Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

RFMP: An Iterative Best Basis Algorithm for Inverse Problems in the Geosciences

Volker Michel*
Geomathematics Group, University of Siegen, Siegen, Germany

1 Abstract

We summarize a recently introduced greedy algorithm for inverse problems in geomathematics. The algorithm is able to combine heterogeneous systems of trial functions to construct a stable approximation to the solution of the given ill-posed problem. The representation of this approximation with respect to the trial functions of mixed types is sparse in the sense that essentially less trial functions than available are used. Some new theoretical results about the method are also proved here.

2 Introduction

Numerous systems of trial functions are available today for the resolution of problems on the sphere or the ball. For the sphere itself, the spherical harmonics are available as a very popular system of orthogonal polynomials (see, e.g., Müller 1966; Heiskanen and Moritz 1981; Freeden et al. 1998; Michel 2013). Its features and algorithms for its applications are well-known. For instance, a fast Fourier transform which is applicable also to non-equispaced data is available in Keiner et al. (2009). Whereas spherical harmonics have particular advantages, they are also connected to drawbacks which become evident, e.g., if spatial irregularities in the data (like regionally varying noise levels or strongly scattered data grids) occur or if the data are only given or shall only be analyzed regionally. For such applications, localized trial functions have been developed in the last decades. Examples of methods of this kind are a spline method based on reproducing kernels (see Freeden 1981a, b; Freeden et al. 1998) and numerous wavelet methods. A small and incomplete selection for the latter kind can be found in Schröder and Sweldens (1995), Freeden and Windheuser (1996), Holschneider (1996), Freeden and Schreiner (1998), Antoine and Vandergheynst (1999), and Gerhards (2011). Moreover, the development of spherical Slepian functions (see Wieczorek and Simons (2005), Simons and Dahlen (2006), Simons et al. (2006), Wieczorek and Simons (2007), and Dahlen and Simons (2008) as well as FJ Simons' article in this handbook) yielded the possibility to construct trial functions which are tailored for a particular spherical subregion of personal interest.

Also for applications on the ball, global orthogonal trial functions (based on Dufour 1977; Ballani et al. 1993; Tscherning 1996) are available as well as localized trial functions, which can be used for a spline interpolation and approximation (see Fengler et al. 2006; Amirbekyan and Michel 2008; Berkel 2009; Berkel and Michel 2010; Berkel et al. 2011) as well as a wavelet analysis (see Michel 2002, 2005a, b). Note that, in particular, ill-posed inverse problems have also been handled

*E-mail: michel@mathematik.uni-siegen.de, www.geomathematics-siegen.de

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

with such tools. For a survey on these methods, see the textbook Michel (2013) and the author's other article in this handbook.

All these methods have their own advantages and disadvantages. There is, certainly, no such thing as a "perfect" method. For instance, global basis functions are ideal to represent global trends by low degree polynomials, but high-frequent structures or perturbations which are restricted to subdomains can cause problems. Moreover, not all localized trial functions can be used for all kinds of problems. Some trial functions are isotropic and are, therefore, a good choice to solve differential or integral equations represented by isotropic operators. However, non-isotropic operators or non-isotropic noise can be handled better with other approaches, which are, themselves not ideal for isotropic problems. Moreover, localized trial functions usually provide us with the possibility to vary their localization, i.e., the size of the "hat" of the hat function. However, not all numerical algorithms associated to such localized functions allow the combination of different levels of localization, although this would certainly bring some advantages, since not all local phenomena are of the same spatial size. Furthermore, some numerical algorithms for the calculation of an expansion of the unknown solution in a selected basis are based on the resolution of a system of linear equations. Large data sizes yield, consequently, limits for the numerical realization since the inversion of the system becomes too time-consuming or too unstable in such cases. For some methods, sophisticated numerical algorithms which compensate for such problems are known (see, e.g., the approach in Gutting (2012) for a spherical spline method), whereas such techniques are unknown for some other methods at present. Moreover, there exist methods which can do without the resolution of a system of linear equations, but require a numerical integration. Numerical integration methods on the sphere are well-known for regular data grids. For irregular data grids on the sphere, a promising method was developed in Gräf et al. (2009). However, at present, the largest polynomial exactness which could be reached is given by degree 1,024, whereas some problems in gravitational field modeling and analysis require higher exactness degrees. Furthermore, the development of quadrature rules for subdomains of the sphere is still in its infancy (see Mhaskar (2004a, b) and Beckmann et al. (2012) for first approaches).

In Mallat and Zhang (1993), a greedy algorithm which is called a Matching Pursuit (MP) is developed for the interpolation of data and is applied to problems in a Euclidean setting. This algorithm is further elaborated and extended to the use of kernel-based trial functions in Vincent and Bengio (2002). In Fischer (2011) and Fischer and Michel (2012), we used the ideas of this Matching Pursuit to develop an algorithm which has three additional features: first, it allows the resolution of inverse problems (represented by a set of functionals), i.e., the data and the solution need not be in the same space anymore and they are connected by equations. Second, a Tikhonov regularization was included to stabilize ill-posed problems. Third, in the numerical implementation, we used trial functions which are relevant for geoscientific problems (more precisely, a three-dimensional ball was used as a domain of the unknown function as it occurs for tomographic inverse problems; for the treatment of a sphere, i.e., the surface of a ball, as the domain, see Michel and Telschow (2014) and Telschow (2014)). This new algorithm, which contains the classical Matching Pursuit as a particular case, is, indeed, applicable to ill-posed inverse problems in the geosciences as we demonstrated in the numerical experiments in Fischer (2011) and Fischer and Michel (2012, 2013a, b). The particular advantages of this novel technique, which is called the Regularized Functional Matching Pursuit (RFMP), are as follows:

- Trial functions of different types (e.g., global trial functions such as polynomials and localized hat-like functions with different hat widths) can be combined to represent the solution.

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

This allows the combination of the advantages that different kinds of basis functions provide. The algorithm itself chooses a best basis (in a particular sense explained further below) out of a large set of possible trial functions. This set is called a dictionary and typically is constructed as a union of different basis systems, i.e., it is overcomplete.

- The approximation is calculated iteratively, i.e., step by step the next element of the best basis is selected by the RFMP. As a consequence, intermediate results can also be analyzed in the sense of a multiresolution.
- It is neither necessary to solve a system of linear equations nor to use a quadrature rule. Therefore, the method is more robust with respect to the use of large data grids (connected to high expectations on the accuracy of the result) and with respect to the handling of strongly irregular data grids. The latter is, in particular, also achieved by taking the least square error as a basis for the search for the solution. Hence, data points which are very close to each other simply cause similar summands in the calculation of the error functional but do not cause almost identical rows in a matrix to be inverted.

The purpose of this article is to summarize the previous publications on the RFMP, to explain the algorithm, to prove the theoretical results, and to show some examples of previous numerical experiments. These objectives are represented in the outline of this paper. In Sect. 3, we describe the general setting of those inverse problems which can be solved with the RFMP. In Sect. 4, we derive the formulae of the RFMP and formulate the algorithm. In Sect. 5, we prove the theoretical properties of the RFMP, where we are able to partially extend the previous knowledge about the properties of our algorithm. In Sect. 6, we investigate the conditions of the convergence theorem proved in Sect. 5 in the case of particular examples of trial functions and geomathematical inverse problems. Furthermore, in Sect. 7, we recapitulate some examples of numerical results obtained with the RFMP and report our further experiences from numerical experiments. Finally, in Sect. 8, we conclude the paper with a summary and an outlook on forthcoming publications.

3 Constellation of the Problem

We consider **inverse problems** of the form

$$\mathcal{F}^k F = y_k, \quad k = 1, \ldots, l,$$

where $y = (y_1, \ldots, y_l)^{\mathrm{T}} \in \mathbb{R}^l$ is given, each $\mathcal{F}^k : \mathrm{L}^2(D) \to \mathbb{R}$ $(k = 1, \ldots, l)$ is a linear and continuous **functional**, $F \in \mathrm{L}^2(D)$ is unknown, and $D \subset \mathbb{R}^d$ is a measurable domain (in particular, D could be the **unit sphere** $\Omega \subset \mathbb{R}^3$ or the **ball** $\mathcal{B} \subset \mathbb{R}^3$ with radius $\beta > 0$). Furthermore, $\mathrm{L}^2(D)$ stands for the usual Hilbert space of (almost everywhere identical) square-integrable functions on D. For some examples of (tomographic) inverse problems in the case $D = \mathcal{B}$, see the author's other article in this handbook. We will summarize the l functionals $\mathcal{F}^k$ in the form of a vector in the **operator** $\mathcal{F} : \mathrm{L}^2(D) \to \mathbb{R}^l$.

