Siegen Preprints
on Geomathematics

A First Approach to Learning a Best Basis for Gravity Field Modelling

V. Michel, N. Schneider

Geomathematics Group
Department of Mathematics
University of Siegen, Germany

www.geomathematics-siegen.de

21

A first approach to learning a best basis for gravity field modelling

Volker Michel and Naomi Schneider*

Geomathematics Group Siegen

Abstract

Gravitational field modelling is an important tool for inferring past and present dynamic processes of the Earth. Functions on the sphere such as the gravitational potential are usually expanded in terms of either spherical harmonics or radial basis functions (RBFs). The (Regularized) Functional Matching Pursuit ((R)FMP) and its variants use an overcomplete dictionary of diverse trial functions to build a best basis as a sparse subset of the dictionary and compute a model, for instance, of the gravity field, in this best basis. Thus, one advantage is that the dictionary may contain spherical harmonics and RBFs. Moreover, these methods represent a possibility to obtain an approximative and stable solution of an ill-posed inverse problem, such as the downward continuation of gravitational data from the satellite orbit to the Earth's surface, but also other inverse problems in geomathematics and medical imaging. A remaining drawback is that in practice, the dictionary has to be finite and, so far, could only be chosen by rule of thumb or trial-and-error. In this paper, we develop a strategy for automatically choosing a dictionary by a novel learning approach. We utilize a nonlinear constrained optimization problem to determine best-fitting RBFs (Abel–Poisson kernels). For this, we use the Ipopt software package with an HSL subroutine. Details of the algorithm are explained and first numerical results are shown.

Keywords: dictionary learning, downward continuation, greedy algorithm, inverse problem, matching pursuit, nonlinear optimization, radial basis functions, spherical harmonics

AMS: 31B20, 41A45, 65D15, 65J20, 65K10, 65N20, 65R32, 68T05, 86A22

1 Introduction

The gravitational potential is an important observable in the geosciences as it is used as a reference for multiple static and dynamic phenomena of the complex Earth system. The EGM2008 gives us a high-precision model in spherical harmonics, i.e. polynomials, up to degree 2190 and order 2159, see [36, 39]. From satellite missions like GRACE or its successor GRACE-FO, we have time-dependent models of the potential, see, for example, [10, 35, 45, 47]. These data enable a visualization of mass transports on the Earth such as seasonal short-term phenomena like the wet season in the Amazon basin as well as long-term phenomena like the climate change. Therefore, gravity field modelling and especially the downward continuation of satellite data is one of the major important mathematical problems in physical geodesy, see, for instance, [2, 22].

---

*Corresponding author: naomi.schneider@mathematik.uni-siegen.de

From a mathematical point of view, the gravitational potential F on the approximately spherical Earth's surface can be modelled as a Fourier expansion in a suitable basis, for example in the mentioned spherical harmonics $Y_{n,j}$, $n \in \mathbb{N}_0$, $j = -n, ..., n$. If we assume the Earth to be a closed unit ball, we obtain, for $\sigma > 1$, a pointwise representation of the potential as

$$V(\sigma\eta) = (\mathcal{T}F)(\sigma\eta) = \sum_{n=0}^{\infty} \sum_{j=-n}^{n} \langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)} \sigma^{-n-1} Y_{n,j}(\eta) \tag{1}$$

for the unit sphere Ω and $\eta \in \Omega$, see, for example [2, 13, 34, 48]. This gives us the potential in the outer space including a satellite orbit. The inverse problem of the downward continuation of this potential is given as follows: if data values $V(\sigma\eta) = (\mathcal{T}F)(\sigma\eta)$, $\sigma > 1$, are known, determine the function F on Ω. For more details on inverse problems in general, see the classical literature, for example, [5, 26, 42]. The occurring mathematical challenges of the downward continuation are well-known. First of all, the operator $\mathcal{T}$ has exponentially decreasing singular values due to $\sigma > 1$ in (1). Thus, the inverse operator which we need for the downward continuation has exponentially increasing singular values. For this reason, the inverse problem is called exponentially ill-posed. In particular, it violates the third characteristic of a well-posed problem according to Hadamard (continuous dependence on the data). Furthermore, the existence of F is only ensured if V is in the range of $\mathcal{T}$. However, if F exists, then it is unique.

Therefore, sophisticated algorithms need to be used to solve the problem of the downward continuation of satellite data. Previous studies showed that the (Regularized) Functional Matching Pursuit ((R)FMP), the (Regularized) Orthogonal Functional Matching Pursuit ((R)OFMP) as well as the latest (Regularized) Weak Functional Matching Pursuit ((R)WFMP) are possible approaches for this and other inverse problems, see, for instance, [3, 6, 7, 8, 9, 17, 19, 21, 20, 29, 30, 31, 32, 48]. In the sequel, we will write Inverse Problem Matching Pursuit (IPMP) if we refer to either one of the mentioned algorithms. Although the core routine of these algorithms is well established by now, there are still possibilities to improve their performance.

One of these possibilities is given due to the following circumstances. The IPMPs are based on a finite dictionary $\mathcal{D}$ of suitable trial functions from which they build a best basis and eventually the approximate solution in terms of this best basis. Originally, matching pursuits utilize a dictionary consisting of vectors from finite-dimensional spaces. The first development of a matching pursuit was done by S.G. Mallat and Z. Zhang (1993). The ROFMP is additionally based on works of P. Vincent and Y. Bengio (2002) and Y.C. Pati et al. (1993). The RWFMP inherits ideas from V.N. Temlyakov (2000). The idea of using dictionaries instead of finding a representation of a signal in an a-priori given basis can be summed up as follows [27]: the human language gives us nearly infinite possibilities to describe the very same thing in the real world. However, these descriptions vary in length and rigour. This idea can be transferred to mathematics, for instance, gravity field modelling. We can model the gravitational potential in spherical harmonics as given in (1). However, if we look for a model in a best basis from a dictionary $\mathcal{D}$, the representation of the signal might be sparser and/or more precise. In particular, the reduction to those basis functions which are essential increases the interpretability of the obtained model. Further, numerical experiments showed that the obtained solution is more accurate and stable. These aims can be achieved by the IPMPs.

Further characteristics of the IPMPs can be summed up from previous publications as follows: they represent a regularization for ill-posed inverse problems; they can combine different kinds of trial functions, e.g. global and localized ones, in their solution; they can

be used for pure interpolation / approximation as well as for linear and non-linear ill-posed inverse problems; they work with single-source data as well as are capable of a joint-inversion of multiple-source data; the data can be given on different geometries, like a sphere, a ball, or an interval of the real line; they yield an approximative function and not a discretized approximation; they build this solution iteratively without the need to invert a matrix or solve a large linear system of equations; the orthogonal variant yields a linear combination of orthogonal trial functions; the runtime can be improved with the use of preprocessing of certain data; or they can be combined with a weak strategy to cut runtime without significant loss of accuracy; the implementation is easy and they can be parallelized very well.

In the practical tests for diverse applications, very good approximations could be achieved not only for the downward continuation but also for other ill-posed inverse problems, for instance the (linear as well as the non-linear) inverse gravimetric problem or the inversion of MEG- and EEG-data, see, for instance, [6, 7, 19, 21, 23, 24].

However, the experiments also revealed a sensitivity of the result regarding the choice of the dictionary, for example concerning the runtime and the convergence behaviour. Therefore, the main focus of this paper is on a first dictionary learning strategy for the downward continuation of gravitational potential data.

Previous works on dictionary learning considered discretized approximation problems. In this case, the dictionary can be interpreted as a matrix. The approaches aimed to obtain a solution of the approximation problem and a sparse dictionary matrix simultaneously. For more details, see, for example, [4, 40, 43].

However, a particular feature of the IPMPs is that their solution is a linear combination of established trial functions. Neither do we want to discretize the dictionary elements, i.e. the trial functions, nor do we want to modify them. In the latter case, the comparability with traditional models in these trial functions would be lost. Furthermore, with the use of scaling functions and / or wavelets in the dictionary, the IPMP generates a solution in a multiscale basis. This allows a multiresolution analysis of the obtained model revealing hidden local detail structures as it was shown in, for example, [6, 7, 8, 9, 31, 48]. Moreover, we do not only consider interpolation / approximation problems, but also ill-posed inverse problems. Thus, a dictionary matrix would not contain the basis elements themselves, but, for example, their upward continued values. Applying previous strategies, like, for instance, MOD or K-SVD, would only alter the upward continued values and leave us with the question of how to downward continue them. All in all, this shows that learning a dictionary for the IPMPs requires the development of a different strategy.

For a first approach to learning a dictionary, we concentrate on the RFMP as the basic IPMP in this paper. For this algorithm, we develop a procedure to determine a best basis for the gravitational potential from different types of infinitely many trial functions. We choose to learn dictionary elements from spherical harmonics and Abel–Poisson kernels as radial basis functions. In particular, while previously a discrete grid of centres of the RBFs had to be chosen a-priori, which could have put a bias in the obtained numerical result, we now allow every point on the unit sphere to be a centre of an RBF. Equally, the localization parameter of the Abel–Poisson kernel is now determined from an interval instead of a finite set. Our continuous, i.e. non-discrete, learning ansatz produces a 'best-dictionary' with which the RFMP can be run. We call this procedure the Learning (Regularized) Functional Matching Pursuit (L(R)FMP). The results show that the use of a learnt dictionary in the RFMP gives us a higher sparsity and better results with less storage demand.

This paper is structured as follows. On the way to a detailed description of our learning approach, we define some fundamental basics in Section 2. We introduce the trial functions under investigation as well as a general form of a dictionary and the basic principles of

the RFMP. With these aspects explained, we state our learning strategy in Section 3. We motivate its idea and explain how this is embedded into the established theory of dictionary learning. Then we define its routine and give necessary derivatives. These are derived in Appendix A. We end Section 3 by introducing some additional learning features that guide the learning process positively in practice. In Section 4, we describe experiments for which we learn a dictionary and compare the results of the learnt dictionary with the results which a manually chosen dictionary yields. At last, we conclude this paper in Section 5 with an outlook of how we want to further develop this first learning approach.

2 Some mathematical tools for learning a dictionary

2.1 Trial functions under consideration for dictionaries

First of all, to develop a learning strategy, we have to define what trial functions we want to consider as possible dictionary elements, i.e. what the learnt dictionary shall consist of. Our aim is to determine which well-known trial functions are most suitable for a problem at hand. For gravity field modelling, it is sensible to determine suitable spherical harmonics as well as Abel–Poisson kernels. Examples of those trial functions are given in Figure 1.
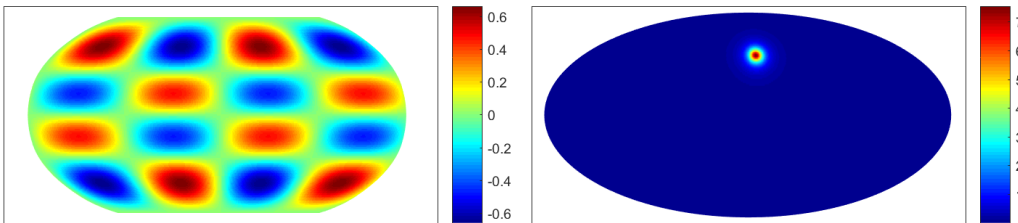

Figure 1: Examples of trial functions. Left: spherical harmonic. Right: Abel–Poisson kernel for a fixed centre ξ (location of the maximum).

