\begin{aligned}\left\| f^\delta_{\lambda(\delta),K(\delta)} - f^\delta_{\lambda(\delta),\infty} \right\|_{\mathcal{X}} &\leq C_3\, (m_1\, (\delta/\tau)^{2/3})^{-1/2} \left(m_2\, \delta^{-(4+2\varrho)/\varrho} \right)^{-\varrho/(4+2\varrho)} \\ &= C_4\, \delta^{-1/3}\, \delta = C_4\, \delta^{2/3},\end{aligned}$$

where $C_4 := C_3\, m_1^{-1/2}\, m_2^{-\varrho/(4+2\varrho)}\, \tau^{1/3}$.

In conclusion, we have by the triangle inequality that

$$\begin{aligned}\left\| f^\delta_{\lambda(\delta),K(\delta)} - f^+ \right\|_{\mathcal{X}} &\leq \left\| f^\delta_{\lambda(\delta),K(\delta)} - f^\delta_{\lambda(\delta),\infty} \right\|_{\mathcal{X}} + \left\| f^\delta_{\lambda(\delta),\infty} - f^+ \right\|_{\mathcal{X}} \\ &\leq C_4\, \delta^{2/3} + C_1 \delta^{2/3} = (C_4 + C_1)\, \delta^{2/3},\end{aligned}$$

which proves the claim for $C := C_4 + C_1$. ∎

The preceding theorem is important, since in the implementation of the RWFMP algorithm only a finite number of iterations can be realized, but results about the convergence of Tikhonov regularization can only be applied directly if we allow for an infinite number of iterations. The theorem shows that by stopping the RWFMP iteration after $K(\delta) \sim \delta^{-(4+2\varrho)/\varrho}$ steps, one is able to conserve the convergence rate $\delta^{2/3}$ for $\delta \to 0$ of Tikhonov regularization even with only a finite number of iterations.

9.3. Numerical Example

In this section, we will present a proof of concept that the *weak* approach is not only advantageous from the theoretical perspective as already mentioned, but also offers the opportunity to speed up the iteration of the RFMP.

The main difference of the RFMP and the RWFMP is the selection criterion for the next dictionary element d_{k+1}. For the RFMP, it is given by

$$|\langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1}\rangle_{\mathcal{X}}| = \sup_{d\in\mathcal{D}} |\langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}| \tag{9.30}$$

and for the RWFMP the criterion is

$$|\langle r_k, \mathcal{T} d_{k+1}\rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1}\rangle_{\mathcal{X}}| \geq \varrho \sup_{d\in\mathcal{D}} |\langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}|. \tag{9.31}$$

Here, we adopted the normalization $\|\mathcal{T} d\|_{\mathcal{Y}}^2 + \lambda \|d\|_{\mathcal{X}}^2 = 1$ for the dictionary elements $d \in \mathcal{D}$ from the RWFMP also for the RFMP. On a computer, we can only realize a finite dictionary, that is, $\#\mathcal{D} = J < \infty$, $J \in \mathbb{N}$. At the first sight, it seems that the implementation of the RWFMP has no advantage over the RFMP, since for the determination of a d_{k+1} that fulfills Eq. (9.31), the supremum has to be known. That is, one would have to go through all of the finitely many dictionary elements to determine the maximum. The maximizer, which can of course be stored, fulfills Eq. (9.30) such that an application of the RWFMP would have no advantage.

For this reason, we pursue a different approach. We observe that

$$\begin{aligned}
\sup_{d\in\mathcal{D}} |\langle r_k, \mathcal{T} d\rangle_{\mathcal{Y}} - \lambda \langle f_k, d\rangle_{\mathcal{X}}| &= \sup_{d\in\mathcal{D}} \left| \left\langle \begin{pmatrix} r_k \\ -\sqrt{\lambda} f_k \end{pmatrix}, \begin{pmatrix} \mathcal{T} d \\ \sqrt{\lambda} d \end{pmatrix} \right\rangle_{\mathcal{Y}\times\mathcal{X}} \right| \\
&\leq \left\| \begin{pmatrix} r_k \\ -\sqrt{\lambda} f_k \end{pmatrix} \right\|_{\mathcal{Y}\times\mathcal{X}} \sup_{d\in\mathcal{D}} \left\| \begin{pmatrix} \mathcal{T} d \\ \sqrt{\lambda} d \end{pmatrix} \right\|_{\mathcal{Y}\times\mathcal{X}} \\
&= \sqrt{\|r_k\|_{\mathcal{Y}}^2 + \lambda \|f_k\|_{\mathcal{X}}^2} \sup_{d\in\mathcal{D}} \sqrt{\|\mathcal{T} d\|_{\mathcal{Y}}^2 + \lambda \|d\|_{\mathcal{X}}^2}
\end{aligned}$$

$$= \sqrt{\|r_k\|_{\mathcal{Y}}^2 + \lambda \|f_k\|_{\mathcal{X}}^2}$$

by the Cauchy-Schwarz inequality and the normalization of the dictionary.

In consequence, if we choose $d_{k+1} \in \mathcal{D}$ such that

$$|\langle r_k, \mathcal{T} d_{k+1} \rangle_{\mathcal{Y}} - \lambda \langle f_k, d_{k+1} \rangle_{\mathcal{X}}| \geq \varrho \sqrt{\|r_k\|_{\mathcal{Y}}^2 + \lambda \|f_k\|_{\mathcal{X}}^2}, \tag{9.32}$$

then it also fulfills Eq. (9.31) and, thus, can be chosen by the RWFMP. Since the term on the right-hand side does no longer depend on $d \in \mathcal{D}$, it can be computed without going through the whole dictionary.

This enables us to improve the search procedure for a next dictionary element. Instead of going through all of the dictionary elements and storing the index of the maximizer, we can abort this search as soon as a dictionary element fulfills Eq. (9.32). Since, in most of the cases, this will not be the last element in the dictionary, there arises an improvement in computation time from this procedure. This will, in particular, be of importance if an a-posteriori parameter choice method is applied, where the algorithm has to be applied several times for a sequence of regularization parameters. Of course, it is possible that there exists no dictionary element that fulfills Eq. (9.32). In this case, one would simply store the dictionary element for which the term on the left-hand side of Eq. (9.32) is maximal and thus perform an RFMP iteration.

In the following section, we will present a one-dimensional model problem, for which we will present numerical results later on.

9.3.1. A one-dimensional model problem

To give a proof of concept for the improved computation time that the RWFMP offers, we restrict to a simple one-dimensional model problem, which we adopt from Rieder [153, Beispiel 3.2.2].

Consider the boundary-value problem

$$\begin{aligned} -g''(x) &= f(x) && \text{for all } x \in (0,1), \\ g(0) &= g(1) = 0 \end{aligned} \tag{9.33}$$

for given $f \in \mathrm{C}((0,1))$, where Eq. (9.33) is the one-dimensional Poisson equation. By applying the concept of a Green's function and switching to spaces of

square-integrable functions, we can deduce the integral operator $\mathcal{T}\colon \mathrm{L}^2((0,1)) \to \mathrm{L}^2((0,1))$,

$$\mathcal{T}f(x) = \int_0^1 k(x,y)\, f(y)\, \mathrm{d}y,$$

where

$$k(x,y) = \begin{cases} x\,(1-y), & x \leq y, \\ y\,(1-x), & x > y, \end{cases}$$

such that $\mathcal{T}f = g$ if and only if f fulfills the weak formulation of the given boundary value problem.

In our implementation, we assume that we are given evaluations $(\mathcal{T}f(x_j))_{j=1,\dots,J}$ for $X = (x_1, \dots, x_J) \in [0,1]^J$, which are denoted by $(g_j)_{j=1,\dots,J} \in \mathbb{R}^J$. We assume that $x_1 < x_2 < \cdots < x_J$. We denote the corresponding operator by $\mathcal{T}_X\colon \mathrm{L}^2([0,1]) \to \mathbb{R}^J$. Note that it is only well-defined if the evaluation is well-defined for all points in the vector X, which is not clear in the space $\mathrm{L}^2([0,1])$, but we will ignore this difficulty, since we will only use continuous dictionary elements in the following.

We will consider two different dictionaries $\mathcal{D}_1, \mathcal{D}_2 \subseteq \mathrm{L}^2([0,1])$.

The first dictionary is motivated by the singular-value decomposition of the operator. Rieder [153] states that a singular system is given by $(1/(\pi\, n)^2, u_n, u_n)_{n\in\mathbb{N}}$, where $u_n(x) = \sqrt{2}\, \sin(\pi\, n\, x)$. We, therefore, first define

$$\tilde{\mathcal{D}}_1 := \{\, \sin(\pi\, n\, \cdot) \mid n = 1, \dots, N \,\}$$

for some $N \in \mathbb{N}$ to obtain the dictionary

$$\mathcal{D}_1^\lambda := \left\{ \frac{d}{\sqrt{\|\mathcal{T}_X d\|_{\mathbb{R}^J}^2 + \lambda \|d\|_{\mathrm{L}^2([0,1])}^2}} \;\middle|\; d \in \tilde{\mathcal{D}}_1 \right\}$$

if the regularization parameter is $\lambda > 0$ to obtain the correct normalization of the elements from the dictionary.

The second dictionary is motivated by one-dimensional finite element methods (see, for example, Johnson [91, Section 1.2]). For this purpose, given a vector $Y = (y_1, \dots, y_S) \in [0,1]^S$ of nodes, which fulfill $y_1 < y_2 < \cdots < y_S$, we define the functions

$$\varphi_s(y) = \begin{cases} 0, & y \leq y_{s-1}, \\ \frac{y - y_{s-1}}{y_s - y_{s-1}}, & y_{s-1} < y \leq y_s, \\ \frac{y_{s+1} - y}{y_{s+1} - y_s}, & y_s < y \leq y_{s+1}, \\ 0, & y_{s+1} < y, \end{cases}$$

for $s = 1, \dots, S$ and $y \in [0,1]$. These functions are *hat functions*, that is, they are piecewise linear between the nodes in Y, and they fulfill $\varphi_s(y_j) = \delta_{sj}$. Using these functions we first define

$$\tilde{\mathcal{D}}_2 \coloneqq \{ \varphi_s \mid s = 1, \dots, S \},$$

and obtain, for the regularization parameter $\lambda > 0$, the dictionary

$$\mathcal{D}_2^\lambda \coloneqq \left\{ \frac{d}{\sqrt{\|\mathcal{T}_X d\|_{\mathbb{R}^J}^2 + \lambda \|d\|_{\mathrm{L}^2([0,1])}^2}} \;\middle|\; d \in \tilde{\mathcal{D}}_2 \right\}$$

by normalization.

9.3.2. Numerical results

To test the improvement of the computation time, when applying the search strategy presented above, we prescribed the solution $f(y) = 1$ for all $y \in [0,1]$ such that $\mathcal{T} f(x) = \frac{1}{2} x\,(1 - x) = g(x)$, which can be easily checked using the boundary value problem or by integration.

We set $J = 1000$ to sample g at 1000 evenly distributed data points in $[0,1]$ and apply 1% of deterministic noise. Furthermore, we chose $N = S = 400$ to obtain 400 elements in both of the dictionaries $\mathcal{D}_1^\lambda, \mathcal{D}_2^\lambda$.

We implemented a simple a-posteriori parameter choice method, where we applied 100 iterations of the RWFMP using 100 different regularization parameters and chose the regularized solution where the approximation error is the smallest. The regularization parameters were chosen equally spaced on a logarithmic scale between 1 and 10^{-14}. Of course, a different parameter choice method must be used if the solution to the problem is not known. For the RFMP, several of these methods were compared by Gutting et al. [70]. For all of them, one needs to execute the algorithm for many regularization parameters such that the saving in computation time will also apply for other parameter choice methods.

Moreover, to show the power of the algorithm for arbitrary dictionaries, we did not use the dictionaries in the order in which they were defined, but instead used 20 random permutations of the dictionary elements to eliminate the bias of dictionaries whose order is by chance very well suited or very badly suited to the problem at hand. Finally, we tested the algorithm for 10 different values of the weakness parameter ϱ.

Table 9.1.: Results for the dictionary $\mathcal{D}_1^\lambda$ of sine functions. The columns are: the weakness parameter ϱ, the mean optimal regularization parameter λ and its standard deviation, the average RMSE and its standard deviation, the mean and standard deviation of the computation time for one execution of the RWFMP, and the percentage of the computation time with respect to the first line. The minimal RMSE and computation time are set in a bold font.