For the approximation of F, we have a set of trial functions $\mathcal{D} \subset \mathrm{L}^2(D)$ available, which we call a **dictionary**. Without loss of generality, we assume that $0 \notin \mathcal{D}$. In the case $D = \Omega$, $\mathcal{D}$ could, for example, contain spherical harmonics, spline basis functions, wavelets, and Slepian functions. In the case $D = \mathcal{B}$, $\mathcal{D}$ could, for instance, contain the corresponding counterparts of the global

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

(i.e., polynomial) and localized trial functions known on the sphere. For further details on trial functions on the sphere and the ball, see the textbook Michel (2013) and the references therein.

4 The Algorithm

The basic idea of the algorithm is to iteratively construct a sequence of approximations $(F_n)_n$ to the unknown function F by consecutively adding further summands to the approximation. In other words, we start with $F_0 := 0$ (or some initial approximation) and continue with $F_{n+1} := F_n + \alpha_{n+1} d_{n+1}$ for all $n \in \mathbb{N}_0$. The new summand $\alpha_{n+1} d_{n+1}$ is chosen such that

$$\|y - \mathcal{F}(F_n + \alpha_{n+1} d_{n+1})\|_{\mathbb{R}^l}^2 + \lambda \|F_n + \alpha_{n+1} d_{n+1}\|_{\mathrm{L}^2(D)}^2$$

is minimized, where $\lambda \in \mathbb{R}_0^+$ is a regularization parameter. We first introduce some basic notations.

Definition 1. For a sequence of approximations $(F_n)_n \subset \operatorname{span} \mathcal{D} \subset \mathrm{L}^2(D)$, we define the sequence of **residuals** $(R^n)_n \subset \mathbb{R}^l$ by $R^n := y - \mathcal{F} F_n$. Moreover, the family of mappings $J_\lambda : \mathbb{R}^l \times \mathrm{L}^2(D) \times \mathcal{D} \times \mathbb{R} \to \mathbb{R}$, $\lambda \in \mathbb{R}_0^+$, is defined by

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

For the mapping J_λ, we, obviously, get (note that $\mathcal{F}$ is linear)

$$\begin{aligned} J_\lambda(y, F, d, \alpha) = {} & \|y - \mathcal{F}F\|_{\mathbb{R}^l}^2 - 2\alpha \langle y - \mathcal{F}F, \mathcal{F}d \rangle_{\mathbb{R}^l} + \alpha^2 \|\mathcal{F}d\|_{\mathbb{R}^l}^2 \\ & + \lambda \|F\|_{\mathrm{L}^2(D)}^2 + 2\lambda\alpha \langle F, d \rangle_{\mathrm{L}^2(D)} + \lambda\alpha^2 \|d\|_{\mathrm{L}^2(D)}^2 . \end{aligned}$$

Hence, if we already know the first $n+1$ approximations $F_0, \ldots, F_n$, then

$$\begin{aligned} J_\lambda(y, F_n, d, \alpha) = {} & \|R^n\|_{\mathbb{R}^l}^2 - 2\alpha \langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} + \alpha^2 \|\mathcal{F}d\|_{\mathbb{R}^l}^2 \\ & + \lambda \|F_n\|_{\mathrm{L}^2(D)}^2 + 2\lambda\alpha \langle F_n, d \rangle_{\mathrm{L}^2(D)} + \lambda\alpha^2 \|d\|_{\mathrm{L}^2(D)}^2 , \end{aligned} \tag{1}$$

which has to be minimized to find $\alpha_{n+1} \in \mathbb{R}$ and $d_{n+1} \in \mathcal{D}$. As a consequence, we get

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial \alpha} J_\lambda(y, F_n, d_{n+1}, \alpha) \right) \Big|_{\alpha = \alpha_{n+1}} \\ &= -2 \langle R^n, \mathcal{F}d_{n+1} \rangle_{\mathbb{R}^l} + 2\alpha_{n+1} \|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 \\ &\quad + 2\lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(D)} + 2\lambda\alpha_{n+1} \|d_{n+1}\|_{\mathrm{L}^2(D)}^2 \end{aligned}$$

such that

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

We insert this result in (1) and obtain

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

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

This motivates the following algorithm.

Algorithm 1 (Regularized Functional Matching Pursuit, RFMP). *Let $y \in \mathbb{R}^l$, a linear and continuous operator $\mathcal{F} : \mathrm{L}^2(D) \to \mathbb{R}^l$, and an initial approximation $F_0 \in \mathrm{L}^2(D)$ be given.*

1. *Set $n := 0$ and $R^0 := y - \mathcal{F} F_0$, choose a stopping criterion (e.g., require $\| R^{n+1} \| < \varepsilon$ for a given $\varepsilon > 0$ or require $n + 1 \leq N$ for a given $N \in \mathbb{N}$), and choose a regularization parameter $\lambda \in \mathbb{R}_0^+$ (e.g., with the L-curve method).*
2. *Determine*

$$
d_{n+1} := \operatorname*{arg\,max}_{d \in \mathcal{D}} \frac{\left(\langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(D)} \right)^2}{\left\| \mathcal{F} d \right\|_{\mathbb{R}^l}^2 + \lambda \left\| d \right\|_{\mathrm{L}^2(D)}^2} , \qquad (3)
$$

$$
\alpha_{n+1} := \frac{\langle R^n, \mathcal{F} d_{n+1} \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(D)}}{\left\| \mathcal{F} d_{n+1} \right\|_{\mathbb{R}^l}^2 + \lambda \left\| d_{n+1} \right\|_{\mathrm{L}^2(D)}^2} \qquad (4)
$$

   and set $F_{n+1} := F_n + \alpha_{n+1} d_{n+1}$ and $R^{n+1} := R^n - \alpha_{n+1} \mathcal{F} d_{n+1}$.
3. *If the stopping criterion is fulfilled, then F_{n+1} is the output. Otherwise, increase n by 1 and go to step 2.*

Note that the choice of the optimal dictionary element d_{n+1} is the most expensive part of the algorithm since every iteration step $n + 1$ requires the calculation of the fraction in (3) for each dictionary element $d \in \mathcal{D}$. For an efficient implementation, one should compute $\mathcal{F} d$, $\| \mathcal{F} d \|_{\mathbb{R}^l}$, and $\| d \|_{\mathrm{L}^2(D)}$ for all $d \in \mathcal{D}$ in the preprocessing and store the corresponding scalars and vectors, respectively. Moreover, the formulae

$$
\begin{aligned}
\left\langle R^{n+1}, \mathcal{F} d \right\rangle_{\mathbb{R}^l} &= \langle R^n, \mathcal{F} d \rangle_{\mathbb{R}^l} - \alpha_{n+1} \langle \mathcal{F} d_{n+1}, \mathcal{F} d \rangle_{\mathbb{R}^l} , \\
\langle F_{n+1}, d \rangle_{\mathrm{L}^2(D)} &= \langle F_n, d \rangle_{\mathrm{L}^2(D)} + \alpha_{n+1} \langle d_{n+1}, d \rangle_{\mathrm{L}^2(D)}
\end{aligned}
$$

allow a fast update of the required inner products. Provided that enough memory is available, one can further reduce the computational expenses by calculating the inner products $\langle \mathcal{F} d, \mathcal{F} \tilde{d} \rangle_{\mathbb{R}^l}$ and $\langle d, \tilde{d} \rangle_{\mathrm{L}^2(D)}$ for all pairs $(d, \tilde{d}) \in \mathcal{D}^2$ of dictionary elements in the preprocessing and by using them in these updates.

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

5 Properties of the Algorithm

In the following, we derive some theorems for the algorithm. We start with theorems which are valid in the regularized case ($\lambda > 0$) as well as the unregularized case ($\lambda = 0$).

Theorem 1. *The sequence* $(\|R^n\|^2_{\mathbb{R}^l} + \lambda\|F_n\|^2_{\mathrm{L}^2(D)})_n$ *produced by the algorithm is monotonically decreasing and convergent.*

Proof. The monotonicity is a direct consequence of (2). Moreover, since the monotonically decreasing sequence is bounded from below, it is convergent. ■

Theorem 1 in combination with (2) and (3), also shows that the strategy of the algorithm is to reduce the value $\|R^n\|^2_{\mathbb{R}^l} + \lambda\|F_n\|^2_{\mathrm{L}^2(D)}$ as much as possible in each iteration step. This means that, in this respect, an optimal progress in the Tikhonov regularized data misfit is achieved.