Spherical harmonics or fully normalized spherical harmonics for practical purposes are global trial functions, see for instance, [12, 14, 15, 28, 33]. An example is given on the left-hand side of Figure 1. They are defined for a unit vector $\xi \in \Omega$ as

$$Y_{n,j}\left(\xi(\varphi,t)\right) := \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-|j|)!}{(n+|j|)!}}P_{n,|j|}(t)\begin{cases} \sqrt{2}\cos(j\varphi), & j < 0, \\ 1, & j = 0, \\ \sqrt{2}\sin(j\varphi), & j > 0, \end{cases} \tag{2}$$

where $\xi(\varphi,t)$ is the representation of $\xi \in \Omega$ in polar coordinates (φ,t), where $t = \cos\vartheta$ and ϑ is the latitude. Further, the definition uses associated Legendre functions given by

$$P_{n,j}(t) := \left(1 - t^2\right)^{j/2}\frac{\mathrm{d}^j}{\mathrm{d}t^j}P_n(t), \qquad t \in [-1, 1], \tag{3}$$

where P_n denotes the n-th Legendre polynomial.

Abel–Poisson kernels are defined for a particular unit vector $\xi \in \Omega$ and a scaling parameter $h \in [0, 1)$ as (with $x = h\xi$)

$$P(x, \eta) := \frac{1 - |x|^2}{4\pi(1 + |x|^2 - 2x \cdot \eta)^{3/2}} = \sum_{n=0}^{\infty}\frac{2n+1}{4\pi}|x|^n P_n\left(\frac{x}{|x|}\cdot\eta\right) \tag{4}$$

4

for any unit vector $\eta \in \Omega$, see, for example, [15, pp. 108-112] or [14, p. 103 and 441]. These kernels are more localized than polynomials as one can see on the right-hand side of Figure 1. It is visible that they are radial basis functions, that means they have one maximum whose descent depends on the distance to the centre $\xi = x/|x|$ of the extremum. In that way, they are zonal functions and can be viewed as 'hat'-functions. Dependent on the parameter $h = |x|$, the size of the extremum or 'hat' varies in size. Thus, the functions have different scales of localization. For more details and examples, see, for instance, [15, p. 111] or [28, p. 117].

In this paper, we consider dictionaries consisting of spherical harmonics and Abel–Poisson kernels. We introduce here a notation for building blocks of spherical dictionaries.

Definition 1. *Let* $N \subset \mathcal{N} \coloneqq \{(n,j) \mid n \in \mathbb{N}_0, j = -n, ..., n\}$ *and* $K \subseteq \mathring{\mathbb{B}}_1(0)$ *for the open unit ball* $\mathring{\mathbb{B}}_1(0)$. *Then we set*

$$[N]_{\mathrm{SH}} \coloneqq \{Y_{n,j} \mid (n,j) \in N\}$$

for spherical harmonics $Y_{n,j}$ *and*

$$[K]_{\mathrm{APK}} \coloneqq \{P(r\xi, \cdot) \mid r\xi \in K\}$$

for Abel–Poisson kernels $P(r\xi, \cdot)$. *We define a dictionary as*

$$\mathcal{D} \coloneqq [N]_{\mathrm{SH}} + [K]_{\mathrm{APK}} \coloneqq [N]_{\mathrm{SH}} \cup [K]_{\mathrm{APK}}.$$

We call $[\cdot]_*$ *a trial function class.*

Note that N and K may be finite or infinite.

2.2 Basic principles of linear ill-posed inverse problems

This subsection is mainly based on [5, 19, 28, 42]. First of all, we recall the definition of a linear inverse problem.

Definition 2. *Let* $\mathcal{X}$, $\mathcal{Y}$ *be Hilbert spaces. Further, let* $\mathcal{T} \colon \mathcal{X} \to \mathcal{Y}$ *be a linear and continuous operator between them. At last, let* $y \in \mathcal{Y}$ *denote the data and* $F \in \mathcal{X}$ *the desired solution. Then a problem of the form*

$$\mathcal{T}F = y$$

is called a linear inverse problem.

Naturally, some fundamental mathematical questions about the problem arise. Those questions lead to the definition of well- and ill-posedness of the linear inverse problem by Hadamard.

Definition 3. *A linear inverse problem* $\mathcal{T}F = y$ *is called well-posed if it fulfils the following three properties.*
 (a) For each y, *there exists a solution* F.
 (b) The solution F *is unique.*
 (c) The inverse operator $\mathcal{T}^{-1}$ *is continuous, i.e. the solution* F *is stable.*
If any of these properties is violated, the problem is called ill-posed.

As we explained in the introduction, the problem of the downward continuation of satellite data for gravity field modelling is a linear ill-posed inverse problem. As most inverse problems from practice are ill-posed, there exists a large theory on how to still solve these problems. In the sequel, we will use the approach by Tikhonov.

Its idea is to find the best approximate solution instead of the true solution. The ill-posedness is treated with an additional penalty term. The best approximate solution is the minimizer of the Tikhonov functional for a vanishing regularization parameter. The functional is given as follows.

Definition 4. *Let* $\mathcal{T}F = y$ *be a linear inverse problem. The Tikhonov functional* $\mathcal{J} \colon \mathcal{X} \to \mathbb{R}$ *is given by*

$$\mathcal{J}(F; \mathcal{T}, \lambda, y) \coloneqq \|y - \mathcal{T}F\|_{\mathcal{Y}}^2 + \lambda \|F\|_{\mathcal{X}}^2.$$

In the case of satellite data, we set $\mathcal{Y} = \mathbb{R}^\ell$, i.e. we are only given data on ℓ discrete points of the unit sphere. For the space $\mathcal{X}$, we propose to use a Sobolev space. The reasons for this choice are, for instance, that this space enforces more smoothness of the solution than, e.g. the $\mathrm{L}^2(\Omega)$-space. This has proven to yield better results. Further, it was shown that the IPMPs also profit theoretically from the use of this space, see, [19]. Specifically, we will use the Sobolev space $\mathcal{H}_2$.

Definition 5. *On the set* $\widetilde{\mathcal{H}}\big((n+0.5)^2; \Omega\big)$ *of all functions* $F \in \mathrm{C}^{(\infty)}(\Omega, \mathbb{R})$ *that fulfil*

$$\sum_{n=0}^{\infty} \sum_{j=-n}^{n} (n+0.5)^4 \langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)}^2 < \infty,$$

we define an inner product via

$$\langle F, G \rangle_{\mathcal{H}_2} \coloneqq \sum_{n=0}^{\infty} \sum_{j=-n}^{n} (n+0.5)^4 \langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)} \langle G, Y_{n,j} \rangle_{\mathrm{L}^2(\Omega)}.$$

The completion of $\widetilde{\mathcal{H}}\big((n+0.5)^2; \Omega\big)$ *with respect to* $\langle \cdot, \cdot \rangle_{\mathcal{H}_2}$ *is called the Sobolev space* $\mathcal{H}_2$.

For practical purposes, we will give a short overview of the main principles of the RFMP algorithm, which is a regularization method for ill-posed linear problems, next. For more details on any IPMP, see, for example, [6, 7, 8, 9, 19, 20, 21, 30, 31, 32, 48]. For theoretical discussions of the IPMPs, we refer to this literature. Here, we will concentrate on the practical aspects of the RFMP.

The underlying idea of this matching pursuit is to build a solution as a linear combination of dictionary elements by iteratively minimizing a Tikhonov functional. In theory, we consider the linear inverse problem $\mathcal{T}F = V$, for instance, as given in (1). In practice, we consider the particular case $\mathcal{Y} = \mathbb{R}^\ell$. Then the problem formulates as follows. We have a relative satellite height $\sigma > 1$, a set of grid points $\{\eta^{(i)}\}_{i=1,\dots,\ell} \in \Omega$ and data values y_i for each grid point $\eta^{(i)}$. The operator $\mathcal{T}$ is exchanged by a finite system of related functionals $\mathcal{T}_{\daleth}^i$ for which $\mathcal{T}_{\daleth}^i F = (\mathcal{T}F)(\sigma \eta^{(i)}) = y_i$ holds for $i = 1, \dots, \ell$. We use the Hebrew letter Dalet $\daleth$ to emphasize that the functionals $\big(\mathcal{T}_{\daleth}^i\big)_{i=1,\dots,\ell}$ represent a discretization of the operator $\mathcal{T}$. Summarized, we consider the linear inverse problem $\mathcal{T}_{\daleth}F = y$ for the operator $\mathcal{T}_{\daleth} = (\mathcal{T}_{\daleth}^i)_{i=1,\dots,\ell}$ and $y \in \mathbb{R}^\ell$.

Further, we assume that $F \in \mathcal{H}_2$. A regularization parameter is denoted by λ. Additionally, we need an a-priori defined dictionary $\mathcal{D}$ as given in Definition 1. Then the aim of the RFMP is to iteratively minimize the Tikhonov functional

$$\mathcal{J}(F_n + \alpha d; \mathcal{T}_{\daleth}, \lambda, y) \coloneqq \|y - \mathcal{T}_{\daleth}(F_n + \alpha d)\|_{\mathbb{R}^\ell}^2 + \lambda \|F_n + \alpha d\|_{\mathcal{H}_2}^2 \tag{5}$$

for an element $d \in \mathcal{D}$ of the dictionary, a real coefficient α and a current approximation F_n. In practice, this means we start with an initial approximation F_0, e.g. $F_0 \equiv 0$, and iteratively determine $F_{n+1} \coloneqq F_n + \alpha_{n+1} d_{n+1}$ via

$$(\alpha_{n+1}, d_{n+1}) \coloneqq \underset{(\alpha, d) \in \mathbb{R} \times \mathcal{D}}{\arg\min} \; \mathcal{J}(F_n + \alpha d; \mathcal{T}_{\neg}, \lambda, y). \tag{6}$$

It can be shown, see, for example, [6, 29, 31] that the minimization with respect to α and d of the Tikhonov functional (6) is equivalent to a maximization of a certain quotient with respect to d. For the RFMP, this quotient is given by

$$d_{n+1} \coloneqq \underset{d \in \mathcal{D}}{\arg\max} \; \frac{\left(\langle R^n, \mathcal{T}_{\neg} d \rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d \rangle_{\mathcal{H}_2} \right)^2}{\|\mathcal{T}_{\neg} d\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_2}^2}.$$

in the n-th iteration step, where $R^n \coloneqq y - \mathcal{T}_{\neg} F_n$ is the residual.

3 The learning approach

In this section, we refer to a linear inverse problem $\mathcal{T}_{\neg} F = y$ for $y \in \mathbb{R}^\ell$ as described above. The term F_n represents the current approximation of the RFMP at iteration step n. A dictionary element is denoted by d, spherical harmonics by $Y_{n,j}$ and Abel–Poisson kernels by $P(x, \cdot)$ as we introduced them in the last section.