ϱ	λ	RMSE	Computation time/s	/% of first line
1.000000	$1.49 \times 10^{-4} \pm 5.40 \times 10^{-4}$	$6.33 \times 10^{-1} \pm 3.84 \times 10^{-1}$	$2.29 \times 10^{2} \pm 1.24 \times 10^{1}$	100.0%
0.464159	$4.54 \times 10^{-7} \pm 2.91 \times 10^{-7}$	$2.02 \times 10^{-1} \pm 3.65 \times 10^{-2}$	$2.25 \times 10^{2} \pm 1.18 \times 10^{1}$	98.2%
0.215443	$4.32 \times 10^{-7} \pm 2.06 \times 10^{-7}$	$1.94 \times 10^{-1} \pm 2.40 \times 10^{-2}$	$2.25 \times 10^{2} \pm 1.04 \times 10^{1}$	98.0%
0.100000	$2.70 \times 10^{-7} \pm 1.39 \times 10^{-7}$	$1.74 \times 10^{-1} \pm 1.07 \times 10^{-2}$	$2.22 \times 10^{2} \pm 1.23 \times 10^{1}$	97.1%
0.046416	$2.20 \times 10^{-7} \pm 1.08 \times 10^{-7}$	$1.70 \times 10^{-1} \pm 9.34 \times 10^{-3}$	$2.01 \times 10^{2} \pm 8.57$	87.6%
0.021544	$2.70 \times 10^{-7} \pm 9.41 \times 10^{-8}$	$\mathbf{1.62 \times 10^{-1} \pm 5.28 \times 10^{-3}}$	$1.72 \times 10^{2} \pm 8.58$	75.2%
0.010000	$2.70 \times 10^{-7} \pm 9.41 \times 10^{-8}$	$1.63 \times 10^{-1} \pm 1.60 \times 10^{-3}$	$1.59 \times 10^{2} \pm 7.74$	69.6%
0.004642	$2.98 \times 10^{-7} \pm 9.56 \times 10^{-8}$	$1.62 \times 10^{-1} \pm 1.03 \times 10^{-3}$	$1.47 \times 10^{2} \pm 7.70$	64.2%
0.002154	$3.83 \times 10^{-7} \pm 5.43 \times 10^{-23}$	$1.63 \times 10^{-1} \pm 4.90 \times 10^{-4}$	$1.39 \times 10^{2} \pm 5.97$	60.9%
0.001000	$3.83 \times 10^{-7} \pm 5.43 \times 10^{-23}$	$1.63 \times 10^{-1} \pm 1.70 \times 10^{-4}$	$\mathbf{1.33 \times 10^{2} \pm 5.98}$	**58.2**%

Table 9.2.: Results for the dictionary $\mathcal{D}_2^\lambda$ of hat functions. The columns are: the weakness parameter ϱ, the mean optimal regularization parameter λ and its standard deviation, the average RMSE and its standard deviation, the mean and standard deviation of the computation time for one execution of the RWFMP, and the percentage of the computation time with respect to the first line. The minimal RMSE and computation time are set in a bold font.

ϱ	λ	RMSE	Computation time/s	/% of first line
1.000000	$1.08 \times 10^{-2} \pm 3.56 \times 10^{-3}$	$8.55 \times 10^{-1} \pm 2.59 \times 10^{-2}$	$2.25 \times 10^{2} \pm 6.55$	100.0%
0.464159	$9.95 \times 10^{-3} \pm 2.67 \times 10^{-3}$	$8.53 \times 10^{-1} \pm 2.25 \times 10^{-2}$	$2.23 \times 10^{2} \pm 6.36$	98.8%
0.215443	$1.08 \times 10^{-2} \pm 3.56 \times 10^{-3}$	$8.41 \times 10^{-1} \pm 1.40 \times 10^{-2}$	$1.94 \times 10^{2} \pm 1.07 \times 10^{1}$	85.9%
0.100000	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.48 \times 10^{-1} \pm 1.64 \times 10^{-2}$	$6.64 \times 10^{1} \pm 9.95$	29.4%
0.046416	$1.51 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$\mathbf{8.32 \times 10^{-1} \pm 4.41 \times 10^{-3}}$	$3.98 \times 10^{1} \pm 2.49$	17.6%
0.021544	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.38 \times 10^{-1} \pm 6.39 \times 10^{-3}$	$3.05 \times 10^{1} \pm 2.70$	13.4%
0.010000	$9.95 \times 10^{-3} \pm 2.67 \times 10^{-3}$	$8.48 \times 10^{-1} \pm 1.25 \times 10^{-2}$	$2.01 \times 10^{1} \pm 2.95$	8.9%
0.004642	$1.12 \times 10^{-2} \pm 3.86 \times 10^{-3}$	$8.44 \times 10^{-1} \pm 1.48 \times 10^{-2}$	$1.64 \times 10^{1} \pm 3.37$	7.2%
0.002154	$1.16 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$8.46 \times 10^{-1} \pm 1.43 \times 10^{-2}$	$1.64 \times 10^{1} \pm 3.40$	7.2%
0.001000	$1.16 \times 10^{-2} \pm 4.08 \times 10^{-3}$	$8.50 \times 10^{-1} \pm 1.35 \times 10^{-2}$	$\mathbf{1.53 \times 10^{1} \pm 3.23}$	**6.8**%

The results for the dictionary $\mathcal{D}_1^\lambda$ consisting of sine functions are given in Table 9.1. For the dictionary $\mathcal{D}_2^\lambda$ of hat functions, the results can be found in Table 9.2.

In these tables, the first column corresponds to the chosen weakness parameter ϱ. We chose 10 values of ϱ, which are logarithmically distributed between 1 and 0.001. The second column shows the regularization parameter λ, which had been chosen from the 100 prescribed parameters to minimize the approximation error. The values shown are the mean value of λ over all 20 random permutations of the dictionary and the corresponding standard deviation. The mean and the standard deviation are also shown in the third column for the approximation error, which we computed in the form of the *root mean squared error (RMSE)* by the evaluation on a grid in this case. For the fourth column, we computed the average computation time for 100 iterations of the RWFMP and its standard deviation. Finally, the last column shows the ratio of the average computation time to the time needed in the case $\varrho = 1$, that is, the RFMP.

We will first discuss the results for the dictionary consisting of sine functions $\mathcal{D}_1^\lambda$. Concerning the regularization parameter λ, we observe that the chosen optimal parameter—apart from the case $\varrho = 1$—always has the same order of magnitude. This is what we would expect, since all of the algorithms converge to the minimizer of the same Tikhonov functional such that in theory the same parameter should be chosen for all ϱ. The choice of the parameter might be affected by the fixed number of iterations, which explains the differences in the chosen parameters. The results for the RMSE are surprising. From the proved convergence rate we would expect that the approximation error for a fixed number of iterations is worse for smaller values of ϱ. From the given data, this is obviously not the case. Looking at the subsequent decimal places, we obtain the minimal approximation error for $\varrho = 0.021\,544$, but the RMSE values are nearly all the same for $\varrho \leq 0.1$. Concerning the computation time, we observe that lower values of ϱ result in a lower computation time, as one would expect due to the optimization of the search strategy.

For the dictionary $\mathcal{D}_1^\lambda$, we can conclude: on average, we can save over 40% of computation time by applying the RWFMP with the parameter $\varrho = 0.001$ instead of the RFMP. We even obtain a smaller RMSE value if we do so. We have to admit that this example is special, since we are using an orthogonal basis as dictionary and additionally, the sine functions are arising in the singular system of the considered operator. Furthermore, we try to approximate a constant function by sine functions, which might lead to additional errors induced by Gibb's phenomenon. We conjecture that the results are affected by these facts. Therefore, we will also discuss the results that we achieved by using the dictionary $\mathcal{D}_2^\lambda$, which is not specifically connected to the inverse problem at hand.

The results for the chosen regularization parameter when using $\mathcal{D}_2^\lambda$ are similar to the results presented above. In this case, even all of the average chosen parameters are nearly the same, as we would expect. Additionally, also the RMSE values are nearly the same for all values of ϱ. In this case, the optimal approximation error is achieved for $\varrho = 0.046\,416$, but all of the errors are not larger than 2.7% of this minimal error. In comparison to the numbers presented for $\mathcal{D}_1^\lambda$, the trend in the computation times is very different. For large values of ϱ the times are very similar to the first example. Starting with $\varrho = 0.1$, the weak approach shows its potential for improving the performance of the algorithm. Although the approximation error does not change dramatically, the computation times drops from 225 s for $\varrho = 1$ down to 15.3 s for $\varrho = 0.001$.

We can, thus, conclude for the dictionary $\mathcal{D}_2^\lambda$ of hat functions that the RWFMP has the potential to outperform the RFMP drastically if the presented search strategy for the next dictionary element is used. For a fixed number of iterations, saving over 90% of computation time is possible leading to nearly no change in the approximation error.

As a general conclusion for both of the given dictionaries, we can say that the implementation of a weak greedy algorithm for inverse problems may give a large improvement in the efficiency of the algorithm. As already stated above, one can save up to 90% of computation time without losing the accuracy of the approximation. For problems from the geoscientific applications, where one may have more than 10 000 data points and a dictionary consisting of thousands of functions, this speed-up is very promising. The improvement of computation time makes it possible to even put more different kinds of functions into the dictionary, which might by better adapted to the solution. This may lead to an improved approximation quality, while one can obtain the same overall computation time as one has for the RFMP with a smaller dictionary.

Chapter 10.

RFMP for nonlinear inverse problems (RFMP_NL)

The RFMP as presented in Chapter 8 is an algorithm for solving linear inverse and ill-posed problems. In this chapter, we will develop a similar algorithm for nonlinear inverse problems, such as the nonlinear inverse gravimetric problem (see Problem 7.16). In general, the problem to be solved is consequently Problem 6.23. Similar to the RFMP, we will consider the case $\mathcal{Y} = \mathbb{R}^\ell$ for $\ell \in \mathbb{N}$. As already mentioned in Chapter 8, this would allow for the joint inversion of different types of data.

10.1. Derivation of the algorithm

Gauß-Newton methods are popular methods for the solution of Problem 6.23. Examples include the Levenberg-Marquardt method and the iteratively regularized Gauß-Newton method (cf. Kaltenbacher et al. [95, Chapter 4]). The basic idea of these methods is the iterative minimization of the linearized Tikhonov functional

$$\left\| g - \mathcal{S}[f_k] - \mathcal{S}'[f_k](f_{k+1} - f_k) \right\|_{\mathbb{R}^\ell}^2 + \lambda_{k+1} \| f_{k+1} - f_k^\circ \|_{\mathcal{X}}^2, \tag{10.1}$$

for $f_{k+1} \in \mathcal{X}$, given $g \in \mathbb{R}^\ell$ and $f_k, f_k^\circ \in \mathcal{X}$. For the linearization of operators, see Theorem 2.12.

Note that by choosing $f_k^\circ = 0$ and bearing in mind that $\mathcal{T}'[f](h) = \mathcal{T}(h)$ for a linear operator $\mathcal{T}$ one obtains the same Tikhonov functional that has been used in the derivation of the FMP and the RFMP in Chapter 8. For nonlinear problems, incorporating the term f_k° takes into account that the zero element in $\mathcal{X}$ plays no special role, in contrast to linear inverse problems, where $\mathcal{T}(0) = 0$ if $\mathcal{T}$ is linear.

Gauß-Newton methods solve the minimization of Eq. (10.1) by finding a solution of the corresponding normal equation, since Eq. (10.1) can be interpreted as a least

squares problem. By applying Theorem 6.10 this leads to iteration formulas of the form

$$\big(\mathcal{S}'[f_k]^*\mathcal{S}'[f_k] + \lambda\mathcal{I}\big)(f_{k+1} - f_k) = \mathcal{S}'[f_k]^*(g - \mathcal{S}[f_k]) + \lambda_{k+1}(f_k^\circ - f_k).$$

In this setting, the Levenberg-Marquardt method and the iteratively regularized Gauß-Newton method correspond to $f_k^\circ = f_k$ and $f_k^\circ = f_0$, respectively.