We can certainly not expect that the algorithm produces a sequence $(F_n)_n$ which converges to an exact solution F of $\mathcal{F}F = y$, if $\lambda > 0$. This is also something that we do not desire, if we try to solve an inverse problem which is ill-posed due to an unstable solution since F discontinuously depends on y, i.e., F is unstable with respect to small perturbations of the data vector y. Moreover, we certainly also cannot expect that the sequence $(\|R^n\|^2_{\mathbb{R}^l} + \lambda\|F_n\|^2_{\mathrm{L}^2(D)})_n$ converges to 0 since this would imply that $(F_n)_n$ converges to 0 and $y = 0$.

We now prove a new theorem, which extends a result which we have previously published for the unregularized case only (see Theorem 4.5 in Fischer and Michel 2012).

Theorem 2 (Convergence Theorem). *Let the dictionary $\mathcal{D}$ satisfy the following properties:*

1. *"Semi-frame condition": There exists a constant $c > 0$ such that, for all expansions $H = \sum_{k=1}^{\infty} \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k are not necessarily pairwise distinct but $\{j \in \mathbb{N} \mid d_j = d_k\}$ is a finite set for each $k \in \mathbb{N}$, the following inequality is valid:*

$$c\|H\|^2_{\mathrm{L}^2(D)} \leq \sum_{k=1}^{\infty} \beta_k^2 .$$

2. $C_1 := \inf_{d\in\mathcal{D}}(\|\mathcal{F}d\|^2_{\mathbb{R}^l} + \lambda\|d\|^2_{\mathrm{L}^2(D)}) > 0.$

If the sequence $(F_n)_n$ is produced by the RFMP and no dictionary element is chosen infinitely often, then $(F_n)_n$ converges in $\mathrm{L}^2(D)$ to $F_\infty := \sum_{n=1}^{\infty} \alpha_n d_n \in \mathrm{L}^2(D)$. Moreover, the following holds true:

(a) If $\overline{\operatorname{span}\mathcal{D}}^{\|\cdot\|_{\mathrm{L}^2(D)}} = \mathrm{L}^2(D)$, $C_2 := \sup_{d\in\mathcal{D}} \|d\|_{\mathrm{L}^2(D)} < +\infty$, and $\lambda \in \mathbb{R}_0^+$ is an arbitrary parameter, then F_∞ solves

$$(\mathcal{F}^*\mathcal{F} + \lambda I)F_\infty = \mathcal{F}^* y ,$$

where $\mathcal{F}^*$ is the adjoint operator corresponding to $\mathcal{F}$ and I is the identity operator on $\mathrm{L}^2(D)$. In other words,

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

$$\|y - \mathcal{F}F_\infty\|_{\mathbb{R}^l}^2 + \lambda \|F_\infty\|_{\mathrm{L}^2(D)}^2 = \min_{F \in \mathrm{L}^2(D)} \left(\|y - \mathcal{F}F\|_{\mathbb{R}^l}^2 + \lambda \|F\|_{\mathrm{L}^2(D)}^2\right),$$

where the minimizer is unique, if $\lambda > 0$.

(b) If $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$ *and* $\lambda = 0$, *then* F_∞ *solves* $\mathcal{F}F_\infty = y$.

Proof. According to (4) and (2), we have

$$\begin{aligned}
\alpha_n^2 &= \left(\|\mathcal{F}d_n\|_{\mathbb{R}^l}^2 + \lambda \|d_n\|_{\mathrm{L}^2(D)}^2\right)^{-1} \frac{\left(\langle R^{n-1}, \mathcal{F}d_n\rangle_{\mathbb{R}^l} - \lambda \langle F_{n-1}, d_n\rangle_{\mathrm{L}^2(D)}\right)^2}{\|\mathcal{F}d_n\|_{\mathbb{R}^l}^2 + \lambda \|d_n\|_{\mathrm{L}^2(D)}^2} \\
&= \left(\|\mathcal{F}d_n\|_{\mathbb{R}^l}^2 + \lambda \|d_n\|_{\mathrm{L}^2(D)}^2\right)^{-1} \\
&\quad \times \left[\|R^{n-1}\|_{\mathbb{R}^l}^2 + \lambda \|F_{n-1}\|_{\mathrm{L}^2(D)}^2 - \left(\|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(D)}^2\right)\right]. \qquad (5)
\end{aligned}$$

Hence, condition 2 of the theorem to be proved as well as Theorem 1 yield

$$\begin{aligned}
\sum_{n=N}^{\infty} \alpha_n^2 &\le \frac{1}{C_1} \sum_{n=N}^{\infty} \left[\|R^{n-1}\|_{\mathbb{R}^l}^2 + \lambda \|F_{n-1}\|_{\mathrm{L}^2(D)}^2 - \left(\|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(D)}^2\right)\right] \\
&= \frac{1}{C_1} \left[\|R^{N-1}\|_{\mathbb{R}^l}^2 + \lambda \|F_{N-1}\|_{\mathrm{L}^2(D)}^2 - \lim_{n\to\infty} \left(\|R^n\|_{\mathbb{R}^l}^2 + \lambda \|F_n\|_{\mathrm{L}^2(D)}^2\right)\right]. \qquad (6)
\end{aligned}$$

Obviously, for $N \to \infty$, the right-hand side converges to 0 and, consequently, also the left-hand side. Thus, condition 1 of the theorem to be proved yields that $F_\infty := \sum_{n=1}^{\infty} \alpha_n d_n$ is a convergent series in $\mathrm{L}^2(D)$:

$$\lim_{N\to\infty} \|F_\infty - F_{N-1}\|_{\mathrm{L}^2(D)}^2 = \lim_{N\to\infty} \left\|\sum_{n=N}^{\infty} \alpha_n d_n\right\|_{\mathrm{L}^2(D)}^2 \le \frac{1}{c} \lim_{N\to\infty} \sum_{n=N}^{\infty} \alpha_n^2 = 0.$$

As a consequence, the sequence of summands must converge to 0, i.e., with (4), we get

$$\alpha_{n+1}^2 = \left(\frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1}\rangle_{\mathrm{L}^2(D)}}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 + \lambda \|d_{n+1}\|_{\mathrm{L}^2(D)}^2}\right)^2 \longrightarrow 0 \quad \text{as } n \to \infty.$$

Moreover, with the continuity of $\mathcal{F}$, the boundedness of the dictionary in the case (a), and (3), we obtain, for all $d \in \mathcal{D}$, the inequality

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

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

$$\geq \frac{1}{(\|\mathcal{F}\|_{\mathcal{L}}^2 + \lambda)C_2^2} \frac{\left(\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d\rangle_{\mathrm{L}^2(D)}\right)^2}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(D)}^2},$$

where $\|\mathcal{F}\|_{\mathcal{L}}$ is the usual operator norm given by

$$\|\mathcal{F}\|_{\mathcal{L}} := \sup_{\substack{F \in \mathrm{L}^2(D) \\ F \neq 0}} \frac{\|\mathcal{F}F\|_{\mathbb{R}^l}}{\|F\|_{\mathrm{L}^2(D)}}$$

for continuous operators. Hence,

$$\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d\rangle_{\mathrm{L}^2(D)} = \langle \mathcal{F}^* R^n - \lambda F_n, d\rangle_{\mathrm{L}^2(D)} \longrightarrow 0 \quad \text{as } n \to \infty$$

for all $d \in \mathcal{D}$ and, due to the first requirement on $\mathcal{D}$ in the case (a), also for all $d \in \mathrm{L}^2(D)$. Thus, $(\mathcal{F}^* R^n)_n$ weakly converges to λF_∞. However, since $\mathcal{F}^* R^n = \mathcal{F}^* y - \mathcal{F}^* \mathcal{F} F_n$, the continuity of $\mathcal{F}$ and $\mathcal{F}^*$ implies that $(\mathcal{F}^* R^n)_n$ strongly converges to $\mathcal{F}^* y - \mathcal{F}^* \mathcal{F} F_\infty$. Consequently,

$$\mathcal{F}^* y - \mathcal{F}^* \mathcal{F} F_\infty = \lambda F_\infty .$$

Hence,

$$\mathcal{F}^* y = (\mathcal{F}^* \mathcal{F} + \lambda I)\, F_\infty . \tag{7}$$

