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

Volker Michel¹ · Naomi Schneider¹

Received: 14 January 2019 / Accepted: 14 January 2020 / Published online: 18 February 2020
© The Author(s) 2020

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 and its variants use an overcomplete dictionary of diverse trial functions to build a best basis as a sparse subset of the dictionary. They also compute a model, for instance, of the gravitational field, in this best basis. Thus, one advantage is that the best basis can be built as a combination of spherical harmonics and RBFs. Moreover, these methods represent a possibility to obtain an approximative and stable solution of an ill-posed inverse problem. The applicability has been practically proven for 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 non-linear 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

✉ Naomi Schneider
naomi.schneider@mathematik.uni-siegen.de

Volker Michel
michel@mathematik.uni-siegen.de

¹ Geomathematics Group, Department of Mathematics, University of Siegen, Walter-Flex-Straße 3, 57068 Siegen, Germany

Mathematics Subject Classification 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 National Geospatial-Intelligence Agency, Office of Geomatics (SN), EGM Development Team (2008); Pavlis et al. (2012). From satellite missions like GRACE or its successor GRACE-FO, we have monthly data provided by the JPL, GFZ and CSR. These are time-dependent models of the potential, see, for example, Flechtner et al. (2014), NASA Jet Propulsion Laboratory (2018), Schmidt et al. (2008), Tapley et al. (2004) and, thus, 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, gravitational field modelling and especially the downward continuation of satellite data is one of the major important mathematical problems in physical geodesy, see, for instance, Baur (2014), Kusche (2015). Details of the used data are given in Sect. 4.

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 $Y_{n,j}$, $n \in \mathbb{N}_0$, $j = -n, \dots, n$. If we assume the Earth to be a closed unit ball, we obtain, for $\sigma > 1$, a pointwise representation of the potential as

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

for the unit sphere Ω and $\eta \in \Omega$, see, for example Baur (2014), Freedden and Michel (2004), Moritz (2010), Telschow (2014), where $\sigma\eta$ is the observation point of the potential V . 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, Engl et al. (1996), Louis (1989), Rieder (2003). 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 of the gravitational potential. 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, Berkel et al. (2011), Fischer (2011), Fischer and Michel (2013a, b), Gutting et al. (2017), Kontak (2018), Kontak and Michel (2018, 2019), Michel (2015), Michel and Orzłowski (2017), Telschow (2014) as well as Fischer and Michel (2012), Leweke (2018), Michel and Telschow (2014), Michel and Telschow (2016), where the latter can also be consulted for a comparison to other methods. These routines are greedy algorithms which iteratively construct stable approximations to the pseudoinverse. 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 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. From the approach it is expected that 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.

While very good approximations could be obtained for the considered applications in Earth sciences and medical imaging (see the references above) so far, the experiments also revealed a sensitivity of the result regarding the choice of the dictionary, for example concerning the runtime and the convergence behaviour. Moreover, this choice is additionally critical since, up to now, the dictionary needed to be finite. Therefore, the main focus of this paper is on a first dictionary learning strategy for the downward continuation of gravitational potential data. This will also allow the use of an infinite dictionary with a free choice of the continuous parameters of an RBF.

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, Bruckstein et al. (2009), Prünke (2008), Rubinstein et al. (2010).

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, Fischer (2011), Fischer and Michel (2012, 2013a, b), Michel and Telschow (2014), Telschow (2014). 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.

In parallel to our research project, a stochastic non-greedy approach for finding an optimal basis of RBFs has been developed (Schall 2019). While adding and removing centres of the RBFs, this stochastic approach is also able to determine an optimal number of basis functions. On the other hand, only one fixed system of RBFs is used there, while our approach is able to learn an optimal localization parameter for each single RBF. Future research is expected to reveal further pros and cons of both approaches.

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 presented results show that the use of a learnt dictionary in the RFMP gives us a higher sparsity and better results with less storage demand.

An overview of the structure of the learning algorithm is shown in Fig. 1. 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

Fig. 1 Schematic representation of the basic learning algorithm

preprocessing and setting of parameters [similarly as described, for example, in Telschow (2014)] 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 (as usually done in the RFMP) 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 Fig. 1, this is shown by the boxes ‘spherical harmonics’ and ‘radial basis functions’ which lead to ‘set of candidates’. 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 Telschow (2014). Next, we check the termination criteria for the learning algorithm. If they are not fulfilled, then we search for the next element d_2 in the same manner as we found d_1 . Otherwise we stop the RFMP and, thus, the learning of the dictionary. We obtain an approximation of the given signal. Additionally, the learnt dictionary is defined as the set of all chosen elements in this approximation.

This paper is structured as follows. On the way to a detailed description of our learning approach, we define some fundamental basics in Sect. 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 Sect. 3. We define its routine and give necessary derivatives. Some of the formulae are elaborated in “Appendix A”. We end Section 3 by introducing some additional learning features that guide the learning process positively. 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. We also explain why we choose this comparison. At last, we conclude this paper in Sect. 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

We consider spherical harmonics as well as Abel–Poisson kernels. Spherical harmonics are global trial functions, see for instance, Freeden et al. (1998), Freeden and Gutting (2013), Freeden and Schreiner (2009), Michel (2013), Müller (1966). They are defined for a unit vector $\xi \in \Omega$ as

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

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

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

where P_n denotes the n -th Legendre polynomial. Abel–Poisson kernels are defined for a particular unit vector $\xi \in \Omega$ (the centre of the RBF) and a scaling parameter $h \in [0, 1)$ as (with $x = h\xi$) the functions