In the following, we adapt the idea of iteratively minimizing Eq. (10.1) to obtain a greedy algorithm for nonlinear inverse problems. In analogy to Chapter 8, we choose a dictionary $\mathcal{D} \subseteq \mathcal{X}$, a fixed regularization parameter $\lambda > 0$, and an initial approximation $f_0 \in \mathcal{X}$. We then iteratively define a sequence $(f_k)_{k\in\mathbb{N}_0}$ of approximations to f^* by

$$f_{k+1} \coloneqq f_k + \alpha_{k+1}d_{k+1}.$$

It is the aim of the following considerations to determine how $\alpha_{k+1} \in \mathbb{R}$ and $d_{k+1} \in \mathcal{D}$ have to be chosen to minimize the linearized Tikhonov functional

$$\widetilde{\mathcal{A}}_\lambda\,[g, f_k, f_k^\circ, d, \alpha] = \big\|g - \mathcal{S}[f_k] - \alpha\mathcal{S}'[f_k](d)\big\|_{\mathbb{R}^\ell}^2 + \lambda\|(f_k - f_k^\circ) + \alpha d\|_{\mathcal{X}}^2,$$

for given $g \in \mathcal{Y}$ and $f_k, f_k^\circ \in \mathcal{X}$. Using the technique from Fischer [45], Fischer and Michel [46], and Michel [123], we first observe that

$$\begin{aligned}
\widetilde{\mathcal{A}}_\lambda\,[g, f_k, f_k^\circ, d, \alpha] &= \|r_k\|_{\mathcal{Y}}^2 - 2\alpha\big\langle r_k, \mathcal{S}'[f_k](d)\big\rangle_{\mathcal{Y}} + \alpha^2\big\|\mathcal{S}'[f_k](d)\big\|_{\mathcal{Y}}^2 \\
&\quad + \lambda\Big(\|f_k - f_k^\circ\|_{\mathcal{X}}^2 + 2\alpha\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}} + \alpha^2\|d\|_{\mathcal{X}}^2\Big) \\
&= \Big(\|r_k\|_{\mathcal{Y}}^2 + \lambda\|f_k - f_k^\circ\|_{\mathcal{X}}^2\Big) \\
&\quad - 2\alpha\big(\big\langle r_k, \mathcal{S}'[f_k](d)\big\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\big) \\
&\quad + \alpha^2\Big(\big\|\mathcal{S}'[f_k](d)\big\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2\Big),
\end{aligned} \tag{10.2}$$

where $r_k \coloneqq g - \mathcal{S}[f_k]$. For fixed $d \in \mathcal{D}$, a necessary condition for the minimization of $\widetilde{\mathcal{A}}_\lambda$ is

$$\begin{aligned}
0 &= \frac{\partial}{\partial\alpha}\widetilde{\mathcal{A}}_\lambda\,[g, f_k, f_k^\circ, d, \alpha] \\
&= -2\big(\big\langle r_k, \mathcal{S}'[f_k](d)\big\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\big) \\
&\quad + 2\alpha\Big(\big\|\mathcal{S}'[f_k](d)\big\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2\Big),
\end{aligned}$$

which is for the minimizer $\alpha = \alpha_{k+1}$ equivalent to

$$\alpha_{k+1} = \frac{\big\langle r_k, \mathcal{S}'[f_k](d)\big\rangle_{\mathcal{Y}} - \lambda\big\langle f_k - f_k^\circ, d\big\rangle_{\mathcal{X}}}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2}.$$

Inserting the latter into Eq. (10.2) yields

$$\begin{aligned}\widetilde{\mathcal{A}}_\lambda\left[g, f_k, f_k^\circ, d, \alpha_{k+1}\right] &= \left(\|r_k\|_{\mathcal{Y}}^2 + \lambda\|f_k - f_k^\circ\|_{\mathcal{X}}^2\right) \\ &\quad - 2\frac{\left(\langle r_k, \mathcal{S}'[f_k](d)\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2} \\ &\quad + \frac{\left(\langle r_k, \mathcal{S}'[f_k](d)\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2} \\ &= \left(\|r_k\|_{\mathcal{Y}}^2 + \lambda\|f_k - f_k^\circ\|_{\mathcal{X}}^2\right) \\ &\quad - \frac{\left(\langle r_k, \mathcal{S}'[f_k](d)\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2}.\end{aligned}$$

Thus, the pair $(\alpha_{k+1}, d_{k+1}) \in \mathbb{R} \times \mathcal{D}$ is a minimizer of $\widetilde{\mathcal{A}}_\lambda\left[g, f_k, f_k^\circ, \cdot, \cdot\right]$ if and only if

$$\begin{aligned}d_{k+1} &= \operatorname*{argmax}_{d\in\mathcal{D}} \frac{\left(\langle r_k, \mathcal{S}'[f_k](d)\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2}, \\ \alpha_{k+1} &= \frac{\langle r_k, \mathcal{S}'[f_k](d_{k+1})\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d_{k+1}\rangle_{\mathcal{X}}}{\|\mathcal{S}'[f_k](d_{k+1})\|_{\mathcal{Y}}^2 + \lambda\|d_{k+1}\|_{\mathcal{X}}^2},\end{aligned}$$

which will be the key ingredient in the following algorithm.

Algorithm 10.1 (RFMP for Nonlinear Problems, RFMP_NL)**.** Let $\mathcal{S}$ and g be given as in Problem 6.23. Choose a dictionary $\mathcal{D} \subseteq \mathcal{X} \setminus \{0\}$, an initial approximation $f_0 \in \mathcal{X}$, and a regularization parameter $\lambda > 0$. Furthermore specify the type of regularization by choosing the sequence $f_k^\circ \in \mathcal{X}$, for example, as one of the options stated above.

1. Set $k := 0$, define the residual $r_0 := g - \mathcal{S}[f_0]$ and choose a stopping criterion.
2. Find
$$d_{k+1} = \operatorname*{argmax}_{d\in\mathcal{D}} \frac{\left(\langle r_k, \mathcal{S}'[f_k](d)\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d\rangle_{\mathcal{X}}\right)^2}{\|\mathcal{S}'[f_k](d)\|_{\mathcal{Y}}^2 + \lambda\|d\|_{\mathcal{X}}^2}$$
and set
$$\alpha_{k+1} := \frac{\langle r_k, \mathcal{S}'[f_k](d_{k+1})\rangle_{\mathcal{Y}} - \lambda\langle f_k - f_k^\circ, d_{k+1}\rangle_{\mathcal{X}}}{\|\mathcal{S}'[f_k](d_{k+1})\|_{\mathcal{Y}}^2 + \lambda\|d_{k+1}\|_{\mathcal{X}}^2},$$
as well as $f_{k+1} := f_k + \alpha_{k+1}d_{k+1}$ and $r_{k+1} := g - \mathcal{S}[f_{k+1}]$.
3. If the stopping criterion is fulfilled, then f_{k+1} is the output. Otherwise, increase k by 1 and return to step 2.

First, we want to discuss the differences in the implementation of the RFMP and the RFMP_NL. In Chapter 8, we highlighted that one can make extensive use of a preprocessing procedure when implementing the RFMP. This was due to the fact that all terms in the maximization step are a combination of several inner products, which do not change throughout the iteration. Unfortunately, this is no longer the case for the RFMP_NL. In particular, the Gâteaux derivative $\mathcal{S}'[f_k](d)$ depends on the current approximation f_k and, consequently, has to be computed anew in every iteration. Since, in general, the Gâteaux derivative does not depend linearly on the approximation, it is also not possible to update this term in any way. The computation of the Gâteaux derivative can be very expensive from the computational point of view. For example, when applying the algorithm to the nonlinear inverse gravimetric problem (see Problem 7.16), where the Gâteaux derivative is given by Eq. (7.19), a numerical integration formula on the sphere has to be used in every iteration. Furthermore, also the residual r_k can no longer be obtained by simply updating the residual from the previous iteration with a precomputed value, but also has to be computed again in every iteration. This makes the algorithm much more expensive. However, this is also true for most of the other established algorithms for nonlinear inverse problems. Since most of them are also based on a linearization of the operator, there is a need of a new computation of the Gâteaux or Fréchet derivative. Furthermore, several of these methods use the adjoint operator of the Fréchet derivative, which then also changes in every iteration. In the RFMP_NL, there is no need to know and compute the adjoint operator.

Secondly, there are also several similarities in the implementation of the RFMP and the RFMP_NL, as one would expect. Both algorithms require a dictionary, which should ideally be suitable to approximate the solution of the given inverse problem. Furthermore, both algorithms incorporate a Tikhonov regularization and there is a need to choose the regularization parameter in a proper way. Finally, the choice of a stopping criterion is a crucial ingredient of the algorithm.

10.2. Comparison to other methods

In general, methods for nonlinear inverse problems are iterative (for an overview, see the book by Kaltenbacher et al. [95]). This includes the already mentioned Gauß-Newton methods, where a linearized Tikhonov functional (therefore *Newton*) is minimized by solving a normal equation (therefore *Gauß*). The linearized Tikhonov functional itself is obtained by applying a Tikhonov regularization to the linearized equation that would be solved by a pure Newton method. Representatives of this category of methods are the Levenberg-Marquardt method (see Levenberg [109]

and Marquardt [115]) and the iteratively regularized Gauß-Newton method (see Bakushinsky [10]). Another classical category of methods are gradient type methods, in particular, the Landweber method (see, for linear problems, Landweber [106] and Engl et al. [42, Section 6.1]; for nonlinear problems, Hanke et al. [75]). Additionally, (direct) Tikhonov regularization methods (see Tikhonov and Glasko [171]), multilevel methods (see Kaltenbacher et al. [95, Chapter 5]), and sequential subspace optimization methods (see Wald and Schuster [181]) have been developed for nonlinear inverse problems. Especially, we want to mention level set methods (see the survey by Burger and Osher [24]), since these are methods that are often used for problems like the nonlinear inverse gravimetric problem, where a domain is the unknown. Although the latter possess several advantageous properties, for example, one is able to recover domains that are not star-shaped or even unconnected, one does not directly get an explicit representation of the surface of the unknown domain (although one can try to obtain such a representation in a post-processing step). In contrast, we directly obtain such a representation in spherical coordinates, which is desirable from the geophysical perspective. Furthermore, the assumptions that the Earth's interior (and also the part of the Earth that is inside a boundary layer like the Mohorovičić discontinuity) is connected or even star-shaped is a good model of the reality.

In the following, we want to compare the newly developed RFMP_NL with existing methods for nonlinear problems from the algorithmic point of view. First, there are several similarities to other methods. The method is iterative, like most of the established methods. Even if these are not iterative by themselves, like the application of a Tikhonov regularization, often iterative optimization algorithms have to be used inside these algorithms. The derivation of the algorithm was actually motivated by Gauß-Newton methods, which shows the clear similarity to these methods.

Secondly, there are also several differences to other methods, both advantages and disadvantages. A disadvantage of the RFMP_NL is the fact that we currently do not have an accurate theoretical analysis of the method, that is, no convergence or regularization result. These results have already been very elaborate to obtain in the case of linear inverse problems (see Chapter 9), and we expect it to be much more difficult in the nonlinear case. Moreover, as in the linear case, since the dictionary does not need to form a basis of the underlying Hilbert space, the solution provided by the RFMP_NL might not be unique. It may be represented by a different linear combination of dictionary elements (in a certain sense) in a better way. However, the RFMP_NL also has several advantages in comparison to the existing methods. Most of these are given in a pure, infinite-dimensional, Hilbert (or Banach) space formulation. In consequence, it is not possible to implement these methods directly

on a computer. Of course, at the first sight, this is also the case for the RFMP_NL. The difference is that in most of the other methods one has to choose a specific basis system in order to implement the method. This is necessary to ensure the regularity of the arising systems of linear equations, for example, in the Levenberg-Marquardt method. Using ansatz functions for the solution that are linearly dependent may lead to singular matrices. In contrast, the RFMP_NL can handle very diverse types of basis functions (for example, global and localized functions) and will choose those functions that are best adapted to the structure of the solution. This is especially important in the applications, for example, in geophysics, where global structures like the Earth's ellipticity must be distinguished from local structures like mountains. Furthermore, there even is no need at all to solve linear systems of equations, like in most of the other methods. Thus, one does not need to care about the condition of arising matrices, which may require a stabilization, and there is also no need to apply iterative solvers, which is often the most efficient way when implementing the other methods. Another advantage of the RFMP_NL is that there is no need to know the adjoint operator of the Fréchet derivative. The operator itself even does not need to be Fréchet differentiable. Gâteaux differentiability is enough in this context.

10.3. Application to the nonlinear inverse gravimetric problem

In this section, we will apply the RFMP_NL to the nonlinear inverse gravimetric problem from Section 7.2.2.

For this purpose, we will first discuss several topics concerning the implementation of the method and the chosen dictionaries. Then, we will present numerical results for several synthetic scenarios with different dictionaries, different prescribed solutions, and different levels of noise.

10.3.1. Implementation details of the algorithm

The algorithm was implemented in C (Kerningham and Ritchie [98]) using the GNU Scientific Library (Galassi et al. [62]), and a parallelization with OpenMP (Dagum and Menon [30] and OpenMP Architecture Review Board [139]).