Let now F_∞ be a fixed solution of (7). It is well-known (see, e.g., Louis 1989, p. 89) that (note that $\mathcal{F}^* \mathcal{F} + \lambda I$ is self-adjoint)

$$\begin{aligned}
&\|y - \mathcal{F}F\|_{\mathbb{R}^l}^2 + \lambda \|F\|_{\mathrm{L}^2(D)}^2 \\
&= \|y\|_{\mathbb{R}^l}^2 - 2\langle y, \mathcal{F}F\rangle_{\mathbb{R}^l} + \langle \mathcal{F}F, \mathcal{F}F\rangle_{\mathbb{R}^l} + \lambda \langle F, F\rangle_{\mathrm{L}^2(D)} \\
&= \|y\|_{\mathbb{R}^l}^2 - 2\,\langle \mathcal{F}^* y, F\rangle_{\mathrm{L}^2(D)} + \langle (\mathcal{F}^* \mathcal{F} + \lambda I)\, F, F\rangle_{\mathrm{L}^2(D)} \\
&= \|y\|_{\mathbb{R}^l}^2 - 2\,\langle \mathcal{F}^* y, F\rangle_{\mathrm{L}^2(D)} + \langle (\mathcal{F}^* \mathcal{F} + \lambda I)\,(F - F_\infty)\,, F - F_\infty\rangle_{\mathrm{L}^2(D)} \\
&\quad + 2\,\langle (\mathcal{F}^* \mathcal{F} + \lambda I)\, F_\infty, F\rangle_{\mathrm{L}^2(D)} - \langle (\mathcal{F}^* \mathcal{F} + \lambda I)\, F_\infty, F_\infty\rangle_{\mathrm{L}^2(D)} \\
&= \|y\|_{\mathbb{R}^l}^2 + \langle (\mathcal{F}^* \mathcal{F} + \lambda I)\,(F - F_\infty)\,, F - F_\infty\rangle_{\mathrm{L}^2(D)} - \langle \mathcal{F}^* y, F_\infty\rangle_{\mathrm{L}^2(D)}
\end{aligned} \tag{8}$$

is minimal if $F = F_\infty$, since $\mathcal{F}^* \mathcal{F} + \lambda I$ is positive semi-definite:

$$\langle (\mathcal{F}^* \mathcal{F} + \lambda I)\, F, F\rangle_{\mathrm{L}^2(D)} = \|\mathcal{F}F\|_{\mathbb{R}^l}^2 + \lambda \|F\|_{\mathrm{L}^2(D)}^2 \geq 0 \quad \text{for all } F \in \mathrm{L}^2(D)\,.$$

Moreover, in the case $\lambda > 0$, (8) is minimal if and only if $F = F_\infty$, since $\mathcal{F}^* \mathcal{F} + \lambda I$ is positive definite. Note that this also implies that

$$F_\infty = (\mathcal{F}^* \mathcal{F} + \lambda I)^{-1} \mathcal{F}^* y$$

is uniquely determined, if $\lambda > 0$.

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

In the unregularized case (case (b), i.e., $\lambda = 0$), Theorem 1 yields that $(\|R^n\|_{\mathbb{R}^l})_n$ converges. Hence, there exists a convergent subsequence $(R^{n_j})_j$ with limit $R^\infty \in \mathbb{R}^l$. Moreover, (3), (5), and the convergence of the series in (6) yield that, for all $d \in \mathcal{D}$,

$$0 \leq \frac{\langle R^\infty, \mathcal{F}d\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2} = \lim_{j\to\infty} \frac{\langle R^{n_j}, \mathcal{F}d\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2} \leq \lim_{j\to\infty} \frac{\left\langle R^{n_j}, \mathcal{F}d_{n_j+1}\right\rangle_{\mathbb{R}^l}^2}{\|\mathcal{F}d_{n_j+1}\|_{\mathbb{R}^l}^2} = 0\,.$$

Due to the requirement that $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$, we get $R^\infty = 0$. Since $(\|R^n\|_{\mathbb{R}^l})_n$ is monotonically decreasing and every convergent subsequence converges to 0, the sequence $(\|R^n\|_{\mathbb{R}^l})$ converges to 0 and, thus, $(R^n)_n$ converges to $0 \in \mathbb{R}^l$. Finally, we use the continuity of $\mathcal{F}$ and conclude that

$$\mathcal{F}F_\infty = \lim_{n\to\infty} \mathcal{F}F_n = \lim_{n\to\infty} (y - R^n) = y\,.$$

■

Note that the requirements of (b) are easier to achieve than the requirements of (a) due to the finite dimensions of $\mathbb{R}^l$ and the infinite dimensions of $\mathrm{L}^2(D)$.

The results of Theorem 2 show that the algorithm, indeed, converges to the desired result. In the absence of a regularization, the limit F_∞ of the algorithm is an exact solution, whereas the regularized case yields a solution of the Tikhonov regularized normal equation, i.e., we, indeed, obtain the minimal (regularized) data misfit, as we desired.

Since the RFMP contains the MP from Mallat and Zhang (1993) and Vincent and Bengio (2002) as a particular case, the convergence rate of the MP, which was proved in Mallat and Zhang (1993), can analogously also be proved for the RFMP in the case $\lambda = 0$, as we observed in Fischer (2011) and Fischer and Michel (2012). We state this result here. Note that the RFMP was called the **FMP** (**Functional Matching Pursuit**) in the case $\lambda = 0$ in Fischer (2011) and Fischer and Michel (2012).

Definition 2. Let a dictionary $\mathcal{D}$ be given. The corresponding **correlation ratio** is given, for each $v \in \mathbb{R}^l \setminus \{0\}$, by

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

Furthermore, we define the **worst case correlation ratio** by

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

The following properties of the correlation ratio are obvious:

- The correlation ratio $\tau(v)$ and, consequently, also $I(\tau)$ are bounded from below by 0 and, due to the Cauchy-Schwarz inequality, bounded from above by 1.
- The correlation ratio $\tau(v)$ is independent of $\|v\|$, i.e., $\tau(rv) = \tau(v)$ for all $r \in \mathbb{R} \setminus \{0\}$ and all $v \in \mathbb{R}^l \setminus \{0\}$.
- In the case $v = R^n$ and $\lambda = 0$, the correlation ratio $\tau(R^n)$ is obtained for $d = d_{n+1}$ (see (3)).

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

Theorem 3 (Convergence Rate of the FMP). *Let a function $F \in \mathrm{L}^2(D)$, the corresponding data vector $y := \mathcal{F}F \in \mathbb{R}^l$, and a dictionary $\mathcal{D}$ with* $\operatorname{span}\{\mathcal{F}d \mid d \in \mathcal{D}\} = \mathbb{R}^l$ *be given and let the sequence $(R^n)_n$ be produced by the RFMP with $\lambda = 0$ (i.e., the FMP). Then the residual converges exponentially to* 0. *More precisely,*

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

for all $n \in \mathbb{N}$, where $I(\tau) > 0$.

Moreover, the choice of a Tikhonov regularization term for the penalty term in J_λ automatically inherits the well-known nice properties of the Tikhonov regularization (see, e.g., Rieder (2003) and the reference therein). These are the stability (i.e., the continuous dependence of the solution with respect to the data vector) and the convergence with respect to the regularization parameter. We start with the following theorem, which (slightly) extends our previous result on the stability of the solution (see Fischer 2011, Theorem 4.7; Fischer and Michel 2012, Theorem 5.4).

Theorem 4 (Stability of the Solution). *Let the dictionary satisfy conditions 1, 2, and (a) of Theorem 2 and let $\lambda > 0$. Moreover, let $(y^k)_k \subset \mathbb{R}^l$ be a convergent sequence with limit $y \in \mathbb{R}^l$ and let, for each $k \in \mathbb{N}_0$, $F_{\infty,k} \in \mathrm{L}^2(D)$ be the corresponding limit produced by the RFMP for the data vector y^k. Then $(F_{\infty,k})_k$ converges to the limit F_∞ produced by the RFMP for the data vector y.*

Proof. Due to Theorem 2, each $F_{\infty,k}$ is the unique minimizer of

$$\mathrm{L}^2(D) \ni F \mapsto \left\|y^k - \mathcal{F}F\right\|_{\mathbb{R}^l}^2 + \lambda \|F\|_{\mathrm{L}^2(D)}^2 . \tag{9}$$

From the well-known theory of Tikhonov regularization (see, e.g., Engl et al. 1989, Theorem 2.1; Rieder 2003, p. 241; Seidman and Vogel 1989, Theorem 2), we know that the minimizer of (9) converges to the minimizer of

$$\mathrm{L}^2(D) \ni F \mapsto \|y - \mathcal{F}F\|_{\mathbb{R}^l}^2 + \lambda \|F\|_{\mathrm{L}^2(D)}^2 \tag{10}$$

as $k \to \infty$. Moreover, the minimization of (10) is equivalent to the equation

$$(\mathcal{F}^*\mathcal{F} + \lambda I)\, F = \mathcal{F}^* y ,$$

which is uniquely solved by the limit F_∞ of the RFMP in the case of the data vector y, as we showed in the proof of Theorem 2. This completes the proof of Theorem 4. ■

The second theorem regarding the regularization addresses the convergence of the solution with respect to the regularization parameter.