3.1 About learning dictionaries for inverse problems

We motivate our learning algorithm as follows. We consider an infinite set of trial functions from which we want to learn dictionary elements in the LRFMP. We set

$$\mathcal{D}^{\inf} = [\mathcal{N}]_{\mathrm{SH}} + \left[\mathring{\mathbb{B}}_1(0) \right]_{\mathrm{APK}}.$$

Thus, for $F_n = \sum_{i=1}^n \alpha_i d_i$, we consider

$$\underset{(\alpha_{n+1}, d_{n+1}) \in \mathbb{R} \times \mathcal{D}^{\inf}}{\arg\min} \; \left(\|y - \mathcal{T}_{\neg}(F_n + \alpha d)\|_{\mathbb{R}^\ell}^2 + \lambda \|F_n + \alpha d\|_{\mathcal{H}_2}^2 \right). \tag{7}$$

If we chose the dictionary element d_i and its coefficient α_i in this greedy manner in each iteration step, in a perfect world and for $n \to \infty$, then this would be equal to

$$\underset{F_\infty \in \overline{\mathrm{span}\, \mathcal{D}^{\inf}}}{\min} \; \left(\|y - \mathcal{T}_{\neg} F_\infty\|_{\mathbb{R}^\ell}^2 + \lambda \|F_\infty\|_{\mathcal{H}_2}^2 \right), \qquad F_\infty = \sum_{i=1}^\infty \alpha_i d_i.$$

Unfortunately, this is practically impossible to solve for several reasons. First of all, the problem is of the type of the travelling salesman problem which is known to be NP-hard, see, for instance, [16, p. 114]. Thus, we simply cannot be sure that picking trial functions in a greedy manner also yields a true greedy algorithm and, thus, the optimal solution F_∞. Secondly, in practice, we can only compute a finite linear combination, i.e. an optimal best-n-term approximation if at all. And thirdly, if we work with infinitely many trial functions, we cannot preprocess any data. However, it is an expensive feature to compute everything needed on the fly.

However, if we consider the RFMP, we know that its solution is a good approximation of

$$\min_{F_n \in \ \mathrm{span}\, \mathcal{D}} \left(\|y - \mathcal{T}\lceil F_n\|_{\mathbb{R}^\ell}^2 + \lambda \|F_n\|_{\mathcal{H}_2}^2 \right), \qquad F_n = \sum_{i=1}^n \alpha_i d_i$$

for a finite dictionary $\mathcal{D}$. Note that the approximation is also a n-term approximation. That means, the structure of the RFMP yields a good approximation of the best-n-term solution dependent on a fixed finite dictionary. Therefore, if we extend the RFMP to the infinite set of functions $\mathcal{D}^{\mathrm{inf}}$ in the way defined in (7), we are able to approximate a solution F_n of

$$\min_{F_n \in \ \mathrm{span}\, \mathcal{D}^{\mathrm{inf}}} \left(\|y - \mathcal{T}\lceil F_n\|_{\mathbb{R}^\ell}^2 + \lambda \|F_n\|_{\mathcal{H}_2}^2 \right), \qquad F_n = \sum_{i=1}^n \alpha_i d_i. \tag{8}$$

This is the main objective of the LRFMP in practice. How can we learn a finite dictionary $\mathcal{D}$ from this solution? If we have approximated the solution by F_n, we know which basis elements we need for this approximation. In this way, we know which dictionary elements should be at least in the learnt finite dictionary $\mathcal{D}$ for its use in the RFMP.

Therefore, a learnt dictionary should contain at least these n elements. More generally, we can set an upper bound $D > n$ for the size of the learnt dictionary $\mathcal{D}$. Furthermore, the approximation of F_n should be in the span of the learnt dictionary $\mathcal{D}$, such that the solution of (8) can be reproduced. Moreover, the dictionary $\mathcal{D}$ naturally needs to be a subset of $\mathcal{D}^{\mathrm{inf}}$. Taking all things into consideration, we can write the objective of the dictionary learning process as

$$\min_{\substack{\mathcal{D} \subset \mathcal{D}^{\mathrm{inf}}, \\ |\mathcal{D}| \leq D}} \ \min_{F_n \in \ \mathrm{span}\, \mathcal{D}} \left(\|y - \mathcal{T}\lceil F_n\|_{\mathbb{R}^\ell}^2 + \lambda \|F_n\|_{\mathcal{H}_2}^2 \right).$$

This is a doubled minimization problem for a predefined dictionary size D. From another point of view, the task it presents is to find a dictionary and build an approximation from its elements that minimizes the Tikhonov functional simultaneously. A doubled minimization problem is a common approach in the field of dictionary learning, see, for instance, [1, 4, 40, 43]. Further, referring to the criteria for dictionary learning from Aharon et. al. (2006), this can be thought of as a well-defined goal of the learning approach.

Thus, the starting point for our learning algorithm is (7). We choose this generalization for the learning approach because we want to learn the dictionary itself as well as maintain the unique characteristic of the RFMP to combine various but established trial functions. Moreover, for this minimization, we choose to follow the structure of the RFMP for two reasons. First of all, as we pointed out above, the RFMP is built to solve the same minimization problem but over a finite set of functions. Secondly, we search for a well-working dictionary for the RFMP. It would be intuitive to expect better results when applying the learnt dictionary if we included the behaviour of the RFMP in the learning process. In this way, we use as much information and structure as we have to obtain an efficient learning routine. Hence, the idea of our learning algorithm, the LRFMP, is to iteratively minimize a Tikhonov functional over an infinite set of trial functions similarly to the RFMP in order to model the solution in a best basis. The chosen basis elements form the learnt dictionary which can be applied in runs of the RFMP.
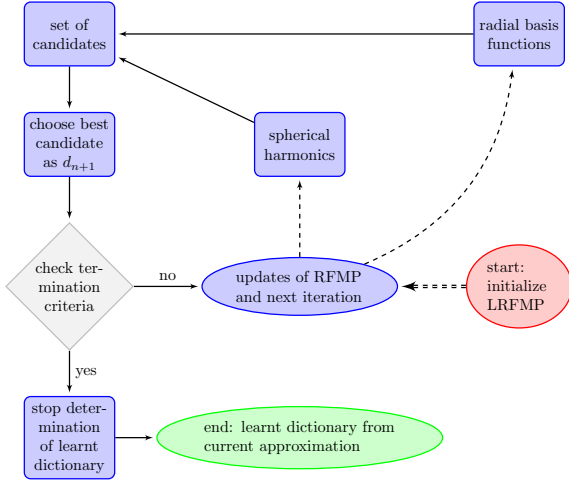
3.2 A first learning algorithm

Figure 2: Schematic representation of the basic learning algorithm.

An overview of the structure of the learning algorithm is shown in Figure 2. We start in the red circle ('start') where we initialize the LRFMP similarly to the initialization which the RFMP needs. This means, the initialization includes the necessary preprocessing and setting of parameters (similarly as described, for example, in [48]) as well as setting parameters for the learning. The latter learning parameters include, most importantly, a starting dictionary and smoothing properties.

Then we step into the first iteration in which we want to minimize the Tikhonov functional in order to find d_1. As we also want to learn a dictionary, the steps up to choosing d_1 differ from the established RFMP: we choose d_1 from (in the case of the RBFs) infinitely many trial functions instead of a finite a-priori selection of trial functions. This is done by first computing a candidate for d_1 among each trial function class we consider. In Figure 2, this is shown by the boxes 'spherical harmonics' and 'radial basis functions' which lead to 'set of candidates'. Mathematically, in this step, we seek

$$\left(\alpha_1^{\mathcal{C}}, d_1^{\mathcal{C}}\right) = \underset{(\alpha,d)\,\in\,\mathcal{C}\times\mathbb{R}}{\arg\min} \left(\|y - \mathcal{T}_{\daleth}(F_0 + \alpha d)\|_{\mathbb{R}^\ell}^2 + \lambda\|(F_0 + \alpha d)\|_{\mathcal{H}_2}^2\right),$$

where $\mathcal{C}$ denotes one trial function class, i.e. $\mathcal{C} = [\mathcal{N}]_{\mathrm{SH}}$ the set of spherical harmonics or $\mathcal{C} = [\mathring{\mathbb{B}}_1(0)]_{\mathrm{APK}}$ the set of Abel–Poisson kernels. Then we have again a finite (but optimized) set of trial functions and can choose d_1 from this set of candidates in the common fashion of the RFMP by comparing how well each one minimizes the Tikhonov functional. The candidate that minimizes the Tikhonov functional among all candidates is chosen as d_1 ('choose best candidate as d_{n+1}').

Then we compute the necessary updates of the RFMP-routine as described, for example, in [48]. Next, we check the termination criteria for the learning algorithm. We adopt the termination criteria of the RFMP which are up to now the norm of the residual or the size of the currently chosen coefficient (this would be α_1 at this stage) being smaller than a given threshold or a maximal number of iterations. Either they are not fulfilled, then we search for the next element d_2 in the same manner as we found d_1. Or we have a fulfilled termination criterion. In this case, we stop the RFMP and, thus, the learning of the dictionary. We obtain the same output as in a non-learning RFMP which is an approximation of the given signal. Additionally, the learnt dictionary is defined as the set of all chosen elements in this approximation. For the sake of completeness, note that for an arbitrary iteration step n, the objective to seek α_{n+1} and d_{n+1} is given by

$$\left(\alpha_{n+1}^{\mathcal{C}}, d_{n+1}^{\mathcal{C}}\right) = \underset{(\alpha,d)\,\in\,\mathbb{R}\times\mathcal{C}}{\arg\min} \left(\|y - \mathcal{T}_{\daleth}(F_n + \alpha d)\|_{\mathbb{R}^\ell}^2 + \lambda\|(F_n + \alpha d)\|_{\mathcal{H}_2}^2\right),$$

9

for

$$\mathcal{C} = [\mathcal{N}]_{\mathrm{SH}} \text{ and } \mathcal{C} = \left[\mathring{\mathbb{B}}_1(0)\right]_{\mathrm{APK}},$$

respectively.

Before we can state the learning algorithm itself, we first have to define an objective function for the determination of the candidates.

Definition 6. *For the Sobolev space $\mathcal{H}_2$, we define the objective function of the RFMP in the n-th iteration step as*

$$\mathrm{RFMP}(d; n) \coloneqq \frac{\left(\langle R^n, \mathcal{T}_\neg d\rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d\rangle_{\mathcal{H}_2}\right)^2}{\|\mathcal{T}_\neg d\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_2}^2}, \tag{9}$$

where d is a trial function, R^n is the current residual, F_n the current approximation and $\mathcal{T}, \lambda$ depend on the linear inverse problem.

Theorem 7. *The minimization of the Tikhonov functional in the n-th step of the RFMP with respect to a trial function d and a real coefficient α as seen in (5) is equivalent to the maximization of $\mathrm{RFMP}(\cdot; n)$ with respect to a trial function d.*

Proof. The proof is analogous to the corresponding proofs in [6, 29]. $\qquad\square$

All in all, we can now state the learning algorithm. We will explain the steps of this algorithm in further detail below. Note that we determine a preliminarily optimal Abel–Poisson kernel from a discretely parametrized dictionary $[\hat{K}]_{\mathrm{APK}}$ and use this as a starting point for the optimization procedure for a continuously parametrized ansatz which uses $[\mathring{\mathbb{B}}_1(0)]_{\mathrm{APK}}$.

Algorithm 8. *We obtain a learnt dictionary for the RFMP as follows. Let $\mathcal{T}_\neg F = y$ be the linear inverse problem under investigation, $\mathcal{H}_2$ the Sobolev space from Definition 5 and λ the regularization parameter.*

(S0) initialize: at least one termination criterion (maximal number of iterations I and/or prescribed accuracy of the relative data error and/or minimal size of the coefficients); data vector y; initial approximation F_0; sets $\hat{N}$ as well as $\hat{K}$ as in Definition 1 and starting dictionary $\mathcal{D} = [\hat{N}]_{\mathrm{SH}} + [\hat{K}]_{\mathrm{APK}}$

(S1) set $R^0 \coloneqq y - \mathcal{T}_\neg F_0$ and compute $\langle F_0, d_i\rangle_{\mathcal{H}_2}$ for $d_i \in \mathcal{D}$.

