Sparse regularization of inverse gravimetry—case study: spatial and temporal mass variations in South America

2012 Inverse Problems 28 065012

(http://iopscience.iop.org/0266-5611/28/6/065012)

IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **28** (2012) 065012 (34pp) doi:10.1088/0266-5611/28/6/065012

Sparse regularization of inverse gravimetry—case study: spatial and temporal mass variations in South America

D Fischer and V Michel

Geomathematics Group, Department of Mathematics, University of Siegen, Germany

E-mail: fischer@mathematik.uni-siegen.de and michel@mathematik.uni-siegen.de

Received 30 January 2012, in final form 19 April 2012
Published 15 May 2012
Online at stacks.iop.org/IP/28/065012

Abstract

Sparse regularization has recently experienced high popularity in the inverse problems community. In this paper, we show that a sparse regularization technique can also be developed for linear geophysical tomography problems. For this purpose, we adapt a known matching pursuit algorithm. The main theoretical features (existence, stability, and convergence) of the new method are given. We also show further properties of some trial functions which we use. Moreover, the algorithm is applied to a static and a monthly varying gravitational field of South America which yields spatial and temporal variations in the mass distribution. The new approach represents essential progress in comparison to a corresponding wavelet method, which is not flexible enough for the use of heterogeneous data, and a respective spline method, where the resolution cannot exceed approximately 10^4 basis functions due to experienced numerical problems with the ill-conditioned and dense matrix. The novel sparse regularization technique does not require homogeneous data and is not limited in the number of basis functions due to its iterative algorithm.

(Some figures may appear in colour only in the online journal)

1. Introduction

Every new satellite mission concerned with the gravitational potential of the Earth allows us to construct more precise models of, e.g., the mass density distribution of the Earth. Missions like the Gravity Recovery and Climate Experiment (GRACE) (see [42]) allow us to reconstruct temporal variations of the mass density distribution as well. Thus, it is possible to get a more global overview over climate changes than with conventional Earth-bound methods. Nonetheless, satellite missions only measure the gravitational potential and not the density distribution itself. Newton's law states the link between the gravitational potential V and the density distribution ρ as

$$V(x) = \gamma \int_{\mathcal{B}} \frac{\rho(y)}{|x-y|}\,\mathrm{d}y, \quad x \in \mathbb{R}^3 \setminus \mathcal{B}, \tag{1.1}$$

where $\mathcal{B}$ is a ball representing the Earth and γ is the gravitational constant. This problem is known as the inverse gravimetric problem and the corresponding data can be represented as the values of a functional applied to the target function (see, e.g., [9]).

Note that the solution to this problem is not unique. In this work, we will use a harmonicity constraint to obtain a unique solution, i.e. we require the Laplace equation $\Delta\rho = 0$. Such an *a priori* condition lacks a physical interpretation but can be motivated by some mathematical arguments. At present, no uniqueness constraint with a satisfactory physical interpretation is known. For further details, see the survey [53].

Needless to say, the inverse gravimetric problem has been solved with a wide variety of methods. Among them are classical methods such as a truncated singular value decomposition (see, e.g., [71, 73]), domain subdivisions in block-like structures (see, e.g., [45, 72]), and an approximation by point-masses (see, e.g., [39, 64]). Within the last decade, advanced approximation methods for this problem have been developed. These are spline methods (see, e.g., [32, 54]) and wavelet methods (see, e.g., [48, 50, 51]). However, most of these methods only allow us to use one kind of predefined trial function on a point grid that is determined by the given data and, thus, mostly equidistributed. Furthermore, the number of points that can be used as data is limited by most methods. In this paper, we want to remedy both these restrictions. We present a new method that allows us to use all different kinds of trial functions imaginable and many more data points than commonly used up to now. Furthermore, the solution generated by our new algorithm will be additionally adapted to the structure of the signal instead of the data structure only.

Note that our novel technique is an essentially different algorithmic approach in comparison to free-positioned point mass modelling (see [6, 7]). The latter method positions point-masses shortly below the surface of the Earth stepwise to best match the gravitational potential given by the data, as well. However, it shows several disadvantages such as numerical instabilities (see [16]).

Research in sparse regularization is concerned with solving under-determined or ill-conditioned systems of linear equations with respect to the sparsity of the solution. Of course, there exists a rather large number of different approaches. The main research areas seem to be combinatorial algorithms (see, e.g., [37, 38]) and convex relaxation (see, e.g., [11, 15, 20, 44]), to name a few. However, these methods either need a very large number of measurements or they are very slow. Iterative greedy algorithms (see, e.g., [19, 25, 47, 57]) seem to be efficient with respect to the computational effort as well as the number of measurements needed (see [57]).

However, most of these methods require prior knowledge of the sparsity of the solution. In our case, as well as in most applications, we do not possess this knowledge. Furthermore, the trial functions on the ball, such as orthogonal polynomials or scaling functions, have different properties than their Euclidean counterparts. As a consequence, the existing methods and algorithms cannot be used directly to solve our problem. In this paper, we introduce a new approach that allows us to exploit the advantages of iterative greedy algorithms on the ball without prior knowledge of the sparsity.

For further details on the general Hilbert space setting as well as sparsity with respect to inverse problems, we refer to, e.g., [5, 18, 21, 24, 70].

Another feature of a matching pursuit is the possibility to combine trial functions of different characters. In our case of approximating a solution on the ball, we combine global basis functions (orthogonal polynomials), which have known pros and cons, with localized trial functions. The latter are based on spline and wavelet methods, which have been developed by the Geomathematics Groups at the University of Kaiserslautern and the University of Siegen. They allow us to minimize the effects of data gaps or differences in the data density. In this

work, we exploit the localizing character of these functions to reconstruct the detail structure of the density distribution of the Earth. We also fill a previous gap in the theory by proving for some system that it forms a basis as well. For further theoretical details of these trial functions and their applications mostly to geophysical problems, we refer to [4, 9, 31, 32, 34–36, 43, 48, 54, 58].

Certainly, further localized trial functions could be considered here. For example, Slepian functions are a locally and globally orthogonal system of functions on the sphere that is optimally suited for a local reconstruction in areas of interest to minimize the effects of data gaps (see [1, 55, 66–68] for theoretical results and applications). However, this concept is, up to now, limited to the sphere, i.e. the surface of a ball.

We also observe another feature of the new method. It allows us to initially use a much denser grid for the centres of the localized kernel functions, where only significant ones are chosen out of this grid by the algorithm. This corresponds to the sparsity of the solution.

The outline of this paper is as follows.

In section 2, we summarize a few important but well-known notations and fundamentals. Furthermore, we explain the global and localized trial functions that we use in the numerical tests later on. We also prove that the chosen localized functions represent a basis system.

In section 3, we give an overview over the inverse gravimetric problem and a representation of the data as the values of functionals applied to the target function. Section 4 is concerned with the development of the functional matching pursuit (FMP) as well as some theoretical results regarding the convergence of the algorithm and its convergence rate. We, hereby, develop an adaptive and iterative greedy algorithm to solve approximation problems where the data is given directly by a linear and continuous operator. The novelties of this method, in comparison to the matching pursuit in [47], are its applicability to inverse problems and the practical studies for functions on the ball. We also study aspects of computational optimization and parallelization of the numerical implementation.

Since the inverse gravimetric problem is an ill-posed problem, we need to enhance this algorithm to become a regularization method. This process is explained in section 5, where the regularized FMP (RFMP) is developed and theoretical results are given. Moreover, the main requirements for a regularization method, i.e. the existence and stability of the solution and the convergence, are addressed there as well.

In section 6, we demonstrate the power of the novel method on some numerical applications. First of all, we reconstruct the mass density variations of South America. As data input, we use the Earth Gravitational Model 2008 (EGM2008, [59]) developed by the National Geospatial Intelligence Agency (NGA). Note that we are now able to use much more data points than in comparable works. After an extensive case study, we compare our new method to some previously developed methods in inverse gravimetry viz splines and wavelets.

Secondly, we concern ourselves with the reconstruction of the mass transport in the Amazon area in the year 2008. Here, we use the monthly solutions which were collected by the GRACE satellite mission and preprocessed by the Jet Propulsion Laboratory (JPL, see [42]).

Finally, in section 7, we summarize the results gathered in this work and give an outlook for further research opportunities.

2. Preliminaries

In this paper, $\mathbb{N}$ represents the set of all positive integers, where $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Moreover, $\mathbb{R}$ is the set of all real numbers. Furthermore, the closed ball with radius $a > 0$ is denoted by $\mathcal{B} := \{x \in \mathbb{R}^3 | \, |x| \leqslant a\}$.

There are two known complete orthonormal systems in the space $\mathrm{L}^2(\mathcal{B})$ denoted by $G^{\mathrm{I}}_{m,n,j}$ and $G^{\mathrm{II}}_{m,n,j}$ (see, e.g., [52] and the references therein). Here, $G^{\mathrm{I}}_{m,n,j}$ is used, which is given by

$$G^{\mathrm{I}}_{m,n,j}(x) := \sqrt{\frac{4m+2n+3}{a^3}}\, P_m^{(0,n+1/2)}\left(2\frac{|x|^2}{a^2}-1\right)\left(\frac{|x|}{a}\right)^n Y_{n,j}\left(\frac{x}{|x|}\right), \tag{2.1}$$

where $m, n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$ and $x \in \mathcal{B}$. Here, $P_m^{(\alpha,\beta)}$, $m \in \mathbb{N}_0$, $\alpha, \beta > -1$, are Jacobi polynomials and $Y_{n,j}$, $n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$, are elements of the orthonormal $\mathrm{L}^2(\Omega)$-basis system of real spherical harmonics [56] where Ω is the unit sphere in $\mathbb{R}^3$. The system $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,\, j=1,\ldots,2n+1}$ can be divided into a harmonic and an anharmonic part, where the basis functions with $m = 0$ correspond to the harmonic part, i.e. the functions $G^{\mathrm{I}}_{m,n,j}$ with $m > 0$ represent a basis for the orthogonal complement of the subspace of all harmonic functions in $\mathrm{L}^2(\mathcal{B})$ viz the space of anharmonic functions. Note that every function is a polynomial of cartesian coordinates x_1, x_2, and x_3, which is well defined in $x = 0$.

Furthermore, we use localized kernel functions in $\mathrm{L}^2(\mathcal{B} \times \mathcal{B})$ based on [2, 8, 10, 52] to reconstruct detail structures of the target function. They are reproducing kernels of Sobolev spaces and can be used to generate an $\mathrm{L}^2(\mathcal{B})$-basis as we prove in this section.

Definition 2.1. *Let the given (non-trivial) sequence* $(K^\wedge(m,n))_{m,n\in\mathbb{N}_0}$ *be I-summable:*

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[K^\wedge(m,n)]^2 n(2m+n)\frac{\left(n+m+\frac{1}{2}\right)^{2m}}{(m!)^2} < \infty. \tag{2.2}$$

The function F is an element of the Hilbert (Sobolev) space $\mathcal{H} = \mathcal{H}((K^\wedge(m,n)), \mathcal{B})$ if F is a function in $\mathrm{L}^2(\mathcal{B})$ with

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

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

The scalar product is for $F_1, F_2 \in \mathcal{H}$ defined as

$$\langle F_1, F_2\rangle_{\mathcal{H}} := \sum_{\substack{m,n=0\\ K^\wedge(m,n)\neq 0}}^{\infty} [K^\wedge(m,n)]^{-2} \sum_{j=1}^{2n+1} \left\langle F_1, G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2(\mathcal{B})}\left\langle F_2, G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2(\mathcal{B})}. \tag{2.3}$$

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

Theorem 2.2 (Sobolev lemma)**.** *Every space $\mathcal{H}$ in definition 2.1 is a subspace of* $\mathrm{C}(\mathcal{B})$.

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

$$\|F\|^2_{\mathrm{C}(\mathcal{B})} \leqslant \left(\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[K^\wedge(m,n)]^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3}\binom{m+n+\frac{1}{2}}{m}^2\right)\|F\|^2_{\mathcal{H}}, \tag{2.4}$$

where the norm on $\mathrm{C}(\mathcal{B})$ is the supremum norm. Moreover, it is known that the summability condition implies the convergence of the series.

Let us now introduce reproducing kernels. For further details on reproducing kernels in general we refer to [22]. For details on the setting of a three-dimensional ball see, e.g., [2, 3, 8].

Definition 2.3 (Reproducing kernel)**.** *Let H be a Hilbert space of real functions on the subset $X \subset \mathbb{R}^n$. $K_H : X \times X \to \mathbb{R}$ is a reproducing kernel of H if*

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

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

Theorem 2.4 (Reproducing kernel of $\mathcal{H}$)**.** *Every $\mathcal{H}((K^\wedge(m,n)),\mathcal{B})$ in definition 2.1 is a reproducing kernel Hilbert space. The corresponding reproducing kernel is the product series associated with the sequence $((K^\wedge(m,n))^2)_{m,n\in\mathbb{N}_0}$:*

$$K^{\mathrm{I}}_{\mathcal{H}}(x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}[K^\wedge(m,n)]^2G^{\mathrm{I}}_{m,n,j}(x)G^{\mathrm{I}}_{m,n,j}(y) \tag{2.5}$$

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

These reproducing kernels can be used to construct basis systems for functions on $\mathcal{B}$, as we show in the following new results, where we refer to a reproducing kernel Hilbert space $\mathcal{H}=\mathcal{H}((K^\wedge(m,n)),\mathcal{B})$ as it was defined in definition 2.1. Note that the kernels have previously been used as scaling functions or as generators of spline bases. The ideas of the following proofs are based on those of the spherical setting.

Theorem 2.5 (Closed system in $\mathcal{H}$)**.** *Let $X\subset\mathcal{B}$ be countable and dense. Then the system $\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}$ is closed (in the sense of the approximation theory) in $(\mathcal{H},\|\cdot\|_{\mathcal{H}})$, i.e.*

$$\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}=\mathcal{H}. \tag{2.6}$$

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

$$\mathcal{H}=\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}\oplus\Big(\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}\Big)^{\perp_{\mathcal{H}}}, \tag{2.7}$$

where $\perp_{\mathcal{H}}$ refers to the orthogonal complement. Let us start with an arbitrary function

$$F\in\Big(\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}\Big)^{\perp_{\mathcal{H}}}\subset\mathcal{H}\subset\mathrm{C}(\mathcal{B}). \tag{2.8}$$

Obviously, we obtain for all functions $G\in\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}$ that the scalar product vanishes, i.e. $\langle F,G\rangle_{\mathcal{H}}=0$. In particular, it follows that $\langle F,K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\rangle_{\mathcal{H}}=0$ for all $x\in X$. Since $K^{\mathrm{I}}_{\mathcal{H}}$ is a reproducing kernel, we conclude that $F(x)=0$ for all $x\in X$.

Since F is a continuous function and X is a dense subset of $\mathcal{B}$, it follows that F is identical to 0. Hence, $\mathcal{H}=\overline{\operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}}^{\|\cdot\|_{\mathcal{H}}}$. □

Note that the concept of 'closed' in the sense of approximation theory differs from the concept of 'closed' in the topological sense.

Theorem 2.6 (Closed system in $\mathrm{C}(\mathcal{B})$)**.** *Let $X\subset\mathcal{B}$ be countable and dense and let $K^\wedge(m,n)\neq0$ for all $m,n\in\mathbb{N}_0$. Then the system $\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot)\mid x\in X\}$ is closed (in the sense of the approximation theory) in $(\mathrm{C}(\mathcal{B}),\|\cdot\|_{\mathrm{C}(\mathcal{B})})$.*

Proof. First of all, we know that $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,\,j=1,\dots,2n+1}\subset\mathcal{H}\subset\mathrm{C}(\mathcal{B})$, since $K^\wedge(m,n)\neq0$ for all $m,n\in\mathbb{N}_0$.