Besides the implementation of the algorithm itself, which we will discuss later in this section, we need an implementation of the nonlinear operator that is associated to the nonlinear inverse gravimetric problem, as well as its derivative.

Remember that the operator $\mathcal{S}_\varrho$ of the nonlinear inverse gravimetric problem is given as

$$\mathcal{S}_\varrho[\sigma](y) = \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} \frac{\varrho(r\,\xi)}{|r\,\xi - y|} r^2 \,\mathrm{d}r\,\mathrm{d}\omega(\xi), \qquad y \in S$$

and its Fréchet and Gâteaux derivatives possess the form

$$\mathcal{S}'_\varrho[\sigma](\tau)(y) = \int_{\mathbb{S}^2} \frac{\varrho(\sigma(\xi)\,\xi)}{|\sigma(\xi)\,\xi - y|} (\sigma(\xi))^2\,\tau(\xi)\,\mathrm{d}\omega(\xi), \qquad y \in S.$$

In all of the numerical examples, we restrict ourselves to a constant density $\varrho = 1$. For the sake of readability, as in Section 7.2.3, we define the function $k\colon S^{\mathrm{int}} \times S \to \mathbb{R}$,

$$k(x,y) := \frac{1}{|x-y|}|x|^2$$

such that

$$\mathcal{S}_1[\sigma](y) = \int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y)\,\mathrm{d}r\,\mathrm{d}\omega(\xi), \tag{10.3}$$

$$\mathcal{S}'_1[\sigma](\tau)(y) = \int_{\mathbb{S}^2} k(\sigma(\xi)\xi, y)\,\tau(\xi)\,\mathrm{d}\omega(\xi). \tag{10.4}$$

We first deal with a numerical integration method on $\mathbb{S}^2$, since it is needed in both Eq. (10.3) and Eq. (10.4). The method, which we will use, was developed by Driscoll and Healy [39] and is based on the following point grid.

Definition 10.2 (cf. Michel [122, Theorem 7.33])**.** Let $m \in \mathbb{N}$. Let the points $\eta_{p,q} = \eta(\varphi_q, t_p) \in \mathbb{S}^2$ be defined by the polar coordinates

$$\varphi_q = \frac{2\pi q}{m+1}, \qquad q = 0, \ldots, m,$$
$$t_p = \cos\frac{\pi p}{m+1}, \qquad p = 0, \ldots, m.$$

Then $\{\, \eta_{p,q} \mid p,q = 0,\ldots,m \,\}$ is called the *Driscoll-Healy grid*.

For spherical harmonics up to degree m, there exists an exact integration formula using the Driscoll-Healy grid.

Theorem 10.3 (cf. Michel [122, Theorem 7.33])**.** Let $f \in \mathrm{Harm}_{0\ldots m}(\mathbb{S}^2)$ for an odd number $m \in \mathbb{N}$ and let $\{\, \eta_{p,q} \mid p,q = 0,\ldots,m \,\}$ be the Driscoll-Healy grid. Let the weights $a_0, \ldots, a_m \in \mathbb{R}$ be defined as

$$a_p := \frac{4}{m+1} \sin\left(\frac{\pi p}{m+1}\right) \sum_{s=0}^{(m+1)/2-1} \frac{1}{2s+1} \sin\left((2s+1)\frac{\pi p}{m+1}\right)$$

for $p = 0, \ldots, m$.

Then,

$$\int_{\mathbb{S}^2} f(\eta) \, \mathrm{d}\omega(\eta) = \frac{2\pi}{m+1} \sum_{p=0}^{m} a_p \sum_{q=0}^{m} F(\eta_{p,q}).$$

For all of the arising integrals over the sphere, we use exactly this method with the parameter $m = 99$. It remains to choose an integration method for the radial integral in Eq. (10.3). We apply a simple partitioned trapezoidal rule with 100 points.

An alternative for the integral in Eq. (10.3) is to transform the integral to the ball $\mathbb{B}_1$ such that

$$\begin{aligned}\int_{\mathbb{S}^2} \int_0^{\sigma(\xi)} k(r\xi, y) \, \mathrm{d}r \, \mathrm{d}\omega(\xi) &= \int_{\mathbb{S}^2} \int_0^1 k(t\sigma(\xi)\xi, y)\sigma(\xi) \, \mathrm{d}t \, \mathrm{d}\omega(\xi) \\ &= \int_{\mathbb{B}_1} k\left(|x|\sigma\left(\frac{x}{|x|}\right)\frac{x}{|x|}, y\right)\sigma\left(\frac{x}{|x|}\right)\frac{1}{|x|^2} \, \mathrm{d}x\end{aligned}$$

and apply an integration method for the ball, as for example derived by Amna and Michel [6], but for the simplicity of the implementation, we do not pursue this approach.

The general procedure for the presented numerical experiments will be as follows.

1. Prescribe a solution $\sigma_{\mathrm{sol}} \colon \mathbb{S}^2 \to (0, \infty)$ and use numerical integration to compute synthetic data $g_j := \mathcal{S}_1[\sigma_{\mathrm{sol}}](y_j)$, $j = 1, \ldots, J$, where $(y_j)_{j=1,\ldots,J} \subseteq S$ are points on a sphere with radius R, $S := \mathbb{S}^2_R := \{ x \in \mathbb{R}^3 \mid |x| = R \}$. Ensure that $R > \sup_{\xi \in \mathbb{S}^2} \sigma(\xi)$.
2. Apply 1% of noise.
3. Choose a function space $\mathcal{X}(\mathbb{S}^2) \subseteq \mathrm{L}^2(\mathbb{S}^2)$ and a dictionary $\mathcal{D} \subseteq \mathcal{X}(\mathbb{S}^2) \setminus \{ 0 \}$.
4. Define a set $(\lambda_s)_{s=1,\ldots,S}$ of regularization parameters and define the sequence $(f_k^\circ)_k$ of functions, which determines the type of regularization.
5. For each regularization parameter, run the RFMP_NL using the noisy data and the predefined dictionary.
6. Choose the regularization parameter, which yields the lowest approximation error.

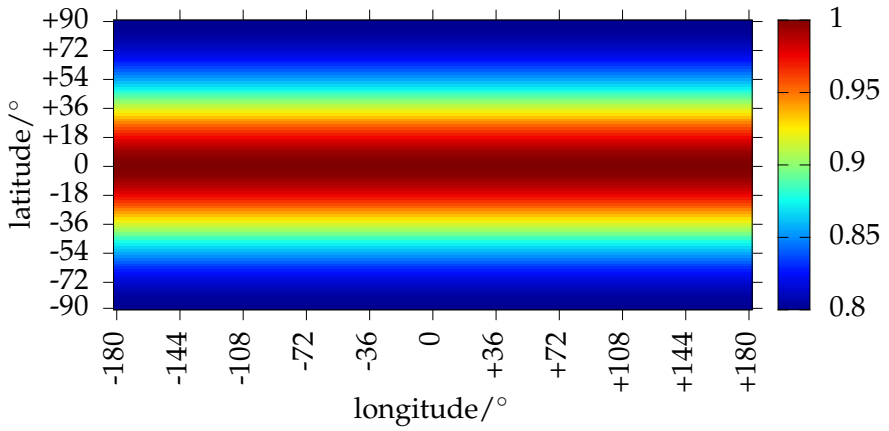


Figure 10.1.: Prescribed solution in Example 1: describes an ellipsoid of revolution with aspect ratio 0.8

In general, we will look at two scenarios. The first one will use a contrived solution with purely global structures and a dictionary of spherical harmonics for a specific combination of the regularization term and the function space. The second one will combine both global and local features in the prescribed solution and correspondingly, the dictionary will consist of spherical harmonics and Abel-Poisson kernels. For the second scenario, we will compare the different alternatives for the regularization term and the function space. For the combination that yields the lowest approximation error we will present concrete results for two different prescribed solutions.

10.3.2. Example 1: ellipsoid of revolution

The prescribed solution is given by

$$f(\xi) := \sigma_{\text{sol}}(\xi) := \frac{1}{\sqrt{\xi_1^2 + \xi_2^2 + \left(\frac{\xi_3}{0.8}\right)^2}}$$

such that $\Sigma^{\text{int}} = \left\{ r\xi \in \mathbb{R}^3 \;\middle|\; \xi \in \mathbb{S}^2, r < \sigma(\xi) \right\}$ is an ellipsoid of revolution with the aspect ratio 0.8. The solution is depicted in Figure 10.1 in spherical coordinates. We measured the gravitational potential in 10 000 points on a Driscoll-Healy grid on a sphere of radius 1.1.

Since the solution describes a global structure, we used a dictionary that consists only of global functions, namely, spherical harmonics. We defined the dic-

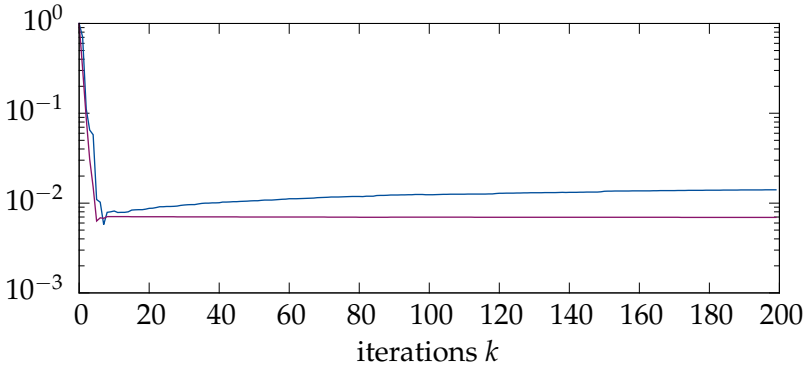
tionary

$$\mathcal{D} := \left\{ Y_{n,j} \;\middle|\; n = 0, \dots, 25, \quad j = -n, \dots, n \right\}$$

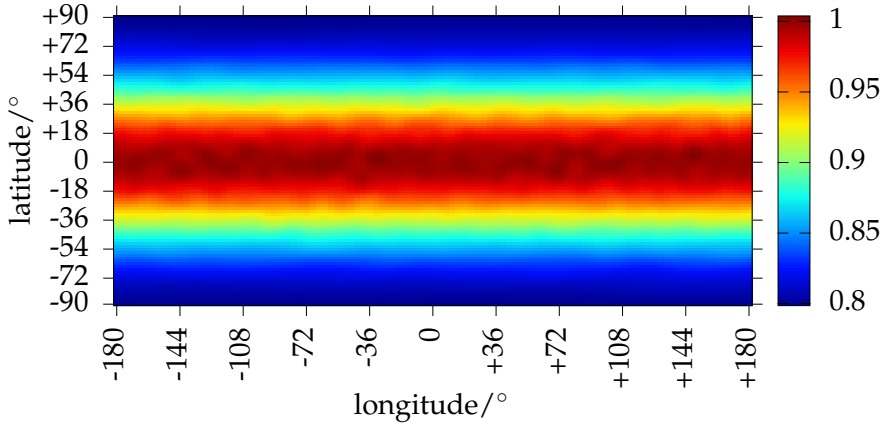
of spherical harmonics up to degree 25 such that $\#\mathcal{D} = 676$. We chose the function space $\mathcal{X}(\mathbb{S}^2) = \mathrm{L}^2(\mathbb{S}^2)$, and set $f_k^\circ := f_0$ for all $k \in \mathbb{N}_0$ such that the regularization term corresponds to the term that is also used in the iteratively regularized Gauß-Newton method. As the initial approximation, we chose $f_0 \equiv 0.8$, corresponding to the sphere with radius 0.8. The best regularization parameter that we found using the strategy presented above, is $\lambda = 1.585 \times 10^{-2}$. We performed 200 iterations of the RFMP_NL. The development of the relative residual, that is, $\|g - \mathcal{S}[f_k]\|_{\mathcal{Y}} / \|g\|_{\mathcal{Y}}$, and the relative approximation error, that is, $\|f - f_k\|_{\mathcal{X}} / \|f\|_{\mathcal{X}}$, during the iteration can be found in Figure 10.2a. The approximation after 200 iterations is depicted in Figure 10.2b. Moreover, the pointwise difference of the approximation after 200 iterations and the solution is shown in Figure 10.2c.