(S2) compute $\mathcal{T}_\neg d$ for each $d \in \mathcal{D}$ evaluated at the data points of y and $\langle d_i, d_j\rangle_{\mathcal{H}_2}$ for each two d_i, $d_j \in \mathcal{D}$.

(S3) while (termination not fulfilled)

    (S3.1) compute candidate

$$d_{n+1}^{\mathrm{SH}} = \arg\max\left\{\mathrm{RFMP}(d; n) \,\Big|\, d \in \left[\hat{N}\right]_{\mathrm{SH}}\right\}$$

    (S3.2) compute local solution

$$d_{n+1}^{\mathrm{APK,loc}} = \arg\max\left\{\mathrm{RFMP}(d; n) \,\Big|\, d \in \left[\hat{K}\right]_{\mathrm{APK}}\right\}$$

    (S3.3) compute optimal solution

$$d_{n+1}^{\mathrm{APK}} = \arg\max\left\{\mathrm{RFMP}(d; n) \,\Big|\, d \in \left[\mathring{\mathbb{B}}_1(0)\right]_{\mathrm{APK}}\right\}$$

10

(S3.4) choose

$$d_{n+1} = \arg\max \left\{ \mathrm{RFMP}\left(d_{n+1}^{\mathrm{SH}}; n\right), \right.$$
$$\left. \mathrm{RFMP}\left(d_{n+1}^{\mathrm{APK,loc}}; n\right), \ \mathrm{RFMP}\left(d_{n+1}^{\mathrm{APK}}; n\right) \right\}$$

(S3.5) compute

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{T}\!\urcorner d_{n+1}\rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d_{n+1}\rangle_{\mathcal{H}_2}}{\|\mathcal{T}\!\urcorner d_{n+1}\|_{\mathbb{R}^\ell}^2 + \lambda \|d_{n+1}\|_{\mathcal{H}_2}^2}$$

(S3.6) set $R^{n+1} := R^n - \alpha_{n+1}\mathcal{T}\!\urcorner d_{n+1}$
(S3.7) for $d_i \in \mathcal{D}$ compute $\langle F_{n+1}, d_i\rangle_{\mathcal{H}_2} = \langle F_n, d_i\rangle_{\mathcal{H}_2} + \alpha_{n+1}\langle d_{n+1}, d_i\rangle_{\mathcal{H}_2}$
(S3.8) increase n by 1
(S4) result: approximation $F_M = \sum_{i=1}^M \alpha_i d_i$ after iteration step M at termination; learnt dictionary

$$\mathcal{D}^* = [N^*]_{\mathrm{SH}} + [K^*]_{\mathrm{APK}},$$
$$N^* = \{(n_i, j_i) \mid \text{there exists } i \in \{1, ..., M\} \text{ such that } Y_{n_i, j_i} = d_i\},$$
$$K^* = \{(r\xi)^{(i)} \mid \text{there exists } i \in \{1, ..., M\} \text{ such that } P((r\xi)^{(i)}, \cdot) = d_i\}$$

3.3 Determination of candidates

For practice, the question remains how the candidates in each trial function class under consideration are determined. We need to explain what is done in S3.1 to S3.3.

First of all, we consider the determination of a spherical harmonic candidate. We want to seek the best-fitting function among all spherical harmonics up to a certain degree $\nu \in \mathbb{N}$. We have to learn the size of ν. It is defined in Algorithm 8 which specific spherical harmonics with a degree up to ν we insert into the learnt dictionary.

The idea is to allow the choice of spherical harmonics up to a degree $\widetilde{N} \in \mathbb{N}$ (i.e. $\hat{N} = \{(n, j) \mid n \in \mathbb{N}_0, \ n \leq \widetilde{N}, \ j = -n, ..., n\}$) which is probably not chosen in practice. For example, the data resolution can provide a threshold up to which a resolution appears to be realistic (as it is also done for other gravitational models like EGM or models based on CHAMP, GRACE and GOCE data). If the LRFMP chooses only spherical harmonics with a lower degree ν, we have a truly learnt bound $\nu < \widetilde{N}$. However, note that the higher we choose $\widetilde{N}$, the more expensive is the preprocessing of the algorithm. The candidate d_{n+1}^{SH} which the LRFMP chooses in each iteration step can be chosen as in the RFMP by comparing $\mathrm{RFMP}(Y_{m,k}; n)$ for all spherical harmonics up to degree $\widetilde{N}$.

Unfortunately, for meaningful results with respect to learning spherical harmonics, we are lacking appropriate data sets. Both the EGM2008 as well as the GRACE data are given to us only as coefficients for a representation in spherical harmonics. Thus, if we allow the LRFMP to choose an optimal spherical harmonic with an arbitrarily high degree, naturally, the algorithm tends to choose more spherical harmonics than other trial functions. If we were able to work with data not given in this polynomial representation, we would assume that this effect diminishes.

Next, we consider the determination of the candidate d_{n+1}^{APK} from the Abel–Poisson kernels in S3.3. In this case, the minimization of the Tikhonov functional is modelled as a non-linear constrained optimization problem. The solution of this problem yields the respective candidate. Note that we do not seek the minimizer of a function, but the minimizer of a functional among a set of functions. Therefore, we have to define the trial functions dependent on their characteristics as we did in (4).

The model of the optimization problem is given as follows. Due to Theorem 7, in the n-th step of the LRFMP, we consider the optimization problem

$$\mathrm{RFMP}(P(x,\cdot);n) \to \max!$$

for learning a dictionary for the RFMP. The optimization with respect to a trial function d can be modelled by an optimization with respect to the characteristics of each trial function. However, these characteristics yield a constraint for the optimization problem. Abel–Poisson kernels are given as $K_h(\xi\cdot) = P(h\xi,\cdot)$, $h \in [0,1)$ and $\xi \in \Omega$ in (4), see, for example, [11, p. 132] or [13, p. 52]. Here, ξ is the centre of the radial basis function and h is the parameter which controls the localization. Therefore, the kernels are well-defined only in the interior of the unit ball and the constraint is given by $\|x\|_{\mathbb{R}^3}^2 < 1$ for $x = h\xi$.

Definition 9. *The optimal candidate* d_{n+1}^{APK} *among the set of Abel–Poisson kernel*

$$\left\{ P(x,\cdot) \mid x \in \mathring{\mathbb{B}}_1(0) \right\}$$

is given by the solution of the optimization problem

$$\mathrm{RFMP}(P(x,\cdot);n) \to \max! \qquad s.t. \qquad \|x\|_{\mathbb{R}^3}^2 < 1.$$

Note that the maximizer is not necessarily unique. In this case, we use one representative among the maximizers.

We prefer a gradient-based approach. Thus, we have to compute the derivatives of (9) with respect to $x \in \mathbb{R}^3$. In general, this can be done by applying the quotient rule and computing the derivatives of the inner products and norms in the de-/nominator separately. We state the results of this derivation at this point. A detailed derivation is given in Appendix A.

Theorem 10. *We define some abbreviation terms and state their derivatives. Let R^n be the current residual of size ℓ and F_n the current approximation, i.e. $F_n = \sum_{i=1}^{n} \alpha_i d_i$ for d_i being a spherical harmonic Y_{n_i,j_i} or an Abel–Poisson kernel $P\left(x^{(i)},\cdot\right)$. Moreover, let $\mathcal{T}_{\daleth}$ be the upward continuation operator and σ the respective satellite orbit. Further, let the data be given on a point grid $\{\sigma\eta^{(i)}\}_{i=1,\dots,\ell}$, $\eta_i \in \Omega$. We consider the Tikhonov functional with a penalty term dependent on the norm of the Sobolev space $\mathcal{H}_2$. At last, P_n denotes a Legendre polynomial and ε^r, ε^φ, ε^t represents the common local orthonormal basis vectors (up, East and North) in $\mathbb{R}^3$, see (20). Then we have for an Abel–Poisson kernel $P(x,\cdot)$ and $x(r,\varphi_x,t_x) \in \mathring{\mathbb{B}}_1(0)$ the terms*

$$a_1(P(x,\cdot)) := \langle R^n, \mathcal{T}_{\daleth} P(x,\cdot) \rangle_{\mathbb{R}^\ell} = \sum_{i=1}^{\ell} \frac{R_i^n}{4\pi} \frac{\sigma^2 - |x|^2}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^{3/2}}, \tag{10}$$

$$a_2(P(x,\cdot)) := \langle F_n, P(x,\cdot) \rangle_{\mathcal{H}_2} = \sum_{i=1}^{n} \alpha_i \begin{cases} T_1, & d_i = Y_{n_i,j_i} \\ T_2, & d_i = P\left(x^{(i)},\cdot\right), \end{cases} \tag{11}$$

$$T_1 := (n_i + 0.5)^4 |x|^{n_i} Y_{n_i,j_i}\left(\frac{x}{|x|}\right)$$

$$T_2 := \sum_{n=0}^{\infty} (n + 0.5)^4 \left(|x_i||x|\right)^n \frac{2n+1}{4\pi} P_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right)$$

$$b_1(P(x,\cdot)) := \|\mathcal{T}_{\daleth} P(x,\cdot)\|_{\mathbb{R}^\ell}^2 = \sum_{i=1}^{\ell} \frac{\left(\sigma^2 - |x|^2\right)^2}{16\pi^2 \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^3}, \tag{12}$$

and

$$b_2(P(x,\cdot)) := \|P(x,\cdot)\|_{\mathcal{H}_2}^2 = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi}(n+0.5)^4 |x|^{2n}. \tag{13}$$

Their partial derivatives with respect to x_j are given by

$$\partial_{x_j} a_1(P(x,\cdot))$$
$$= -\sum_{i=1}^{\ell} \frac{R_i^n}{4\pi} \left[\frac{2x_j}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^{3/2}} + \frac{3\left(\sigma^2 - |x|^2\right)\left(x_j - \sigma \eta_j^{(i)}\right)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^{5/2}} \right], \tag{14}$$

$$\partial_{x_j} a_2(P(x,\cdot)) = \sum_{i=1}^{n} \alpha_i \begin{cases} \partial_{x_j} T_1, & d_i = Y_{n_i,j_i} \\ \partial_{x_j} T_2, & d_i = P\left(x^{(i)}, \cdot\right), \end{cases} \tag{15}$$

$$\partial_{x_j} T_1 = (n_i + 0.5)^4 \partial_{x_j}\left(|x|^{n_i} Y_{n_i,j_i}\left(\frac{x}{|x|}\right)\right)$$

$$\partial_{x_j} T_2 = \sum_{n=0}^{\infty} (n + 0.5)^4 \left(|x_i|\right)^n \frac{2n+1}{4\pi} \partial_{x_j}\left(|x|^n P_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right)\right)$$

$$\partial_{x_j} b_1(P(x,\cdot))$$
$$= -\frac{1}{16\pi^2}\sum_{i=1}^{\ell} \left[\frac{4x_j\left(\sigma^2 - |x|^2\right)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^3} + \frac{6\left(\sigma^2 - |x|^2\right)^2\left(x_j - \sigma\eta_j^{(i)}\right)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^4} \right], \tag{16}$$

and

$$\partial_{x_j} b_2(P(x,\cdot)) = \sum_{n=0}^{\infty} \frac{2n^2 + n}{2\pi}(n+0.5)^4 |x|^{2n-2} x_j. \tag{17}$$

With respect to the derivative of a_2, we have further

$$\nabla\left(|x|^n Y_{n,j}\left(\frac{x}{|x|}\right)\right)$$
$$= \varepsilon^r n |x|^{n-1} Y_{n,j}\left(\frac{x}{|x|}\right) + |x|^{n-1}\varepsilon^\varphi \frac{j}{\sqrt{1-t^2}} Y_{n,-j}\left(\frac{x}{|x|}\right)$$
$$+ |x|^{n-1}\varepsilon^t \sqrt{1-t_x^2}\sqrt{\frac{2n+1}{4\pi}\frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t_x) \begin{cases} \sqrt{2}\cos(j\varphi_x), & j < 0 \\ 1, & j = 0 \\ \sqrt{2}\sin(j\varphi_x), & j > 0 \end{cases} \tag{18}$$

and

$$\nabla\left(|x|^n P_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right)\right)$$
$$= |x|^{n-1}\left[\varepsilon^r n P_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right) + \varepsilon^\varphi P'_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right)\left(\frac{x_i}{|x_i|} \cdot \varepsilon^\varphi\right)\right.$$
$$\left. + \varepsilon^t P'_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right)\left(\frac{x_i}{|x_i|} \cdot \varepsilon^t\right)\right]. \tag{19}$$

For a proof, see Appendix A.