Theorem 5 (Convergence of the Regularization). *Let $y \in \mathcal{F}(\mathrm{L}^2(D))$ be a given (exact) data vector and $(y^\varepsilon)_{\varepsilon>0} \subset \mathbb{R}^l$ be a family of given perturbed data vectors with $\|y - y^\varepsilon\|_{\mathbb{R}^l} \leq \varepsilon$. Moreover, let F^+ be the minimum-norm solution of $\mathcal{F}F = y$, i.e.,*

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

$$\left\| F^+ \right\|_{\mathrm{L}^2(D)} = \min \left\{ \|F\|_{\mathrm{L}^2(D)} \middle| \; F \in \mathrm{L}^2(D) \text{ and } \mathcal{F}F = y \right\} ,$$

which is obtained by the Moore-Penrose pseudoinverse $\mathcal{F}^+$ as $F^+ = \mathcal{F}^+ y$. Furthermore, let the regularization parameter $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ be chosen in dependence on the noise level ε in the sense that

$$\lim_{\varepsilon \to 0+} \lambda(\varepsilon) = 0 = \lim_{\varepsilon \to 0+} \frac{\varepsilon^2}{\lambda(\varepsilon)} .$$

If $(F_{\infty,\varepsilon})_{\varepsilon>0}$ denotes the family of limits produced by the RFMP in the case of the data vector y^ε and the regularization parameter $\lambda(\varepsilon)$ for each $\varepsilon > 0$, provided that again conditions 1, 2, and (a) of Theorem 2 are satisfied, then

$$\lim_{\varepsilon \to 0+} \left\| F_{\infty,\varepsilon} - F^+ \right\|_{\mathrm{L}^2(D)} = 0 .$$

Proof. According to Theorem 2, the limit $F_{\infty,\varepsilon}$ uniquely minimizes

$$\mathrm{L}^2(D) \ni F \mapsto \|y^\varepsilon - \mathcal{F}F\|^2_{\mathbb{R}^l} + \lambda(\varepsilon)\|F\|^2_{\mathrm{L}^2(D)} . \tag{11}$$

From, e.g., Engl et al. (1996, Theorem 5.2), we know that the limit $\varepsilon \to 0+$ and the conditions stated above yield that the family of minimizers of (11) has the required convergence property. ■

6 Discussion of Some Dictionaries

We discuss here some examples of inverse problems and dictionaries that are relevant for geomathematical problems. The list of examples is certainly not complete and not all theoretical problems that occur here have been solved so far. For reasons of simplicity, we assume here, without loss of generality, that all dictionary elements have been normalized, i.e., $\|d\|_{\mathrm{L}^2(D)} = 1$ for all $d \in \mathrm{L}^2(D)$. This has also been used in the majority of our numerical experiments since this assumption simplifies the calculation of the denominators of (3) and (4) in Algorithm 1. The primary objective of this section is the investigation of the conditions stated in Theorem 2. The assumption on a normalized dictionary has the consequence that the requirement on C_1 becomes trivial in the regularized case ($\lambda > 0$), whereas it (always) reduces to the requirement that $\inf_{d \in \mathcal{D}} \|\mathcal{F}d\|_{\mathbb{R}^l} > 0$ in the unregularized case ($\lambda = 0$). Moreover, we, obviously, have $C_2 = 1$ in the case of this assumption.

We focus here on two particular domains: the unit sphere $\Omega \subset \mathbb{R}^3$ and the ball $\mathcal{B} \subset \mathbb{R}^3$ with radius $\beta > 0$. For both domains, a series of trial functions is available, which we will only briefly summarize here. For further details, see, e.g., the textbook Michel (2013) and the author's other article in this handbook. We can, roughly, subdivide the trial functions into two types: global and localized trial functions.

Examples of **global trial functions** are orthogonal polynomials such as the so-called **spherical harmonics** $\{Y_{n,j}\}_{n \in \mathbb{N}_0;\, j=1,\ldots,2n+1}$ (see the first column of Fig. 1) on the unit sphere Ω as well as the systems $\{G^{\mathrm{I}}_{m,n,j}\}_{m,n \in \mathbb{N}_0;\, j=1,\ldots,2n+1}$ and $\{G^{\mathrm{II}}_{m,n,j}\}_{m,n \in \mathbb{N}_0;\, j=1,\ldots,2n+1}$ on the ball $\mathcal{B}$, where it should be mentioned that the latter functions on $\mathcal{B}$ are not polynomials in Cartesian coordinates. All three

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

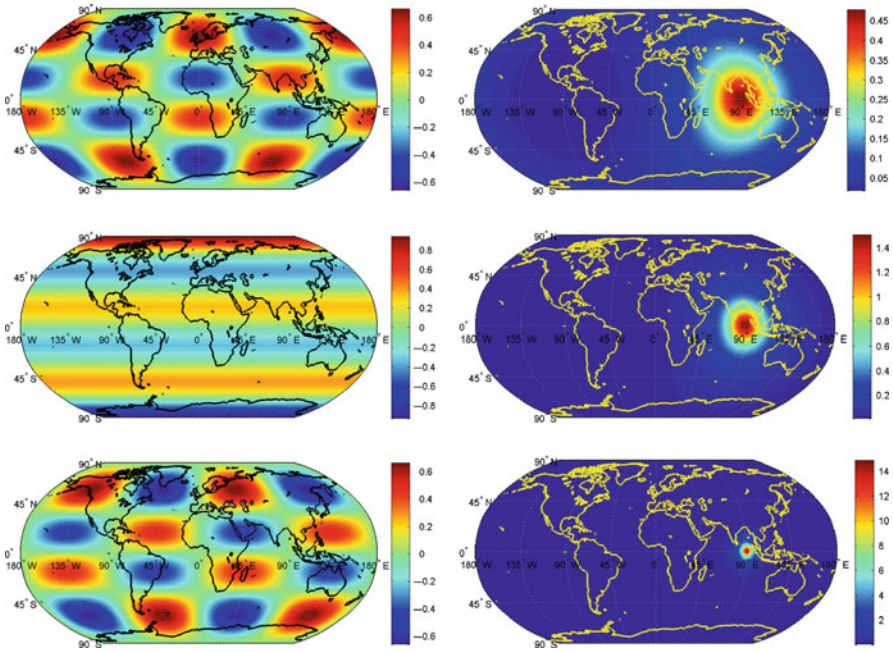

Fig. 1 Examples of functions which can be used as elements of a dictionary on the sphere are spherical harmonics (see the first column for examples of degree $n = 5$) as global trial functions and Abel-Poisson kernels (see the second column, where $h = 0.5$ (*top*), $h = 0.7$ (*middle*), and $h = 0.9$ (*bottom*) and, in each case, $\xi = (0, 1, 0)^{\mathrm{T}}$ were used), which are localized trial functions and are often used as spline or wavelet basis functions on the sphere

systems are orthonormal basis systems in $\mathrm{L}^2(\Omega)$ and $\mathrm{L}^2(\mathcal{B})$, respectively, and, therefore, trivially satisfy the semi-frame condition (condition 1 of Theorem 2) as well as condition (a) of Theorem 2.

Some **localized trial functions** are based on reproducing kernels $K_h(\cdot, \cdot)$ in the sense that $K_h(x, \cdot)$ is a **hat function**, where x is the **center of the hat** and the parameter h controls the localization, i.e., the **"hat width"**. They can be used for spline interpolation/approximation as well as for a wavelet analysis. A celebrated example of a localized trial function on the sphere Ω is generated by the **Abel-Poisson kernel** in the sense that

$$K_h(\xi \cdot) : \Omega \ni \eta \mapsto K_h(\xi \cdot \eta) := \frac{1 - h^2}{4\pi \left(1 + h^2 - 2h\, \xi \cdot \eta\right)^{3/2}} ,$$

see the second column of Fig. 1. Further examples on the sphere and the ball are known.