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

of the unit vector $\eta \in \Omega$, see, for example, Freeden et al. (1998), pp. 108–112 or Freeden and Schreiner (2009), p. 103 and 441. These kernels 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 scaling 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, Freeden et al. (1998), p. 111 or Michel (2013), 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 $N \subset \mathcal{N} := \{(n, j) \mid n \in \mathbb{N}_0, j = -n, \dots, n\}$ and $K \subseteq \mathring{\mathbb{B}}_1(0)$ for the open unit ball $\mathring{\mathbb{B}}_1(0)$. Then we set

$$[N]_{\text{SH}} := \{Y_{n,j} \mid (n, j) \in N\}$$

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

$$[K]_{\text{APK}} := \{P(r\xi, \cdot) \mid r\xi \in K\}$$

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

$$\mathcal{D} := [N]_{\text{SH}} + [K]_{\text{APK}} := [N]_{\text{SH}} \cup [K]_{\text{APK}}.$$

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

For the theory of inverse problems $\mathcal{T}: \mathcal{X} \rightarrow \mathcal{Y}$, we refer the reader to Engl et al. (1996), Kontak (2018), Michel (2013), Rieder (2003). For the domain $\mathcal{X}$, we propose to use a

Sobolev space because, for instance, it enforces more smoothness of the solution than, e.g. the $L^2(\Omega)$ -space. This has proven to yield better results. Specifically, we will use the Sobolev space $\mathcal{H}_2$.

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

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

we define an inner product via

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

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

We give a short overview of the main principles of the RFMP algorithm as far as they are needed for this paper. For further details, we refer to the literature listed in the introduction. 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. We reformulate the linear inverse problem $\mathcal{T}: \mathcal{H}_2 \rightarrow \mathcal{Y}$, $\mathcal{T}F = V$, as given in (1), to the case of discrete grid-based data where $\mathcal{Y} = \mathbb{R}^\ell$.

We have a relative satellite height $\sigma > 1$, a set of grid points $\{\eta^{(i)}\}_{i=1,\dots,\ell} \subset \Omega$ and associated data values $\{y_i\}_{i=1,\dots,\ell}$. 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 $(\mathcal{T}_\daleth^i)_{i=1,\dots,\ell}$ 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 a given vector $y \in \mathbb{R}^\ell$.

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

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

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 $F_{n+1} := F_n + \alpha_{n+1} d_{n+1}$ via

$$(\alpha_{n+1}, d_{n+1}) := \arg \min_{(\alpha, d) \in \mathbb{R} \times \mathcal{D}} \mathcal{J}(F_n + \alpha d; \mathcal{T}_\daleth, \lambda, y). \quad (5)$$

It can be shown, see, for example, Fischer (2011), Michel (2015), Michel and Telschow (2014) that the minimization with respect to α and d of the Tikhonov functional (5) is equivalent to determining α_{n+1} and d_{n+1} via

$$\begin{aligned} d_{n+1} &:= \arg \max_{d \in \mathcal{D}} \frac{(\langle R^n, \mathcal{T} \gamma d \rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d \rangle_{\mathcal{H}_2})^2}{\|\mathcal{T} \gamma d\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_2}^2}, \\ \alpha_{n+1} &:= \frac{\langle R^n, \mathcal{T} \gamma d_{n+1} \rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d_{n+1} \rangle_{\mathcal{H}_2}}{\|\mathcal{T} \gamma d_{n+1}\|_{\mathbb{R}^\ell}^2 + \lambda \|d_{n+1}\|_{\mathcal{H}_2}^2} \end{aligned} \quad (6)$$

where $R^n := y - \mathcal{T} \gamma F_n$ is the residual.

3 The learning approach

3.1 A first learning algorithm

We described the learning routine in Sect. 1. With the notation just introduced, we formulate the minimization task of each trial function class mathematically. For an arbitrary iteration step n , the objective to seek α_{n+1} and d_{n+1} is given by

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

where the superscript refers to the trial function class

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

respectively. In the sequel, we use an equivalent formulation.

Definition 3 Due to (6), we define the objective function of the RFMP in the n -th iteration step as

$$\text{RFMP}(d; n) := \frac{(\langle R^n, \mathcal{T} \gamma d \rangle_{\mathbb{R}^\ell} - \lambda \langle F_n, d \rangle_{\mathcal{H}_2})^2}{\|\mathcal{T} \gamma d\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_2}^2}, \quad (7)$$

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

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

Algorithm 4 We obtain a learnt dictionary for the RFMP as follows. Let $\mathcal{T}\lrcorner F = y$ be the linear inverse problem under investigation, $\mathcal{H}_2$ the Sobolev space from Definition 2 and λ the regularization parameter.

- (S0) initialize: termination criterion; data vector y ; initial approximation F_0 ; sets $\hat{N}$ as well as $\hat{K}$ as in Definition 1 and starting dictionary $\mathcal{D} = [\hat{N}]_{\text{SH}} + [\hat{K}]_{\text{APK}}$
 (S1) set $R^0 := y - \mathcal{T}\lrcorner F_0$ and compute $\langle F_0, d \rangle_{\mathcal{H}_2}$ for all $d \in \mathcal{D}$.
 (S2) compute $\mathcal{T}\lrcorner d$ for each $d \in \mathcal{D}$ and $\langle d, \tilde{d} \rangle_{\mathcal{H}_2}$ for each pair $d, \tilde{d} \in \mathcal{D}$.
 (S3) while (termination criterion not fulfilled)

(S3.1) compute candidate

$$d_{n+1}^{\text{SH}} := \arg \max \left\{ \text{RFMP}(d; n) \mid d \in [\hat{N}]_{\text{SH}} \right\}$$