Theorem 11. *With the abbreviations and derivatives from Theorem 10, the partial derivatives ∂_{x_j}, $j = 1, 2, 3$, of $\mathrm{RFMP}(\cdot; n)$ are given by*

$$\partial_{x_j} \mathrm{RFMP}(P(x, \cdot); n)$$
$$= [b_1(P(x, \cdot)) + \lambda b_2(P(x, \cdot))]^{-2} \left[2[a_1(P(x, \cdot)) - \lambda a_2(P(x, \cdot))] \right.$$
$$\times [\partial_{x_j} a_1(P(x, \cdot)) - \lambda \partial_{x_j} a_2(P(x, \cdot))][b_1(P(x, \cdot)) + \lambda b_2(P(x, \cdot))]$$
$$\left. - [a_1(P(x, \cdot)) - \lambda a_2(P(x, \cdot))]^2 [\partial_{x_j} b_1(P(x, \cdot)) + \lambda \partial_{x_j} b_2(P(x, \cdot))] \right].$$

Proof. We only apply the common rules for derivatives. ∎

Thus, for Abel–Poisson kernels, we determined the gradient of the modelled objective functions $\mathrm{RFMP}(\cdot; n)$, $n \in \mathbb{N}$, analytically such that we are able to use a gradient-based optimization method. We use the primal-dual interior point filter line search method IPOPT. To enable parallelization, we installed the linear solver `ma97` from the HSL package. For more details on the IPOPT algorithm and the HSL package, see [18, 37, 50, 52, 53, 54]. In practice, we set a few options manually which we explain in the necessary contexts later. In all other cases, we use the default option values. However, note that due to [50], we can only expect to obtain local solutions of the optimization problem. Furthermore, this algorithm uses a starting point to start its procedure towards an optimal Abel–Poisson kernel. This starting point is denoted by $d_{n+1}^{\mathrm{APK,loc}}$ in Algorithm 8 and is computed in S3.2. We obtain $d_{n+1}^{\mathrm{APK,loc}}$ by comparing the objective value $\mathrm{RFMP}(P(x, \cdot); n)$ for a selection of kernels given to the algorithm in the discrete dictionary by $[\hat{K}]_{\mathrm{APK}}$ and choosing the one with the highest value. As we have computed this kernel, we make use of it in S3.4 as well in case the optimization failed to find a better kernel.

3.4 Additional features for practice

The previously presented algorithm gives us a first and basic learning technique. During its development, we faced several problems dependent on the choice of the data and the given inverse problem. In order to overcome these difficulties, we introduced some additional features, which will be explained in the following.

First of all, the data of the monthly variation of the gravitational potential provided by GRACE attains very small values. In our experiments, these values lie in the interval [-0.1,0.1]. When inserting this data into the objective function for determining an optimal Abel–Poisson kernel, the IPOPT solver fails to find a solution at first. Thus, we set the option obj_scaling_factor to -10^{10}. For the EGM-data, we only use -1 to perform a maximization instead of a minimization. Note that the scaled objective is only seen internally to support the optimizer.

Next, to cut runtime and possible round-off errors, we implemented a restart method. We initiate a new run of the algorithm by resetting F_E to zero after a previously chosen iteration number E. Note that, in contrast to the restart procedure of the ROFMP, see, for example, [48], we also reset the regularization term $\lambda \|F_E\|_{\mathcal{H}_2}^2$ to zero. In our experiments, we used $E = 250$.

Furthermore, we saw that the learnt dictionary heavily depends on the regularization parameter. Thus, we have to use the same regularization parameter as we used during learning when applying the learnt dictionary. Moreover, in contrast to previous works on the IPMPs, see, for instance, [48], the use of a non-stationary regularization parameter is necessary when learning a dictionary and applying it. In the previous works, the idea of a non-stationary, decreasing regularization parameter was explained in order to emphasize

accuracy instead of smoothness of the obtained approximation. However, the improvement of the results did not justify the additional computational expenses of choosing a parameter and a decrease routine. Nonetheless, we reconsidered this idea since the main aim of the LRFMP is to learn a dictionary and a decreasing parameter appears to guide the learning process positively. Thus, we use a non-stationary regularization parameter $\lambda_n = (\lambda_0 \|R_0\|_{\mathbb{R}^\ell}/(n+1))$, where n is the current iteration number and λ_0 is an optimal regularization parameter. Our experiments show that with a decreasing regularization parameter, we determine a dictionary which yields a better approximation when applied. Hence, with a learnt dictionary, the use of a non-stationary regularization parameter has an impact on the result of the RFMP.

Next, for the case of a high satellite orbit or the seasonal variation in the gravitational potential obtained via GRACE, we developed a dynamic dictionary approach. Note that in previous literature on dictionary learning, see, for example, [40], it was mentioned that the structure of the input data could or even should be considered when learning a dictionary. We developed two strategies to learn a dynamic dictionary whose combination works well in the experiments under considerations.

First of all, we demand that the first 250 learnt dictionary elements are spherical harmonics. It seems that after these 250 iterations, the current residual R_n has a rougher structure than the initial residual R_0. Then the optimization routine finds more easily a more sensible solution and, therefore, learns a better dictionary. Additionally, we allow to only choose from the first $n+1$ learnt dictionary elements in the n-th iteration step of the RFMP when we use the learnt dictionary. In this case, the order of chosen dictionary elements from the LRFMP has to be preserved. In this way, the optimized trial function which was chosen in the n-th step of the LRFMP can be chosen in the RFMP as well. Additionally, we save runtime as the dictionary size is small at the beginning. At last, this also treats possible complications which might otherwise arise from the non-stationary regularization parameter. The fact that we decrease this parameter yields the choice of very localized trial functions in later iterations of the LRFMP. If we allowed the RFMP to choose them prematurely this would only lead to a reduced data error and not a better approximation as we have seen in practice. If we allow the RFMP to only choose from the first n learnt dictionary elements, we can prevent it to choose very localized kernels prematurely.

With these features included in the LRFMP, we currently run our experiments.

4 Numerical results

4.1 Setting of the experiments

In this section, we present the results of the learnt dictionary compared to a manually chosen dictionary applied in the standard RFMP. We will use data from the EGM2008 as well as GRACE satellite data. The EGM2008 data is evaluated up to degree 1500. For the GRACE data, we computed the meanfield from 2003 to 2013 averaged from the release 5 products from JPL, GFZ and CSR as was proposed in [44]. This meanfield was subtracted from the data corresponding to May 2008. Note that in May, traditionally, the wet season in the Amazon basin is about to end such that we can expect a concentration of masses in this region. Additionally, we smoothed the data with a Cubic Polynomial Scaling Function (see, for instance, [46] and [15, p. 295]) of scale 5.

In all experiments, we compute the data on an equidistributed Reuter grid of 12684 data points, see, for instance, [41] and [28, p. 137]. We choose the constant regularization parameter λ of the RFMP and starting regularization parameter λ_0 of the LRFMP (see Subsection 3.4) such that they yield the lowest relative approximation error after 2000

iterations. In detail, we choose $\lambda = 10^{-2}$ for both experiments and $\lambda_0 = 10^{-4}$ for the experiment with EGM data and $\lambda_0 = 10^{-1}$ for the experiment with GRACE data.

The arbitrarily chosen dictionary with which we compare the learnt dictionary is chosen similarly to [48]. We choose

$$\mathcal{D}^{\mathrm{m}} := [N^{\mathrm{m}}]_{\mathrm{SH}} + [K^{\mathrm{m}}]_{\mathrm{APK}}$$

with

$$N^{\mathrm{m}} := \{(n, j) \mid n = 0, ..., 25; \ j = -n, ..., n\},$$
$$K^{\mathrm{m}} := \{r\xi \mid r \in R^{\mathrm{m}}, \ \xi \in X^{\mathrm{m}}\},$$
$$R^{\mathrm{m}} := \{0.75, 0.80, 0.85, 0.89, 0.91, 0.93, 0.94, 0.95, 0.96, 0.97\}$$

and an equidistributed Reuter grid X^{m} of 4551 grid points on the sphere. All in all, the manually chosen dictionary contains 46186 dictionary elements.

The LRFMP needs a starting dictionary as well to provide the spherical harmonics and starting points for the optimization problem. For an equidistributed Reuter grid X^{s} of 1129 grid points, we use

$$\mathcal{D}^{\mathrm{s,EGM}} := \left[N^{\mathrm{s,EGM}}\right]_{\mathrm{SH}} + [K^{\mathrm{s}}]_{\mathrm{APK}}$$

and

$$\mathcal{D}^{\mathrm{s,GRACE}} := \left[N^{\mathrm{s,GRACE}}\right]_{\mathrm{SH}} + [K^{\mathrm{s}}]_{\mathrm{APK}},$$

respectively, with

$$N^{\mathrm{s,EGM}} := \{(n, j) \mid n = 0, ..., 100; \ j = -n, ..., n\},$$
$$N^{\mathrm{s,GRACE}} := \{(n, j) \mid n = 0, ..., 60; \ j = -n, ..., n\},$$
$$K^{\mathrm{s}} := \{r\xi \mid r = 0.94, \ \xi \in X^{\mathrm{s}}\},$$

From experience, we know that it is better to cut down the number of scales r than the number of centres ξ of the Abel–Poisson kernels. Thus, the starting dictionary contains 11330 dictionary elements in the case of EGM2008 data and 4850 dictionary elements in the case of GRACE data.

We use the IPOPT optimizer with the linear solver `ma97` from HSL. For EGM data, we demand a desired and acceptable tolerance of 10^{-4} of the optimal solution. For GRACE data, we demand these tolerances to be only 10^0 due to the scaling of the objective function explained in Section 3.4. As we need to ensure that the constraint is not violated during the optimization process, we set $r^2 < 0.98999999^2$ as well as the options theta_max_fact 10^{-2} and watchdog_shortened_iter_triggered 0 in practice. For details on these options, see the IPOPT documentation [50].

We terminate the LRFMP as well as the RFMP in all cases when one of the following termination criteria is fulfilled. Either the relative data error is less than 10^{-8} or we reached 3000 iterations. We also terminate the LRFMP when the last chosen coefficient α_{n+1} is less than 10^{-5}. We experienced that otherwise the additionally learnt dictionary elements do not improve the solution and we can easily stop the learning process at this stage. Note that in stopping after at most 3000 iterations, we also limit the learnt dictionary to at most 3000 learnt trial functions. This means, that the learnt dictionary is much smaller than the manually chosen dictionary with which we compare the learnt dictionary. Further, note

16

that, due to its size, the manually chosen dictionary obviously has a much larger storage demand.