Secondly, $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,\,j=1,\dots,2n+1}$ is a basis of the space of all polynomials on $\mathcal{B}$ (see [48] for the anharmonic part, where the results for the harmonic part can be shown analogously). With the help of the Weierstraß approximation theorem we conclude that $\overline{\operatorname{span}\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,\,j=1,\dots,2n+1}}^{\|\cdot\|_{\mathrm{C}(\mathcal{B})}}=\mathrm{C}(\mathcal{B})$ and, thus, $\overline{\mathcal{H}}^{\|\cdot\|_{\mathrm{C}(\mathcal{B})}}=\mathrm{C}(\mathcal{B})$. Now, we conclude for $F\in\mathrm{C}(\mathcal{B})$ and $\varepsilon>0$ that there exists a function $G\in\mathcal{H}$ such that $\|F-G\|_{\mathrm{C}(\mathcal{B})}<\frac{\varepsilon}{2}$.

From theorem 2.5, we know that there exists a function $H \in \operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}$ such that

$$\|G - H\|_{\mathcal{H}} \tag{2.9a}$$

$$< \frac{\varepsilon}{2} \left(\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[K^{\wedge}(m,n)\right]^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3} \binom{m+n+\frac{1}{2}}{m}^2 \right)^{-1/2}. \tag{2.9b}$$

With the Sobolev lemma and (2.4), we conclude that

$$\|G - H\|_{\mathrm{C}(\mathcal{B})} \tag{2.10a}$$

$$\leqslant \left(\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[K^{\wedge}(m,n)\right]^2 \frac{(2n+1)(4m+2n+3)}{4\pi a^3} \binom{m+n+\frac{1}{2}}{m}^2 \right)^{1/2} \|G - H\|_{\mathcal{H}} \tag{2.10b}$$

$$< \frac{\varepsilon}{2}. \tag{2.10c}$$

Let us summarize these results:

$$\|F - H\|_{\mathrm{C}(\mathcal{B})} \leqslant \|F - G\|_{\mathrm{C}(\mathcal{B})} + \|G - H\|_{\mathrm{C}(\mathcal{B})} < \varepsilon. \tag{2.11}$$

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

Theorem 2.7 (Closed system in $\mathrm{L}^2(\mathcal{B})$)**.** *Let $X \subset \mathcal{B}$ be countable and dense and let $K^{\wedge}(m,n) \neq 0$ for all $m,n \in \mathbb{N}_0$. Then the system $\left\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\right\}$ is closed (in the sense of the approximation theory) in $(\mathrm{L}^2(\mathcal{B}), \|\cdot\|_{\mathrm{L}^2(\mathcal{B})})$.*

Proof. First, let us prove that $\overline{\mathrm{C}(\mathcal{B})}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}} = \mathrm{L}^2(\mathcal{B})$: Let $\mathcal{B}'$ be an open ball that includes $\mathcal{B}$. All continuous functions F defined on $\mathcal{B}$ can be extended continuously into $\mathcal{B}'$, e.g. by $F(r\xi) := F(a\xi)$ for all $r > a$ and unit vectors $\xi \in \mathbb{R}^3$. Theorem 3.2.2 of [75] yields that $\{\varphi \in \mathrm{C}^{\infty}(\mathcal{B}') \mid \operatorname{supp}\varphi \subset \mathcal{B}' \text{ is compact}\} \subset \mathrm{C}(\mathcal{B}')$ is dense in $\mathrm{L}^2(\mathcal{B}')$. Thus, $\mathrm{C}(\mathcal{B}')|_{\mathcal{B}}$ is dense in $\mathrm{L}^2(\mathcal{B})$, too, where $\mathrm{C}(\mathcal{B}')|_{\mathcal{B}} := \{F|_{\mathcal{B}} \mid F \in \mathrm{C}(\mathcal{B}')\}$. Since any continuous function on $\mathcal{B}'$ is continuous on the subset $\mathcal{B}$, we conclude that $\mathrm{C}(\mathcal{B})$ is dense in $\mathrm{L}^2(\mathcal{B})$.

For all $F \in \mathrm{C}(\mathcal{B})$, we obtain that $\|F\|_{\mathrm{L}^2(\mathcal{B})} \leqslant \sqrt{\frac{4\pi a^3}{3}} \|F\|_{\mathrm{C}(\mathcal{B})}$. Let $F \in \mathrm{L}^2(\mathcal{B})$ and $\varepsilon > 0$. Since $\mathrm{C}(\mathcal{B})$ is dense in $\mathrm{L}^2(\mathcal{B})$, there exists a function $G \in \mathrm{C}(\mathcal{B})$ such that $\|F - G\|_{\mathrm{L}^2(\mathcal{B})} < \frac{\varepsilon}{2}$. With theorem 2.6 we obtain that there exists a function $H \in \operatorname{span}\{K^{\mathrm{I}}_{\mathcal{H}}(x,\cdot) \mid x \in X\}$ such that $\|G - H\|_{\mathrm{C}(\mathcal{B})} < \frac{\varepsilon}{2}\sqrt{\frac{3}{4\pi a^3}}$. As a summary of the above estimates, we conclude that

$$\|F - H\|_{\mathrm{L}^2(\mathcal{B})} \leqslant \|F - G\|_{\mathrm{L}^2(\mathcal{B})} + \|G - H\|_{\mathrm{L}^2(\mathcal{B})} \tag{2.12a}$$

$$\leqslant \|F - G\|_{\mathrm{L}^2(\mathcal{B})} + \sqrt{\frac{4\pi a^3}{3}} \|G - H\|_{\mathrm{C}(\mathcal{B})} \tag{2.12b}$$

$$< \varepsilon \tag{2.12c}$$

and, thus, we have a basis system of $\mathrm{L}^2(\mathcal{B})$ constructed out of a reproducing kernel function. □

As a particular case for our numerical applications, we use certain parameter-dependent kernels, i.e. we set

$$\tilde{K}^{\mathrm{I}}_h(x,y) := \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \left(A^h_{m,n}\right)^2 G^{\mathrm{I}}_{m,n,j}(x) G^{\mathrm{I}}_{m,n,j}(y), \tag{2.13}$$

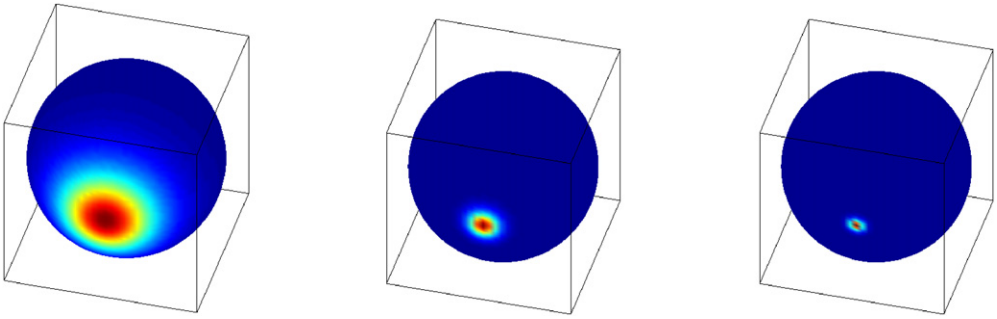

Figure 1. Kernel function $K_h^{\mathrm{I}}(\cdot, x)$ on the surface of the ball $\mathcal{B}$ with $h = 0.5$ (left), 0.8 (middle) and 0.9 (right).

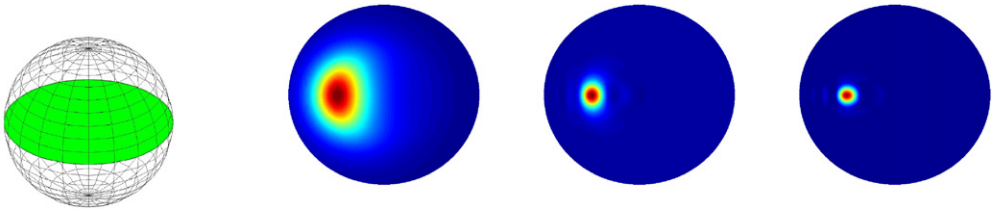

Figure 2. Equatorial cut through $\mathcal{B}$ (left) and kernel function $K_h^{\mathrm{I}}(\cdot, x)$ on this cut at $x = (0, 0.5, 0)^{\mathrm{T}}a$ with $h = 0.5$ (middle left), 0.8 (middle right) and 0.9 (right).

where $x, y \in \mathcal{B}$ and every fixed $h \in]0, 1[$ yields one reproducing kernel for one Sobolev space $\mathcal{H}$. Note that we change the notation of the kernel here.

In this paper, we only need the harmonic localized kernel functions where $m = 0$. Note that, for numerical reasons, we have to truncate the series at some $E_n \in \mathbb{N}$. The parameter-dependent symbol $A_{m,n}^h$ which we have used here is given by

$$A_{m,n}^h := \delta_{m0} \begin{cases} h^{n/2}, & n \leqslant E_n \\ 0, & \text{otherwise} \end{cases} \tag{2.14}$$

for $m, n \in \mathbb{N}_0$ and for fixed $h \in]0, 1[$, where δ_{m0} is the Kronecker delta. Thus, h is a parameter to influence the localizing character of the kernel function. The hat-width decreases for h getting closer to 1 (see figures 1 and 2). The peak of $y \mapsto \tilde{K}_h(x, y)$ is centred at x.

Now the localized kernel function can be represented as

$$\tilde{K}_h^{\mathrm{I}}(x, y) = \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} h^n G_{0,n,j}^{\mathrm{I}}(x) G_{0,n,j}^{\mathrm{I}}(y). \tag{2.15}$$

Let us compute the norm of $\tilde{K}_h^{\mathrm{I}}(\cdot, x)$. The Parseval identity yields

$$\left\| \tilde{K}_h^{\mathrm{I}}(\cdot, x) \right\|_{\mathrm{L}^2(\mathcal{B})}^2 = \int_{\mathcal{B}} \left[\tilde{K}_h^{\mathrm{I}}(y, x) \right]^2 \mathrm{d}y = \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} h^{2n} \left[G_{0,n,j}^{\mathrm{I}}(x) \right]^2 \tag{2.16a}$$

$$= \sum_{n=0}^{E_n} \sum_{j=1}^{2n+1} h^{2n} \frac{2n+3}{a^3} \left[Y_{n,j}\left(\frac{x}{|x|} \right) \right]^2 \left(\frac{|x|}{a} \right)^{2n}, \tag{2.16b}$$

since $P_0^{(0,n+1/2)} \equiv 1$. If we now use the addition theorem for spherical harmonics, we obtain the identity

$$\left\|\tilde{K}_h^{\mathrm{I}}(\cdot,x)\right\|^2_{\mathrm{L}^2(\mathcal{B})} = \sum_{n=0}^{E_n} h^{2n} \frac{2n+3}{a^3} \frac{2n+1}{4\pi} \left(\frac{|x|}{a}\right)^{2n}. \tag{2.17}$$

Note that the norm depends on h and the radial distance $|x|$ only.

In the following, we always consider normalized kernel functions and denote them with

$$K_h^{\mathrm{I}}(\cdot,x) := \frac{\tilde{K}_h^{\mathrm{I}}(\cdot,x)}{\left\|\tilde{K}_h^{\mathrm{I}}(\cdot,x)\right\|_{\mathrm{L}^2(\mathcal{B})}}, \quad x \in \mathcal{B}. \tag{2.18}$$

3. The inverse gravimetric problem

It is well known that Newton's law of gravitation

$$V(x) = \gamma \int_{\mathcal{B}} \frac{\rho(y)}{|x-y|}\,\mathrm{d}y, \quad x \in \mathbb{R}^3 \setminus \mathcal{B}, \tag{3.1}$$

where γ is the gravitational constant, represents the relation between the gravitational potential V and the mass density distribution ρ. Note that we can only reconstruct the harmonic part of ρ from the gravitational potential, since the operator $T : \mathrm{L}^2(\mathcal{B}) \to T(\mathrm{L}^2(\mathcal{B}))$, where

$$(TF)(x) := \int_{\mathcal{B}} \frac{F(y)}{|x-y|}\,\mathrm{d}y, \quad x \in \mathbb{R}^3 \setminus \mathcal{B}, \tag{3.2}$$

has the anharmonic subspace of $\mathrm{L}^2(\mathcal{B})$ as its null-space (see [27, 46, 60, 61, 76]). Furthermore, the determination of a (harmonic) solution F from given TF is instable, such that this inverse problem is ill-posed.

Let us consider the functionals $\mathcal{F}_G^k$, where the subscript G stands for gravitation, that map the density to the gravitational potential

$$\mathcal{F}_{\mathrm{G}}^k \rho := \int_{\mathcal{B}} \frac{\rho(y)}{|x_k-y|}\,\mathrm{d}y \tag{3.3}$$

where $\rho \in \mathrm{L}^2(\mathcal{B})$ is mapped to $\mathbb{R}$ and $x_k \in \mathbb{R}^3 \setminus \mathcal{B}$. In [49], the series representation

$$\mathcal{F}_{\mathrm{G}}^k\,\rho = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{4\pi}{2n+1} \sqrt{\frac{a^3}{2n+3}}\, \left\langle \rho, G^{\mathrm{I}}_{0,n,j} \right\rangle_{\mathrm{L}^2(\mathcal{B})} \left(\frac{a}{|x_k|}\right)^n \frac{1}{|x_k|}\, Y_{n,j}\left(\frac{x_k}{|x_k|}\right) \tag{3.4}$$

is derived.

Since there already exist approximate models of the mass density distribution of the Earth—for example PREM (see [28])—it is useful to apply the functionals to the deviation $\delta\rho$ which is the difference of the mass density and the density of a reference model, i.e. $\delta\rho = \rho - \rho_{\mathrm{M}}$. This means that we need to calculate the gravitational potential V_δ associated to this deviation. For instance, in a radially symmetric model, we have

$$V_\delta(x) = \gamma \int_{\mathcal{B}} \frac{\delta\rho(y)}{|x-y|}\,\mathrm{d}y = \gamma \int_{\mathcal{B}} \frac{\rho(y) - \rho_{\mathrm{M}}(|y|)}{|x-y|}\,\mathrm{d}y \tag{3.5a}$$

$$= V(x) - \gamma \frac{4\pi}{|x|} \int_0^a r^2 \rho_{\mathrm{M}}(r)\,\mathrm{d}r \tag{3.5b}$$

for all $x \in \mathbb{R}^3 \setminus \mathcal{B}$. For PREM as a reference model, we obtain $V_\delta(x) = V(x) - \gamma \frac{4\pi a^3}{3|x|}\, 5.5134 \frac{\mathrm{g}}{\mathrm{cm}^3}$ (see [48]).

At last, let us apply the functional $\mathcal{F}_{\mathrm{G}}^k$ to the two kinds of dictionary functions that we want to use in our numerical applications, i.e. the $\mathrm{L}^2(\mathcal{B})$-basis functions $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n\in\mathbb{N}_0,\,j=1,\ldots,2n+1}$ and the normalized and localized kernel functions $\{K^{\mathrm{I}}_h(\cdot,x) \mid h \in]0,1[, x \in \mathcal{B}\}$. For the $\mathrm{L}^2(\mathcal{B})$-basis, we obtain

$$\mathcal{F}_{\mathrm{G}}^k G^{\mathrm{I}}_{m,n,j} = \delta_{m0}\frac{4\pi}{2n+1}\sqrt{\frac{a^3}{2n+3}}\left(\frac{a}{|x_k|}\right)^n \frac{1}{|x_k|}\,Y_{n,j}\left(\frac{x_k}{|x_k|}\right). \tag{3.6}$$

Note that we use the harmonic basis functions $G_{0,n,j}$ only, since the other functions are elements of the null-space of $\mathcal{F}_{\mathrm{G}}^k$.