(S3.2) compute starting point for the optimization

$$d_{n+1}^{\text{APK}, \text{start}} := \arg \max \left\{ \text{RFMP}(d; n) \mid d \in [\hat{K}]_{\text{APK}} \right\}$$

(S3.3) compute optimal solution

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

(S3.4) choose

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

(S3.5) compute

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

(S3.6) set $R^{n+1} := R^n - \alpha_{n+1} \mathcal{T}\lrcorner d_{n+1}$

(S3.7) for each $d \in \mathcal{D}$ compute $\langle F_{n+1}, d \rangle_{\mathcal{H}_2} = \langle F_n, d \rangle_{\mathcal{H}_2} + \alpha_{n+1} \langle d_{n+1}, d \rangle_{\mathcal{H}_2}$

(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

$$\mathcal{D}^* = [N^*]_{\text{SH}} + [K^*]_{\text{APK}},$$

$$N^* = \{(n_i, j_i) \mid \text{there exists } i \in \{1, \dots, M\} \text{ such that } Y_{n_i, j_i} = d_i\},$$

$$K^* = \{(r\xi)^{(i)} \mid \text{there exists } i \in \{1, \dots, M\} \text{ such that } P((r\xi)^{(i)}, \cdot) = d_i\}$$

In our implementation, we use the following truncation criterion: we stop if $n > I$ or $\|R^n\|_{\mathbb{R}^\ell} < \varepsilon$ or $|\alpha_n| < \delta$ for preliminarily chosen values $I \in \mathbb{N}$ and $\varepsilon, \delta \in \mathbb{R}^+$.

3.2 Determination of candidates

The question remains how the candidates in each trial function class under consideration are determined, i. e. what is done in (S3.1)–(S3.3).

First, 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 4 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 $\tilde{N} \in \mathbb{N}$ (i.e. $\hat{N} = \{(n, j) \mid n \in \mathbb{N}_0, n \leq \tilde{N}, j = -n, \dots, n\}$) 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 up to a lower degree ν , we have a truly learnt bound $\nu < \tilde{N}$. However, note that the higher we choose $\tilde{N}$, 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 $\text{RFMP}(Y_{m,k}; n)$ for all spherical harmonics up to degree $\tilde{N}$. Note that this optimization problem is discrete by nature.

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 optimization problem

$$\text{RFMP}(P(x, \cdot); n) \rightarrow \max!$$

can be modelled by an optimization with respect to the characteristics $x = h\xi$ of each RBF. 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, Freeden and Gerhards (2013), p. 132 or Freeden and Michel (2004), 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 $\|x\|_{\mathbb{R}^3}^2 < 1$.

Definition 5 The optimal candidate d_{n+1}^{APK} among the set of Abel–Poisson kernels $\{P(x, \cdot) \mid x \in \mathring{\mathbb{B}}_1(0)\}$ is given by the solution of the optimization problem

$$\text{RFMP}(P(x, \cdot); n) \rightarrow \max! \quad \text{subject to} \quad \|x\|_{\mathbb{R}^3}^2 < 1.$$

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 (7) 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 6 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 $P(x^{(i)}, \cdot)$. Moreover, let $\mathcal{T}_\uparrow$ be the upward continuation operator and σ the respective satellite orbit height. 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 (18). Then we have for an Abel–Poisson kernel $P(x, \cdot)$ and the representation of $x \in \mathbb{B}_1(0)$ in polar coordinates (r_x, φ_x, t_x) the terms

$$a_1(P(x, \cdot)) := \langle R^n, \mathcal{T}_\uparrow P(x, \cdot) \rangle_{\mathbb{R}^\ell} = \sum_{i=1}^{\ell} \frac{R_i^n}{4\pi} \frac{\sigma^2 - |x|^2}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^{3/2}}, \quad (8)$$

$$a_2(P(x, \cdot)) := \langle F_n, P(x, \cdot) \rangle_{\mathcal{H}_2} = \sum_{i=1}^n \alpha_i \begin{cases} T_1, & d_i = Y_{n_i, j_i} \\ T_2, & d_i = P(x^{(i)}, \cdot), \end{cases} \quad (9)$$

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

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

$$b_1(P(x, \cdot)) := \|\mathcal{T}_\uparrow P(x, \cdot)\|_{\mathbb{R}^\ell}^2 = \sum_{i=1}^{\ell} \frac{(\sigma^2 - |x|^2)^2}{16\pi^2 (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^3}, \quad (10)$$

and

$$b_2(P(x, \cdot)) := \|P(x, \cdot)\|_{\mathcal{H}_2}^2 = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (n + 0.5)^4 |x|^{2n}. \quad (11)$$

Their partial derivatives with respect to x_j are given by

$$\begin{aligned} & \partial_{x_j} a_1(P(x, \cdot)) \\ &= - \sum_{i=1}^{\ell} \frac{R_i^n}{4\pi} \left[\frac{2x_j}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^{3/2}} + \frac{3(\sigma^2 - |x|^2)(x_j - \sigma \eta_j^{(i)})}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^{5/2}} \right], \end{aligned} \quad (12)$$

$$\partial_{x_j} a_2(P(x, \cdot)) = \sum_{i=1}^n \alpha_i \left\{ \begin{array}{l} \partial_{x_j} T_1, d_i = Y_{n_i, j_i} \\ \partial_{x_j} T_2, d_i = P(x^{(i)}, \cdot) \end{array} \right., \quad (13)$$