In Figure 10.2a, we first consider the residual. We observe that the relative residual drops rapidly in the first few iterations below 1%. Since we used a noise level of 1% in the data space, this is what we would expect. This also shows that the algorithm works as it should, since (ignoring the regularization) it was derived as a minimization algorithm for the (linearized) residual. Considering the relative approximation error, we observe that the final value after 200 iterations is approximately 1.4%. Due to the ill-posedness and the nonlinearity of the inverse problem, it is not surprising that the approximation error is larger than the residual and the noise level. Indeed, a factor of 1.4 between the approximation error and the noise level is a good result for an algorithm for ill-posed inverse problems. Unfortunately, we also observe that the error is even lower in several of the earlier iterations. It would, consequently, be even more efficient to stop the algorithm, when the error is minimal. Since, in general, we do not know the solution, this cannot be achieved. Looking at the results that the RFMP_NL yields for all of the other regularization parameters, we can say that for these parameters the difference between the minimal error and the error after 200 iterations is even larger such that, currently, the results presented in Figure 10.2 are the best we could achieve. Of course, one could also think about stopping the iteration earlier, for example, using a discrepancy principle, which is a common procedure for iterative regularization methods. From methods like the Levenberg-Marquardt algorithm, it is known that one would have to choose a different regularization parameter in every iteration, which is even more difficult. This is why we postpone this subject to our research in the future.

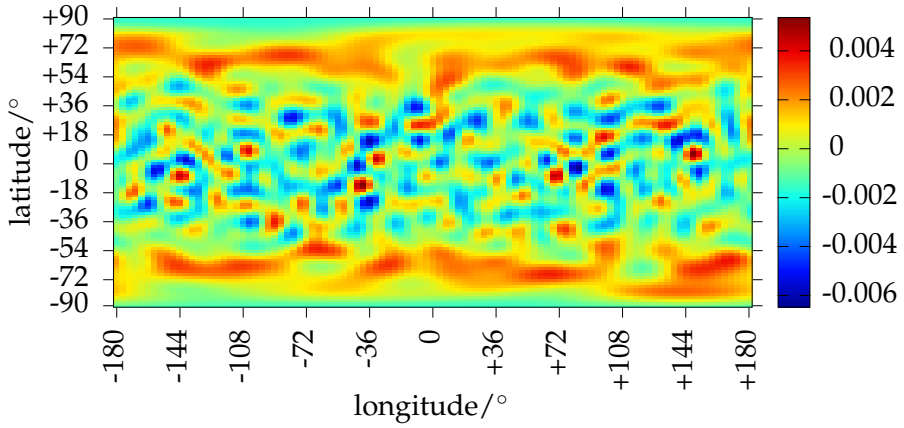
The approximation itself and the approximation error as a function in spherical coordinates are shown in Figure 10.2b and Figure 10.2c, respectively. In addition to the plot of the approximation error in spherical coordinates in Figure 10.2c, we



(a) Development of the relative residual (purple) and the relative approximation error (blue) during the iteration of the RFMP_NL.



(b) Approximation generated by the RFMP_NL.



(c) Difference of the solution and the approximation generated by the RFMP_NL.

Figure 10.2.: Results from the application of the RFMP_NL in Example 1.

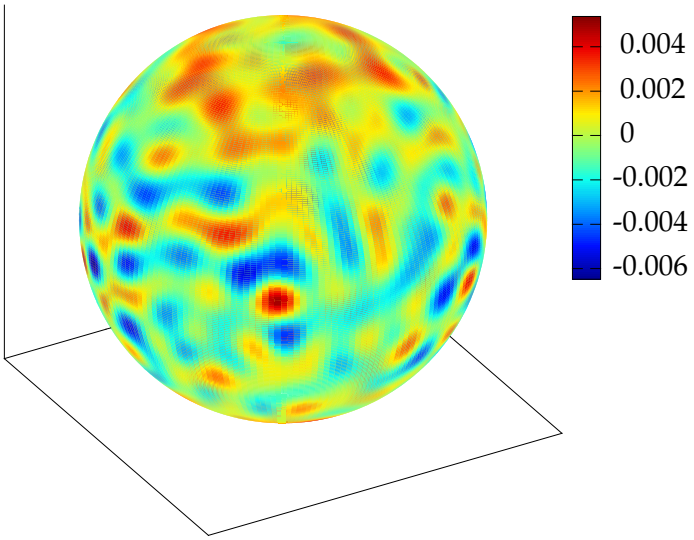


Figure 10.3.: Three-dimensional plot of the approximation error after 200 iterations of the RFMP_NL in Example 1 as a function on the sphere.

also provide a three-dimensional plot on the sphere in Figure 10.3 to account for the misperception of structures around the poles that might arise due to the cartographic projection. It can be seen that the error is very small everywhere on the sphere. Its absolute value is nowhere larger than 0.01, where the surface of the body of mass is between 0.8 and 1 units away from the center. We find that there are several small artifacts in the approximation error distributed over the whole sphere. From the application of the RFMP (and other regularization methods) to linear inverse problems, we know that such artifacts normally arise if the regularization parameter is chosen too low. As already said, we observed larger approximation errors for larger values of the regularization parameter such that we do not believe that the artifacts arise from under-regularization.

Nevertheless, we can say that, for this example, the approximations generated by the RFMP_NL after 200 iterations are very good, since the relative approximation error is only 1.4% for a noise level of 1%.

10.3.3. Example 2: ellipsoid and added Abel-Poisson kernels

In the derivation of every greedy algorithm that arises in this work, we have stressed that these algorithms are able to combine different types of basis functions.

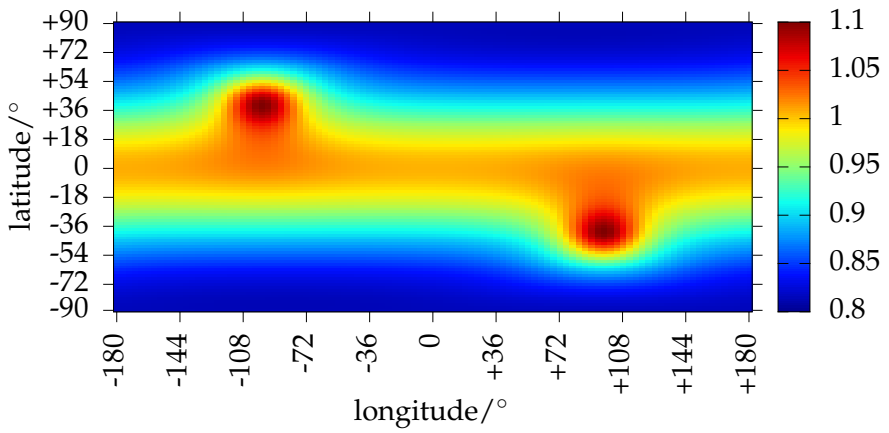


Figure 10.4.: Prescribed solution in Example 2: the sum of an ellipsoid of revolution and two Abel-Poisson kernels.

Therefore, in this section we will discuss an example, which will show that this is also true in practice for the RFMP_NL.

We will first compare different possibilities for the chosen function spaces and the regularization term.

For this purpose, we chose a dictionary $\mathcal{D}$ that consists of spherical harmonics up to degree 9 and Abel-Poisson kernels with parameter $h = 0.7$, which are centered on a Driscoll-Healy grid with parameter $m = 25$, yielding 100 spherical harmonics and 652 Abel-Poisson kernels to obtain $\#\mathcal{D} = 752$ in total. The prescribed solution consisted of a sum of the function used in the first example corresponding to an ellipsoid of revolution and two Abel-Poisson kernels with parameter $h = 0.7$, which are centered at 41°N, 96°W and 41°S, 96°E. The solution is depicted in Figure 10.4.

In this case, we measured the gravitational potential at 10 000 points on a Driscoll-Healy grid on the sphere with radius 1.2.

In Table 10.1, we gather the approximation errors for two different types of regularization terms and three different function spaces $\mathcal{X}(\mathbb{S}^2)$. For every combination, we executed 200 iterations of the RFMP_NL with several regularization parameters. The values in the table are the lowest approximation errors that we obtained among all of the regularization parameters. As already mentioned before, choosing $f_k^\circ = f_0$ in the RFMP_NL is analogous to the iteratively regularized Gauß-Newton method, and $f_k^\circ = f_k$ is analogous to the Levenberg-Marquardt method. This is the reason why we compare these two choices for the regularization term. Apart from the space

Table 10.1.: Approximation errors for different types of regularization and different function spaces in Example 2. The minimum approximation error is set in a bold font.

Function space	$f_k^\circ = f_0$	$f_k^\circ = f_k$
$L^2(\mathbb{S}^2)$	**3.10%**	3.20%
$H^1(\mathbb{S}^2)$	3.26%	3.23%
$H^2(\mathbb{S}^2)$	3.30%	3.22%

$L^2(\mathbb{S}^2)$, we also considered the inverse problem in the Sobolev spaces $H^1(\mathbb{S}^2)$ and $H^2(\mathbb{S}^2)$ (for a definition see, e. g., Freeden et al. [53, Section 5.1]).

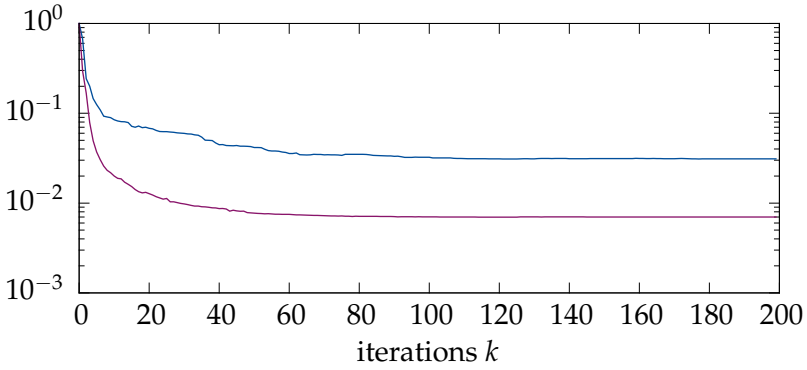
We observe that the combination of $f_k^\circ = f_0$ and the space $L^2(\mathbb{S}^2)$ yields the best results, although all of the other results have the same order of magnitude. We, thus, stick to this combination in the further analysis of the results. This also has the advantage that $L^2(\mathbb{S}^2)$-norms of the functions in the dictionary can be easily computed (see Example 2.37 for the Abel-Poisson kernel), which is not the case for the Sobolev spaces. There, one has to compute a truncated Legendre series, which is much more expensive from the computational point of view.

In Figure 10.5, we present the results after 200 iterations of the RFMP_NL for this combination for the optimal regularization parameter.

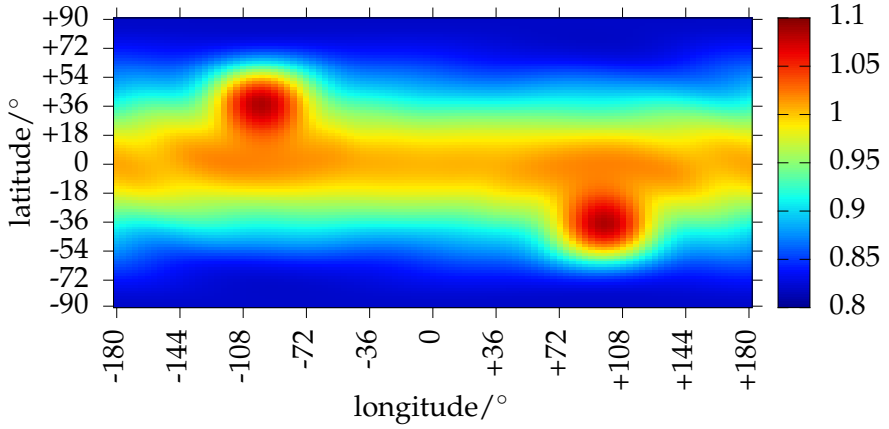
Considering the approximation error and the residual in Figure 10.5a, we observe again that the relative residual drops rapidly in the first few iterations to a value of 0.7%, which is below the noise level. The relative approximation error attains a value of 3.10% after 200 iterations. This is again a very good result for the applied noise level of 1%.

The approximation in Figure 10.5b and the approximation error in Figure 10.5c are larger, where the local structures of the solution can be found. In particular, the maximal error is located in the centers of the Abel-Poisson kernels that are present in the prescribed solution. Interestingly, in contrast to the first example, we do not see such a big amount of artifacts in the error. A look at the absolute values of the approximation error again shows that the obtained results are very good, since it never exceeds 0.02, whereas the surface of the body of mass is between 0.8 and 1.1 units away from the center.