Since $\mathcal{F}_{\mathrm{G}}^k$ is a linear functional, we obtain

$$\mathcal{F}_{\mathrm{G}}^k K^{\mathrm{I}}_h(\cdot,x) = \frac{\mathcal{F}_{\mathrm{G}}^k \tilde{K}^{\mathrm{I}}_h(\cdot,x)}{\left\|\tilde{K}^{\mathrm{I}}_h(\cdot,x)\right\|_{\mathrm{L}^2(\mathcal{B})}}. \tag{3.7}$$

With the same argument as before, we only use the kernel functions corresponding to the harmonic basis functions $G^{\mathrm{I}}_{0,n,j}$, $n \in \mathbb{N}_0$, $j = 1,\ldots,2n+1$. Subsequently, we obtain for $x \in \mathcal{B}$ with the help of the addition theorem for spherical harmonics

$$\mathcal{F}_{\mathrm{G}}^k \tilde{K}^{\mathrm{I}}_h(\cdot,x) = \sum_{n=0}^{E_n}\sum_{j=1}^{2n+1} h^n\,\mathcal{F}_{\mathrm{G}}^k\big(G^{\mathrm{I}}_{0,n,j}\big)\,G^{\mathrm{I}}_{0,n,j}(x) \tag{3.8a}$$

$$= \sum_{n=0}^{E_n}\sum_{j=1}^{2n+1} h^n \frac{4\pi}{2n+1}\sqrt{\frac{a^3}{2n+3}}\left(\frac{a}{|x_k|}\right)^n \sqrt{\frac{2n+3}{a^3}}\left(\frac{|x|}{a}\right)^n \times \frac{1}{|x_k|}\,Y_{n,j}\left(\frac{x_k}{|x_k|}\right) Y_{n,j}\left(\frac{x}{|x|}\right) \tag{3.8b}$$

$$= \sum_{n=0}^{E_n} h^n \left(\frac{|x|}{|x_k|}\right)^n \frac{1}{|x_k|}\,P_n\left(\frac{x_k}{|x_k|}\cdot\frac{x}{|x|}\right), \tag{3.8c}$$

where P_n is the Legendre polynomial of degree n.

4. Functional matching pursuit

The idea to develop a solution adaptively and iteratively is not a new one. A corresponding algorithm was introduced as a matching pursuit in [47]. An enhanced version was constructed in [74], where kernel functions were introduced into the setting. The hitherto existing matching pursuits intrinsically require that the unknown function F is given in terms of grid-based data $F(x_i)$, $i = 1,\ldots,l$. For instance, in [47], the projection of F on every single trial function is calculated out of the data. However, this is not applicable if an inverse problem is to be solved.

In section 4.1, we enhance the present concept for our purposes in order to include the resolution of inverse problems. Furthermore, we give results in regard to the convergence of the algorithm and its convergence rate in section 4.2.

4.1. The algorithm

The type of problem that we want to solve is as follows:

Let l (noisy) observations $y_1,\ldots,y_l \in \mathbb{R}$ of linear and continuous functionals $\mathcal{F}^1,\ldots,\mathcal{F}^l$ applied to the target function $F \in \mathrm{L}^2(\mathcal{B})$ be given:

$$y_i = \mathcal{F}^i F, \quad i = 1,\ldots,l. \tag{4.1}$$

We want to find an approximation to F.

Furthermore, let a dictionary $\mathcal{D} \subset \{d \in \mathrm{L}^2(\mathcal{B}) \mid \|d\|_{\mathrm{L}^2(\mathcal{B})} = 1\}$ be chosen beforehand. Note that we do not put any further restrictions on the elements of this dictionary. Even the restriction to normalized functions is not necessary and merely simplifies a step in the (regularized) algorithm.

The basic strategy of our algorithm (as of every matching pursuit) is to look for an approximation to the target function F as a linear combination of selected dictionary elements $F_n = \sum_{k=1}^{n} \alpha_k d_k$, where n is the number of dictionary functions in the expansion, $d_1, \ldots, d_n \in \mathcal{D}$ are the chosen dictionary functions and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ are the corresponding coefficients.

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

Here we are looking for a function as a solution to our problem instead of a vector as usual in sparse regularization. Furthermore, the data (here in $\mathbb{R}^l$) and the target function (here in $\mathrm{L}^2(\mathcal{B})$) may exist in different spaces. Nonetheless, we use the idea of the matching pursuit, i.e. we choose expansion functions $d_1, \ldots, d_n \in \mathcal{D}$ and coefficients $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ stepwise such that they minimize the norm of the residual

$$\|R^n\|_{\mathbb{R}^l}^2 = \|y - \mathcal{F}F_n\|_{\mathbb{R}^l}^2 = \sum_{i=1}^{l} (y_i - \mathcal{F}^i F_n)^2, \quad F_n = \sum_{k=1}^{n} \alpha_k d_k. \tag{4.2}$$

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

Let us develop the method for the step from n to $n+1$ chosen expansion functions. We assume that F_n is given. Since $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ and the operator $\mathcal{F}$ is linear, we need to look for a combination $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$ that minimizes $\|y - \mathcal{F}(F_n + \alpha d)\|_{\mathbb{R}^l}^2 = \|y - \mathcal{F}F_n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2 = \|R^n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2$, i.e.

$$(d_{n+1}, \alpha_{n+1}) = \mathrm{argmin}_{d \in \mathcal{D}, \alpha \in \mathbb{R}} \|R^n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2. \tag{4.3}$$

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

$$0 = \frac{\partial \|R^n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2}{\partial \alpha} = -2\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} + 2\alpha \|\mathcal{F}d\|_{\mathbb{R}^l}^2 \tag{4.4}$$

and, thus,

$$\alpha = \frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2}. \tag{4.5}$$

Let us insert this into $\|R^n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2$:

$$\|R^n - \alpha \mathcal{F}d\|_{\mathbb{R}^l}^2 = \|R^n\|_{\mathbb{R}^l}^2 - 2\alpha \langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} + \alpha^2 \|\mathcal{F}d\|_{\mathbb{R}^l}^2 \tag{4.6a}$$

$$= \|R^n\|_{\mathbb{R}^l}^2 - 2\frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2} \langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} + \left(\frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2} \right)^2 \|\mathcal{F}d\|_{\mathbb{R}^l}^2 \tag{4.6b}$$

$$= \|R^n\|_{\mathbb{R}^l}^2 - \left(\frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} \right)^2. \tag{4.6c}$$

As a consequence, a dictionary element d minimizes $\|R^n - \alpha\mathcal{F}d\|^2_{\mathbb{R}^l}$ if and only if it maximizes $\left|\frac{\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}}\right|$. Furthermore, we only need to update the residual at every step, since we obtain that

$$R^{n+1} = y - \mathcal{F}F_{n+1} = y - \mathcal{F}F_n - \alpha_{n+1}\mathcal{F}d_{n+1} = R^n - \alpha_{n+1}\mathcal{F}d_{n+1} \tag{4.7}$$

due to the linearity of the operator $\mathcal{F}$.

Let us state the whole algorithm.

Algorithm 4.1 (Basic FMP)**.**

Start with $F_0 = 0$ *and* $R^0 = y$.

Given F_n.

Build $F_{n+1} = F_n + \alpha_{n+1}d_{n+1}$ *such that*

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

Set $R^{n+1} = R^n - \alpha_{n+1}\mathcal{F}d_{n+1}$.

Note that the same dictionary element may be chosen at different steps of the algorithm. Moreover, this algorithm does not provide us with the best match to the target function F. Since we determine the expansion functions and the corresponding coefficients stepwise, the expansion with n elements is possibly not optimal at step $n+1$. To remedy this inaccuracy we can do a back-projection in analogy to [23]. That means we choose the function d_{n+1} as in the original algorithm but recompute the optimal set of coefficients in each step:

$$\left(\alpha_1^{n+1}, \ldots, \alpha_{n+1}^{n+1}\right) = \operatorname{argmin}_{(\alpha_1,\ldots,\alpha_{n+1})\in\mathbb{R}^{n+1}} \|y - \mathcal{F}F_{n+1}\|^2_{\mathbb{R}^l}. \tag{4.8}$$

This extension of the algorithm gives us a better approximation while the computation time is increased. To obtain an even more accurate result we may use pre-fitting, again in analogy to [23], where we directly optimize

$$\left(d_{n+1}, \alpha_1^{n+1}, \ldots, \alpha_{n+1}^{n+1}\right) = \operatorname{argmin}_{(d,\alpha_1,\ldots,\alpha_{n+1})\in\mathcal{D}\times\mathbb{R}^{n+1}} \|y - \mathcal{F}F_{n+1}\|^2_{\mathbb{R}^l}. \tag{4.9}$$

It can easily be seen that this is the most time-consuming version of the three but it gives the best-fitting solution as well. Nonetheless, in this work we only use the original version of algorithm 4.1 (FMP) to reduce the computational effort.

Let us now consider algorithm 4.1 (FMP) from a computational point of view. We may compute $\mathcal{F}d$ and $\|\mathcal{F}d\|_{\mathbb{R}^l}$ beforehand for all $d \in \mathcal{D}$ and store them. This is an essential part of the algorithm that can be parallelized to reduce the computation time. All in all, in step $n+1$, we just need to search for the optimal dictionary element d_{n+1} and update the residual R^{n+1}.

4.2. Theoretical results

Although most proofs may be adapted from [47], we lose some properties by considering inverse problems instead of the setting pursued therein. We, therefore, show here how the convergence of the FMP can be proved.

To treat theoretical questions, let us first rewrite the expression for the residual R^n. For $n = 0$, we obtain $R^0 = y = \mathcal{F}F$. For $n > 0$, we obtain due to (4.7) that

$$R^n = R^{n+1} + \alpha_{n+1}\mathcal{F}d_{n+1} = R^{n+1} + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l}}\mathcal{F}d_{n+1}. \tag{4.10}$$

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

$$\frac{|\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}} \geqslant \sup_{d\in\mathcal{D}} \frac{|\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l}|}{\|\mathcal{F}d\|_{\mathbb{R}^l}}. \tag{4.11}$$

Let us derive some important properties of the residual R^n to prove the convergence of the algorithm later on.

Lemma 4.2 (Convergence of $(\|R^n\|_{\mathbb{R}^l})_n$)**.** *Let $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ be a linear operator. If the dictionary functions and the corresponding coefficients are chosen according to algorithm 4.1 (FMP) then the $\mathbb{R}^l$-norm of the residual $(\|R^n\|_{\mathbb{R}^l})_n$, where the residual R^n is given in (4.10), converges for n tending to infinity. Furthermore, the following equalities hold:*

$$\|R^n\|_{\mathbb{R}^l}^2 = \|R^{n+1}\|_{\mathbb{R}^l}^2 + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} \tag{4.12}$$

$$0 = \langle R^{n+1}, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}. \tag{4.13}$$

Proof. The proof is analogous to a corresponding proof in [47]. □

Lemma 4.3. *Let $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ be a linear operator and let all dictionary functions and the corresponding coefficients be chosen according to algorithm 4.1 (FMP). Then*

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} = 0. \tag{4.14}$$

Proof. Let us display the norm of R^0 by means of a telescoping sum. Due to lemma 4.2, we obtain

$$\|R^0\|_{\mathbb{R}^l}^2 = \sum_{n=0}^{m-1} \left(\|R^n\|_{\mathbb{R}^l}^2 - \|R^{n+1}\|_{\mathbb{R}^l}^2\right) + \|R^m\|_{\mathbb{R}^l}^2 \tag{4.15a}$$

$$= \sum_{n=0}^{m-1} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} + \|R^m\|_{\mathbb{R}^l}^2. \tag{4.15b}$$

For m tending to infinity, we obtain with the convergence of $\|R^m\|_{\mathbb{R}^l}$ (see lemma 4.2) that

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} = 0. \tag{4.16}$$

□

With lemmas 4.2 and 4.3, we are now able to prove that algorithm 4.1 (FMP) provides us with the right solution, i.e that the residual in itself tends to 0.

Theorem 4.4 (Convergence of algorithm 4.1 (FMP))**.** *Let $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ be a linear operator and let the dictionary $\mathcal{D}$ be large enough such that* span $\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$*. Furthermore, let all dictionary functions and the corresponding coefficients be chosen according to algorithm 4.1 (FMP). Then the residual R^n converges to* 0 *for n tending to infinity.*

Proof. We already know from lemma 4.2 that the sequence of the norm of the residuals $(\|R^n\|_{\mathbb{R}^l})_n$ converges. Thus, the sequence of residuals $(R^n)_n$ is bounded. Hence, there exists a convergent subsequence $(R^{n_j})_j$ in $\mathbb{R}^l$ with limit R^∞ for j tending to infinity.

Due to lemma 4.3, we already know that

$$\lim_{n\to\infty} \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2} = 0 \tag{4.17}$$

and, consequently,

$$\lim_{j\to\infty} \frac{\langle R^{n_j}, \mathcal{F}d_{n_j+1}\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d_{n_j+1}\|_{\mathbb{R}^l}} = 0. \tag{4.18}$$

In combination with (4.11), we conclude for all dictionary elements $d \in \mathcal{D}$ that

$$\lim_{j\to\infty} \frac{\langle R^{n_j}, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = 0. \tag{4.19}$$

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

$$0 = \frac{\langle \lim_{j\to\infty} R^{n_j}, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} = \frac{\langle R^{\infty}, \mathcal{F}d\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d\|_{\mathbb{R}^l}} \tag{4.20}$$

and, since we chose the dictionary $\mathcal{D}$ large enough such that span $\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$, we conclude that $R^{\infty} = 0$. Now we know that the subsequence of the residual $(R^{n_j})_j$ tends to 0 for j tending to infinity. Due to the monotonicity of $(\|R^n\|_{\mathbb{R}^l})_n$, see (4.12), we obtain

$$\lim_{n\to\infty} \|R^n\|_{\mathbb{R}^l} = 0. \tag{4.21}$$

□

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

Theorem 4.5. *Let $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ be a linear and continuous operator and let $y \in \mathbb{R}^l$ be the given data. Moreover, let the dictionary satisfy the following properties:*

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

(2) 'semi-frame condition':
There exists a constant $c > 0$ such that for all expansions $H = \sum_{k=0}^{\infty} \beta_k d_k$, $\beta_k \in \mathbb{R}$, $d_k \in \mathcal{D}$ (not necessarily pairwise distinct), where no dictionary element is chosen infinitely often, the following inequality holds:

$$c\|H\|^2_{\mathrm{L}^2(\mathcal{B})} \leqslant \sum_{k=0}^{\infty} \beta_k^2. \tag{4.22}$$

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

Let all dictionary elements and the corresponding coefficients be chosen according to algorithm 4.1 (FMP), where no dictionary element is chosen infinitely often. Then the sequence $(F_n)_n$ of the algorithm converges to a function $F \in \mathrm{L}^2(\mathcal{B})$ with $\mathcal{F}F = y$.

Proof. Due to algorithm 4.1 (FMP), (4.12) and theorem 4.4, we obtain

$$\sum_{k=N}^{\infty} \alpha_k^2 \leqslant \frac{1}{C^2} \sum_{k=N}^{\infty} \left(\|R^{k-1}\|^2_{\mathbb{R}^l} - \|R^k\|^2_{\mathbb{R}^l}\right) \tag{4.23a}$$

$$= \frac{1}{C^2} \left(\|R^{N-1}\|^2_{\mathbb{R}^l} - \lim_{k\to\infty} \|R^k\|^2_{\mathbb{R}^l}\right) \tag{4.23b}$$

$$= \frac{1}{C^2} \|R^{N-1}\|^2_{\mathbb{R}^l} \overset{N\to\infty}{\longrightarrow} 0. \tag{4.23c}$$

Hence, the conditions on the dictionary imply that

$$\lim_{N\to\infty} \|F - F_{N-1}\|_{\mathrm{L}^2(\mathcal{B})} = \lim_{N\to\infty} \left\| \sum_{k=N}^{\infty} \alpha_k d_k \right\|_{\mathrm{L}^2(\mathcal{B})} = 0. \tag{4.24}$$

Since $\mathcal{F}$ is continuous, we finally obtain $\mathcal{F}F = \lim_{n\to\infty} \mathcal{F}F_n = y$ from theorem 4.4. □

Note that condition (1) means that all possible combinations of real data $y_1, \ldots, y_l$ can be achieved by taking appropriate linear combinations of dictionary elements. In other words, for given $y \in \mathbb{R}^l$, an algorithm could (theoretically) find a finite linear combination $F_n = \sum_{k=1}^{n} \alpha_k d_k$, where $n = l$ is possible, such that $\mathcal{F}F_n = y$.