For many inverse problems in geomathematics, **singular value decompositions** are known. This can, e.g., occur in the sense that the operator $\mathcal{F}$ satisfies

$$\mathcal{F} Y_{n,j} = \left(\tau_n Y_{n,j}\left(\eta^k\right)\right)_{k=1,\ldots,l} ,$$

where $(\tau_n)_n$ converges to 0 and $\{\eta^k\}_{k=1,\ldots,l}$ is a grid of pairwise distinct points on Ω. An example is the **downward continuation problem**, where $\mathcal{F}$ maps a harmonic potential (such as the gravitational potential) from the surface of the Earth to a higher altitude (e.g., a satellite altitude). The inverse problem consists of the recovery of the potential at the surface from given data at the point grid $\{\eta^k\}_{k=1,\ldots,l}$ at the high altitude. In this example (see also the results in Telschow (2014) for the application of an enhanced version of the RFMP to the downward continuation problem),

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

we have

$$\tau_n = c\,\sigma^n \text{ for constants } c > 0 \text{ and } 0 < \sigma < 1. \tag{12}$$

If we include all spherical harmonics basis functions $\{Y_{n,j}\}_{n\in\mathbb{N}_0;\, j=1,\ldots,2n+1}$ in the dictionary $\mathcal{D}$, then the well-known maximum-norm estimate for spherical harmonics $\max_{\zeta\in\Omega} |Y_{n,j}(\zeta)| \le \sqrt{\frac{2n+1}{4\pi}}$ (see, e.g., Freeden et al. 1998, Lemma 3.1.5 or Müller 1966, Lemma 8) yields

$$\begin{aligned}
\left\|\mathcal{F}Y_{n,j}\right\|^2_{\mathbb{R}^l} &= \sum_{k=1}^{l} \tau_n^2 \left(Y_{n,j}\left(\eta^k\right)\right)^2 \\
&\le \sum_{k=1}^{l} \tau_n^2 \frac{2n+1}{4\pi} \\
&= l\,\tau_n^2 \frac{2n+1}{4\pi} \quad \underset{n\to\infty}{\longrightarrow} \quad 0\,.
\end{aligned}$$

Hence, $\inf_{d\in\mathcal{D}} \|\mathcal{F}d\|_{\mathbb{R}^l} = 0$ in this case. This violation of condition 2 of Theorem 2 (in the unregularized case) cannot be compensated by renormalizing the dictionary elements for the following reason: let us replace the normalized basis $\{Y_{n,j}\}_{n\in\mathbb{N}_0;\, j=1,\ldots,2n+1} \subset \mathcal{D}$ by a system $\{\beta_n Y_{n,j}\}_{n\in\mathbb{N}_0;\, j=1,\ldots,2n+1} \subset \mathcal{D}$. We have to take into account the (possible) arbitrariness of the point grid $\{\eta^k\}_{k=1,\ldots,l}$ such that an assumption like $\max_{k=1,\ldots,l} |Y_{n,j}(\eta^k)| \ge \gamma\sqrt{\frac{2n+1}{4\pi}}$ for all $n \in \mathbb{N}_0$ and all $j = 1,\ldots,2n+1$, where γ is a fixed constant with $0 < \gamma < 1$, appears to be reasonable (note the maximum-norm estimate mentioned above). In this case, condition 2 of Theorem 2 is only satisfied (in the unregularized case), if

$$\begin{aligned}
\inf_{\substack{n\in\mathbb{N}_0\\ j=1,\ldots,2n+1}} \|\mathcal{F}(\beta_n Y_{n,j})\|^2_{\mathbb{R}^l} &= \inf_{\substack{n\in\mathbb{N}_0\\ j=1,\ldots,2n+1}} \sum_{k=1}^{l} \tau_n^2 \left(\beta_n\, Y_{n,j}\left(\eta^k\right)\right)^2 \\
&\ge \gamma^2 \inf_{n\in\mathbb{N}_0} \left(\tau_n^2 \beta_n^2 \frac{2n+1}{4\pi}\right) \\
&> 0\,.
\end{aligned}$$

As a consequence, we have to choose $(\beta_n)_n$ as an exponentially diverging sequence due to (12). This, however, violates both condition (a) of Theorem 2 since

$$C_2 = \sup_{d\in\mathcal{D}} \|d\|_{\mathrm{L}^2(\Omega)} \ge \sup_{\substack{n\in\mathbb{N}_0\\ j=1,\ldots,2n+1}} \|\beta_n Y_{n,j}\|_{\mathrm{L}^2(\Omega)} = \sup_{n\in\mathbb{N}_0} \beta_n = +\infty$$

and the semi-frame condition (condition 1 of Theorem 2) since

$$\left\| \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{1}{\beta_n} \left(\beta_n Y_{n,j}\right) \right\|^2_{\mathrm{L}^2(\Omega)} = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} 1 = +\infty\,,$$

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

but

$$\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\frac{1}{\beta_n^2} < +\infty\,.$$

As a consequence, the convergence to a solution in the unregularized case cannot be guaranteed in the case of spherical harmonics. This might have the consequence that numerical instabilities occur, if relatively low regularization parameters λ are used in the combination with a dictionary with high degree spherical harmonics.

Let us discuss the following alternative: we insert Abel-Poisson kernels (with a grid of pairwise distinct points $\{\xi^i\}_{i\in\mathbb{N}_0}$ which is dense in Ω)

$$\begin{aligned}
K_h\left(\xi^i\cdot\right) &= \frac{1-h^2}{4\pi\left(1+h^2-2h\,\xi^i\cdot\right)^{3/2}}\\
&= \sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\,h^n\,P_n\left(\xi^i\cdot\right)\\
&= \sum_{n=0}^{\infty}h^n\sum_{j=1}^{2n+1}Y_{n,j}\left(\xi^i\right)Y_{n,j}
\end{aligned}$$

in the dictionary, where we used the addition theorem for spherical harmonics (see, e.g., Freeden et al. 1998, Theorem 3.1.3 or Müller 1966, Theorem 2) in this representation and P_n is a Legendre polynomial of degree n. We now verify the existence of C_1 and C_2. Since $\mathcal{F}$ is linear and continuous, we get

$$\begin{aligned}
\left\|\mathcal{F}K_h\left(\xi^i\cdot\right)\right\|_{\mathbb{R}^l}^2 &= \left\|\left(\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}h^n\,Y_{n,j}\left(\xi^i\right)\,\tau_n\,Y_{n,j}\left(\eta^k\right)\right)_{k=1,\ldots,l}\right\|_{\mathbb{R}^l}^2\\
&= c^2\left\|\left(\sum_{n=0}^{\infty}h^n\,\sigma^n\,\frac{2n+1}{4\pi}\,P_n\left(\xi^i\cdot\eta^k\right)\right)_{k=1,\ldots,l}\right\|_{\mathbb{R}^l}^2\\
&= c^2\sum_{k=1}^{l}\left(\sum_{n=0}^{\infty}h^n\,\sigma^n\,\frac{2n+1}{4\pi}\,P_n\left(\xi^i\cdot\eta^k\right)\right)^2\\
&= c^2\sum_{k=1}^{l}K_{h\sigma}\left(\xi^i\cdot\eta^k\right)^2\\
&\geq c^2\,l\,K_{h\sigma}(-1)^2\\
&> 0
\end{aligned}$$

for all i such that $\inf_{i\in\mathbb{N}_0}\|\mathcal{F}K_h(\xi^i\cdot)\|_{\mathbb{R}^l} > 0$. Moreover, the fact that $P_n(1) = 1$ for all $n \in \mathbb{N}_0$ implies

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

$$\begin{aligned}\left\| K_h\left(\xi^i\cdot\right)\right\|^2_{\mathrm{L}^2(\Omega)} &= \left\langle K_h\left(\xi^i\cdot\right), K_h\left(\xi^i\cdot\right)\right\rangle_{\mathrm{L}^2(\Omega)}\\ &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} h^{2n}\left(Y_{n,j}\left(\xi^i\right)\right)^2\\ &= \sum_{n=0}^{\infty}\frac{2n+1}{4\pi}h^{2n}\,P_n\left(\xi^i\cdot\xi^i\right)\\ &= K_{h^2}(1)\end{aligned}$$

such that $\sup_{i\in\mathbb{N}_0}\|K_h(\xi^i\cdot)\|^2_{\mathrm{L}^2(\Omega)} = K_{h^2}(1) < +\infty$. Hence, the Abel-Poisson kernel guarantees the existence of C_1 and C_2 in the unregularized and the regularized case. It remains, however, an open problem to verify the semi-frame condition for this particular dictionary (note that, in Freeden and Schreiner (1995), it is shown that the Abel-Poisson kernel constitutes a frame on the sphere; however, we use the word "frame" here in a different context). At least, it is well-known that the choice of a dense point grid $\{\xi^i\}_{i\in\mathbb{N}_0}$ in Ω yields a basis in $\mathrm{L}^2(\Omega)$ (see Freeden et al. 1998, Corollary 6.4.3), i.e.,

$$\overline{\operatorname{span}\left\{K_h\left(\xi^i\cdot\right)\right\}_{i\in\mathbb{N}_0}}^{\|\cdot\|_{\mathrm{L}^2(\Omega)}} = \mathrm{L}^2(\Omega),$$

which is required by condition (a) of Theorem 2.