$$\begin{aligned} \partial_{x_j} T_1 &= (n_i + 0.5)^4 \partial_{x_j} \left(|x|^{n_i} Y_{n_i, j_i} \left(\frac{x}{|x|} \right) \right) \\ \partial_{x_j} T_2 &= \sum_{n=0}^{\infty} (n + 0.5)^4 (|x_i|)^n \frac{2n+1}{4\pi} \partial_{x_j} \left(|x|^n P_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right) \right) \\ \partial_{x_j} b_1(P(x, \cdot)) &= -\frac{1}{16\pi^2} \sum_{i=1}^{\ell} \left[\frac{4x_j (\sigma^2 - |x|^2)}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^3} + \frac{6(\sigma^2 - |x|^2)^2 (x_j - \sigma \eta_j^{(i)})}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^4} \right], \end{aligned} \quad (14)$$

and

$$\partial_{x_j} b_2(P(x, \cdot)) = \sum_{n=0}^{\infty} \frac{2n^2 + n}{2\pi} (n + 0.5)^4 |x|^{2n-2} x_j. \quad (15)$$

With respect to the derivative of a_2 , we have further

$$\begin{aligned} &\nabla \left(|x|^n Y_{n,j} \left(\frac{x}{|x|} \right) \right) \\ &= \varepsilon^r n |x|^{n-1} Y_{n,j} \left(\frac{x}{|x|} \right) + |x|^{n-1} \varepsilon^\varphi \frac{j}{\sqrt{1-t^2}} Y_{n,-j} \left(\frac{x}{|x|} \right) \\ &\quad + |x|^{n-1} \varepsilon^t \sqrt{1-t_x^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t_x) \begin{cases} \sqrt{2} \cos(j\varphi_x), & j < 0 \\ 1, & j = 0 \\ \sqrt{2} \sin(j\varphi_x), & j > 0 \end{cases} \end{aligned} \quad (16)$$

and

$$\begin{aligned} &\nabla \left(|x|^n P_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right) \right) \\ &= |x|^{n-1} \left[\varepsilon^r n P_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right) + \varepsilon^\varphi P'_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right) \left(\frac{x_i}{|x_i|} \cdot \varepsilon^\varphi \right) \right. \\ &\quad \left. + \varepsilon^t P'_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right) \left(\frac{x_i}{|x_i|} \cdot \varepsilon^t \right) \right]. \end{aligned} \quad (17)$$

For a proof, see “Appendix A”.

Theorem 7 *With the abbreviations and derivatives from Theorem 6, the partial derivatives ∂_{x_j} , $j = 1, 2, 3$, of $\text{RFMP}(\cdot; n)$ are given by*

$$\begin{aligned} \partial_{x_j} \text{RFMP}(P(x, \cdot); n) &= [b_1(P(x, \cdot)) + \lambda b_2(P(x, \cdot))]^{-2} [2[a_1(P(x, \cdot)) - \lambda a_2(P(x, \cdot))] \\ &\quad \times [\partial_{x_j} a_1(P(x, \cdot)) - \lambda \partial_{x_j} a_2(P(x, \cdot))][b_1(P(x, \cdot)) + \lambda b_2(P(x, \cdot))] \\ &\quad - [a_1(P(x, \cdot)) - \lambda a_2(P(x, \cdot))]^2 [\partial_{x_j} b_1(P(x, \cdot)) + \lambda \partial_{x_j} b_2(P(x, \cdot))]] \end{aligned}$$

Proof We only apply the common rules for derivatives. $\square$

Thus, for Abel–Poisson kernels, we determined the gradient of the modelled objective functions $\text{RFMP}(\cdot; n)$, $n \in \mathbb{N}$, 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 Vigerske et al. (2016), HSL (2018), Nocedal et al. (2008), Wächter and Biegler (2005a, b, 2006). 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 Vigerske et al. (2016), 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 $d_{n+1}^{\text{APK,loc}}$ in Algorithm 4 and is computed in S3.2. We obtain $d_{n+1}^{\text{APK,loc}}$ by comparing the objective value $\text{RFMP}(P(x, \cdot); n)$ for a selection of kernels given to the algorithm in the discrete dictionary by $[\hat{K}]_{\text{APK}}$ 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 the case that the optimization failed to find a better kernel.

3.3 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 these 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 function 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, Telschow (2014), 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, Telschow (2014), 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 a preliminarily chosen regularization parameter (see below). 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, Prünke (2008), 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

We consider two scenarios with different data sets: the EGM2008 as well as one month of GRACE satellite data. These data sets vary strongly with respect to their structure and, thus, yield two different challenges as numerical tests.

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 Sakumura et al. (2014). This meanfield was subtracted from the data corresponding to May 2008. Note that in May, usually, 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 of scale 5, see, for instance, Schreiner (1996) and Freeden et al. (1998), p. 295. This is a simple and common smoothing filter based on a spherical wavelet method. Certainly, there are other filters which were specifically designed for destriping GRACE data such as Davis et al. (2008), Klees et al. (2008), Kusche (2007). However, here, we are concerned with testing the learning algorithm and not with detecting climatological phenomena. In both tests (EGM2008 and GRACE), we present results for the downward continuation from 500 km satellite height with 5% white noise.

In all experiments, we compute the data on an equidistributed Reuter grid of 12684 data points, see, for instance, Reuter (1982) and Michel (2013), p. 137. We choose the constant regularization parameter λ of the RFMP and the initial regularization parameter λ_0 of the LRFMP (see Sect. 3.3) 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.