Lastly, the norm of the residuals decays exponentially as shown in [47]. The decay of the norm of the residuals depends on the correlation between the residual and the dictionary elements. We denote the correlation ratio of a vector $v \in \mathbb{R}^l \setminus \{0\}$ with respect to the dictionary $\mathcal{D}$ with

$$\tau(v) = \sup_{\substack{d \in \mathcal{D} \\ \mathcal{F}d \neq 0}} \frac{|\langle v, \mathcal{F}d \rangle_{\mathbb{R}^l}|}{\|v\|_{\mathbb{R}^l} \|\mathcal{F}d\|_{\mathbb{R}^l}}. \tag{4.25}$$

Because of the Cauchy–Schwarz inequality, $\tau(v)$ cannot exceed 1. The maximal value 1 is obtained, for instance, if there exists a sequence $(d_k)_k \subset \mathcal{D}$ such that $\mathcal{F}d_k$ converges to a vector in $\mathbb{R}^l$ which is collinear with respect to v. Thus, $\tau(v)$ describes how well the data v can be matched by the dictionary $\mathcal{D}$. Remember the criterion for selecting the optimal dictionary element in algorithm 4.1 (FMP). Clearly, the dictionary element that yields the supremum in the definition of $\tau(R^n)$ is chosen.

Furthermore, to quantify the 'worst case', we denote the infimum of the correlation ratio with

$$I(\tau) = \inf_{v \in \mathbb{R}^l \setminus \{0\}} \tau(v). \tag{4.26}$$

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

Theorem 4.6 (Exponential decay of $\|R^n\|_{\mathbb{R}^l}$)**.** *Let $F \in \mathrm{L}^2(\mathcal{B})$ with $\mathcal{F}F = y \in \mathbb{R}^l$, where $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ is a linear operator. Furthermore, let all dictionary elements and the corresponding coefficients be chosen according to algorithm 4.1 (FMP). Then, for all $m \in \mathbb{N}$,*

$$\|R^m\|_{\mathbb{R}^l} \leqslant \|y\|_{\mathbb{R}^l} [1 - I^2(\tau)]^{m/2}. \tag{4.27}$$

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

Proof. The theorem can be proven in analogy to [47]. □

Note that the decay rate decreases if the correlation ratio $\tau(v)$ decreases. Moreover, if the data includes coherent structures with respect to $\mathcal{F}d$, $d \in \mathcal{D}$, then $\tau(R^n)$ is large for all n, i.e. we get a faster decay of the norm of the residuals. Thus, it is an important step in the preprocessing to choose a well-matched dictionary with respect to the structure of the solution. If we have some idea about the structure of the target function we may impose this information on the choice of the dictionary. Otherwise, we recommend to use a dictionary with more general functions of different kinds to get a faster convergence of the algorithm.

5. Regularized functional matching pursuit

The ill-posedness of the inverse gravimetric problem requires the use of a regularization technique. In this work, we use a Tikhonov regularization, i.e. we try to achieve a trade-off between fitting the data and reducing the norm of the solution. The regularization parameter λ correlates both terms.

In section 5.1, we introduce a regularized version of algorithm 4.1 (FMP) where the penalty term is concerned with the smoothness of the solution, i.e. its $\mathrm{L}^2(\mathcal{B})$-norm. Although

the use of, e.g., the $\mathrm{L}^1(\mathcal{B})$-norm is closer to the usual setting in sparse regularization, our approach is due to the Hilbert space structure of $\mathrm{L}^2(\mathcal{B})$ advantageous in the implementation and yields, in contrast to the L^1-version, an analytical expression for the coefficient α.

In section 5.2, we address the main requirements for regularization methods, i.e. we give results about the existence and stability of the solution as well as the convergence of the regularized solution in the limit for the regularization parameter for both exact and noisy given data.

5.1. The algorithm

Using a Tikhonov regularization, we need to find $(d_{n+1}, \alpha_{n+1}) \in \mathcal{D} \times \mathbb{R}$ such that they minimize

$$\|R^n - \alpha \mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|F_n + \alpha d\|_{\mathrm{L}^2(\mathcal{B})}^2 \tag{5.1}$$

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

The minimizing α fulfils

$$0 = \frac{\partial}{\partial \alpha} \left(\|R^n - \alpha \mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|F_n + \alpha d\|_{\mathrm{L}^2(\mathcal{B})}^2 \right) \tag{5.2a}$$

$$= -2 \langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} + 2\alpha \|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \left(2 \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})} + 2\alpha \|d\|_{\mathrm{L}^2(\mathcal{B})}^2 \right) \tag{5.2b}$$

and, thus,

$$\alpha = \frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2}. \tag{5.3}$$

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

$$\begin{aligned} &\|R^n - \alpha \mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|F_n + \alpha d\|_{\mathrm{L}^2(\mathcal{B})}^2 \\ &\quad = \|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2 + 2\alpha \left(-\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} + \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})} \right) \\ &\qquad + \alpha^2 \left(\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2 \right) \end{aligned} \tag{5.4a}$$

$$\begin{aligned} &\quad = \|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2 \\ &\qquad + 2 \frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2} \left(-\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} + \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})} \right) \\ &\qquad + \left(\frac{\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2} \right)^2 \left(\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2 \right) \end{aligned} \tag{5.4b}$$

$$= \|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2 - \frac{\left(\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})} \right)^2}{\|\mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathcal{B})}^2}. \tag{5.4c}$$

And, consequently, a dictionary element d minimizes $\|R^n - \alpha \mathcal{F} d\|_{\mathbb{R}^l}^2 + \lambda \|F_n + \alpha d\|_{\mathrm{L}^2(\mathcal{B})}^2$ if and only if it maximizes

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

Thus, we get the following algorithm for the regularized version of algorithm 4.1 (FMP).

Algorithm 5.1 (RFMP)**.**

Start with $F_0 = 0$ *and* $R^0 = y$.

Given F_n.

Build $F_{n+1} = F_n + \alpha_{n+1} d_{n+1}$ *such that*

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

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

Set $R^{n+1} = R^n - \alpha_{n+1}\mathcal{F}d_{n+1}$.

Here, too, we may improve the approximation quality of the method at the expense of the computation time by introducing back-projection or pre-fitting. Furthermore, we can use the same ideas on preprocessing and parallelization to reduce the computational costs as in the unregularized version (see the end of section 4.1).

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

5.2. Theoretical results

Let, in this section, $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \supset D(\mathcal{F}) \to \mathbb{R}^l$ be an arbitrary operator and let $D(\mathcal{F}) \subset \overline{\text{span } \mathcal{D}}$, where $D(\mathcal{F})$ denotes the domain of $\mathcal{F}$.

We give results on all three important topics when dealing with regularization methods—the existence and the stability of the solution as well as the convergence with respect to the regularization parameter. The results follow the ideas of [29, 30, 62, 65].

Let us denote the expression that is to be minimized by

$$J_\lambda(F, y) := \|\mathcal{F}F - y\|^2_{\mathbb{R}^l} + \lambda \|F\|^2_{\mathrm{L}^2(\mathcal{B})}, \tag{5.5}$$

where $\lambda > 0$ is the regularization parameter. Here, F is a series expansion in dictionary elements, i.e. $F = \sum_{k=1}^{\infty} \alpha_k d_k$. Of course, regularization techniques are mostly used to deal with data from measurements, i.e. noisy data y^ε, where $\|y - y^\varepsilon\|_{\mathbb{R}^l} \leqslant \varepsilon$, if the data $y \in \mathbb{R}^l$ is given exactly.

We first give a result concerned with the existence of a solution of the regularized optimization problem, where noisy data is considered.

Theorem 5.2 (Existence of a solution)**.** *Let the operator* $\mathcal{F}$ *be weakly sequentially closed. Furthermore, let the noisy data* $y^\varepsilon \in \mathbb{R}^l$ *be given. Then, there always exists a solution* $F^{\lambda,\varepsilon} \in D(\mathcal{F})$ *such that*

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

Proof. See, e.g., [62] and the references therein for the proof. □

We know that a linear and continuous operator that maps into a space with finite dimensions is compact, i.e. the operator considered in this paper is compact as well. Since every linear and compact operator is strongly continuous (confer, e.g., [77]), we conclude that the operator corresponding to our application is strongly continuous, too, and therefore weakly sequentially closed. Thus, it fulfils the condition imposed in theorem 5.2. Note that this is the only serious restriction occurring in our theoretical considerations.

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

$$J_\lambda(F^\lambda, y) = \min_{F \in D(\mathcal{F})} J_\lambda(F, y). \tag{5.6}$$

Let us consider the sequence

$$\left(\|\mathcal{F}F_n - y\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2\right)_n = \left(\|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2\right)_n \tag{5.7}$$

and prove its convergence. Note that

$$R^n = R^{n+1} + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 + \lambda \|d_{n+1}\|_{\mathrm{L}^2(\mathcal{B})}^2} \mathcal{F}d_{n+1} \tag{5.8}$$

and

$$F_{n+1} = F_n + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 + \lambda \|d_{n+1}\|_{\mathrm{L}^2(\mathcal{B})}^2} d_{n+1}. \tag{5.9}$$

Theorem 5.3. *Let $\mathcal{F}$ be a linear and continuous operator and let all dictionary elements and the corresponding coefficients be chosen according to algorithm 5.1 (RFMP). Then $(\|R^n\|_{\mathbb{R}^l}^2 + \lambda\|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2)_n$, where the residual R^n is defined as $R^n := y^\varepsilon - \mathcal{F}F_n$ and F_n is given in (5.9), converges for n tending to infinity.*

Proof. Let us consider the sequence element F_{n+1}. With (5.4*c*) we obtain

$$\begin{aligned}
&\|R^{n+1}\|_{\mathbb{R}^l}^2 + \lambda\|F_{n+1}\|_{\mathrm{L}^2(\mathcal{B})}^2 \\
&\quad = \|R^n\|_{\mathbb{R}^l}^2 + \lambda\|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2 - \frac{\left(\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(\mathcal{B})}\right)^2}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 + \lambda\|d_{n+1}\|_{\mathrm{L}^2(\mathcal{B})}^2} && (5.10a) \\
&\quad \leqslant \|R^n\|_{\mathbb{R}^l}^2 + \lambda\|F_n\|_{\mathrm{L}^2(\mathcal{B})}^2. && (5.10b)
\end{aligned}$$

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

Theorem 5.4 (Stability of the regularized solution)**.** *Let the operator $\mathcal{F}$ be weakly sequentially closed and let $y^\varepsilon \in \mathbb{R}^l$ be given noisy data. Furthermore, let $(y^{\varepsilon_k})_k \subset \mathbb{R}^l$ be a sequence that converges to y^ε for k tending to infinity. Let $(F^{\lambda,\varepsilon_k})_k$ be a corresponding sequence of minimizing elements of $J_\lambda(\cdot, y^{\varepsilon_k})$. Then there exists at least one convergent subsequence of $(F^{\lambda,\varepsilon_k})_k$ and every convergent subsequence converges to a minimizing element $F^{\lambda,\varepsilon}$ of $J_\lambda(\cdot, y^\varepsilon)$.*

Moreover, if $F^{\lambda,\varepsilon}$ is unique then $(F^{\lambda,\varepsilon_k})_k$ converges to $F^{\lambda,\varepsilon}$ for k tending to infinity.

Proof. See, e.g., [62] and the references therein for the proof. □

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

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

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

$$J_{\lambda(\varepsilon)}(F^{\lambda(\varepsilon),\varepsilon,\eta}, y^\varepsilon) \leqslant \min_{F \in D(\mathcal{F})} J_{\lambda(\varepsilon)}(F, y^\varepsilon) + \eta. \tag{5.11}$$

Note that the regularization parameter λ was replaced by the function λ that depends on the data error ε. We use $\eta = \eta(\lambda(\varepsilon))$ to describe the error that arises when minimizing numerically. With theorem 5.2, we already know that there exists a solution $F^{\lambda(\varepsilon),\varepsilon,\eta}$ of the regularized problem with noisy data input.

Theorem 5.5 (Convergence of the regularization)**.** *Let the operator $\mathcal{F}$ be weakly sequentially closed and let y in the range of $\mathcal{F}$ be exactly given data. Furthermore, let F^+ be a minimum-norm solution of $\mathcal{F}F = y$. Let the family $(y^\varepsilon)_{\varepsilon>0}$ fulfil $\|y - y^\varepsilon\|_{\mathbb{R}^l} \leqslant \varepsilon$. Let us choose $\lambda :]0, \infty[\to]0, \infty[$ such that $\lambda(\varepsilon)$ and $\frac{\varepsilon^2}{\lambda(\varepsilon)}$ tend to 0 for ε tending to 0 from above. Let the error satisfy $\eta(\lambda(\varepsilon)) = o(\lambda(\varepsilon))$ for ε tending to 0 from above, where o is a Landau symbol.*

Then the family $(\mathcal{R}_{\lambda(\varepsilon)}^{\eta(\lambda(\varepsilon))}(y^\varepsilon))_{\varepsilon>0}$ has at least one convergent subsequence (for ε tending to 0 from above) and the limit of each of these subsequences is a minimum-norm solution of $\mathcal{F}F = y$.

If the minimum-norm solution F^+ is additionally unique, then the regularization converges to F^+ in its norm for ε tending to 0 from above:

$$\lim_{\varepsilon \to 0^+} \left\| \mathcal{R}_{\lambda(\varepsilon)}^{\eta(\lambda(\varepsilon))}(y^\varepsilon) - F^+ \right\|_{\mathrm{L}^2(\mathcal{B})} = 0. \tag{5.12}$$

Proof. See, e.g., [62] and the references therein for the proof. □

Note that there is a link between algorithm 5.1 (RFMP) presented here and algorithm 4.1 (FMP) as well as the commonly known matching pursuit discussed in [47]. If one chooses $Y := \mathbb{R}^l \times \mathrm{L}^2(\mathcal{B})$ as the (artificial) data space and sets $\langle (x, F), (y, G) \rangle_Y := \langle x, y \rangle_{\mathbb{R}^l} + \lambda \langle F, G \rangle_{\mathrm{L}^2(\mathcal{B})}$, the problem of solving the (unsolvable) equation $\mathcal{G}F = (y, 0)$ with $\mathcal{G}F = (\mathcal{F}F, F)$, $F \in \mathrm{L}^2(\mathcal{B})$, leads to an algorithm which is analogous to algorithm 5.1 (RFMP). However, e.g., the convergence results in theorem 4.4 can nevertheless not be transferred to algorithm 5.1 (RFMP), since the proof requires the existence of a solution (which means here a solution of $\mathcal{G}F = (y, 0)$ and not of $\mathcal{F}F = y$).

6. Numerical applications

In times of sea-level rise, rapid mass loss in the polar regions, and other climate changes of global impact, the processing of data that is collected by the various satellite missions rapidly gains importance. Satellite missions allow a global or supra-regional overview over events that may not be distinguishable from a local point of view or even from the surface of the Earth itself.

In section 6.1, we recover the mass density variation of South America out of the Earth Gravitational Model 2008 (EGM2008, [59]) developed by the National Geospatial Intelligence Agency (NGA). The model is given in spherical harmonics coefficients up to degree 2190 and order 2159. Thus, right now, it is the gravity model with the highest resolution available. In our applications, we use the EGM2008 model from degree 3 up to degree 2190.