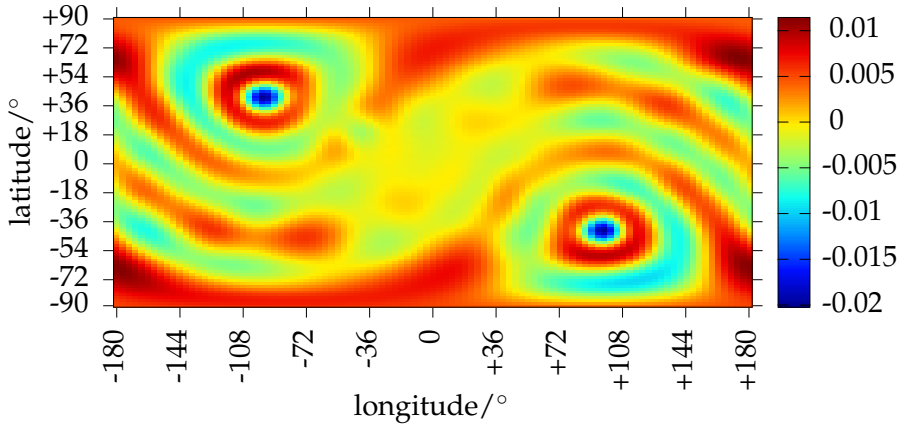
If one considers the chosen dictionary elements, we observe that the algorithm chose spherical harmonics in 187 of the iterations and Abel-Poisson kernels in 13 iterations. Since the algorithm can choose functions from the dictionary multiple times, the solution consisted of 89 distinct spherical harmonics and 13 distinct Abel-Poisson kernels. We have displayed the centers of these Abel-Poisson kernels in Figure 10.6



(a) Development of the relative residual (purple) and the relative approximation error (blue) during the iteration of the RFMP_NL.



(b) Approximation generated by the RFMP_NL.



(c) Difference of the solution and the approximation generated by the RFMP_NL.

Figure 10.5.: Results from the application of the RFMP_NL in Example 2.

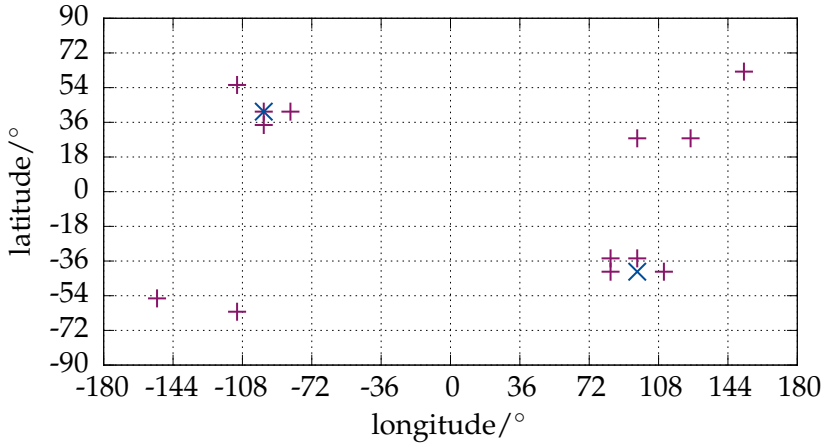


Figure 10.6.: Plot of the centers of the Abel-Poisson kernels that are present in the solution (blue) and the approximation (purple).

alongside the centers of the kernels that are present in the solution. We observe that 8 of the 13 chosen kernels have their centers near to the kernels of the solution. One of the two kernels in the solution is even chosen itself. Due to the use of noisy data, also some kernels are chosen that are centered where there is no kernel in the solution.

In conclusion, we can say that also in this example, the RFMP_NL produces very good approximations of the prescribed solution. In the derivation of the algorithm, we stated that it will be possible to combine different types of basis functions. The presented example shows that this is indeed true in practice and the results are very promising due to the low approximation errors.

Thus, the RFMP_NL seems to be a suitable algorithm for the solution of the nonlinear inverse gravimetric problem. Unfortunately, for practical applications, the computer power that is needed is still a bit too high. In the second example, the computation time was 9.37 h on a node of the HorUS cluster of the University of Siegen, where a parallelization of the code to the 12 kernels of the node was used whenever possible. To invert the gravitational potential, for example, for the topography of the Earth, one would need to use a much larger integration grid in the implementation of the operator due to the finer structures that arise in the solution. Therefore, the examples presented before can currently only serve as a proof of concept that the algorithm successfully solves the nonlinear inverse gravimetric problem. On the one hand, with rising computer power in the future, it is not unrealistic that the RFMP_NL will some day be applied in a more realistic scenario or even to real-world data. On the other hand, one could also think about other optimizations of the algorithm to obtain lower computation times. For example, the strategy that

we have applied to obtain the RWFMP in Chapter 9 could also be applied to the RFMP_NL. The most expensive part of the algorithm is the re-computation of the Fréchet derivatives in step 2. If it would be possible to compute these derivatives analytically instead by using numerical integration, this could also speed up the algorithm.

Part IV.

Final remarks

Chapter 11.

Conclusions and outlook

In this work we presented novel greedy algorithms for two classes of problems.

The first problem was the estimation of a probability density function on the sphere, which was considered in Part II. For this purpose, we derived a greedy algorithm based on the Pure Greedy Algorithm (PGA) from approximation theory. We applied the greedy algorithm to a data set that represents the directions of fibers inside a so-called nonwoven fabric, which is a technical textile that is often used in industrial applications. Since these data were generated by a CT scanner with a very high resolution, we had to deal with nearly 10 million data points. By the application of the greedy algorithm for density estimation, we were able to improve the computation time of a simple simulation algorithm for nonwoven fabrics by a factor of 750 in comparison to the standard non-parametric approach of a kernel density estimator.

Furthermore, we dealt with linear and nonlinear inverse problems in Part III. Here, we generalized the RFMP algorithm that has been developed for the solution of linear inverse problems by the Geomathematics Group of the University of Siegen in recent years in two manners.

First, we applied the idea of the Weak Greedy Algorithm (WGA) to the RFMP to obtain the Regularized Weak Functional Matching Pursuit (RWFMP). By this approach, we could prove convergence of the RWFMP for linear inverse problems between arbitrary (possibly) infinite-dimensional Hilbert spaces, whereas convergence was only proved for the RFMP if the image space is finite-dimensional. Up to now, convergence rates were only known for the unregularized FMP. We provided convergence rates both for the WFMP and the RWFMP, which therefore includes the rate of convergence of the RFMP. It has been stated in the outlooks of the PhD theses by Fischer [45] and Telschow [166] that such a convergence rate would be desirable, which is thus achieved in the work at hand. We were also able to provide an a-priori parameter choice rule for the RWFMP, which yields a convergent regularization. Furthermore, a numerical example in $L^2([0,1])$ showed that the approach

of the RWFMP can be used to reduce the computation time of the RFMP by up to 90%.

Secondly, we presented a greedy algorithm for nonlinear inverse problems, which we called RFMP_NL. We applied the algorithm to the nonlinear inverse gravimetric problem, where one inverts gravitational data for the surface of or boundary layers inside a body of mass like the Earth or any other planetary body. We tested the algorithm using two synthetic examples: on the one hand we proposed a purely global solution and on the other hand, we proposed a solution that is a combination of global and localized functions. For both of these examples, the RFMP_NL yielded very good results: for a noise level of 1%, the approximation error is 1.4% and 3.1%, respectively. Also pointwise, the approximation error attains very low values, which highlights the quality of the approximation. In the introduction, we have already mentioned that we decided to use greedy algorithms to obtain a sparse solution due to the possibility of choosing an overcomplete dictionary. In the second example, we could observe that the approximation that was generated by the RFMP_NL is comprised both of the global functions and the localized functions that we put into the dictionary. We also saw that the localized functions that the algorithm chose are mainly concentrated in those regions where local structures were present in the solution. This supports our view that greedy algorithms possess this property and that they are advantageous for that reason in comparison to other methods that yield sparse approximations.

Of course, there is also much more research to be done.

Concerning the problem of density estimation, we were not able to prove the convergence of the algorithm (or at least asymptotic unbiasedness of the resulting estimator or similar statistical properties). It is desirable that this is done in future research. Furthermore, the simulation algorithm for nonwoven fabrics should be checked for its quality. Unfortunately, this is not so easy to accomplish. We could, of course, compare the distribution of fiber directions in the real nonwoven and the simulated nonwoven. Due to the concept behind the simulation algorithm this is not sensible since the algorithm is designed such that these distributions coincide (at least approximately). For that reason, one needs to find a different method to determine the approximation quality. Moreover, a comparison to the (mainly two-dimensional) mathematical models that already exist would be interesting both from our perspective as developers of the model and the industrial application.

The idea of the RWFMP was to apply the idea of the WGA to the RFMP. Since there exists also a Weak Orthogonal Greedy Algorithm (WOGA) in approximation theory, it would be interesting to see if we can also transfer the idea of the WOGA

to the RFMP to yield a Regularized Weak Orthogonal Functional Matching Pursuit (RWOFMP). Since both the ROFMP and the RWFMP possess several advantages over the RFMP, a combination of both approaches could provide even better approximations of the solutions to linear inverse problems. One could also try to apply the idea of greedy algorithms for the approximation of functions in Banach spaces to the setting of inverse problems.

Finally, the good numerical results suggest that the algorithm converges. This convergence should also be proved theoretically. As already stated in Section 10.2, the convergence analysis has already been complicated for linear inverse problems, since greedy algorithms are nonlinear algorithms. We apprehend that, consequently, the analysis for nonlinear problems is even more complicated (and may be impossible). Of course, this subject should be studied further to provide a rigorous mathematical basis for the application of the algorithm. Furthermore, one could try to apply the strategy of the WGA also to the RFMP_NL. This might be a promising approach to reduce the computation time, which is currently the major issue why we are not able to apply the RFMP_NL to more realistic scenarios or even real data. Of course, the RFMP_NL could also be applied to other nonlinear inverse problems, not necessarily from geophysics and not necessarily on a spherical domain. Areas of application may include medical imaging and industrial applications.

For all of the greedy algorithms, one should deal with the problem of finding an optimal dictionary. As already mentioned, for the RFMP and the ROFMP for linear inverse problems this is the subject of the work-in-progress PhD thesis by Schneider [159]. Also for all of the other methods presented in this thesis, this subject is crucial for the efficiency of the algorithms as, for example, the proved rate of convergence of the RWFMP shows.

Bibliography

[1] R. G. Airapetyan, A. G. Ramm, and A. B. Smirnova. Continuous analog of the Gauß-Newton method. *Math. Models Methods Appl. Sci.* **9** (1999), 463–474.

[2] A. R. A. Aitken. Moho geometry gravity inversion experiment (MoGGIE): A refined model of the Australian Moho, and its tectonic and isostatic implications. *Earth Planet. Sci. Lett.* **297** (2010), 71–83.

[3] E. N. Akimova, D. V. Belousov, and V. E. Misilov. Algorithms for solving inverse geophysical problems on parallel computing systems. *Numer. Anal. Appl.* **6** (2013), 98–110.

[4] E. N. Akimova and V. Vasin. Stable parallel algorithms for solving the inverse gravimetry and magnetometry problems. *Eng. Model.* **17** (2004), 13–19.

[5] W. Albrecht, H. Fuchs, and W. Kittelmann, eds. *Nonwoven Fabrics: Raw Materials, Manufacture, Applications, Characteristics, Testing Processes*. Weinheim: Wiley-VCH, 2003.

[6] I. Amna and V. Michel. Pseudodifferential operators, cubature and equidistribution on the 3D-ball. *Numer. Funct. Anal. Optim.* **38** (2017), 891–910.

[7] D. D. Ang, R. Gorenflo, and L. K. Vy. A uniqueness theorem for a nonlinear integral equation of gravimetry. *Proceedings of the First World Congress of Nonlinear Analysts*. Tampa, Florida, August 19–26, 1992. Ed. by V. Lakshmikantham. Vol. 3. Berlin, New York: de Gruyter, 1996, 2423–2430.

[8] D. D. Ang, R. Gorenflo, and L. K. Vy. Regularization of a nonlinear integral equation of gravimetry. *J. Inverse Ill-Posed Probl.* **5** (1997), 101–116.

[9] D. D. Ang, N. V. Nhan, and D. N. Thanh. A nonlinear integral equation of gravimetry: uniqueness and approximation by linear moments. *Vietnam J. Math.* **27** (1999), 61–67.

[10] A. B. Bakushinsky. The problem of the convergence of the iteratively regularized Gauss-Newton method. *Comput. Math. Math. Phys.* **32** (1992), 1353–1359.

[11] A. B. Bakushinsky and M. Y. Kokurin. On the construction of stable iterative methods for solving ill-posed nonlinear equations with nondifferentiable operators. *J. Inverse Ill-Posed Probl.* **11** (2003), 329–341.