Note that, in previous publications, see Fischer and Michel (2013b), Michel and Telschow (2014), Leweke (2018) it has already been shown that the RFMP is a competitive method in comparison to traditional method such as spline approximation. Therefore, we abstain from this comparison in this paper. Further, it would certainly be valuable to run tests with a large number of random dictionaries. This would allow the verification of the learnt dictionary in comparison to the average of the random dictionaries. However, for a scientific significance, the number of such dictionaries would need to be so large such that the experiments could not be feasible in a reasonable amount of time (see the run-times in the numerical results below). Hence, we compare the results of the learnt dictionary with one manually chosen dictionary which is similar to well-working dictionaries from previous publications, see e. g. Michel and Telschow (2016). This has also the advantage that this dictionary has already been a result of a 'human learning' process via trial-and-error. It might, therefore, be already better than a random choice.

To be specific, we define this manually chosen dictionary by

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

with

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

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

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

and

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

respectively, with

$$\begin{aligned} N^{s, \text{EGM}} &:= \{(n, j) \mid n = 0, \dots, 100; j = -n, \dots, n\}, \\ N^{s, \text{GRACE}} &:= \{(n, j) \mid n = 0, \dots, 60; j = -n, \dots, n\}, \\ K^s &:= \{r\xi \mid r = 0.94, \xi \in X^s\}, \end{aligned}$$

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 Sect. 3.3. 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 (Vigerske et al. 2016).

For the termination criteria of the LRFMP as well as the RFMP (see Sect. 3.1), we set $I = 3000$, $\varepsilon = 10^{-8} \|y\|_{\mathbb{R}^\ell}$ and $\delta = 10^{-5}$ in all cases. 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 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 $\mathcal{D}^m$. In the second case, we first learn a dictionary $\mathcal{D}^{*,\bullet}$ (see Algorithm 4) for $\bullet \in \{\text{EGM}, \text{GRACE}\}$ by using the LRFMP (which requires the starting dictionary $\mathcal{D}^{s,\bullet}$) 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 Fig. 2, 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 exact, unnoisy solution is given for comparison. 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-runtimes 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 the LRFMP includes learning and applying the learnt dictionary, it takes less time than the RFMP. However, these results are to be understood with care because they were not obtained with 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.

Fig. 2 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 m^2/s^2

Table 1 Comparison of RFMP with a manually chosen dictionary (RFMP*) and a learnt dictionary (RFMP**)

	$\#D$	rel. approximation error	rel. data error	CPU-runtime in h
RFMP*	46,186	0.000794	0.065830	299.82
RFMP**	≤ 3000	0.000455	0.047092	41.34
RFMP*	46,186	0.001461	0.057765	295.49
RFMP**	≤ 3000	0.000306	0.046404	7.5

Upper comparison with respect to EGM2008 data. Lower comparison with respect to GRACE data. 3000 iterations allowed in all experiments

5 Conclusion and outlook

We started our investigations by aiming to improve our method, the RFMP, for gravity field modelling. 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 both from EGM2008 as well as 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 for learning a dictionary.

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. 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.

Acknowledgements Open Access funding provided by Projekt DEAL. The authors gratefully acknowledge the financial support by the German Research Foundation (DFG; Deutsche Forschungsgemeinschaft), project MI 655/7-2.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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

In this appendix, we prove Theorem 6.

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 ' $\cdot$ ' 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

$$\varepsilon^r(\varphi, t) := \begin{pmatrix} \sqrt{1-t^2} \cos(\varphi) \\ \sqrt{1-t^2} \sin(\varphi) \\ t \end{pmatrix}, \quad \varepsilon^\varphi(\varphi, t) := \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \\ 0 \end{pmatrix}, \quad \varepsilon^t(\varphi, t) := \begin{pmatrix} -t \cos(\varphi) \\ -t \sin(\varphi) \\ \sqrt{1-t^2} \end{pmatrix}, \quad (18)$$

see, for example, Michel (2013), 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

$$\nabla = (\partial_{x_j})_{j=1,2,3} = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r} \left(\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} \frac{\partial}{\partial \varphi} + \varepsilon^t \sqrt{1-t^2} \frac{\partial}{\partial t} \right), \quad (19)$$

see, for instance, Michel (2013), p. 87. Next, we consider the following recurring inner products.

$$\langle Y_{m,k}, Y_{n,j} \rangle_{L^2(\Omega)} = \delta_{n,m} \delta_{j,k}, \quad (20)$$

$$\langle P(x, \cdot), Y_{n,j} \rangle_{L^2(\Omega)} = |x|^n Y_{n,j} \left(\frac{x}{|x|} \right), \quad (21)$$

see, for instance, Telschow (2014), p. 114. At last, we note one specific property of the fully normalized spherical harmonics. It holds

$$\frac{\partial}{\partial \varphi} Y_{n,j}(\xi(\varphi, t)) = j Y_{n,-j}(\xi(\varphi, t)), \quad (22)$$

see, for instance, Leweke et al. (2018). Now, we can compute the terms in Theorem 6.