In section 6.2, we examine the mass transport in the Amazon area for the year 2008. The satellite mission Gravity Recovery and Climate Experiment (GRACE) has provided us with monthly solutions of the gravitational potential since the mid of 2003, again in the form of spherical harmonics coefficients. These solutions are preprocessed and published by different research groups. We use the solutions provided by the Jet Propulsion Laboratory (JPL, see [42]). Since GRACE solutions provide a monthly global coverage of the gravitational potential they are predestined to be used to reflect temporal changes caused by large changes of ground water levels or the deglaciation.

6.1. Reconstructing the mass density distribution of the Earth

Since it is well known that the harmonicity constraint in particular and gravitational data in general are only appropriate for the determination of mass anomalies in the uppermost layer

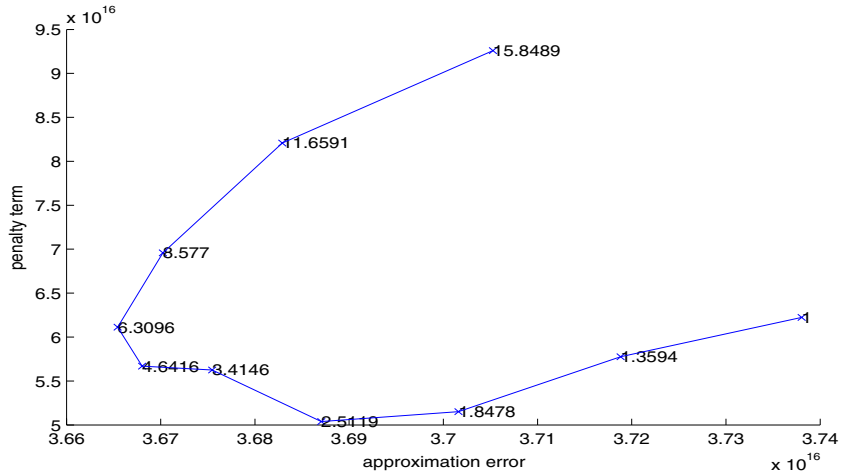

Figure 3. Adapted L-curve where the approximation error $\|R^{100}\|^2_{\mathbb{R}^l}$ is plotted for different regularization parameters λ (given at the corresponding point) against the penalty term $\|F_{100}\|^2_{\mathrm{L}^2(\mathcal{B})}$.

of the Earth (see [53]), we only reconstruct the density close to the surface. Here, a case study for South America is presented.

The solution of algorithm 5.1 (RFMP)

In this section, we use the dictionary

$$\begin{aligned}\mathcal{D} = &\left\{ K_h^{\mathrm{I}}(\cdot, x) \;\middle|\; h \in \{0.95, 0.97, 0.99\}, x \in \mathrm{grid}(\mathcal{B}) \right\} \\ &\cup \left\{ G^{\mathrm{I}}_{0,n,j} \;\middle|\; n = 3, \ldots, 8, j = 1, \ldots, 2n+1 \right\}, \end{aligned} \tag{6.1}$$

where $\mathrm{grid}(\mathcal{B})$ is a nearly quadratic grid which is equiangular each in longitude and latitude. After restriction to a spherical rectangle covering South America we are left with 39 800 grid points. Furthermore, we will stop the summation in the kernel functions at degree 2190 in accordance with the degree of the spherical harmonics coefficients used to compute the data. This dictionary suits our needs, since we may reconstruct global trends with the basis elements $G^{\mathrm{I}}_{0,n,j}$ while the localized kernel functions $K_h^{\mathrm{I}}(\cdot, x)$ are a very good choice to recover detail structures of the target function.

If not denoted otherwise, the data will be given at 25 440 points on an equiangular grid, which is similar to $\mathrm{grid}(\mathcal{B})$ and is located slightly above the Earth's surface at 7 km height. The method will be stopped after 20 000 iterations, i.e. 20 000 (not necessarily pairwise distinct) dictionary functions will be chosen to reconstruct the solution.

Since the inverse gravimetric problem is ill-posed, we use the regularized algorithm 5.1 (RFMP) and need to choose an appropriate regularization parameter λ. However, our computations are too time consuming to use common methods to determine λ. Thus, we adapt the well-known L-curve method to our purpose. After 100 steps, the weighted approximation error $\|R^{100}\|^2_{\mathbb{R}^l}/\|y\|^2_{\mathbb{R}^l}$ is already very small. Thus, we compute an L-curve for different regularization parameters after 100 steps instead of at the end of the computations. This strategy, certainly, has to be regarded critically, since it is far from optimal. However, later considerations with respect to the regularization parameter (see figure 9) show that our choice yields an acceptable balance between noise and smoothing.

We compute the adapted L-curve for parameters $\lambda_i = 10^{-2+4i/30}$, $i = 0, \ldots, 30$, i.e. $\lambda = 0.01, \ldots, 100$ (see figure 3, zoomed in on the knickpoint). This does not yield a proper

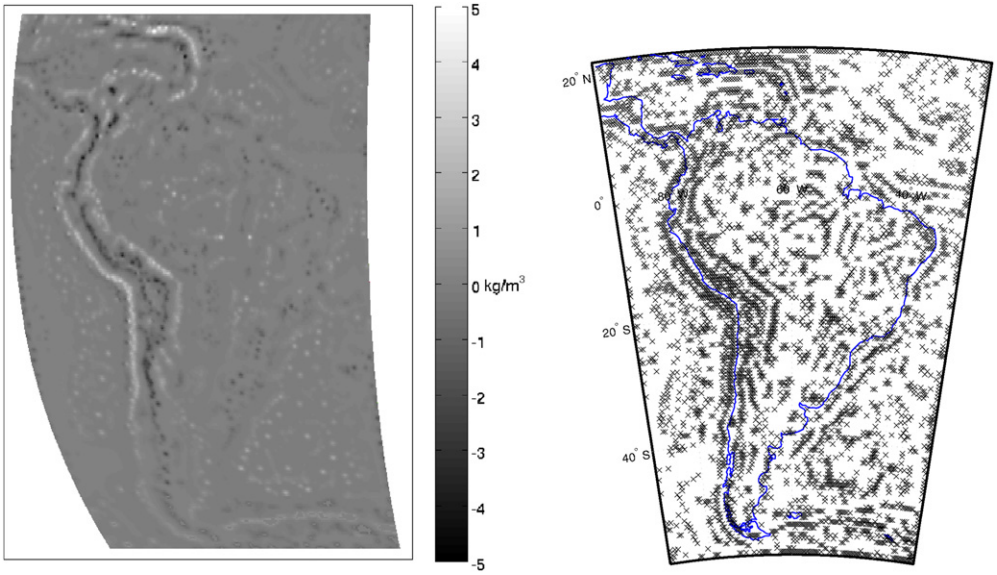

Figure 4. Reconstructed density deviation (left) and centre points x of the chosen expansion functions $K_h^1(\cdot, x)$ (right) computed out of 25 440 data points with expansion functions chosen from dictionary (6.1) in 20 000 iterations (i.e. $F_{20\,000}$ is shown), $\lambda = 4.6416$.

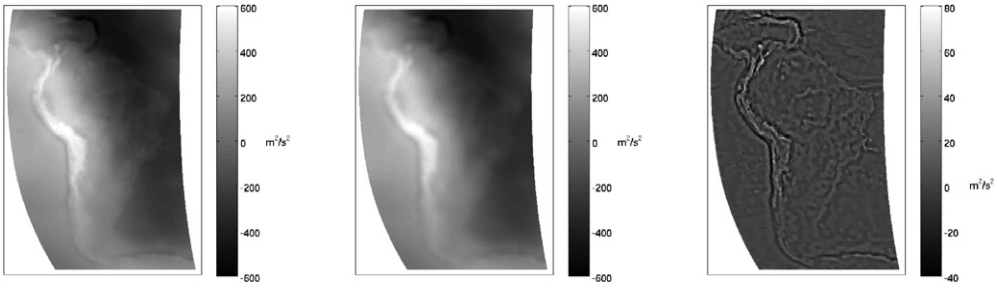

Figure 5. Data input EGM2008 (left), potential generated by the solution $F_{20\,000}$ for $\lambda = 4.6416$ (middle) and difference of both (right) for 25 440 data points.

'L'. However, there is an easy to identify knickpoint, which corresponds to $\lambda = 4.6416$. We, consequently, choose this value as our regularization parameter.

Using this regularization parameter, we obtain the solution $F_{20\,000}$ displayed on the left-hand side of figure 4. We clearly see the outline of the continent and the main topographic structures, i.e. the Andes and the Caribbean. On the right-hand side of figure 4, we display the centre points x of the chosen kernel functions $K_h^1(\cdot, x)$. Note that we artificially include the coast lines of South America in blue as an orientation. Obviously, the localized dictionary functions are chosen mainly in areas where the detail structure needs to be more accurate, i.e. in the Andes or the Caribbean. Thus, algorithm 5.1 (RFMP) clearly recovers a solution that is adapted to the structure of the target function.

In figure 5, we display the potential given by the model EGM2008 at 25 440 data points (left), the approximation of the potential by the solution $F_{20\,000}$ (i.e. $\mathcal{F}F_{20\,000}$) at the same points (middle) and the difference between both (i.e. the residual $R^{20\,000}$). Clearly, the main structures

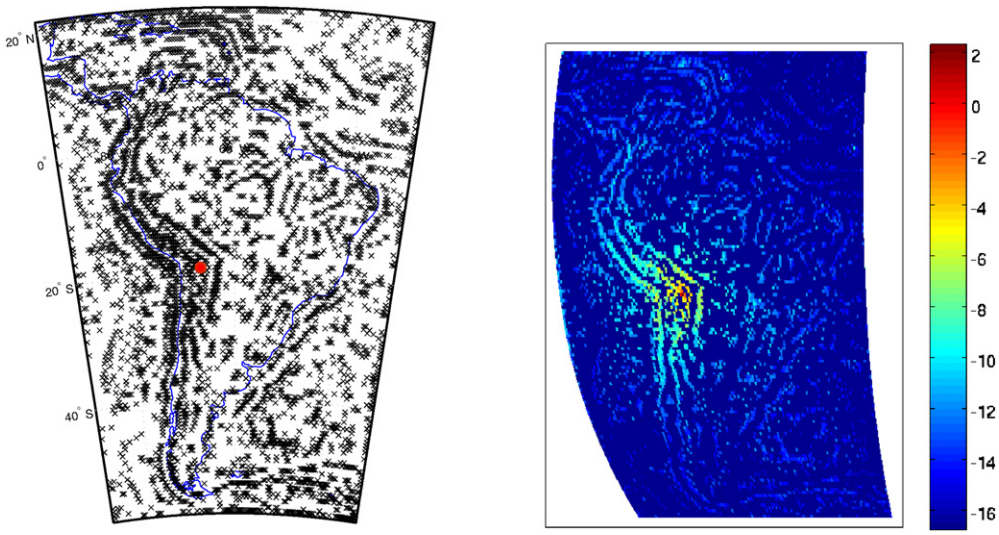

Figure 6. Influence of the expansion functions of the solution $F_{20\,000}$ (right) on the point $x = (0.3784, -0.8661, -0.3267)^{\mathrm{T}}a$ (red dot) located in the Andes (left).

of the potential are approximated well enough while some of the detail structures remain as an error. However, the values of the error are mostly far below 10% of the original data.

Now imagine a data grid that is as large as the one chosen in this application, i.e. 25 440 data points. A spline method would not be able to handle this data grid, since the corresponding (ill-conditioned and dense) matrix is much too large to be handled in the computations. In spline methods, the chosen data points and the centre points x of the basis functions $K_h^{\mathrm{I}}(\cdot, x)$ used to expand the solution are directly connected. Thus, it is not possible to obtain a resolution as high as in the Andes with a spline method. We will study a comparison of the novel method to previous approaches in further detail later in this paper.

The localized character of the solution

In figure 6, we examine the influence of the solution $F_{20\,000}$ as displayed in figure 4, i.e. the influence of the expansion elements, on exactly one point

$$x = (0.3784, -0.8661, -0.3267)^{\mathrm{T}}a \tag{6.2}$$

situated in the Andes. We mark this point in the left plot of figure 6 with a red dot. On the right-hand side of figure 6 we display the influence of the expansion functions of the solution $F_{20\,000}$ on the value at the point x with a logarithmic colourbar, i.e. we display $\log|\alpha_k d_k(x)|$ for all expansion elements.

For a localized expansion function, we display $\log|\alpha_k K_{h_k}^{\mathrm{I}}(x, x_k)|$ as a value at the point x_k. The influence of an expansion function $G_{0,n_k,j_k}^{\mathrm{I}}$ is included as an additive at all points in the plot. Since we use a logarithmic colourbar, the blue colour denotes a very small impact while the red colour denotes a large impact on the value at the point x. Obviously, the value at x is mostly controlled by localized expansion functions. Furthermore, we clearly observe that the influence of the localized expansion functions increases when the distance between x and the centre x_k decreases. This behaviour was expected.

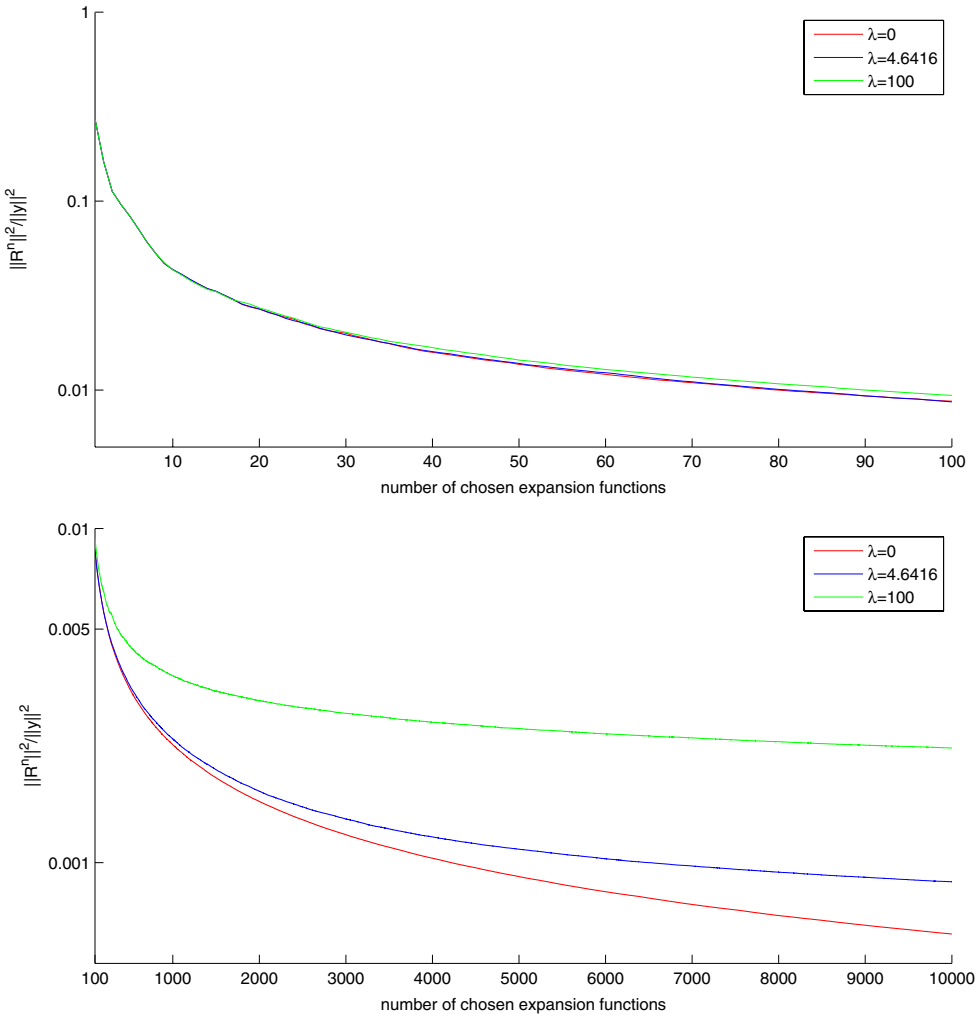

Figure 7. Evolution of the weighted approximation error when reconstructing the density deviation out of 25 440 data points for $n = 1, \ldots, 100$ (top) and $n = 100, \ldots, 10\,000$ (bottom) in the case of the solution computed with algorithm 4.1 (FMP), i.e. with regularization parameter $\lambda = 0$ (red), and the solutions computed with algorithm 5.1 (RFMP) with regularization parameters $\lambda = 4.6416$ (blue) and $\lambda = 100$ (green). Note that the evolution of all three is very similar in the upper plot. The lines are hardly distinguishable.