The results which we will compare here are obtained as follows: in one case, we use the RFMP with the manually chosen dictionary $\mathcal{D}^{\mathrm{m}}$. In the second case, we first learn a dictionary $\mathcal{D}^{*,\bullet}$ (see Algorithm 8) for $\bullet \in \{\mathrm{EGM}, \mathrm{GRACE}\}$ by using the LRFMP (which requires the starting dictionary $\mathcal{D}^{\mathrm{s},\bullet}$) and then run the RFMP with the learnt dictionary $\mathcal{D}^{*,\bullet}$. The major question is: is the learnt dictionary able to yield better results than the manually chosen dictionary?

The plots shown in this paper are done with MATLAB. Note that the colours for the results of the GRACE data are flipped in comparison to the results of the EGM data. This is done in order to emphasize wet regions with blue colour and dryer regions with red colour.

4.2 Results

In Figure 3, the results of the two experiments are shown. The first row shows the results with the EGM data. The second row depicts the results of the GRACE data. In the left-hand column the solution is given. The middle column shows the absolute approximation error of the RFMP with the manually chosen dictionary. The right-hand column depicts this error of the RFMP with the learnt dictionary. We adjusted the scales of the values for a better comparison.

Obviously, in both cases the algorithm is able to construct a good approximation. The relatively low errors occur basically within regions where more local structures are given. These regions are in the case of the EGM data in particular the Andean region as well as the Himalayas and the borders of the tectonic plates in Asia. In the case of monthly GRACE data, the masses in the Amazon basin originating from the ending wet season show the strongest structure. As we only allow 3000 iterations, it can be expected that such regions cannot be approximated perfectly.

Particularly interesting are the results in the right-hand column which were obtained with the learnt dictionary. Clearly, in both scenarios, the approximation error is notably reduced. This is, in particular, also the case in the regions with localized anomalies.

In Table 1, the relative approximation and data errors after 3000 iterations of the RFMP with the manually chosen dictionary and the learnt dictionary are given. Furthermore, the (currently) needed CPU-runtime for the experiments are presented in the last column. We notice that we do not only obtain a smaller relative approximation error with the learnt dictionary but also a smaller relative data error.

We state the CPU-runtime in hours for the sake of completeness. Although we see that we need less time to compute and preprocess the learnt dictionary than to preprocess the manually chosen dictionary, these results are to be understood with care because we

	$\#\mathcal{D}$	rel. approximation error	rel. data error	CPU-runtime in h
RFMP*	46186	0.000794	0.065830	299.82
RFMP**	≤ 3000	0.000455	0.047092	41.34
RFMP*	46186	0.001461	0.057765	295.49
RFMP**	≤ 3000	0.000306	0.046404	7.5

Table 1: Comparison of RFMP with a manually chosen dictionary (RFMP*) and a learnt dictionary (RFMP**). Upper comparison with respect to EGM2008 data. Lower comparison with respect to GRACE data. 3000 iterations allowed in all experiments.
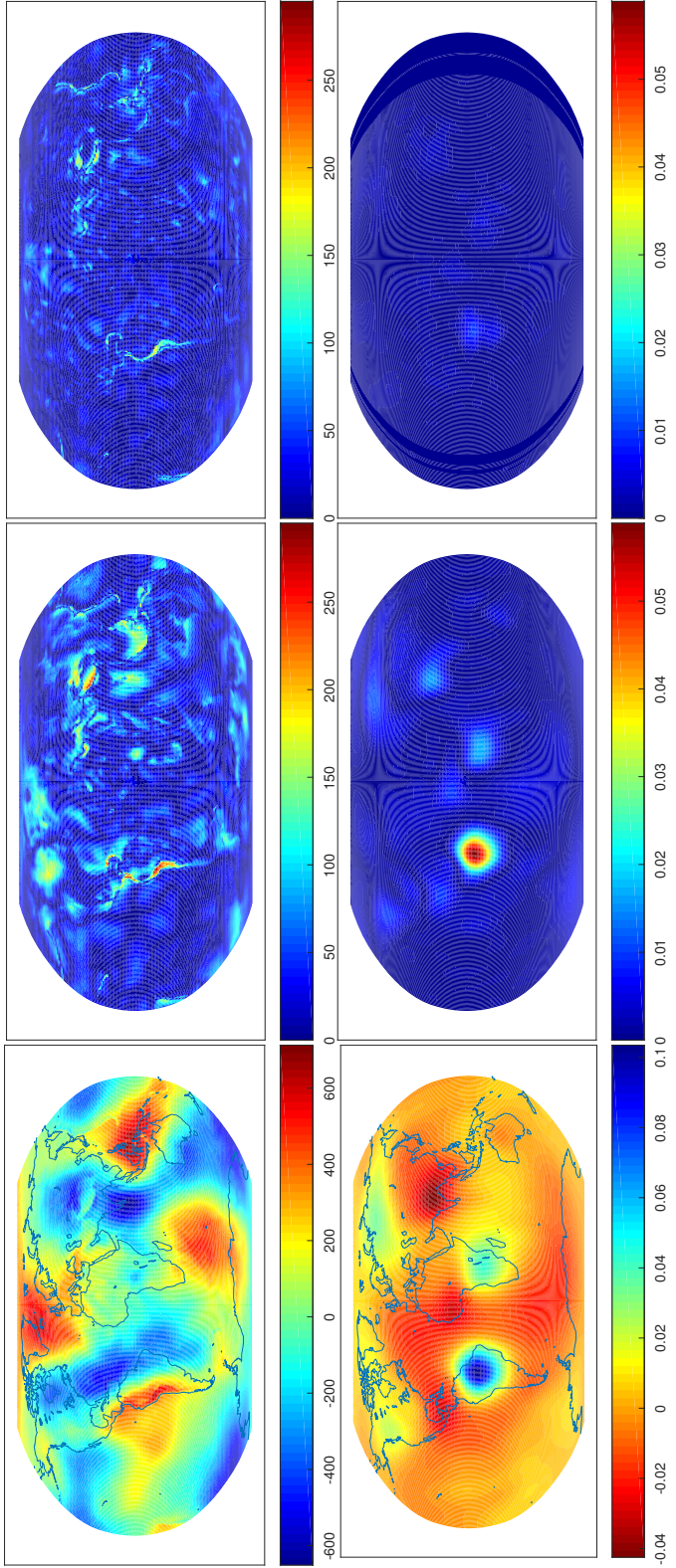
Figure 3: Results of the RFMP with a manually chosen dictionary and the learnt dictionary. Upper row: Results for EGM2008 data. Lower row: Results for GRACE data. Left: Solution. Middle: Absolute approximation error of RFMP with arbitrary dictionary. Right: Absolute approximation error of RFMP with learnt dictionary. 3000 iterations allowed in all experiments. All values in $\mathrm{m^2/s^2}$.

18

currently do not work with an optimized code.

All in all, the results show that we are able to learn a dictionary which yields a smaller data as well as approximation error than a manually chosen dictionary. In addition, we obtain these results with a sparser dictionary, less storage demand and an appropriate CPU-runtime.

5 Conclusion and Outlook

We started our investigations by aiming to improve our methods for gravity field modelling. In practice, we considered the RFMP for now. We expected to reduce the computational demands of the RFMP and the approximation error if a learnt dictionary is used rather than a manually chosen 'rule-of-thumb' dictionary.

In this paper, we presented a first approach to learn a dictionary of spherical harmonics and Abel–Poisson kernels for the downward continuation of gravitational data. In the numerical tests, we used data generated from EGM2008 and GRACE models. The idea of our learning approach is to iteratively minimize a Tikhonov functional over an infinite set of these trial functions. We do so by using non-linear constrained optimization techniques. Our results show that we obtain better results with respect to the relative data and approximation error when applying the learnt dictionary than when using a manually chosen dictionary in the RFMP. Moreover, we obtain these results with a sparser dictionary and less storage demand. Further, even non-optimized code yields satisfactory runtime results.

In the future research, we aim to transfer this learning approach to the ROFMP and enlarge its idea to Abel–Poisson wavelets and Slepian functions. Further, the presented learning algorithm can be viewed as one component of a structured learning technique. The question is whether we can determine an optimized learnt dictionary via learning from the results of the here presented strategy. In this way, we aim to build a learning hierarchy for the dictionary where the present technique represents the second level and the first level still needs to be developed. This could also lead to an automation of the additional features which we explained in Subsection 3.4. An additional objective is to obtain a dictionary for GRACE from a given set of training data to apply it with new test data such that we can provide an optimal dictionary for GRACE-FO satellite data. Further, we need to consider the theoretical aspects of the learning algorithm like determining a quantitative measure for the quality of a dictionary and investigating how the learnt dictionary of the LRFMP is related to that measure. In addition, with respect to practical aspects, we plan on optimizing our code to have more meaningful runtime results.

A Gradient of the objective function with respect to Abel–Poisson kernels

In this appendix, we prove Theorem 10.

First considerations   We discuss the terms for the upward continuation operator as given in (1). Note that the Euclidean inner product of two vectors is emphasized by using a '·' at the particular positions. Additionally, we make use of the following basic aspects.

In geomathematics, a common orthonormal basis in $\mathbb{R}^3$ is given by

$$\varepsilon^r(\varphi,t) := \begin{pmatrix} \sqrt{1-t^2}\cos(\varphi) \\ \sqrt{1-t^2}\sin(\varphi) \\ t \end{pmatrix}, \quad \varepsilon^\varphi(\varphi,t) := \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \\ 0 \end{pmatrix}, \quad \varepsilon^t(\varphi,t) := \begin{pmatrix} -t\cos(\varphi) \\ -t\sin(\varphi) \\ \sqrt{1-t^2} \end{pmatrix}, \quad (20)$$

see, for example, [28, p. 86]. Note that it holds $\partial_\varphi \varepsilon^r = \sqrt{1-t^2}\varepsilon^\varphi$ and $\partial_t \varepsilon^r = \frac{1}{\sqrt{1-t^2}}\varepsilon^t$. For the gradient ∇, we will use a Cartesian definition as well as its decomposition into radial and angular parts. We have

$$\nabla = \left(\partial_{x_j}\right)_{j=1,2,3} = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\left(\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}}\frac{\partial}{\partial \varphi} + \varepsilon^t \sqrt{1-t^2}\frac{\partial}{\partial t}\right), \quad (21)$$

see, for instance, [28, p. 87]. Next, we consider the following recurring inner products.

$$\langle Y_{m,k}, Y_{n,j}\rangle_{\mathrm{L}^2(\Omega)} = \delta_{n,m}\delta_{j,k}, \quad (22)$$

$$\langle P(x,\cdot), Y_{n,j}\rangle_{\mathrm{L}^2(\Omega)} = |x|^n Y_{n,j}\left(\frac{x}{|x|}\right), \quad (23)$$

see, for instance, [48, p. 114]. At last, we note one specific property of the fully normalized spherical harmonics. It holds

$$\frac{\partial}{\partial \varphi}Y_{n,j}(\xi(\varphi,t)) = jY_{n,-j}(\xi(\varphi,t)), \quad (24)$$

see, for instance, [25]. Now, we can compute the terms in Theorem 10.