Similar considerations are possible for the case where the ball $\mathcal{B}$ is the domain of the unknown function. For example, the inverse gravimetric problem consists of the inversion of the gravitational potential for the mass density distribution inside the Earth and at its surface (see the survey article (Michel and Fokas 2008) for further details). In this case, we obtain a linear and continuous operator $\mathcal{F}:\mathrm{L}^2(\mathcal{B})\to\mathbb{R}^l$ which satisfies

$$\mathcal{F}G^{\mathrm{I}}_{m,n,j} = \left(\delta_{m0}\tilde{\tau}_n Y_{n,j}\left(\frac{x^k}{|x^k|}\right)\right)_{k=1,\ldots,l},$$

where $\{x^k\}_{k=1,\ldots,l}$ is a grid of pairwise distinct points at the Earth's surface or outside the Earth where the gravitational potential is given and δ_{m0} is the Kronecker delta. In the case of data given at the surface, $(\tilde{\tau}_n)_n$ converges to 0 at the order $\mathcal{O}(n^{-3/2})$. If the data are given outside the Earth, then $(\tilde{\tau}_n)_n$ exponentially converges to 0, since the downward continuation problem is involved.

7 Numerical Results

We demonstrate here some numerical results which we obtained for the inverse gravimetric problem and which were previously published. The purpose of this section is to show the applicability of the RFMP and its particular features which are visible in the discussion of the numerical results. For each example, we refer to the corresponding papers for further details on the implementation. We also refer to (Telschow 2014) where the downward continuation problem and the treatment of extremely irregular data grids are discussed.

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

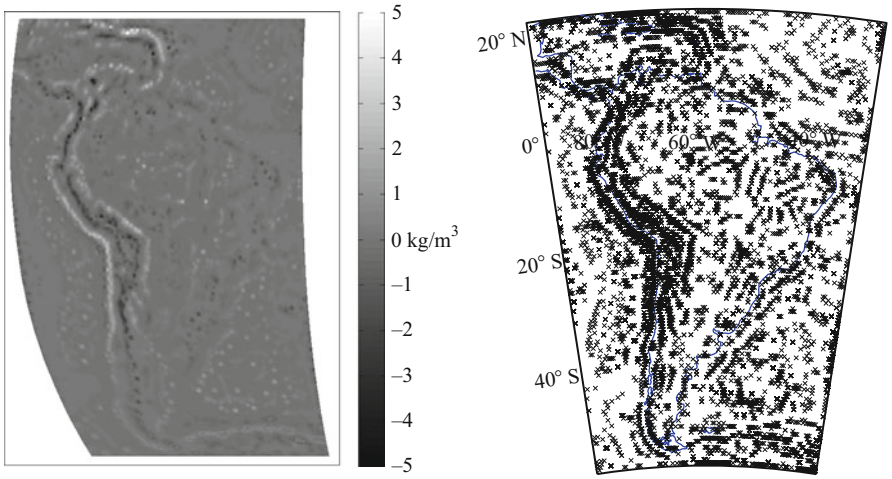

Fig. 2 The RFMP was used to invert EGM2008 data for mass anomalies in South America (*left*). The algorithm prefers localized trial functions which are concentrated to areas with a high detail density as the plot of the chosen centers of the hat functions (*right*) shows; from Fischer and Michel (2012)

In all cases, the unknown function (mass density anomalies) is a function on the ball $\mathcal{B}$. For this purpose, we used dictionaries with combined global trial functions (the orthogonal polynomials $G^{\mathrm{I}}_{m,n,j}$) for low degrees and localized trial functions of different types (i.e., hat functions with different hat widths). It is a particular feature of the RFMP that such a heterogeneous mixture of trial functions can be used for the calculation of the solution. The algorithm selects those trial functions which are (in some sense) the best choice to represent the result (see Algorithm 1). The obtained solution is sparse in the sense that essentially less trial functions than available are eventually used. We do not use any a-priori information on the solution. In particular, the grid of centers of the hat functions in the dictionary is uniformly distributed over the investigated region.

In the first two examples, we consider the (static) gravitational potential model EGM2008 (Earth Gravitational Model 2008, see Pavlis et al. 2008). In both cases, the model is evaluated (starting from polynomial degree 3) on a regular regional point grid, slightly above the surface. The mass density anomaly is computed with the RFMP and plotted at the surface of the Earth in the corresponding region. In the first case (from Fischer and Michel 2012), data at 25,440 points over South America are considered, the dictionary contains approximately 120,000 trial functions, and the RFMP is stopped after 20,000 iterations. The obtained result $F_{20,000}$ is shown on the left-hand side of Fig. 2, where typical mass anomalies like the Andes, the Caribbean, or traces of some tectonic structures can be identified. The right-hand side of the figure shows the centers of those localized trial functions which were chosen by the RFMP. Clearly, the algorithm prefers hat functions which are concentrated to those areas where the solution has a complicated detail structure. This is a reasonable choice and corresponds to the intended sparsity effect since many hat functions are rejected in areas where the solution has a low detail density.

The second example (see Fischer and Michel 2013a) has a very similar constellation. This time, we use the same amount of data but over the Himalayas and India and a slightly larger dictionary (degrees $3, \ldots, 50$ for the polynomials instead of $3, \ldots, 8$) and stop the RFMP again after 20,000 iterations. However, in this example we compare the result with the result obtained for data with artificial noise. Figure 3 shows the result, where the first line corresponds to the noise-free case with $\lambda = 500$ and the second line corresponds to the case where we added 5 % uniformly distributed random noise (relative to the data, i.e. the perturbed data are $\hat{y}_i = y_i + 0.05\rho_i y_i$ for some random

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

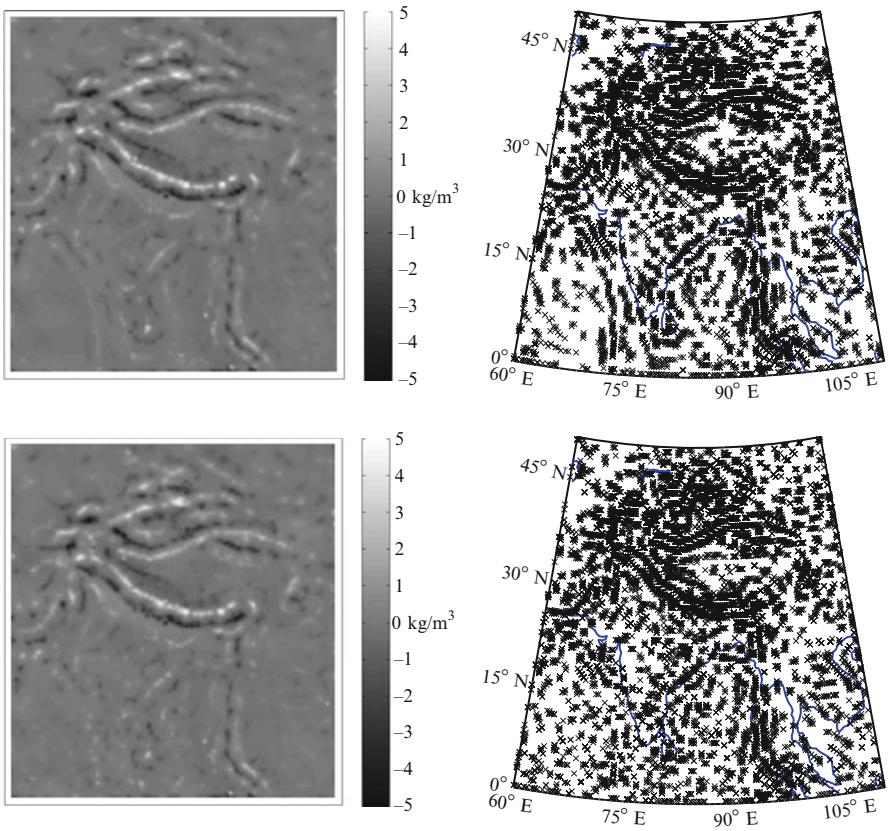

Fig. 3 A test similar to Fig. 2 was applied to the Himalayas and India. The second row shows the result obtained when 5 % uniformly distributed noise relative to the data was added to the data used for the first row. Obviously, the RFMP is, indeed, a regularization, i.e., the result of a data inversion is stable, also if the underlying inverse problem is ill-posed; from Fischer and Michel (2013a)

values $\rho_i \in [-1, 1]$) and used the regularization parameter $\lambda = 600$. Again, the left column shows the approximate solution $F_{20,000}$ and the right column shows the centers of the localized trial functions chosen by the RFMP. These results show the same quality of the algorithm that the application to the data over South America revealed. The solution allows the identification of typical mass anomalies in the corresponding region, and the trial functions are chosen primarily in those areas where the solution has a high detail density. Moreover, the result of this second example shows that the method is, indeed, a regularization, i.e., the solution shows only a low sensitivity with respect to noise (see also Fig. 4, where the absolute difference of the two solutions is shown), though the original problem is ill-posed. Note also that the choice of the centers of the localized trial functions is stable as well, if noise is added.