The iterative character of the method

We require the method to be adaptive and iterative. Thus, we may expect that the solution improves when we choose more dictionary functions for the expansion of the solution. First of all, let us consider the trend of the weighted approximation error:

$$\frac{\|R^n\|^2_{\mathbb{R}^l}}{\|R^0\|^2_{\mathbb{R}^l}} = \frac{\|R^n\|^2_{\mathbb{R}^l}}{\|y\|^2_{\mathbb{R}^l}}. \tag{6.3}$$

In figure 7, we display the evolution of the residual for the first 100 steps as well as the development from step 100 to 10 000 for $\lambda = 4.6416$ in a logarithmic scale for the weighted approximation error.

Clearly, the residual decreases rapidly in the beginning. After choosing only 100 expansion functions, the weighted approximation error is already reduced to less than 1% of the initial error. Furthermore, the weighted approximation error can be reduced to less than 0.5% after about 500 steps and after approximately 10 000 steps we obtain a weighted approximation error of 0.09% for our choice of the regularization parameter $\lambda = 4.6416$. Let us remark that the approximation error is only reduced slightly anymore in the steps from 10 000 to 20 000 iterations. These results are not unexpected. Remember the regularization functional $\|R^n\|^2_{\mathbb{R}^l} + \lambda \|F_n\|^2_{\mathrm{L}^2(\mathcal{B})}$ that is to be minimized. We may expect that, for large n, the penalty term becomes more important in comparison to the residual. Thus, its influence in the choice of the next dictionary function increases as well. We expect the algorithm to concentrate on the reduction of the approximation error in the beginning and shift its focus to the penalty term, i.e. the smoothness of the solution, as the iteration proceeds (depending on the choice of the regularization parameter).

Let us consider the solution after 3000 and 10 000 steps in comparison to the solution after 20 000 steps (see the left-hand column of figure 8). Clearly, the solution improves if we choose 10 000 dictionary functions instead of only 3000. At first sight, we do not see the same kind of improvement if we increase the number of chosen expansion functions from 10 000 to 20 000. However, this is not surprising, since the main structures were obviously already reconstructed after choosing 10 000 expansion functions. The additional 10 000 expansion functions in $F_{20\,000}$ refine the resolution of the details as can be observed, e.g., at the western boundary of the Andes where the edges are not as rough anymore. Note that these observations match remarkably well with our findings from the approximation error. Considering the centre points x of the chosen kernel functions $K^{\mathrm{I}}_h(\cdot, x)$ again for F_{3000}, $F_{10\,000}$ and $F_{20\,000}$, as displayed in the right column of figure 8, reinforces these findings.

The influence of the regularization parameter λ

In figure 7, we display the weighted residuals for the unregularized solution ($\lambda = 0$ in red) as well as for the regularized solutions for $\lambda = 4.6416$ (in blue) and $\lambda = 100$ (in green). In later steps, the effects of the choice of the regularization parameter are more obvious. Choosing the regularization parameter $\lambda = 0$, i.e. a reconstruction with no regard for the smoothness of the solution, provides us, as expected, with the smallest weighted approximation error of about 0.06% after 10 000 iterations. With increasing regularization parameter the importance of the penalty term which forces the smoothness of the solution increases as well, and, thus, the focus shifts from finding the solution with the lowest residual to finding a solution that is smooth, too. As a consequence, the approximation error of the smoothest solution ($\lambda = 100$ in green and see figure 9) is with a value of about 0.2% the largest one.

In figure 9, we compare the solution of the unregularized method (algorithm 4.1 (FMP)) with the solution of the regularized method (algorithm 5.1 (RFMP)) for our regularization parameter $\lambda = 4.6416$ and the larger parameter $\lambda = 100$. In each case, we reconstruct the density deviation out of 25 440 data points by selecting 10 000 expansion functions out of the dictionary (6.1). The left column of figure 9 represents the solutions with the original colourbar while the colourbars of the same solutions in the right column are adapted for comparison to the range of the solution regularized with $\lambda = 4.6416$.

Regarding the plots with adapted colourbar, we clearly see the characteristics of a regularization. With increasing regularization parameter, the influence of the penalty term increases as well. In our case, the penalty term is concerned with the smoothness of the solution, i.e. the solution becomes smoother when we increase the regularization parameter.

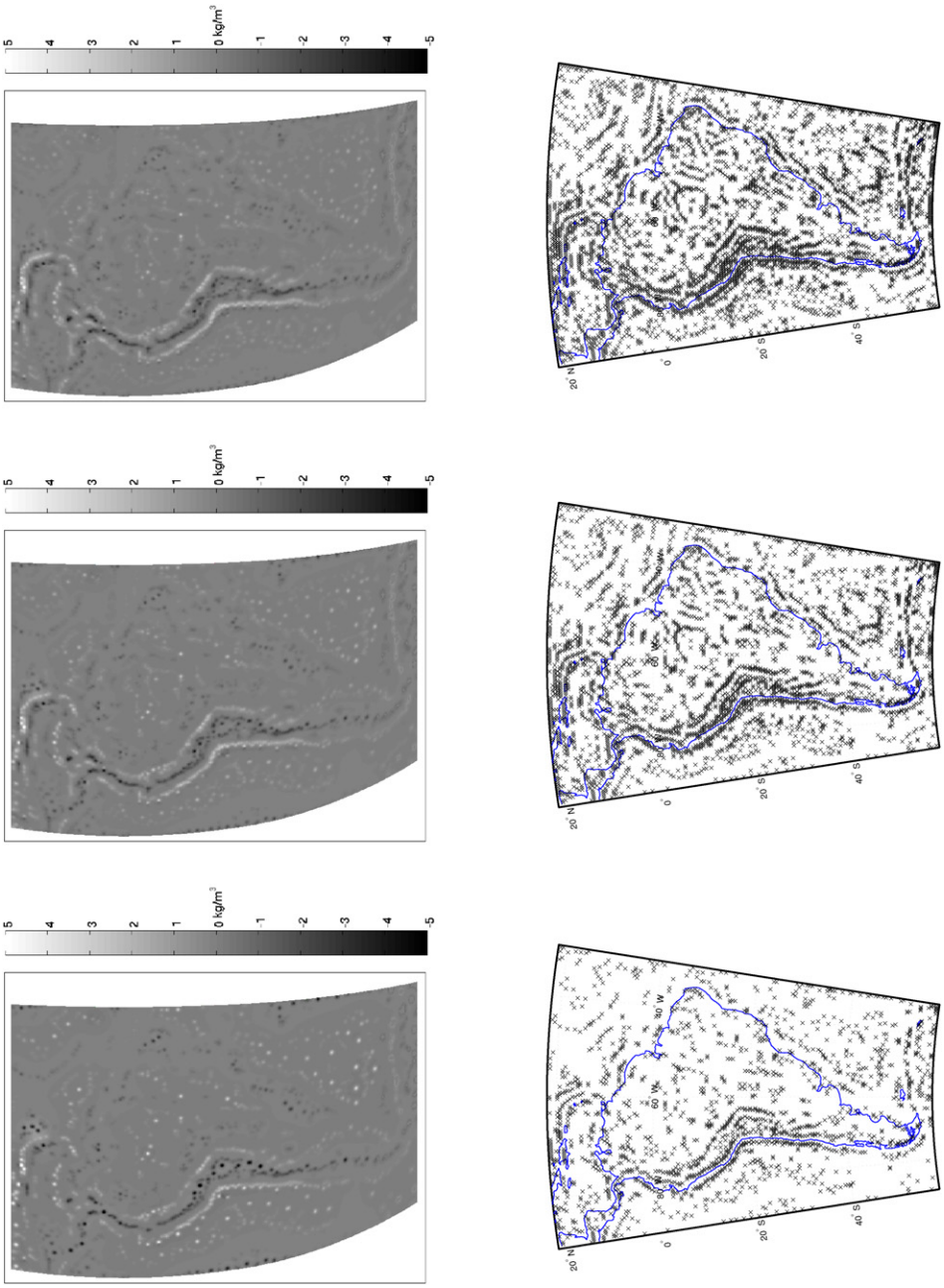

Figure 8. Reconstructed density deviation (left column) and centre points x of the chosen dictionary functions $K_h^1(\cdot, x)$ (right column) computed out of 25 440 data points in the case of 3000 (bottom), 10 000 (middle) and 20 000 (top) iterations, $\lambda = 4.6416$.

If we consider the range of the recovered values for the density deviation we observe another trait of a regularization. Smoothening oftentimes manifests as a change in magnitude of the solution when we have only comparably small variations in the surrounding regions.

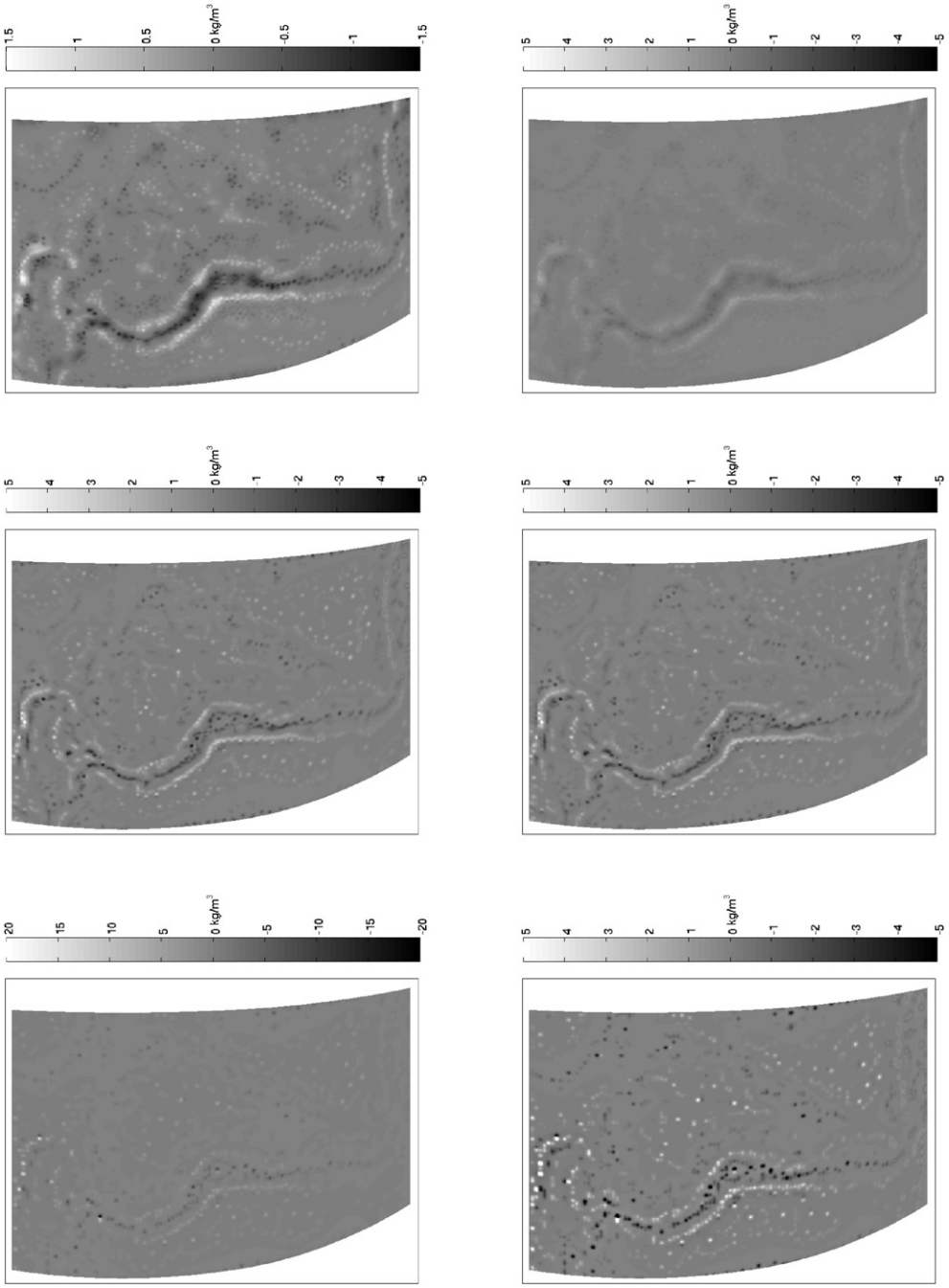

Figure 9. Reconstructed density deviation computed out of 25 440 data points with expansion functions chosen from dictionary (6.1) in 10 000 iterations (i.e. $F_{10\,000}$ is shown) for $\lambda = 0$ (bottom), 4.6416 (middle) and 100 (top) with original (left) and adapted (right) colourbar.

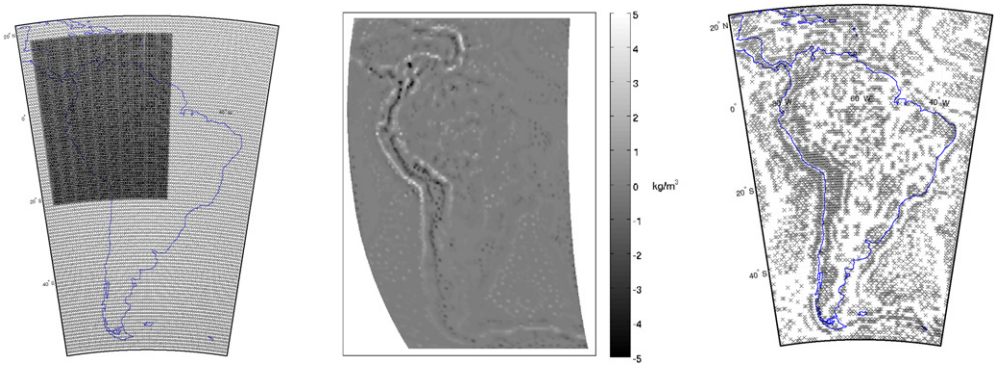

Figure 10. Irregular data grid with 19 800 points (left), reconstructed density deviation $F_{20\,000}$ (middle) and centre points (right) of the chosen localized expansion functions computed out of these data points for $\lambda = 10$, where algorithm 5.1 (RFMP) was stopped after 20 000 iterations.

Anyway, all these considerations demonstrate that algorithm 5.1 (RFMP) is, indeed, a regularization method as was proven in section 5.2.

Reconstruction on an irregular data grid

Let us consider now a data grid that is not equidistributed, i.e. a grid as in the left plot of figure 10. After 20 000 iterations, algorithm 5.1 (RFMP) generates the solution and the centre points as displayed in the middle and right plot of figure 10, respectively.

Although the higher density of data points in the north-west of South America influences the choices of the algorithm, the solution still mostly fits the detail structures of the target function. Clearly, the novel method is as independent of the data grid as can be expected.

Refinement

The iterative character of our method allows us to re-use results from previous computations to, e.g., zoom in on certain parts of the solution. We investigate this property by refining Central America and the Caribbean.