The term $a_1(P(x, \cdot))$ and its derivative Obviously, for the formulation of $a_1(P(x, \cdot))$ in (8), we only have to show that

$$\mathcal{T}P(x, \cdot)(\eta) = \frac{\sigma^2 - |x|^2}{4\pi(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{3/2}} \quad (23)$$

for any $\eta \in \Omega$. We start at the left-hand side of (23).

$$\begin{aligned}
 \mathcal{T}P(x, \cdot)(\eta) &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \langle Y_{n,j}, P(x, \cdot) \rangle_{L^2(\Omega)} \sigma^{-n-1} Y_{n,j}(\eta) \\
 &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \sigma^{-n-1} |x|^n Y_{n,j} \left(\frac{x}{|x|} \right) Y_{n,j}(\eta) \\
 &= \sum_{n=0}^{\infty} |x|^n \sigma^{-n-1} \frac{2n+1}{4\pi} P_n \left(\frac{x}{|x|} \cdot \eta \right) = \frac{1}{\sigma} P \left(\frac{x}{\sigma}, \eta \right) \\
 &= \frac{1}{\sigma} \frac{1 - \left| \frac{x}{\sigma} \right|^2}{4\pi (1 + \left| \frac{x}{\sigma} \right|^2 - 2 \frac{x}{\sigma} \cdot \eta)^{3/2}} \\
 &= \frac{\sigma^{-2} (\sigma^2 - |x|^2)}{4\pi \sigma (\sigma^{-2} (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta))^{3/2}} = \frac{\sigma^2 - |x|^2}{4\pi (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{3/2}}.
 \end{aligned} \tag{24}$$

Its derivative, as used in (12), is obtained via

$$\begin{aligned}
 \partial_{x_j} \mathcal{T}P(x, \cdot)(\eta) &= \frac{\partial}{\partial x_j} \frac{\sigma^2 - |x|^2}{4\pi (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{3/2}} \\
 &= \frac{-2x_j (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{3/2} - 3 (\sigma^2 - |x|^2) (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{1/2} (x_j - \sigma \eta_j)}{4\pi (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^3} \\
 &= -\frac{1}{4\pi} \left[\frac{2x_j}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{3/2}} + \frac{3 (\sigma^2 - |x|^2) (x_j - \sigma \eta_j)}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta)^{5/2}} \right].
 \end{aligned}$$

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 $d_i = P(x^{(i)}, \cdot)$ depending on what element was chosen in the i -th step. Then we can derive the representation (9) of $a_2(P(x, \cdot))$ as follows

$$\begin{aligned}
 a_2(P(x, \cdot)) &:= \langle F_n, P(x, \cdot) \rangle_{\mathcal{H}_2} = \sum_{i=1}^n \alpha_i \langle d_i, P(x, \cdot) \rangle_{\mathcal{H}_2} \\
 &= \sum_{i=1}^n \alpha_i \begin{cases} (n_i + 0.5)^4 |x|^{n_i} Y_{n_i, j_i} \left(\frac{x}{|x|} \right), & d_i = Y_{n_i, j_i} \\
 \sum_{n=0}^{\infty} (n + 0.5)^4 (|x_i| |x|)^n \frac{2n+1}{4\pi} P_n \left(\frac{x_i}{|x_i|} \cdot \frac{x}{|x|} \right), & d_i = P(x^{(i)}, \cdot) \end{cases}
 \end{aligned}$$

due to (20), (21) and the addition theorem for spherical harmonics. The derivative of $a_2(P(x, \cdot))$ as given in (13) is obvious. However, we have to show (16) and (17). 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 (16).

$$\begin{aligned}
& \nabla_x (r^n Y_{n,j}(\xi)) \\
&= \varepsilon^r \left(\frac{\partial}{\partial r} r^n Y_{n,j}(\xi(\varphi, t)) \right) + \frac{1}{r} \varepsilon^\varphi \frac{1}{\sqrt{1-t^2}} \left(r^n \frac{\partial}{\partial \varphi} Y_{n,j}(\xi(\varphi, t)) \right) \\
&\quad + \frac{1}{r} \varepsilon^t \sqrt{1-t^2} \left(r^n \frac{\partial}{\partial t} Y_{n,j}(\xi(\varphi, t)) \right) \\
&= \varepsilon^r n r^{n-1} Y_{n,j}(\xi(\varphi, t)) + r^{n-1} \varepsilon^\varphi \frac{j}{\sqrt{1-t^2}} Y_{n,-j}(\xi(\varphi, t)) \\
&\quad + r^{n-1} \varepsilon^t \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t) \begin{cases} \sqrt{2} \cos(j\varphi), & j < 0, \\ 1, & j = 0, \\ \sqrt{2} \sin(j\varphi), & j > 0, \end{cases} \quad (25)
\end{aligned}$$

where we used (19), (22) and (2). We have to take a closer look at (25) regarding a possible singularity in $t = \pm 1$. The term (25) contains two possibly problematic terms:

$$\frac{j}{\sqrt{1-t^2}} Y_{n,-j}(\xi) \quad \text{and} \quad \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t). \quad (26)$$

We first consider the term on the left-hand side of (26). 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

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

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 (26). With respect to the term on the right-hand side of (26), we have

$$\begin{aligned}
& \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} P'_{n,|j|}(t) \\
&= \sqrt{1-t^2} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} \frac{d}{dt} \left[(1-t^2)^{|j|/2} \frac{d^{|j|}}{dt^{|j|}} P_n(t) \right]
\end{aligned}$$

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

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

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

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 (26) contains no singularity.

At last, for the derivative of $a_2(P(x, \cdot))$, we need to consider (17). We obtain

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

With $\varepsilon^r = \frac{x}{|x|}$, this is the formulation of (17).