The term $a_1(P(x,\cdot))$ and its derivative   Obviously, for the formulation of $a_1(Px,\cdot)$ in (10), we only have to show that

$$\mathcal{T}P(x,\cdot)(\eta) = \frac{\sigma^2 - |x|^2}{4\pi(\sigma^2 + |x|^2 - 2\sigma x\cdot \eta)^{3/2}} \quad (25)$$

for any $\eta \in \Omega$. We start at the left-hand side of (25).

$$\mathcal{T}P(x,\cdot)(\eta) = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\langle Y_{n,j}, P(x,\cdot)\rangle_{\mathrm{L}^2(\Omega)}\sigma^{-n-1}Y_{n,j}(\eta)$$

$$= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\sigma^{-n-1}|x|^n Y_{n,j}\left(\frac{x}{|x|}\right)Y_{n,j}(\eta)$$

$$= \sum_{n=0}^{\infty}|x|^n \sigma^{-n-1}\frac{2n+1}{4\pi}P_n\left(\frac{x}{|x|}\cdot \eta\right) = \frac{1}{\sigma}P\left(\frac{x}{\sigma},\eta\right)$$

$$= \frac{1}{\sigma}\frac{1-|\frac{x}{\sigma}|^2}{4\pi(1+|\frac{x}{\sigma}|^2 - 2\frac{x}{\sigma}\cdot \eta)^{3/2}}$$

$$= \frac{\sigma^{-2}\left(\sigma^2 - |x|^2\right)}{4\pi\sigma\left(\sigma^{-2}\left(\sigma^2 + |x|^2 - 2\sigma x\cdot \eta\right)\right)^{3/2}} = \frac{\sigma^2 - |x|^2}{4\pi\left(\sigma^2 + |x|^2 - 2\sigma x\cdot \eta\right)^{3/2}}. \quad (26)$$

Its derivative, as used in (14), is obtained via

$$\partial_{x_j} \mathcal{T} P(x, \cdot)(\eta)$$

$$= \frac{\partial}{\partial x_j} \frac{\sigma^2 - |x|^2}{4\pi \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^{3/2}}$$

$$= \frac{-2x_j \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^{3/2} - 3\left(\sigma^2 - |x|^2\right)\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^{1/2}(x_j - \sigma \eta_j)}{4\pi \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^3}$$

$$= -\frac{1}{4\pi} \left[\frac{2x_j}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^{3/2}} + \frac{3\left(\sigma^2 - |x|^2\right)(x_j - \sigma \eta_j)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta\right)^{5/2}} \right].$$

The term $a_2(P(x, \cdot))$ and its derivative For the current approximation F_n, we write $F_n = \sum_{i=1}^n \alpha_i d_i$ for dictionary elements $d_i = Y_{n_i, j_i}$ or $d_i = P\left(x^{(i)}, \cdot\right)$ depending on what element was chosen in the i-th step. Then we can derive the representation (11) of $a_2(P(x, \cdot))$ as follows

$$a_2(P(x, \cdot)) := \langle F_n, P(x, \cdot)\rangle_{\mathcal{H}_2} = \sum_{i=1}^n \alpha_i \langle d_i, P(x, \cdot)\rangle_{\mathcal{H}_2}$$

$$= \sum_{i=1}^n \alpha_i \begin{cases} (n_i + 0.5)^4 |x|^{n_i} Y_{n_i, j_i}\left(\frac{x}{|x|}\right), & d_i = Y_{n_i, j_i} \\ \sum_{n=0}^\infty (n + 0.5)^4 \left(|x_i||x|\right)^n \frac{2n+1}{4\pi} P_n\left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|}\right), & d_i = P\left(x^{(i)}, \cdot\right) \end{cases}$$

due to (22), (23) and the addition theorem for spherical harmonics. The derivative of $a_2(P(x, \cdot))$ as given in (15) is obvious. However, we have to show (18) and (19). We will exchange the term $|x|$ by r as well as $\frac{x}{|x|}$ by ξ and $\frac{x^{(i)}}{|x^{(i)}|}$ by $\xi^{(i)}$ for this. Then we obtain the following results. We first consider (18).

$$\nabla_x \left(r^n Y_{n,j}(\xi)\right)$$

$$= \varepsilon^r \left(\frac{\partial}{\partial r} r^n Y_{n,j}(\xi(\varphi, t))\right) + \frac{1}{r}\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} \left(r^n \frac{\partial}{\partial \varphi} Y_{n,j}(\xi(\varphi, t))\right)$$

$$\quad + \frac{1}{r}\varepsilon^t \sqrt{1-t^2} \left(r^n \frac{\partial}{\partial t} Y_{n,j}(\xi(\varphi, t))\right)$$

$$= \varepsilon^r n r^{n-1} Y_{n,j}(\xi(\varphi, t)) + r^{n-1}\varepsilon^\varphi \frac{j}{\sqrt{1-t^2}} Y_{n,-j}(\xi(\varphi, t))$$

$$\quad + r^{n-1}\varepsilon^t \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t) \begin{cases} \sqrt{2}\cos(j\varphi), & j < 0, \\ 1, & j = 0, \quad (27) \\ \sqrt{2}\sin(j\varphi), & j > 0, \end{cases}$$

where we used (21), (24) and (2). We have to take a closer look at (27) regarding a possible singularity in $t = \pm 1$. The term (27) contains two possibly problematic terms:

$$\frac{j}{\sqrt{1-t^2}} Y_{n,-j}(\xi) \qquad \text{and} \qquad \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t). \qquad (28)$$

We first consider the term on the left-hand side of (28). Obviously, if $j = 0$, this is a removable singularity with a zero value. In the case $j \neq 0$, we recall the definition of the

fully normalized spherical harmonics, which we use in practice, from (2) and (3). Thus, for the problematic term, we have

$$
\frac{j}{\sqrt{1-t^2}} Y_{n,-j}(\xi)
$$

$$
= j \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} \left(1-t^2\right)^{(|j|-1)/2} \left(\frac{\mathrm{d}^{|j|}}{\mathrm{d}t^{|j|}} P_n(t) \right) \begin{cases} \sqrt{2}\cos(j\varphi), & j < 0, \\ 1, & j = 0, \\ \sqrt{2}\sin(j\varphi), & j > 0. \end{cases}
$$

Obviously, there exists a problem only for $\frac{|j|-1}{2} < 0 \Leftrightarrow |j| - 1 < 0 \Leftrightarrow |j| < 1 \Leftrightarrow j = 0$. However, this is excluded in this case. Thus, there is no problem in the term of the left-hand side of (28). With respect to the term on the right-hand side of (28), we have

$$
\sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t)
$$

$$
= \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} \frac{\mathrm{d}}{\mathrm{d}t} \left[\left(1-t^2\right)^{|j|/2} \frac{\mathrm{d}^{|j|}}{\mathrm{d}t^{|j|}} P_n(t) \right]
$$

$$
= \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} \left[-jt \left(1-t^2\right)^{(|j|-1)/2} \frac{\mathrm{d}^{|j|}}{\mathrm{d}t^{|j|}} P_n(t) + \left(1-t^2\right)^{(|j|+1)/2} \frac{\mathrm{d}^{|j|+1}}{\mathrm{d}t^{|j|+1}} P_n(t) \right]
$$

with the definition in (3). Obviously, the problematic term is

$$
\tfrac{j}{2} \left(1-t^2\right)^{(|j|-1)/2} .
$$

If $j = 0$, the term vanishes. If $j \neq 0$, the exponents are non-negative. Thus, also the term on the right-hand side of (28) contains no singularity.

At last, for the derivative of $a_2(P(x,\cdot))$, we need to consider (19). We obtain

$$
\nabla_x \left(r^n P_n(\xi^{(i)} \cdot \xi) \right) = \nabla_x \left(r^n P_n(\xi^{(i)} \cdot \varepsilon^r) \right)
$$

$$
= \varepsilon^r n r^{n-1} P_n(\xi^{(i)} \cdot \varepsilon^r) + r^{n-1} \left(\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \frac{\partial}{\partial \varphi} \varepsilon^r \right) \right.
$$

$$
\left. + \varepsilon^t \sqrt{1-t^2} P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \frac{\partial}{\partial t} \varepsilon^r \right) \right)
$$

$$
= \varepsilon^r n r^{n-1} P_n(\xi^{(i)} \cdot \varepsilon^r) + r^{n-1} \left(\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \sqrt{1-t^2} \varepsilon^\varphi \right) \right.
$$

$$
\left. + \varepsilon^t \sqrt{1-t^2} P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \frac{1}{\sqrt{1-t^2}} \varepsilon^t \right) \right)
$$

$$
= r^{n-1} \left(\varepsilon^r n P_n(\xi^{(i)} \cdot \varepsilon^r) + \varepsilon^\varphi P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \varepsilon^\varphi \right) + \varepsilon^t P'_n(\xi^{(i)} \cdot \varepsilon^r) \left(\xi^{(i)} \cdot \varepsilon^t \right) \right).
$$

With $\varepsilon^r = \frac{x}{|x|}$, this is the formulation of (19).

The term $b_1(P(x, \cdot))$ and its derivative The term $b_1(P(x, \cdot))$ as in (12) is obvious when we take (26) into account. The derivative of $b_1(P(x, \cdot))$ as in (16) is obtained by

$$\partial_{x_j} b_1(P(x, \cdot))$$

$$= \frac{\partial}{\partial x_j} \frac{1}{16\pi^2} \sum_{i=0}^{\ell} \frac{\left(\sigma^2 - |x|^2\right)^2}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^3}$$

$$= \frac{1}{16\pi^2} \sum_{i=0}^{\ell} \left(\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^{-6} \left(-4x_j \left(\sigma^2 - |x|^2\right) \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^3 \right.\right.$$

$$\left.\left. -6 \left(\sigma^2 - |x|^2\right)^2 \left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^2 \left(x_j - \sigma\eta_j^{(i)}\right) \right) \right)$$

$$= -\frac{1}{16\pi^2} \sum_{i=0}^{\ell} \left(\frac{4x_j \left(\sigma^2 - |x|^2\right)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^3} + \frac{6 \left(\sigma^2 - |x|^2\right)^2 \left(x_j - \sigma\eta_j^{(i)}\right)}{\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)}\right)^4} \right).$$

The term $b_2(P(x, \cdot))$ and its derivative The formulation as in (13) of $b_2(P(x, \cdot))$ is due to the following considerations which use (23) and $P_n(1) = 1$, see, for instance, [28, p. 49].

$$\|P(x, \cdot)\|_{\mathcal{H}_2}^2 = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} (n + 0.5)^4 \langle P(x, \cdot), Y_{n,j}\rangle_{L^2(\Omega)}^2$$

$$= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} (n + 0.5)^4 |x|^{2n} \left(Y_{n,j}\left(\frac{x}{|x|}\right) \right)^2$$

$$= \sum_{n=0}^{\infty} (n + 0.5)^4 |x|^{2n} \frac{2n+1}{4\pi} P_n\left(\frac{x}{|x|} \cdot \frac{x}{|x|}\right) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (n + 0.5)^4 |x|^{2n}.$$

The gradient with respect to x is then obtained as follows.

$$\nabla_x b_2(P(x, \cdot)) = \sum_{n=0}^{\infty} (n + 0.5)^4 \frac{2n+1}{4\pi} \nabla_x \left(|x|^{2n}\right). \tag{29}$$

With (21), we have

$$\nabla_x \left(|x|^{2n}\right) = \nabla_x r^{2n} = 2n\varepsilon^r r^{2n-1} = 2n\varepsilon^r |x|^{2n-1}.$$

Inserting this result into (29), we obtain

$$\nabla_x b_2(P(x, \cdot)) = \sum_{n=0}^{\infty} \frac{2n^2 + n}{2\pi} (n + 0.5)^4 \varepsilon^r |x|^{2n-1} = \sum_{n=0}^{\infty} \frac{2n^2 + n}{2\pi} (n + 0.5)^4 x |x|^{2n-2}.$$

This is in accordance with (17). Hence, Theorem 10 is proven. $\square$

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[2] O. Baur. Gravity field of planetary bodies. In E. Grafarend, editor, *Encyclopedia of Geodesy*, pages 1–6. Springer International Publishing Switzerland AG, Cham, 2014.