In the refinement, we want to improve the approximation in a certain area only. Thus, it is reasonable to consider a dictionary that consists of functions with a local character exclusively, e.g., the dictionary

$$\mathcal{D} = \left\{ K_h^{\mathrm{I}}(\cdot, x) \;\middle|\; h \in \{0.95, 0.97, 0.99\},\ x \in \mathrm{grid}(\mathcal{B}) \right\}, \tag{6.4}$$

where $\mathrm{grid}(\mathcal{B})$ is an equiangular grid, which is similar to the previous one and restricted to a spherical rectangle covering the area of Central America and the Caribbean with 39 800 grid points. Furthermore, we use 25 440 data points on a grid, which is similar to the previous ones and located slightly above the Earth's surface at 7 km height for this same area of the refinement. We choose the regularization parameter $\lambda = 10$.

As a starting point for the iterations we use the approximate solution $F_{3\,000}$ for Central and South America, which we already computed before (see the left-hand side of figure 11 and figure 8, bottom left). In the middle of figure 11, we display the result after the refinement with 3000 additionally chosen (not necessarily pairwise distinct) expansion functions in the North-West. Remember that dictionary (6.4) consists of localized functions only. When comparing the solution without the refinement (left plot) and the solution including the refinement (middle

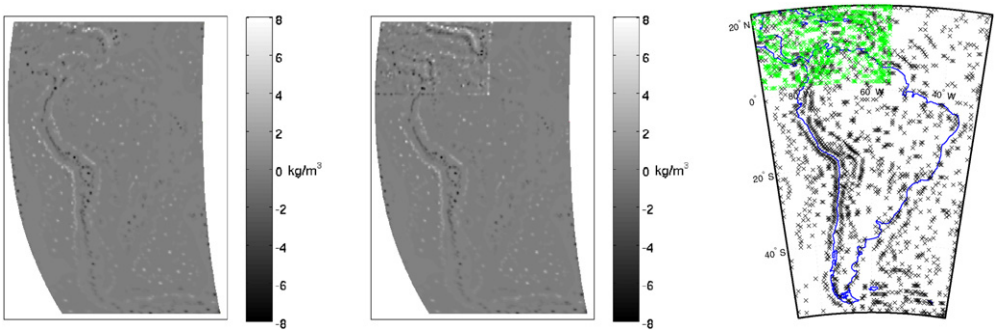

Figure 11. Reconstructed density deviation $F_{3\,000}$ computed out of 25 440 data points, where the expansion functions were selected from dictionary (6.1) for $\lambda = 4.6416$ (left), refined in the North-West for a new set of 25 440 data points located in the North-West and the dictionary (6.4) after additional 3 000 iterations for $\lambda = 10$ (middle), and the centre points x of the chosen localized expansion functions (right), where the centres corresponding to dictionary (6.4) are green.

plot) we clearly see that the resolution of Central America and the Caribbean is by far improved. If we consider the centre points of the chosen expansion functions (right plot), we observe that they are mostly centred in these regions, i.e. Central America and the Caribbean, as well. However, we observe some minor boundary effects. Note that we give some ideas to control these boundary effects in [33].

These results are, of course, expected behaviour as we investigated these properties before. However, it is quite remarkable how well the refinement process works. It allows us to do a coarse investigation of a certain area first. And then, with these results in mind, we may decide where we want to refine this solution. Furthermore, this property allows us to save computational effort when refining.

Dealing with noise

In this section, we want to examine the behaviour of the regularizing algorithm 5.1 (RFMP) when applied to noisy data y^ε. Here ε denotes the noise level where a value $\varepsilon = 0.1$ corresponds to a data input that is disturbed with 10% random noise relative to the exact data y:

$$y_i^\varepsilon = y_i + \varepsilon \, \mathrm{rand}_i \, y_i, \quad i = 1, \ldots, l, \tag{6.5}$$

where rand_i is a random number in the interval $[0, 1]$. In figure 12, we consider the reconstructed density deviations out of 25 440 data points, where dictionary (6.1) was used and algorithm 5.1 (RFMP) with regularization parameter $\lambda = 10$ was stopped after 20 000 iterations.

In comparison to figure 4, where the same setup was considered for exact data and a regularization parameter $\lambda = 4.6416$, we clearly observe that the introduction of noise does not have an overly negative influence on the reconstruction quality of the algorithm, since the main structures are still identified.

Comparison to other methods

Let us compare our new method numerically to previously developed localized methods in inverse gravimetry viz spline and wavelet methods. First we briefly describe the methods used. For a more detailed discussion (in particular, for readers who are not familiar with these techniques), we refer to [8, 9] for the used spline method and to [49, 51, 53] for the used wavelet method.

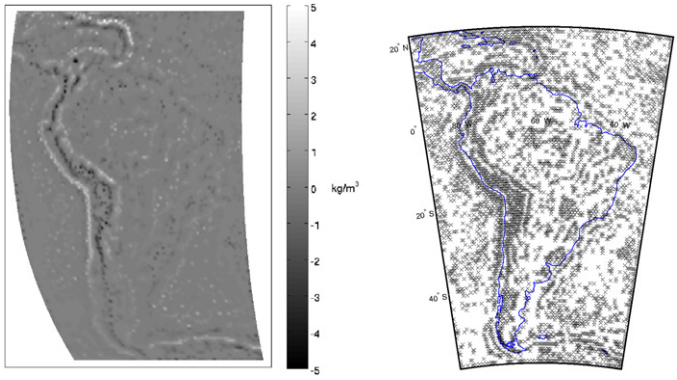

Figure 12. Reconstructed density deviation $F_{20\,000}$ (left) and centre points (right) of the chosen localized expansion functions computed out of 25 440 data points for $\lambda = 10$ out of data with 10% noise ($\varepsilon = 0.1$).

In our setting, a spline $S \in \mathcal{H}$ relative to $\mathcal{F}_{\mathrm{G}}$ is defined as

$$S(x) = \sum_{k=1}^{l} a_k \mathcal{F}_{\mathrm{G}}^k K_h(\cdot, x), \ x \in \mathcal{B}, \ a = (a_1, \ldots, a_l)^{\mathrm{T}} \in \mathbb{R}^l. \tag{6.6}$$

We now have to determine that spline S which fulfills

$$\mathcal{F}_{\mathrm{G}}^i S = \sum_{k=1}^{l} a_k \mathcal{F}_{\mathrm{G}}^i \mathcal{F}_{\mathrm{G}}^k K_h(\cdot, \cdot) = y_i, \ i = 1, \ldots, l, \tag{6.7}$$

i.e. we solve a system of linear equations with the matrix $\left(\mathcal{F}_{\mathrm{G}}^i \mathcal{F}_{\mathrm{G}}^k K_h(\cdot, \cdot)\right)_{i,k=1,\ldots,l}$ which is positive definite. Solving this interpolation problem yields the interpolating function with minimal Sobolev norm. Moreover, in comparison to all other functions of the form (6.6), the spline is the best approximation to the target function. Note that the choice of the kernel function used for the spline is fixed. In this work, we use the kernel $\tilde{K}_h$ with parameter $h = 0.99$. Furthermore, we may regularize the linear equation system of the spline method by adding a constant value to the diagonal elements of the spline matrix, where the constant may depend on the maximal absolute value M of all entries of that matrix. Here we use the regularization parameter $0.1M$.

When using a wavelet method, we compute the integral

$$\int_{\Omega} \Phi_J(\cdot, y) V(y) \, \mathrm{d}y \tag{6.8}$$

numerically, where the kernel function Φ_J is of degree $2^J - 1$. In our case, the cubic polynomial scaling function Φ_J at scale $J = 8$ was chosen (see [53] for its definition and characteristics) and the potential V is of degree 2 190. Thus, to gain the accurate solution, we have two options: Either we use a quadrature formula up to degree 2 445, i.e. when using a Driscoll–Healy grid we would need more than $2\,445^2 = 5978\,025$ data points for accuracy and use the Driscoll–Healy weights (see [26]), or we use a global equidistributed point set which is restricted to South America and weight all points with $4\pi/N$ where N is the number of grid points on the sphere (see, for example, [17]). In our case, the second option is much more attractive, since it is more time-efficient with an acceptably low inaccuracy of the quadrature. Note that there exist more sophisticated quadrature tools on the sphere or subdomains of it (see, e.g., [40]).

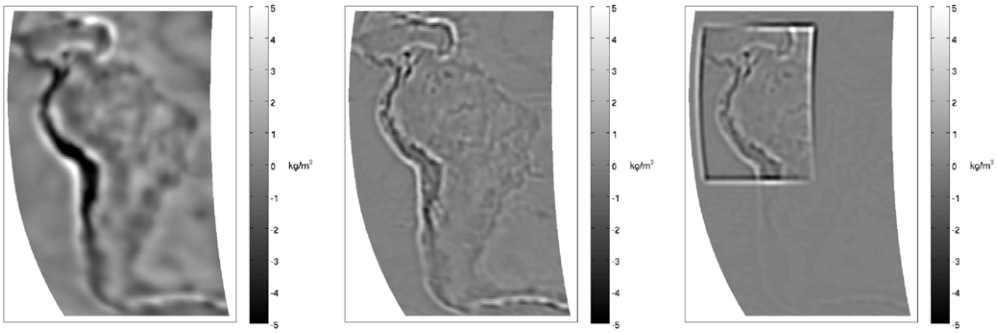

Figure 13. Reconstructed density deviation with the spline method out of 9 900 equidistributed data points regularized with 10% of the absolutely maximal entry of the corresponding matrix (left) and with the wavelet method on a grid of 25 440 data points on an equidistributed grid (middle) and an irregular grid (right).

Note that neither splines nor wavelets (as they were constructed for inverse gravimetry) are able to combine different kernel functions as well as irregular grids to reconstruct a function—in contrast to our novel method.

In figure 13, we compare the best results that may be achieved with the spline and wavelet method, left-hand side and middle respectively, to the result given by our novel method (see figure 4). For the spline method we use a data grid with 9900 data points which is close to the numerical limit for this problem. However, the resolution and detail structure of the solution is obviously worse than with our novel method. The wavelet method applied to the same equidistributed data grid with 25 440 data points as in the case of our method gives a very good result as well. However, it strongly depends on the choice of the data grid as we display in the right-hand side of figure 13, where we use the same irregular data grid as displayed in figure 10. Here we can clearly see artefacts similar to the Gibbs phenomenon at the edges of the denser part of the data grid. This data grid did not pose a problem for algorithm 5.1 (RFMP).

6.2. Reconstructing the mass transport in the Amazon area by using GRACE-Data

The satellite mission Gravity Recovery and Climate Experiment (GRACE) was started in 2002 to gain more information about the Earth's gravitational potential which allows us to detect climate phenomena like water mass transports in the gravitational field (see [12]). The GRACE mission provides us with a monthly global coverage of the gravitational potential such that we are able to investigate temporal variations as well. In this subsection, we will concentrate on detecting the monthly mass transport, i.e. seasonal changes of ground and surface water levels, in the Amazon area for the year 2008. It is very important to observe the mass transport in the Amazon area regularly, since it is one of the largest watersheds on Earth. Thus, it has been observed by other research groups using different techniques and observation periods (see, e.g., [13, 14, 31, 69]).

We use the monthly data provided by the Jet Propulsion Laboratory (JPL, see [42], Release 04). The data, i.e. the spherical harmonics coefficients, are given up to degree and order 120. To analyse the temporal variations, we subtract a mean potential from the monthly solutions and use this difference as an input to our algorithm. We use the available coefficients from July 2004 to June 2009 to compute the mean potential.

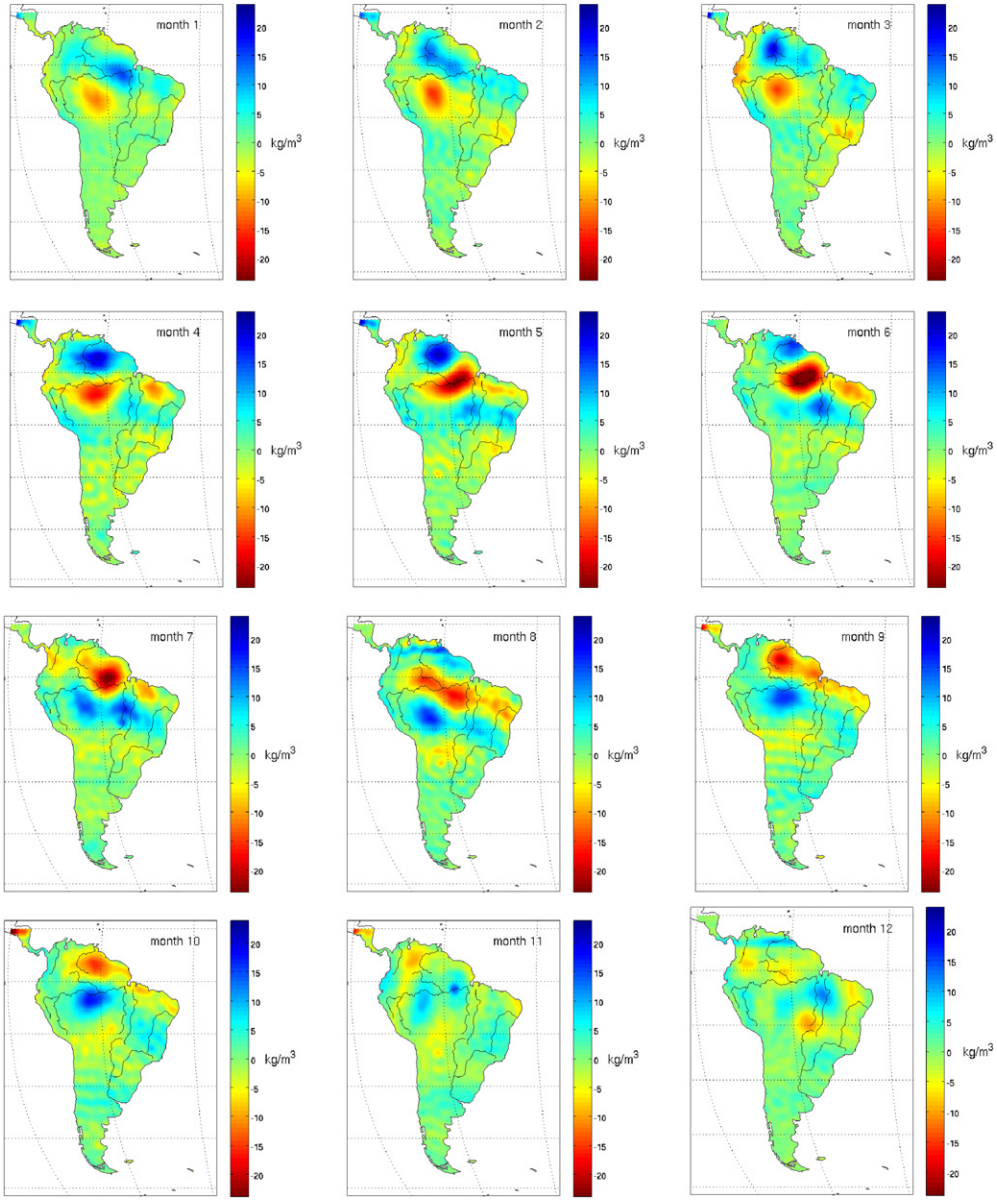

Figure 14. Density deviations for the year 2008 (top left: January, top right: March) computed out of 11 990 data points after 10 000 iterations, $\lambda = 8.7128$.

However, the higher degrees and orders contain noise that needs to be removed from the input with some kind of smoothing. As explained before, smoothing also attenuates the real signal such that we have to expect a change in magnitudes. We use spherical wavelets to analyse variations in the gravitational potential of the Earth as suggested in [31]. The cubic polynomial wavelet filter depends upon a scale J which controls up to which degree and to what extent the spherical harmonics coefficients are considered. An increasing scale J admits more detail information. However, it bears the risk to include errors or artefacts like satellite

tracks as well. Thus, it has to be investigated carefully which filter yields a realistic and useful input.