The term $b_1(P(x, \cdot))$ and its derivative The term $b_1(P(x, \cdot))$ as in (10) is obvious when we take (24) into account. The derivative of $b_1(P(x, \cdot))$ as in (14) is obtained by

$$\begin{aligned} \partial_{x_j} b_1(P(x, \cdot)) &= \frac{\partial}{\partial x_j} \frac{1}{16\pi^2} \sum_{i=0}^{\ell} \frac{(\sigma^2 - |x|^2)^2}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^3} \\ &= \frac{1}{16\pi^2} \sum_{i=0}^{\ell} \left(\left(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)} \right)^{-6} \left(-4x_j (\sigma^2 - |x|^2) (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^3 \right. \right. \\ &\quad \left. \left. - 6 (\sigma^2 - |x|^2)^2 (\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^2 (x_j - \sigma \eta_j^{(i)}) \right) \right) \\ &= -\frac{1}{16\pi^2} \sum_{i=0}^{\ell} \left(\frac{4x_j (\sigma^2 - |x|^2)}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^3} + \frac{6 (\sigma^2 - |x|^2)^2 (x_j - \sigma \eta_j^{(i)})}{(\sigma^2 + |x|^2 - 2\sigma x \cdot \eta^{(i)})^4} \right). \end{aligned}$$

The term $b_2(P(x, \cdot))$ and its derivative The formulation as in (11) of $b_2(P(x, \cdot))$ is due to the following considerations which use (21) and $P_n(1) = 1$, see, for instance,

Michel (2013), p. 49.

$$\begin{aligned}
 \|P(x, \cdot)\|_{\mathcal{H}_2}^2 &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} (n+0.5)^4 \langle P(x, \cdot), Y_{n,j} \rangle_{L^2(\Omega)}^2 \\
 &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} (n+0.5)^4 |x|^{2n} \left(Y_{n,j} \left(\frac{x}{|x|} \right) \right)^2 \\
 &= \sum_{n=0}^{\infty} (n+0.5)^4 |x|^{2n} \frac{2n+1}{4\pi} P_n \left(\frac{x}{|x|} \cdot \frac{x}{|x|} \right) \\
 &= \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (n+0.5)^4 |x|^{2n}.
 \end{aligned}$$

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

$$\nabla_x b_2(P(x, \cdot)) = \sum_{n=0}^{\infty} (n+0.5)^4 \frac{2n+1}{4\pi} \nabla_x \left(|x|^{2n} \right). \quad (27)$$

With (19), 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 (27), 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 (15). Hence, Theorem 6 is proven. $\square$

References

- Baur, O.: Gravity field of planetary bodies. In: Grafarend, E. (ed.) *Encyclopedia of Geodesy*, pp. 1–6. Springer, Cham (2014)
- Berkel, P., Fischer, D., Michel, V.: Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation. *Int. J. Geomath.* **1**(2), 167–204 (2011)
- Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* **51**(1), 34–81 (2009)
- Davis, J.L., Tamisiea, M.E., Elósegui, P., Mitrovica, J.X., Hill, E.M.: A statistical filtering approach for Gravity Recovery and Climate Experiment (GRACE) gravity data. *J. Geophys. Res. Solid Earth* (2008). <https://doi.org/10.1029/2007JB005043>
- Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems. Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht (1996)
- Fischer, D.: *Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies*. PhD thesis, University of Siegen, Verlag Dr. Hut, Munich (2011). <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2012/544/index.html>. Accessed 24 Jan 2020

- Fischer, D., Michel, V.: Sparse regularization of inverse gravimetry—case study: spatial and temporal mass variations in South America. *Inverse Probl.* **28**(6), 065012 (2012)
- Fischer, D., Michel, V.: Automatic best-basis selection for geophysical tomographic inverse problems. *Geophys. J. Int.* **193**(3), 1291–1299 (2013)
- Fischer, D., Michel, V.: Inverting GRACE gravity data for local climate effects. *J. Geod. Sci.* **3**(3), 151–162 (2013)
- Flechtner, F., Morton, P., Watkins, M., Webb, F.: Status of the GRACE follow-on mission. In: Marti U (ed.) *Gravity, Geoid and Height Systems*. International Association of Geodesy Symposia. Springer, Cham, vol. 141, pp. 117–121 (2014)
- Freedeen, W., Gerhards, C.: *Geomathematically Oriented Potential Theory*. Taylor & Francis Group, Boca Raton (2013)
- Freedeen, W., Gutting, M.: *Special Functions of Mathematical (Geo-)Physics*. Springer, Basel (2013)
- Freedeen, W., Michel, V.: *Multiscale Potential Theory with Applications to Geoscience*. Birkhäuser, Boston (2004)
- Freedeen, W., Schreiner, M.: *Spherical Functions of Mathematical Geosciences—A Scalar, Vectorial, and Tensorial Setup*. Springer, Berlin (2009)
- Freedeen, W., Gervens, T., Schreiner, M.: *Constructive Approximation on the Sphere—with Applications to Geomathematics*. Oxford University Press, Oxford (1998)
- Gutting, M., Kretz, B., Michel, V., Telschow, R.: Study on parameter choice methods for the RFMP with respect to downward continuation. *Front. Appl. Math. Stat.* (2017). <https://doi.org/10.3389/fams.2017.00010>
- HSL: A collection of Fortran codes for large scale scientific computation (2018). <http://www.hsl.rl.ac.uk/>. Last Accessed 11 Dec 2018
- Klees, R., Revtova, E.A., Gunter, B.C., Ditmar, P., Oudman, E., Winsemius, H.C., Savenjie, H.H.G.: The design of an optimal filter for monthly GRACE gravity models. *Geophys. J. Int.* **175**(2), 417–432 (2008)
- Kontak, M.: *Novel Algorithms of Greedy-Type for Probability Density Estimation as well as Linear and Nonlinear Inverse Problems*. Ph.D. thesis, University of Siegen, Verlag Dr. Hut, Munich (2018). <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1316/index.html>. Accessed 24 Jan 2020
- Kontak, M., Michel, V.: A greedy algorithm for nonlinear inverse problems with an application to nonlinear inverse gravimetry. *GEM Int. J. Geomath.* **9**(2), 167–198 (2018)
- Kontak, M., Michel, V.: The regularized weak functional matching pursuit for linear inverse problems. *J. Inverse Ill-Posed Probl.* **27**(3), 317–340 (2019)
- Kusche, J.: Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *J. Geod.* **81**(11), 733–749 (2007)
- Kusche, J.: Time-variable gravity field and global deformation of the Earth. In: Freedeen, W., Nashed, M.Z., Sonar, T. (eds.) *Handbook of Geomathematics*, 2nd edn, pp. 321–338. Springer, Berlin (2015)
- Leweke, S.: *The Inverse Magneto-electroencephalography Problem for the Spherical Multiple-shell Modell*. Ph.D. thesis, University of Siegen (2018). <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2019/1396/>. Accessed 24 Jan 2020
- Leweke, S., Michel, V., Schneider, N.: Vectorial Slepian functions on the ball. *Numer. Funct. Anal. Optim.* **39**(11), 1120–1152 (2018)
- Louis, A.K.: *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart (1989)
- Michel, V.: *Lectures on Constructive Approximation—Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball*. Birkhäuser, New York (2013)
- Michel, V.: RFMP—An iterative best basis algorithm for inverse problems in the geosciences. In: Freedeen, W., Nashed, M.Z., Sonar, T. (eds.) *Handbook of Geomathematics*, 2nd edn, pp. 2121–2147. Springer, Berlin (2015)
- Michel, V., Orzlowski, S.: On the convergence theorem for the Regularized Functional Matching Pursuit (RFMP) algorithm. *GEM Int. J. Geomath.* **8**(2), 183–190 (2017)
- Michel, V., Telschow, R.: A non-linear approximation method on the sphere. *GEM Int. J. Geomath.* **5**(2), 195–224 (2014)
- Michel, V., Telschow, R.: The regularized orthogonal functional matching pursuit for ill-posed inverse problems. *SIAM J. Numer. Anal.* **54**(1), 262–287 (2016)
- Moritz, H.: *Classical physical geodesy*. In: Freedeen, W., Nashed, M.Z., Sonar, T. (eds.) *Handbook of Geomathematics*, 2nd edn, pp. 253–289. Springer, Berlin (2010)
- Müller, C.: *Spherical Harmonics*. Springer, Berlin (1966)