[3] P. Berkel, D. Fischer, and V. Michel. Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation. *International Journal on Geomathematics*, 1(2):167–204, 2011.

[4] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse Solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[5] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[6] D. Fischer. *Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies*. PhD thesis, University of Siegen, Verlag Dr. Hut, Munich, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2012/544/index.html, 2011.

[7] D. Fischer and V. Michel. Sparse regularization of inverse gravimetry – case study: spatial and temporal mass variations in South America. *Inverse Problems*, 28(6):065012 (34 pp.), 2012.

[8] D. Fischer and V. Michel. Automatic best-basis selection for geophysical tomographic inverse problems. *Geophysical Journal International*, 193(3):1291–1299, 2013.

[9] D. Fischer and V. Michel. Inverting GRACE gravity data for local climate effects. *Journal of Geodetic Science*, 3:151–162, 2013.

[10] F. Flechtner, P. Morton, M. Watkins, and F. Webb. Status of the GRACE follow-on mission. In U. Marti, editor, *Gravity, Geoid and Height Systems. International Association of Geodesy Symposia*, volume 141, pages 117–121. Springer, Cham, 2014.

[11] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Taylor & Francis Group, Boca Raton, 2013.

[12] W. Freeden and M. Gutting. *Special Functions of Mathematical (Geo-)Physics*. Springer, Basel, 2013.

[13] W. Freeden and V. Michel. *Multiscale Potential Theory with Applications to Geoscience*. Birkhäuser, Boston, 2004.

[14] W. Freeden and M. Schreiner. *Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup*. Springer, Berlin, 2009.

[15] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere – with Applications to Geomathematics*. Oxford University Press, Oxford, 1998.

[16] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman, New York, 2009.

[17] M. Gutting, B. Kretz, V. Michel, and R. Telschow. Study on parameter choice methods for the RFMP with respect to downward continuation. *Frontiers in Applied Mathematics and Statistics*, 3, 2017. Article 10. doi: 10.3389/fams.2017.00010.

[18] HSL. A collection of Fortran codes for large scale scientific computation. `http://www.hsl.rl.ac.uk/`, 2018. last accessed 11 Dec, 2018.

[19] M. Kontak. *Novel Algorithms of Greedy-Type for Probability Density Estimation as well as Linear and Nonlinear Inverse Problems*. PhD thesis, University of Siegen, Verlag Dr. Hut, Munich, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1316/index.html, 2018.

[20] M. Kontak and V. Michel. The regularized weak functional matching pursuit for linear inverse problems. *Journal of Inverse and Ill-Posed Problems*, 2018. published online https://doi.org/10.1515/jiip-2018-0013.

[21] M. Kontak and V. Michel. A greedy algorithm for nonlinear inverse problems with an application to nonlinear inverse gravimetry. *GEM – International Journal on Geomathematics*, 9(2):167–198, 2018.

[22] J. Kusche. Time-variable gravity field and global deformation of the Earth. In W. Freeden, M. Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, pages 321–338. Springer, Berlin, Heidelberg, 2nd edition, 2015.

[23] S. Leweke. *The Inverse Magneto-electroencephalography Problem for the Spherical Multiple-shell Modell*. PhD thesis, University of Siegen, 2018. submitted and accepted 2018.

[24] S. Leweke, V. Michel, and A. S. Fokas. Electro-Magnetoencephalography for the spherical multiple-shell model: Novel integral operators with singular-value decompositions. *Siegen preprints on Geomathematics*, 2018.

[25] S. Leweke, V. Michel, and N. Schneider. Vectorial Slepian functions on the ball. *Numerical Functional Analysis and Optimization*, 39(11):1120–1152, 2018.

[26] A. K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[27] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Tansactions on Signal Processing*, 41(12):3397–3415, 1993.

[28] V. Michel. *Lectures on Constructive Approximation – Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball*. Birkhäuser, New York, 2013.

[29] V. Michel. RFMP – An iterative best basis algorithm for inverse problems in the geosciences. In W. Freeden, M. Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, pages 2121–2147. Springer, Berlin, Heidelberg, 2nd edition, 2015.

[30] V. Michel and S. Orzlowski. On the convergence theorem for the Regularized Functional Matching Pursuit (RFMP) algorithm. *GEM – International Journal on Geomathematics*, 8(2):183–190, 2017.

[31] V. Michel and R. Telschow. A non-linear approximation method on the sphere. *GEM – International Journal on Geomathematics*, 5(2):195–224, 2014.

[32] V. Michel and R. Telschow. The regularized orthogonal functional matching pursuit for ill-posed inverse problems. *SIAM Journal on Numerical Analysis*, 54(1):262–287, 2016.

[33] C. Müller. *Spherical Harmonics*. Springer, Berlin, 1966.

[34] H. Moritz. Classical physical geodesy. In W. Freeden, M. Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, pages 253–289. Springer, Berlin, Heidelberg, 2nd edition, 2010.

[35] NASA Jet Propulsion Laboratory. GRACE Tellus. `https://grace.jpl.nasa.gov/publications/`, 2018. last accessed 11 Dec, 2018.

[36] National Geospatial-Intelligence Agency, Office of Geomatics (SN), EGM Development Team. Earth Gravitational Model 2008. `http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/`, 2008. last accessed 11 Dec, 2018.

[37] J. Nocedal, A. Wächter, and R. A. Waltz. Adaptive barrier strategies for nonlinear interior methods. *SIAM Journal on Optimization*, 19(4):1674–1693, 2008.

[38] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 40–44, 1993.

[39] N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research: Solid Earth*, 117(B4), 2012. Correction in Volume 118, Issue 5.

[40] L. Prünte. *Learning: Wavelet-Dictionaries and Continuous Dictionaries*. PhD thesis, University of Bremen, https://d-nb.info/989563855/34, 2008.

[41] R. Reuter. *Über Integralformen der Einheitssphäre und harmonische Splinefunktionen*. PhD thesis, RWTH Aachen, Veröffentlichung des Geodätisches Institut der RWTH Aachen, vol. 33, 1982.

[42] A. Rieder. *Keine Probleme mit inversen Problemen. Eine Einführung in ihre stabile Lösung*. Vieweg, Wiesbaden, 2003.

[43] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[44] C. Sakumura, S. Bettadpur, and S. Bruinsma. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, 41(5):1389–1397, 2014.

[45] R. Schmidt, F. Flechtner, U. Meyer, K. H. Neumayer, C. Dahle, R. König, and J. Kusche. Hydrological signals observed by the GRACE satellites. *Surveys in Geophysics*, 29(4–5):319–334, 2008.

[46] M. Schreiner. A pyramid scheme for spherical wavelets. *AGTM Report*, (170), 1996. Geomathematics Group, Kaiserslautern.

[47] B. D. Tapley, S. Bettadpur, M. Watkins, and C. Reigber. The gravity recovery and climate experiment: mission overview and early results. *Geophysical Research Letters*, 31(9), 2004.

[48] R. Telschow. *An Orthogonal Matching Pursuit for the Regularization of Spherical Inverse Problems*. PhD thesis, University of Siegen, Verlag Dr. Hut, Munich, 2014.

[49] V. N. Temlyakov. Weak greedy algorithms. *Advances in Computational Mathematics*, 12(2–3):213–227, 2000.

[50] S. Vigerske, A. Wächter, Y. Kawajir, and C. Laird. Introduction to IPOPT: A tutorial for downloading, installing, and using IPOPT. `https://projects.coin-or.org/Ipopt/browser/stable/3.11/Ipopt/doc/documentation.pdf?format=raw`, 2016. last accessed 11 Dec, 2018.

[51] P. Vincent and Y. Bengio. Kernel matching pursuit. *Machine Learning*, 48(1–3):165–187, 2002.

[52] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: Local convergence. *SIAM Journal on Optimization*, 16(1):1–31, 2005.

[53] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):32–48, 2005.

[54] A. Wächter and L. T. Biegler. On the implementation of a primal-dual interior point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

Siegen Preprints
on Geomathematics

The preprint series "Siegen Preprints on Geomathematics" was established in 2010. See www.geomathematics-siegen.de for details and a contact address. At present, the following preprints are available:

1. P. Berkel, D. Fischer, V. Michel: *Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation*, 2010.

2. M. Akram, V. Michel: *Regularisation of the Helmholtz decomposition and its application to geomagnetic field modelling*, 2010.

3. V. Michel: *Optimally Localized approximate identities on the 2-sphere*, 2010.

4. N. Akhtar, V. Michel: *Reproducing kernel based splines for the regularization of the inverse spheroidal gravimetric problem*, 2011.

5. D. Fischer, V. Michel: *Sparse regularization of inverse gravimetry – case study: spatial and temporal mass variations in South America*, 2011.

6. A.S. Fokas, O. Hauk, V. Michel: *Electro-magneto-encephalography for the three-shell model: numerical implementation for distributed current in spherical geometry,* 2011.

7. M. Akram, I. Amna, V. Michel: *A study of differential operators for complete orthonormal systems on a 3D ball*, 2011.

8. D. Fischer, V. Michel: *How to combine spherical harmonics and localized bases for regional gravity modelling and inversion*, 2012.

9. D. Fischer, V. Michel: *Inverting GRACE gravity data for local climate effects,* 2012.

10. V. Michel, R. Telschow: *A non-linear approximation method on the sphere, 2014.*

11. V. Michel, R. Telschow: *The Regularized Orthogonal Functional Matching Pursuit for ill-posed inverse problems,* 2014.

12. V. Michel, S. Orzlowski: *On the null space of a class of Fredholm integral equations of the first kind*, 2015.

13. V. Michel, S. Orzlowski: *On the convergence theorem for the Regularized Functional Matching Pursuit (RFMP) algorithm*, 2016.

14. I. Amna, V. Michel: *Pseudodifferential operators and equidistribution on the 3D-ball - an approach based on orthonormal basis systems*, 2016.

15. S. Gramsch, M. Kontak, V. Michel: *Three-dimensional simulation of nonwoven fabrics using a greedy approximation of the distribution of fiber directions*, 2016.

16. M. Gutting, B. Kretz, V. Michel, R. Telschow: *Study on parameter choice methods for the RFMP with respect to downward continuation*, 2016.

17. V. Michel, S. Orzlowski, N. Schneider: *Vectorial Slepian functions on the ball*, 2017.

18. V. Michel, F.J. Simons: *A general approach to regularizing inverse problems with regional data using Slepian wavelets*, 2017.

19. M. Kontak, V. Michel: *A greedy algorithm for non-linear inverse problems with an application to non-linear inverse gravimetry*, 2018.

20. S. Leweke, V. Michel, A.S. Fokas: *Electro-magnetoencephalography for the spherical multiple-shell model: novel integral operators with singular-value decompositions*, 2018.

21. V. Michel, N. Schneider: *A first approach to learning a best basis for gravity field modelling*, 2019.

Geomathematics Group Siegen
Prof. Dr. Volker Michel

Contact at:
Geomathematics Group
Department of Mathematics
University of Siegen
Walter-Flex-Str. 3
57068 Siegen
www.geomathematics-siegen.de

UNIVERSITÄT
SIEGEN