[12] A. Bakushinsky, A. Smirnova, and M. A. Skinner. Iteratively regularized gradient method with a posteriori stopping rule for 2d inverse gravimetry problem. *J. Integral Equations Appl.* **17** (2005), 375–390.

[13] L. Ballani, J. Engels, and E. Grafarend. Global base functions for the mass density in the interior of a massive body (Earth). *Manuscr. Geodaet.* **18** (1993), 99–114.

[14] A. R. Barron, A. Cohen, W. Dahmen, and R. DeVore. Approximation and learning by greedy algorithms. *Ann. Statist.* **36** (2008), 64–94.

[15] H. Bauer. *Measure and Integration Theory*. Berlin: deGruyter, 2001.

[16] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2** (2009), 183–202.

[17] E. van den Berg and M. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* **31** (2008), 890–912.

[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. *Int. J. Geomath.* **1** (2011), 167–204.

[19] P. Berkel and V. Michel. On mathematical aspects of a combined inversion of gravity and normal mode variations by a spline method. *Math. Geosci.* **42** (2010), 795–816.

[20] H. Bertete-Aguirre, E. Cherkaev, and M. Oristaglio. Non-smooth gravity problem with total variation penalization functional. *Geophys. J. Int.* **149** (2002), 499–507.

[21] C. Blick, W. Freeden, and H. Nutz. Feature extraction of geological signatures by multiscale gravimetry. *Int. J. Geomath.* **8** (2017), 57–83.

[22] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.* **14** (2008), 629–654.

[23] K. Bredies and D. Lorenz. Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.* **14** (2008), 813–837.

[24] M. Burger and S. Osher. A survey on level set methods for inverse problems and optimal design. *European J. Appl. Math.* **16** (2005), 263–301.

[25] E. J. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Stat.* **35** (2007), 2313–2351.

[26] J. Cea. *Lectures on Optimization—Theory and Algorithms*. Berlin, Heidelberg, New York: Springer, 1978.

[27] A. R. Chappell and N. J. Kusznir. Three-dimensional gravity inversion for Moho depth at rifted continental margins incorporating a lithosphere thermal gravity anomaly correction. *Geophys. J. Int.* **174** (2008), 1–13.

[28] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by Basis Pursuit. *SIAM J. Sci. Comput.* **20** (1999), 33–61.

[29] C. Clauser. *Einführung in die Geophysik*. Berlin, Heidelberg: Springer, 2014.

[30] L. Dagum and R. Menon. OpenMP: an industry-standard API for shared-memory programming. *IEEE Comp. Sci. Eng.* **5** (1998), 46–55.

[31] I. Daubechies, M. Defrise, and C. de Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* **57** (2004), 1413–1457.

[32] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.* **5** (1996), 173–187.

[33] L. Devroye. *Non-Uniform Random Variate Generation*. New York: Springer, 1986.

[34] S. J. Dilworth, N. J. Kalton, D. Kutzarova, and V. N. Temlyakov. The thresholding greedy algorithm, greedy bases, and duality. *Constr. Approx.* **19** (2003), 575–597.

[35] M. A. Diniz-Ehrhardt, J. M. Martínez, and S. A. Santos. Parallel projection methods and the resolution of ill-posed problems. *Comput. Math. Appl.* **27** (1994), 11–24.

[36] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theor.* **41** (1995), 613–627.

[37] D. L. Donoho and M. Elad. On the stability of the basis pursuit in the presence of noise. *Signal Process.* **86** (2006), 511–532.

[38] M. R. Drinkwater, R. Floberghagen, R. Haagmans, D. Muzi, and A. Popescu. GOCE: ESA's first Earth explorer core mission. *Space Sci. Rev.* **108** (2003), 419–432.

[39] J. R. Driscoll and D. M. Healy. Computing Fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.* **15** (1994), 202–250.

[40] A. M. Dziewonski and D. L. Anderson. Preliminary Reference Earth Model. *Phys. Earth Planet. Inter.* **25** (1981), 297–356.

[41] J. Elschner and M. Yamamoto. Uniqueness in determining polygonal sound-hard obstacles with a single incoming wave. *Inverse Probl.* **22** (2006), 355–364.

[42] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Dordrecht: Kluwer, 1996.

[43] K. Eriksson, D. Estep, and C. Johnson. *Applied Mathematics: Body and Soul. Derivatives and Geometry in* $\mathbb{R}^3$. Vol. 1. Berlin: Springer, 2004.

[44] M. J. Fengler, D. Michel, and V. Michel. Harmonic spline-wavelets on the 3-dimensional and their application to the reconstruction of the Earth's density distribution from gravitational data at arbitrarily shaped satellite orbits. *Z. Angew. Math. Mech.* **86** (2006), 856–873.

[45] D. Fischer. *Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies*. PhD thesis. Geomathematics Group, University of Siegen, 2011. Published by Dr. Hut, München.

[46] D. Fischer and V. Michel. Sparse regularization of inverse gravimetry—case study: spatial and temporal mass variations in South America. *Inverse Probl.* **28** (2012), 065012.

[47] D. Fischer and V. Michel. Automatic best-basis selection for geophysical tomographic inverse problems. *Geophys. J. Int.* **193** (2013), 1291–1299.

[48] D. Fischer and V. Michel. Inverting GRACE gravity data for local climate effect. *J. Geod. Sci.* **3** (2013), 151–162.

[49] F. Flechtner, P. Morton, and F. Webb. Status of the GRACE follow-on mission. *Gravity, Geoid and Height Systems*. Ed. by U. Marti. Cham: Springer, 2014, 117–121.

[50] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. New York: Springer, 2013.

[51] W. Freeden. Spherical spline interpolation—basic theory and computational aspects. *J. Comp. Appl. Math.* **11** (1984), 367–375.

[52] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Boca Raton: CRC Press, 2013.

[53] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere*. Oxford: Oxford University Press, 1998.

[54] W. Freeden and M. Gutting. *Special Functions of Mathematical (Geo-)physics*. Basel: Birkhäuser, 2013.

[55] W. Freeden and K. Hesse. On the multiscale solution of satellite problems by use of locally supported kernel functions corresponding to equidistributed data on spherical orbits. *Studia Sci. Math. Hungar.* **39** (2002), 37–74.

[56] W. Freeden and V. Michel. *Multiscale Potential Theory. With Applications to Geoscience*. Boston, Basel, Berlin: Birkhäuser, 2004.

[57] W. Freeden and M. Schreiner. Non-orthogonal expansions on the sphere. *Math. Meth. Appl. Sci.* **18** (1995), 83–120.

[58] W. Freeden and M. Schreiner. *Spherical Functions of Mathematical Geosciences. A Scalar, Vectorial, and Tensorial Setup*. Berlin: Springer, 2009.

[59] W. Freeden and U. Windheuser. Combined spherical harmonic and wavelet expansion—a future concept in Earth's gravitational determination. *Appl. Comput. Harmon. Anal.* **4** (1997), 1–37.

[60] J. H. Friedman and W. Stuetzle. Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** (1981), 817–823.

[61] J. H. Friedman, W. Stuetzle, and A. Schroeder. Projection pursuit density estimation. *J. Amer. Statist. Assoc.* **79** (1984), 599–608.

[62] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual*. 3rd ed. Bristol: Network Theory, 2009.

[63] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A framework for high-resolution descretization and fast analysis of data distributed on the sphere. *Astrophys. J.* **622** (2005), 759–771.

[64] L. Grafakos. *Classical Fourier Analysis*. 3rd ed. New York: Springer, 2014.

[65] S. Gramsch, D. Hietel, and R. Wegener. Optimizing spunbond, meltblown, and airlay processes with FIDYST. *Melliand Int.* **21** (2015), 115–117.

[66] S. Gramsch, M. Kontak, and V. Michel. Three-dimensional simulation of nonwoven fabrics using a greedy approximation of the distribution of fiber directions. *Z. Angew. Math. Mech.* **98** (2018), 277–288.

[67] R. Gribonval and M. Nielsen. Approximate weak greedy algorithms. *Adv. Comput. Math.* **14** (2001), 361–378.

[68] M. Grothaus, A. Klar, J. Maringer, P. Stilgenbauer, and R. Wegener. Application of a three-dimensional fiber lay-down model to non-woven production processes. *J. Math. Ind.* **4** (2014), Article No. 4.

[69] J. Guimera, L. Rivero, R. Salas, and A. Casas. Moho depth inferred from gravity and topography in an intraplate area (Iberian Chain). *Tectonophysics* **666** (2016), 134–143.

[70] M. Gutting, B. Kretz, V. Michel, and R. Telschow. Study on parameter choice methods for the RFMP with respect to downward continuation. *Front. Appl. Math. Stat.* **3** (2017), Article 10.

[71] E. Haber and D. Oldenburg. Joint inversion: a structural approach. *Inverse Probl.* **13** (1997), 63–77.

[72] E. Haber and D. Oldenburg. A GCV based method for nonlinear ill-posed problems. *Comput. Geosci.* **4** (2000), 41–63.

[73] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull.* **13** (1902), 49–52.

[74] P. Hall, G. S. Watson, and J. Cabrera. Kernel density estimation with spherical data. *Biometrika* **74** (1987), 751–762.

[75] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* **72** (1995), 21–37.

[76] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Berlin: Springer, 2004.

[77] L. L. Helms. *Potential Theory*. London: Springer, 2009.

[78] F. Hettlich and W. Rundell. Iterative methods for the reconstruction of an inverse potential problem. *Inverse Probl.* **12** (1996), 251–266.

[79] F. Hettlich and W. Rundell. Recovery of the support of a source term in an elliptic differential equation. *Inverse Probl.* **13** (1997), 959–976.

[80] F. Hettlich and W. Rundell. A second degree method for nonlinear inverse problems. *SIAM J. Numer. Anal.* **37** (2000), 587–620.

[81] H. Heuser. *Funktionalanalysis*. 3rd ed. Stuttgart: Teubner, 1992.

[82] B. Hofmann and O. Scherzer. Factors influencing the ill-posedness of nonlinear problems. *Inverse Probl.* **10** (1994), 1277–1297.

[83] B. Hofmann and O. Scherzer. Local ill-posedness and source conditions of operator equations in Hilbert spaces. *Inverse Probl.* **14** (1998), 1189–1206.

[84] M. H. Holmes. *Introduction to the Foundations of Applied Mathematics*. New York: Springer, 2009.

[85] H.-H. Hsieh and H.-Y. Yen. Three-dimensional density structures of Taiwan and tectonic implications based on the analysis of gravity data. *J. Asian Earth Sci.* **124** (2016), 247–259.

[86] P. J. Huber. Projection pursuit. *Ann. Statist.* **13** (1985), 435–475.

[87] K. K. Imomnazarov, P. P. Korovin, and T. T. Rakhmonov. Numerical solution of an inverse problem of gravimetry for a contact surface. *Appl. Math. Lett.* **18** (2005), 187–190.

[88] V. Isakov. *Inverse Source Problems*. Providence: American Mathematical Society, 1990.

[89] V. Isakov. *Inverse Problems for Partial Differential Equations*. 2nd ed. New York: Springer, 2006.

[90] V. Isakov. Inverse obstacle problems. *Inverse Probl.* **25** (2009), 123002.

[91] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Republication. Mineola: Dover, 2009.

[92] L. K. Jones. On a conjecture of Huber concerning the convergence of projection pursuit regression. *Ann. Statist.* **15** (1987), 880–882.

[93] M. Jones, J. Marron, and S. Sheather. A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** (1996), 401–407.

[94] M. I. Kadets and V. M. Kadets. *Series in Banach Spaces, Conditional and Unconditional Convergence*. Basel, Boston, Berlin: Birkhäuser, 1997.

[95] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Berlin: de Gruyter, 2008.

[96] J. B. Keller. Inverse problems. *Amer. Math. Monthly* **83** (1976), 107–118.

[97] O. D. Kellogg. *Foundations of Potential Theory*. New York: Frederick Ungar, 1929.

[98] B. W. Kerningham and D. M. Ritchie. *The C Programming Language*. 2nd ed. Englewood Cliffs: Prentice Hall, 1988.

[99] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. 2nd ed. New York: Springer, 2011.

[100] A. Klar, N. Marheineke, and R. Wegener. Hierarchy of mathematical models for production processes of technical textiles. *Z. Angew. Math. Mech.* **89** (2009), 941–961.

[101] A. Klar, J. Maringer, and R. Wegener. A 3D model for fiber lay-down in nonwoven production processes. *Math. Models Methods Appl. Sci.* **22** (2012), 1250020.

[102] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer, 1992.