- NASA Jet Propulsion Laboratory (2018) GRACE Tellus. <https://grace.jpl.nasa.gov/publications/>. Last Accessed 11 Dec 2018
- National Geospatial-Intelligence Agency, Office of Geomatics (SN), EGM Development Team (2008) Earth Gravitational Model 2008. <http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/>. Last Accessed 11 Dec 2018
- Nocedal, J., Wächter, A., Waltz, R.A.: Adaptive barrier strategies for nonlinear interior methods. *SIAM J. Optim.* **19**(4), 1674–1693 (2008)
- Pavlis, N.K., Holmes, S.A., Kenyon, S.C., Factor, J.K.: The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *J. Geophys. Res. Solid Earth* **117**(B4), correction in Volume 118, Issue 5 (2012)
- Prünte, L.: Learning: Wavelet-Dictionaries and Continuous Dictionaries. Ph.D. thesis, University of Bremen (2008). <https://d-nb.info/989563855/34>. Accessed 24 Jan 2020
- Reuter, R.: Über Integralformen der Einheitssphäre und harmonische Splinesfunktionen. Ph.D. thesis, RWTH Aachen, Veröffentlichung des Geodätischen Instituts der RWTH Aachen, vol. 33 (1982)
- Rieder, A.: Keine Probleme mit inversen Problemen. Eine Einführung in ihre stabile Lösung. Vieweg, Wiesbaden (2003)
- Rubinstein, R., Bruckstein, A.M., Elad, M.: Dictionaries for sparse representation modeling. *Proc. IEEE* **98**(6), 1045–1057 (2010)
- Sakumura, C., Bettadpur, S., Bruinsma, S.: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.* **41**(5), 1389–1397 (2014)
- Schall, J.: Optimization of point grids in regional satellite gravity analysis using a Bayesian approach. Ph.D. thesis, University of Bonn, accepted (2019)
- Schmidt, R., Flechtner, F., Meyer, U., Neumayer, K.H., Dahle, C., König, R., Kusche, J.: Hydrological signals observed by the GRACE satellites. *Surv. Geophys.* **29**(4–5), 319–334 (2008)
- Schreiner, M.: A pyramid scheme for spherical wavelets. AGTM Report (170), Geomathematics Group, Kaiserslautern (1996)
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C.: The gravity recovery and climate experiment: mission overview and early results. *Geophys. Res. Lett.* (2004). <https://doi.org/10.1029/2004GL019920>
- Telschow, R.: An Orthogonal Matching Pursuit for the Regularization of Spherical Inverse Problems. Ph.D. thesis, University of Siegen, Verlag Dr. Hut, Munich (2014)
- Vigerske, S., Wächter, A., Kawajir, Y., Laird, C.: Introduction to IPOPT: A tutorial for downloading, installing, and using IPOPT (2016). <https://projects.coin-or.org/Ipopt/browser/stable/3.11/Ipopt/doc/documentation.pdf?format=raw>. Last Accessed 11 Dec 2018
- Wächter, A., Biegler, L.T.: Line search filter methods for nonlinear programming: local convergence. *SIAM J. Optim.* **16**(1), 1–31 (2005a)
- Wächter, A., Biegler, L.T.: Line search filter methods for nonlinear programming: motivation and global convergence. *SIAM J. Optim.* **16**(1), 32–48 (2005b)
- Wächter, A., Biegler, L.T.: On the implementation of a primal-dual interior point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106**(1), 25–57 (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.