In [31], the problem of choosing the right scale was already considered. As a result, it was proposed to use scale $J = 4$, since scale $J = 5$ already includes errors. At scale $J = 4$, all spherical harmonics coefficients up to degree 31 are considered, while at scale $J = 5$ the degree of considered coefficients is increased up to 63. To get a better input to recover the seasonal changes in the Amazon area we choose a filter where all spherical harmonics coefficients up to degree 49 are considered, i.e. $J = \frac{\ln 25}{\ln 2}$. Now we expect that there are enough details included to recover the desired effects where the contained noise is still suppressed sufficiently.

We use dictionary (6.1), where the series in the kernel functions is terminated at degree 100. Moreover, we stop algorithm 5.1 (RFMP) after 10 000 iterations. As input we use 11 990 data points given on an equiangular grid as before computed 7 km above the Earth's surface. We regularize with $\lambda = 8.7128$ for all months to keep comparability.

In figure 14, we display the resulting density deviations for the year 2008. The colour blue denotes that the humidity is higher than in the mean, i.e. the surface and ground water levels are higher than in the mean, while red denotes that the humidity is lower than in the mean.

Considering, e.g., April 2008 (left-hand side of the second row in figure 14), we conclude that we have a rainy season north of the equator and a dry season south of the equator. These findings conform with meteorological and hydrological observations (see Global Land Data Assimilation System (GLDAS) [41, 63]). In comparison to September 2008 (right-hand side of the third row in figure 14), we clearly see the seasonal changes, since now there is a dry region north of the equator and a wet region south of the equator.

We observe a clear separation of the Amazon watershed and the Orinoco watershed, situated north of the Amazon area, which is a very important feature to be reconstructed, since we do not only have a meteorological separation by the equator but a topographic separation by the Guiana highlands as well.

Overall, the displayed results conform to empiric data from a temporal perspective as well as from a spatial one, i.e. the changes appear in accordance to the seasons in the Amazon area and the equator seems to be a natural interface for the change of conditions.

7. Conclusions and outlook

In this work, we presented a new regularization method to solve the inverse gravimetric problem by using the idea of a sparse regularization based on a matching pursuit. We were able to show theoretical results about the convergence of the developed algorithms as well as the main properties of a regularization method, i.e. the existence and stability of a solution and the convergence of the regularization.

One main feature of this new method is that it provides us with a solution that is primarily matched to the structure of the target function and not only the data structure, which is very advantageous in comparison to the usual wavelet and spline methods. Since the proposed method is iterative, we may directly control the sparsity of the solution and re-use earlier results as an initial solution for refinement or zooming-in. Another advantage is that we may collect all different kinds of functions in a so-called dictionary to reconstruct different structures in the solution, accordingly. In this work, we collected functions with global and local character in our dictionary and demonstrated the power of the new method by reconstructing the density distribution of South America as well as examining the mass transport in the Amazon area.

However, the presented method also allows us to combine different data types. Our main goal is to recover a model of the density distribution of the interior of the Earth as is done

in [9]. Gravitational data only gives information about the harmonic part of the density. The anharmonic part can be partially recovered from seismic data such as normal mode splitting or travel times. Detailed studies of the combined inversion with this new method are currently being investigated and will be published in a forthcoming work. However, first results may be found in [33].

Acknowledgments

We gratefully acknowledge the support by the German Research Foundation (DFG), projects MI 655/2-2 and MI 655/7-1. Moreover, the authors thank Frederik J Simons (Princeton University) for inspiring discussions and competent advices as well as the reviewers for their valuable comments and suggestions.

References

[1] Albertella A, Sansò F and Sneeuw N 1999 Band-limited functions on a bounded spherical domain: the Slepian problem on the sphere *J. Geod.* **73** 436–47
[2] Amirbekyan A 2007 The Application of reproducing kernel based spline approximation to seismic surface and body wave tomography: theoretical aspects and numerical results *PhD Thesis* University of Kaiserslautern, Kaiserslautern, Germany, http://kluedo.ub.uni-kl.de/volltexte/2007/2103/pdf/ThesisAbel.pdf
[3] Amirbekyan A and Michel V 2008 Splines on the three-dimensional ball and their application to seismic body wave tomography *Inverse Problems* **24** 1–25
[4] Amirbekyan A, Michel V and Simons F J 2008 Parameterizing surface-wave tomographic models with harmonic spherical splines *Geophys. J. Int.* **174** 617–28
[5] Barron A R, Cohen A, Dahmen W and DeVore R A 2008 Approximation and learning by greedy algorithms *Ann. Stat.* **36** 64–94
[6] Barthelmes F 1986 *Untersuchungen zur Approximation des äußeren Schwerefeldes der Erde durch Punktmassen mit optimierten Positionen* (Potsdam: Veröffentlichungen des Zentralinstituts Physik der Erde)
[7] Barthelmes F and Dietrich R 1991 Use of point masses on optimized positions for the approximation of the gravity field *Determination of the Geoid: Present and Future* ed R H Rapp and F Sansò (New York: Springer) pp 484–93
[8] Berkel P 2009 Multiscale methods for the combined inversion of normal mode and gravity variations *PhD Thesis* University of Kaiserslautern, Kaiserslautern, Germany (Aachen: Shaker)
[9] Berkel P, Fischer D and Michel V 2011 Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation *Int. J. Geomath.* **1** 167–204
[10] Berkel P and Michel V 2010 On mathematical aspects of a combined inversion of gravity and normal mode variations by a spline method *Math. Geosci.* **42** 795–816
[11] Candès E J, Romberg J and Tao T 2006 Exact signal reconstruction from highly incomplete Fourier information *IEEE Trans. Inf. Theory* **52** 489–509
[12] Center for Space Research, University of Texas Austin, TX, USA, www.csr.utexas.edu/grace/overview.html
[13] Chen J L, Wilson C R and Tapley B D 2006 Satellite gravity measurements confirm accelerated melting of Greenland ice sheet *Science* **313** 1958–60
[14] Chen J L, Wilson C R, Tapley B D, Yang Z L and Niu G Y 2009 2005 drought event in the Amazon river basin as measured by GRACE and estimated by climate models *J. Geophys. Res.* **114** B05404
[15] Chen S S, Donoho D L and Saunders M A 2001 Atomic decomposition by basis pursuit *SIAM Rev.* **43** 129–59
[16] Claessens S J, Featherstone W E and Barthelmes F 2001 Experiences with point-mass gravity field modelling in the Perth region, Western Australia, http://www.cage.curtin.edu.au/will/gra07dec2001.pdf
[17] Cui J and Freeden W 1997 Equidistribution on the sphere *SIAM J. Sci. Stat. Comput.* **18** 595–609
[18] Dahlke S, Fornasier M and Raasch T 2009 Multilevel preconditioning and adaptive sparse solution of inverse problems *Math. Comput.* **81** 419–46
[19] Dai W and Milenkovic O 2008 Subspace pursuit for compressive sensing: closing the gap between performance and complexity, http://dsp.rice.edu/files/cs/SubspacePursuit.pdf
[20] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Commun. Pure Appl. Math.* **57** 1413–57

[21] Daubechies I, Fornasier M and Loris I 2008 Accelerated projected gradient method for linear inverse problems with sparsity constraints *J. Fourier Anal. Appl.* **14** 764–92
[22] Davis P J 1975 *Interpolation and Approximation* (New York: Dover)
[23] Davis G, Mallat S G and Zhang Z 1994 Adaptive time-frequency approximations with matching pursuits *SPIE J. Opt. Eng.* **33** 2183–91
[24] DeVore R A 1998 Nonlinear approximation *Acta Numer.* **7** 51–150
[25] Donoho D L, Tsaig Y, Drori I and Starck J-L 2006 Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit http://www-stat.stanford.edu/~donoho/Reports/2006/
[26] Driscoll J R and Healy R M 1994 Computing Fourier transforms and convolutions on the 2-sphere *Adv. Appl. Math.* **15** 202–50
[27] Dufour H M 1977 Fonctions orthogonales dans la sphère. Résolution théorique du problème du potentiel terrestre *Bull. Geod.* **51** 227–37
[28] Dziewonski A and Anderson D L 1981 The preliminary reference Earth model *Phys. Earth Planet. Inter.* **25** 297–356
[29] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[30] Engl H W, Kunisch K and Neubauer A 1989 Convergence rates for Tikhonov regularization of nonlinear ill-posed problems *Inverse Problems* **5** 523–40
[31] Fengler M J, Freeden W, Kohlhaas A, Michel V and Peters T 2007 Wavelet modeling of regional and temporal variations of the Earth's gravitational potential observed by GRACE *J. Geod.* **81** 5–15
[32] Fengler M J, Michel D and Michel V 2006 Harmonic spline-wavelets on the 3-dimensional ball and their application to the reconstruction of the Earth's density distribution from gravitational data at arbitrarily shaped satellite orbits *Z. Angew. Math. Mech.* **86** 856–73
[33] Fischer D 2011 Sparse regularization of a joint inversion of gravitational data and normal mode anomalies *PhD Thesis* (Munich: Dr Hut Verlag)
[34] Freeden W 1981 On approximation by harmonic splines *Manuscr. Geod.* **6** 193–244
[35] Freeden W 1981 On spherical spline interpolation and approximation *Math. Methods Appl. Sci.* **3** 551–75
[36] Freeden W and Michel V 2004 *Multiscale Potential Theory (With Applications to Geoscience)* (Boston, MA: Birkhäuser)
[37] Gilbert A C, Muthukrishnan S and Strauss M J 2005 Improved time bounds for near-optimal sparse Fourier representation via sampling *Proc. SPIE Wavelets* **11**
[38] Gilbert A C, Strauss M J, Tropp J A and Vershynin R 2007 One sketch for all: fast algorithms for compressed sensing *39th ACM Symp. on Theory of Computing (San Diego, CA)*
[39] Hein G, Sansò F, Strykowsky G and Tscherning C C 1989 On the choice of norm and base functions for the solution of the inverse gravimetric problem *Ricerche di Geodesia Topografia Fotogrammetria* vol 5 (Milan: CLUP) 121–38
[40] Hesse K and Sloan I H 2010 Numerical integration on the sphere *Handbook of Geomathematics* ed W Freeden *et al* (Heidelberg: Springer) pp 1189–217
[41] Hydrological Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD, USA, http://mirador.gsfc.nasa.gov
[42] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, http://podaac.jpl.nasa.gov/grace/index.html
[43] Kammann P and Michel V 2008 Time-dependent Cauchy–Navier splines and their application to seismic wave front propagation *Z. Angew. Math. Mech.* **88** 155–78
[44] Kim S J, Koh K, Lustig M, Boyd S and Gorinevsky D 2007 An interior-point method for large-scale l^1-regularized least squares *IEEE J. Spec. Top. Signal Process.* **1** 606–17
[45] Last B J and Kubik K 1983 Compact gravity inversion *Geophysics* **48** 713–21
[46] Lauricella G 1912 Sulla distribuzione della massa nell'interno dei pianeti *Rend. Accad. Naz. Lincei* **21** 18–26
[47] Mallat S G and Zhang Z 1993 Matching pursuits with time–frequency dictionaries *IEEE Trans. Signal Process.* **41** 3397–415
[48] Michel V 1999 A multiscale method for the gravimetry problem: theoretical and numerical aspects of harmonic and anharmonic modelling *PhD Thesis* University of Kaiserslautern, Kaiserslautern, Germany (Aachen: Shaker)
[49] Michel V 2002 A multiscale approximation for operator equations in separable Hilbert spaces—case study: reconstruction and description of the Earth's interior *Habilitation Thesis* University of Kaiserslautern, Kaiserslautern, Germany (Aachen: Shaker)
[50] Michel V 2002 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. Harmon. Anal.* **12** 77–99

[51] Michel V 2005 Regularized wavelet-based multiresolution recovery of the harmonic mass density distribution from data of the Earth's gravitational field at satellite height *Inverse Problems* **21** 997–1025
[52] Michel V 2010 Tomography: problems and multiscale solutions *Handbook of Geomathematics* ed W Freeden *et al* (Heidelberg: Springer) pp 949–72
[53] Michel V and Fokas A S 2008 A unified approach to various techniques for the non-uniqueness of the inverse gravimetric problem and wavelet-based methods *Inverse Problems* **24** 045091
[54] Michel V and Wolf K 2008 Numerical aspects of a spline-based multiresolution recovery of the harmonic mass density out of gravity functionals *Geophys. J. Int.* **173** 1–16
[55] Miranian L 2004 Slepian functions on the sphere, generalized Gaussian quadrature rule *Inverse Problems* **20** 877–92
[56] Müller C 1966 *Spherical Harmonics* (Berlin: Springer)
[57] Needell D and Tropp J A 2009 CoSaMP: iterative signal recovery from incomplete and inaccurate samples *Appl. Comput. Harmon. Anal.* **26** 301–21
[58] Nutz H and Wolf K 2008 Time-space multiscale analysis by use of tensor product wavelets and its application to hydrology and GRACE data *Stud. Geophys. Geod.* **52** 321–39
[59] Pavlis N K, Holmes S A, Kenyon S C and Factor J K 2008 An Earth gravitational model to degree 2160: EGM2008 *General Assembly of the European Geosciences Union (Vienna)*
[60] Pizzetti P 1909 Corpi equivalenti rispetto alla attrazione Newtoniana esterna *Rend. Accad. Naz. Lincei* **18** 211–5
[61] Pizzetti P 1910 Intorno alle possibili distribuzioni della massa nell'interno della terra *Ann. Mat. Pura Appl.* **17** 225–58
[62] Rieder A 2003 *Keine Probleme mit Inversen Problemen* (Braunschweig: Vieweg)
[63] Rodell M *et al* 2004 The global land data assimilation system *Bull. Am. Meteorol. Soc.* **85** 381–94
[64] Sansò F, Barzaghi R and Tscherning C C 1986 Choice of norm for the density distribution of the Earth *Geophys. J. R. Astron. Soc.* **87** 123–41
[65] Seidmann T I and Vogel C R 1989 Well-posedness and convergence of least-square estimation to ill-posed problems *Inverse Problems* **5** 227–38
[66] Simons F J 2010 Slepian functions and their use in signal estimation and spectral analysis *Handbook of Geomathematics* ed W Freeden (Heidelberg: Springer) pp 891–923
[67] Simons F J and Dahlen F A 2006 Spherical Slepian functions and the polar gap in geodesy *Geophys. J. Int.* **166** 1039–61
[68] Simons F J, Dahlen F A and Wieczorek M A 2006 Spatiospectral concentration on a sphere *SIAM Rev.* **48** 504–36
[69] Tapley B D, Bettadpur S, Ries J C, Thopson P F and Watkins M M 2004 GRACE measurements of mass variability in the Earth system *Science* **305** 503–5
[70] Temlyakov V N 2003 Nonlinear methods of approximation *Found. Comput. Math.* **3** 33–107
[71] Thalhammer M, Ricard Y, Rummel R and Ilk K H 1996 Application of space-borne gravimetry to research on the interior of the Earth *ESA Study—CIGAR 4*
[72] Tscherning C C 1991 Density-gravity covariance functions produced by overlapping rectangular blocks of constant density *Geophys. J. Int.* **105** 771–6
[73] Tscherning C C and Strykowski G 1987 Quasi-harmonic inversion of gravity field data *Model Optimization in Exploration Geophysics 2: Proc. 5th International Mathematical Geophysics Seminar (Berlin)* pp 137–54
[74] Vincent P and Bengio Y 2002 Kernel matching pursuit *Mach. Learn.* **48** 169–91
[75] Voigt A and Wloka J 1975 *Hilberträume und elliptische Differentialoperatoren* (Mannheim: Bibliographisches Institut)
[76] Weck N 1972 Zwei inverse Probleme in der Potentialtheorie *Mitteilungen aus dem Institut für Theoretische Geodäsie (Bonn)* vol 4, pp 27–36
[77] Zeidler E 1985 *Nonlinear Functional Analysis and Its Applications* (Berlin: Springer)