[103] D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation: I. Validation. *Geophys. J. Int.* **149** (2002), 390–412.

[104] Y. Korolev. Making use of a partial order in solving inverse problems. *Inverse Probl.* **29** (2013), 095012.

[105] B. Korte and J. Vygen. *Combinatorial Optimization*. Berlin, Heidelberg: Springer, 2012.

[106] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.* **73** (1951), 615–624.

[107] G. Lauricella. Sulla distribuzione della massa nell'interno dei pianeti. *Rend. Acc. Linei* **XXI** (1912), 18–26.

[108] J. M. Lee. *Manifolds and Differential Geometry*. Providence: American Mathematical Society, 2009.

[109] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.* **2** (1944), 164–168.

[110] S. Leweke. PhD thesis. Geomathematics Group, University of Siegen, 2018. In preparation.

[111] O. Y. Loukianov. Problème inverse de la théorie du potentiel logarithmique pour les lemniscates. *Potential Anal.* **1** (1992), 337–341.

[112] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41** (1993), 3397–3415.

[113] N. Mao and S. Russel. Modelling of nonwoven materials. *Modelling and predicting textile behaviour*. Ed. by X. Chen. Cambridge: Woodhead, 2010, 180–224.

[114] N. Marheineke and R. Wegener. Modeling and application of a stochastic drag for fiber dynamics in turbulent flows. *Int. J. Multiphase Flow* **37** (2011), 136–148.

[115] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* **11** (1963), 431–441.

[116] G. Marsaglia. Choosing a point from the surface of a sphere. *Ann. Math. Stat.* **43** (1972), 645–646.

[117] S. Matousek. The Juno new frontiers mission. *Acta Astronautica* **61** (2007), 932–939.

[118] V. Michel. *A Multiscale Method for the Gravimetry Problem*. PhD Thesis. Geomathematics Group, University of Kaiserslautern, 1999. Published by Shaker, Aachen.

[119] V. Michel. *A Multiscale Approximation for Operator Equations in Separable Hilbert Spaces—Case Study: Reconstruction and Description of the Earth's Interior*. Habilitation thesis. Geomathematics Group, University of Kaiserslautern, 2002. Published by Shaker, Aachen.

[120] 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. *Appl. Comput. Harmonic Analysis* **12** (2002), 77–99.

[121] 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 Probl.* **21** (2005), 997–1025.

[122] V. Michel. *Lectures on Constructive Approximation. Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball*. New York: Birkhäuser, 2013.

[123] V. Michel. RFMP: an iterative best basis algorithm for inverse problems in the geosciences. *Handbook of Geomathematics*. Ed. by W. Freeden, M. Z. Nashed, and T. Sonar. 2nd ed. Berlin, Heidelberg: Springer, 2015, 2121–2147.

[124] 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 Probl.* **24** (2008), 045019.

[125] V. Michel and S. Orzlowski. On the null space of a class of Fredholm integral equations of the first kind. *J. Inverse Ill-Posed Probl.* **24** (2016), 687–710.

[126] V. Michel and S. Orzlowski. On the convergence theorem for the Regularized Functional Matching Pursuit (RFMP) algorithm. *Int. J. Geomath.* **8** (2017), 183–190.

[127] V. Michel and R. Telschow. A non-linear approximation method on the sphere. *Int. J. Geomath.* **5** (2014), 195–224.

[128] V. Michel and R. Telschow. The Regularized Orthogonal Functional Matching Pursuit for ill-posed inverse problems. *SIAM J. Numer. Anal.* **54** (2016), 262–287.

[129] V. Michel and K. Wolf. Numerical aspects of a spline-based multiresolution recovery of the harmonic mass density out of gravity functionals. *Geophys. J. Int.* **173** (2008), 1–16.

[130] S. G. Mikhlin. *Mathematical Physics, An Advanced Course*. Amsterdam, London: North-Holland, 1970.

[131] L. Misici and F. Zirilli. The inverse gravimetry problem: an application to the northern San Francisco craton granite. *J. Optim. Theory. Appl.* **63** (1989), 39–49.

[132] D. S. Mitrinovic. *Analytic Inequalities*. Berlin, Heidelberg, New York: Springer, 1970.

[133] P. J. Mohr, D. B. Newell, and B. N. Taylor. CODATA recommended values of the fundamental physical constants: 2014. *Rev. Mod. Phys.* **88** (2016), 035009.

[134] C. Müller. *Spherical Harmonics*. Berlin, Heidelberg, New York: Springer, 1966.

[135] M. Z. Nashed. A new approach to classification and regularization of ill-posed operator equations. *Inverse and Ill-Posed Problems*. Ed. by H. W. Engl and C. W. Groetsch. Boston: Academic Press, 1987, 53–75.

[136] I. Newton. *Philosophiae Naturalis Principia Mathematica*. 1687.

[137] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM, 1992.

[138] P. Novikov. Sur le problème inverse du potentiel. *Dokl. Akad. Nauk* **18** (1938), 165–168.

[139] OpenMP Architecture Review Board, ed. *OpenMP Application Program Interface*. Version 4.0. 2013. URL: http://www.openmp.org/specifications/.

[140] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** (1962), 1065–1076.

[141] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Asilomar Conference on Signals, Systems and Computers*. Vol. 1. IEEE Conference Publications, 1993, 40–44.

[142] N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *J. Geophys. Res.* **117** (2012), B04406.

[143] B. Pawlowski. Gravity gradiometry in resource exploration. *The Leading Edge* **17** (1998), 51–52.

[144] W. R. Pestman. *Mathematical Statistics*. 2nd ed. Berlin: de Gruyter, 2009.

[145] P. Pizzetti. Corpi equivalenti rispetto alla attrazione newtoniana esterna. *Rom. Acc. L. Rend.* **XVIII** (1909), 211–215.

[146] P. Pizzetti. Intorno alla possibli distribuzioni della massa nell'interno della terra. *Annali di Mat., Milano* **XVII** (1910), 225–258.

[147] A. G. Ramm and A. B. Smirnova. A numerical method for solving nonlinear ill-posed problems. *Numer. Funct. Anal. Optim.* **20** (1999), 317–332.

[148] C. Redenbach, A. Rack, K. Schladitz, O. Wirjadi, and M. Godehardt. Beyond imaging: on the quantitative analysis of tomographic volume data. *Int. J. Mater. Res.* **103** (2012), 217–227.

[149] M. Reguzzoni and D. Sampietro. An inverse gravimetric problem with GOCE data. *Gravity, Geoid and Earth Observation*. Ed. by S. P. Mertikas. Berlin, Heidelberg: Springer, 2010, 451–456.

[150] M. Reguzzoni and D. Sampietro. GEMMA: an Earth crustal model based on GOCE satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **35** (2015), 31–43.

[151] C. Reigber, P. Schwintzer, and H. Lühr. The CHAMP geopotential mission. *Boll. Geof. Teor. Appl.* **40** (1999), 285–289.

[152] M. Richter. *Inverse Problems*. Basel: Birkhäuser, 2016.

[153] A. Rieder. *Keine Probleme mit inversen Problemen*. Wiesbaden: Vieweg, 2003.

[154] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27** (1956), 832–837.

[155] W. Rudin. *Functional Analysis*. 2nd ed. New York: McGraw-Hill, 1991.

[156] F. Sansò. On the regular decomposition of the inverse gravimetric problem in non-L^2 spaces. *Int. J. Geomath.* **5** (2014), 33–61.

[157] F. Sansò, R. Barzaghi, and C. C. Tscherning. Choice of norm for the density distribution of the Earth. *Geophys. J. R. Astron. Soc.* **87** (1986), 123–141.

[158] P. H. Savet. Satellite orbits derived from a gravitational model of the Earth. *Planet. Space Sci.* **7** (1961), 154–163.

[159] N. Schneider. PhD thesis. Geomathematics Group, University of Siegen, 2018+. In preparation.

[160] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization Methods in Banach Spaces*. Berlin, Boston: De Gruyter, 2012.

[161] T. I. Seidman and C. R. Vogel. Well posedness and convergence of some regularisation methods for non-linear ill posed problems. *Inverse Probl.* **5** (1989), 227–238.

[162] A. Sen and M. Srivastava. *Regression Analysis. Theory, Methods, and Applications*. New York: Springer, 1990.

[163] V. S. Shubha, S. George, and P. Jidesh. A derivative free iterative method for the implementation of Lavrentiev regularization method for ill-posed equations. *Numer. Algor.* **68** (2015), 289–304.

[164] D. Stromeyer and L. Ballani. Uniqueness of the inverse gravimetric problem for point mass models. *Manuscr. Geodaet.* **9** (1984), 125–136.

[165] B. D. Tapley, S. Bettadpur, M. Watkins, and C. Reigber. The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.* **31** (2004), L09607.

[166] R. Telschow. *An Orthogonal Matching Pursuit for the Regularization of Spherical Inverse Problems*. PhD thesis. Geomathematics Group, University of Siegen, 2014. Published by Dr. Hut, München.

[167] V. N. Temlyakov. Weak greedy algorithms. *Adv. Comput. Math.* **12** (2000), 213–227.

[168] V. N. Temlyakov. Greedy algorithms in Banach spaces. *Adv. Comput. Math.* **14** (2001), 277–292.

[169] V. N. Temlyakov. *Greedy Approximation*. Cambridge: Cambridge University Press, 2011.

[170] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* **58** (1996), 267–288.

[171] A. N. Tikhonov and V. B. Glasko. Use of the regularization method in non-linear problems. *USSR Comp. Math. Math. Phys.* **5** (1965), 93–107.

[172] C. C. Tscherning and G. Strykowski. Quasi-harmonic inversion of gravity field data. *Proc. 5th Int. Mathematical Geophysics Seminar held at the Free University of Berlin*. Ed. by A. Vogel. Braunschweig, Wiesbaden: Vieweg, 1987.

[173] V. V. Vasin. Iterative processes of Fejér type in ill-posed problems with a priori information. *Russian Math. (Iz. VUZ)* **53** (2009), 1–20.

[174] V. V. Vasin. The Levenberg–Marquardt method for approximation of solutions of irregular operator equations. *Autom. Remote Control* **73** (2012), 440–449.

[175] V. V. Vasin. Modified Newton-type processes generating Fejér approximations of regularized solutions to nonlinear equations. *Proc. Steklov Inst. Math.* **284** (2014), S145–S158.

[176] V. V. Vasin and G. Y. Perestoronina. The Levenberg-Marquardt method and its modified versions for solving nonlinear equations with application to the inverse gravimetry problem. *Proc. Steklov Inst. Math.* **280** (2013), S174–S182.

[177] V. V. Vasin and G. Skorik. Iterative processes of gradient type with applications to gravimetry and magnetometry inverse problems. *J. Inverse Ill-Posed Probl.* **18** (2011), 855–876.

[178] I. Velicogna and J. Wahr. Greenland mass balance from GRACE. *Geophys. Res. Lett.* **32** (2005), L18505.

[179] P. Vincent and Y. Bengio. Kernel matching pursuit. *Mach. Learn.* **48** (2002), 165–187.

[180] K. Vozoff and D. L. B. Jupp. Joint inversion of geophysical data. *Geophys. J. R. Astron. Soc.* **42** (1975), 977–991.

[181] A. Wald and T. Schuster. Sequential subspace optimization for nonlinear inverse problems. *J. Inverse Ill-Posed Probl.* **25** (2017), 99–117.

[182] Y. Wang and Y. Yuan. Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Probl.* **21** (2005), 821–838.

[183] N. Weck. Inverse Probleme der Potentialtheorie. *Appl. Anal.* **2** (1972), 195–204.

[184] R. Wegener, N. Marheineke, and D. Hietel. Virtual production of filaments and fleeces. *Currents in Industrial Mathematics*. Ed. by H. Neunzert and D. Prätzel-Wolters. Berlin: Springer, 2015.

[185] Y. Yokota, K. Koketsu, Y. Fujii, K. Satake, S. Sakai, M. Shinohara, and T. Kanazawa. Joint inversion of strong motion, teleseismic, geodetic, and tsunami datasets for the rupture process of the 2011 Tohoku earthquake. *Geophys. Res. Lett.* **38** (2011), L00G21.

[186] K. Yosida. *Functional Analysis*. 6th ed. Berlin, Heidelberg, New York: Springer, 1980.

[187] M. T. Zuber, D. E. Smith, D. H. Lehman, T. L. Hoffmann, S. W. Asmar, and M. M. Watkins. Gravity recovery and interior laboratory (GRAIL): mapping the lunar interior from crust to core. *Space Sci. Rev.* **178** (2013), 3–24.