In the third example (see Fischer and Michel 2012), we show the applicability of the RFMP to the identification of mass transports in data of the GRACE mission (Gravity Recovery and Climate Experiment, see WWW: CSR). The GRACE mission has provided us with monthly models of the Earth's gravitational field since its launch in 2002. We computed the mean of all monthly potentials from July 2004 to June 2009 provided in Release 04, Level 2 of the Jet Propulsion Laboratory (WWW: JPL). This mean potential was subtracted from each monthly potential of the year 2008. Such differences of GRACE models are usually contaminated with noise, which becomes apparent in North-South oriented stripes in the results. We used the Freeden wavelets

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

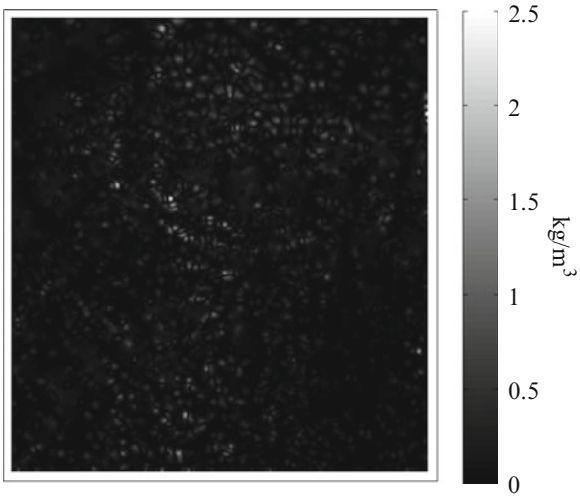

Fig. 4 The absolute difference of the results obtained in the *left column* of Fig. 3 for the inversion of noise-free and noisy data confirms the stability of the regularization algorithm RFMP; from Fischer and Michel (2013a)

of cp-type (see Schreiner 1996; Freeden et al. 1998, pp. 295–296) to denoise the obtained model differences. The Freeden wavelets are general isotropic bandpass filters for spherical signals and are sufficient for our purposes. Note that the result could be further improved by using a more sophisticated filtering technique which has been particularly tailored for the GRACE stripes such as the method introduced in Kusche (2007).

The denoised monthly differences were inverted with the RFMP for mass anomalies. The used dictionary again was a combination of low degree orthogonal polynomials and localized trial functions with different hat widths, in analogy to the previous two examples. The data were given on a regular grid above South America with 11,990 points, the same regularization parameter $\lambda = 8.7128$ was used for all months, and the RFMP was stopped after 10,000 iterations. The results are shown in Figs. 5 and 6. We can clearly identify the mass transports in the Amazon area which are caused by the different rain seasons in the northern and the southern part.

We recommend the cited references Fischer and Michel (2012) and Fischer and Michel (2013a) for further reading, since the numerical results are analyzed in more detail there, and additional results are shown. In particular, the influence of the localized trial functions is visualized, where it becomes clear that local perturbations of the data only have a local influence on the model. Moreover, another example of an identification of mass transports by means of the RFMP is shown in Fischer and Michel (2013b), where some droughts and a flood in South America are recovered from GRACE data.

8 Conclusions

The RFMP was presented as a best basis algorithm which was able to combine trial functions of different kinds (like spherical harmonics and spline/wavelet basis functions) to construct an approximate solution to an (geoscientific) ill-posed inverse problem. This is done by iteratively selecting an additional trial function from a large toolbox of available trial functions (called the dictionary) and adding the chosen trial function to the approximation obtained in the previous iteration step. The criterion for the selection of this next summand in the expansion of the

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

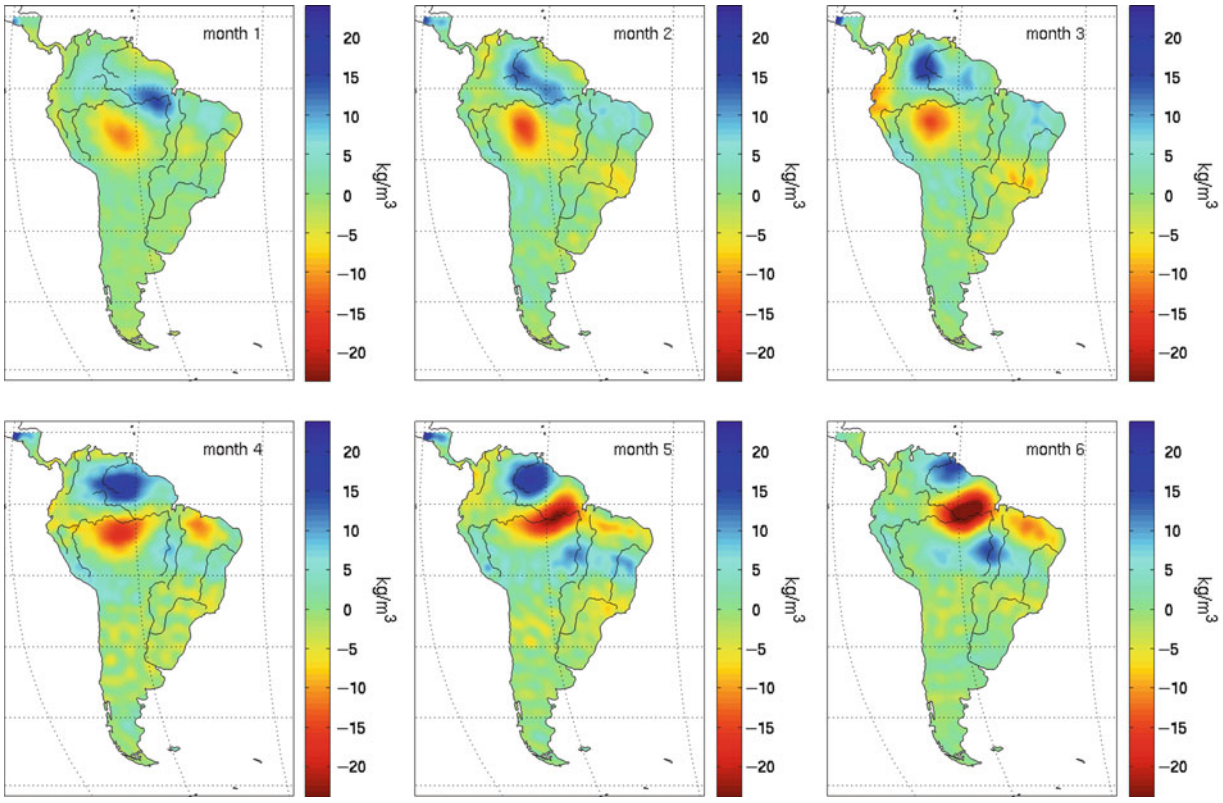

Fig. 5 A long-term mean was subtracted from each monthly model of the gravitational field provided by the GRACE mission for the year 2008. These differences were inverted for mass anomalies with the RFMP. The results are shown for January 2008 (*top left*) via March 2008 (*top right*) until June 2008 (*bottom right*). Typical mass transports due to different rain seasons in the northern part and the southern part of the Amazon are visible in the results. Note, in particular, the mass surplus north of the equator in April 2008 (*bottom left*); from Fischer and Michel (2012)

approximate solution is the minimization of the data misfit plus a Tikhonov regularization term. In this sense, the selected system of trial functions is a "best basis" (note that we call it "best" here, although further enhancements of the RFMP are possible such as in Telschow (2014), where the number of trial functions to be chosen for a given accuracy level is additionally reduced). Numerical experiments show that the algorithm is able to construct a stable and good approximation to the unknown function. Several application scenarios were used for the practical verification of the method. Further numerical experiments performed in our group also showed that often the algorithm starts by predominantly choosing orthogonal polynomials (i.e., global trial functions) in the first iteration steps. Later in the iteration, it chooses localized trial functions more and more often. In other words: At the beginning, a coarse "image" of the main features of the solution is constructed by using global trial functions. Later, the remaining data misfit can be better reduced by locally correcting the solution, which can be achieved best, by taking localized trial functions. Furthermore, in some experiments, we could also show that, with an increasing number of iterations, functions which are more and more localized (i.e., smaller and smaller hats) are selected from the available set of localized trial functions. Thus, the algorithm, indeed, uses the provided possibility to use spline/wavelet basis functions with different levels of localizations. Moreover, some figures in this paper also illustrate the choice of the centers of the hat functions. It is obvious that the algorithm selects much more hat functions with centers in areas where

Handbook of Geomathematics
DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

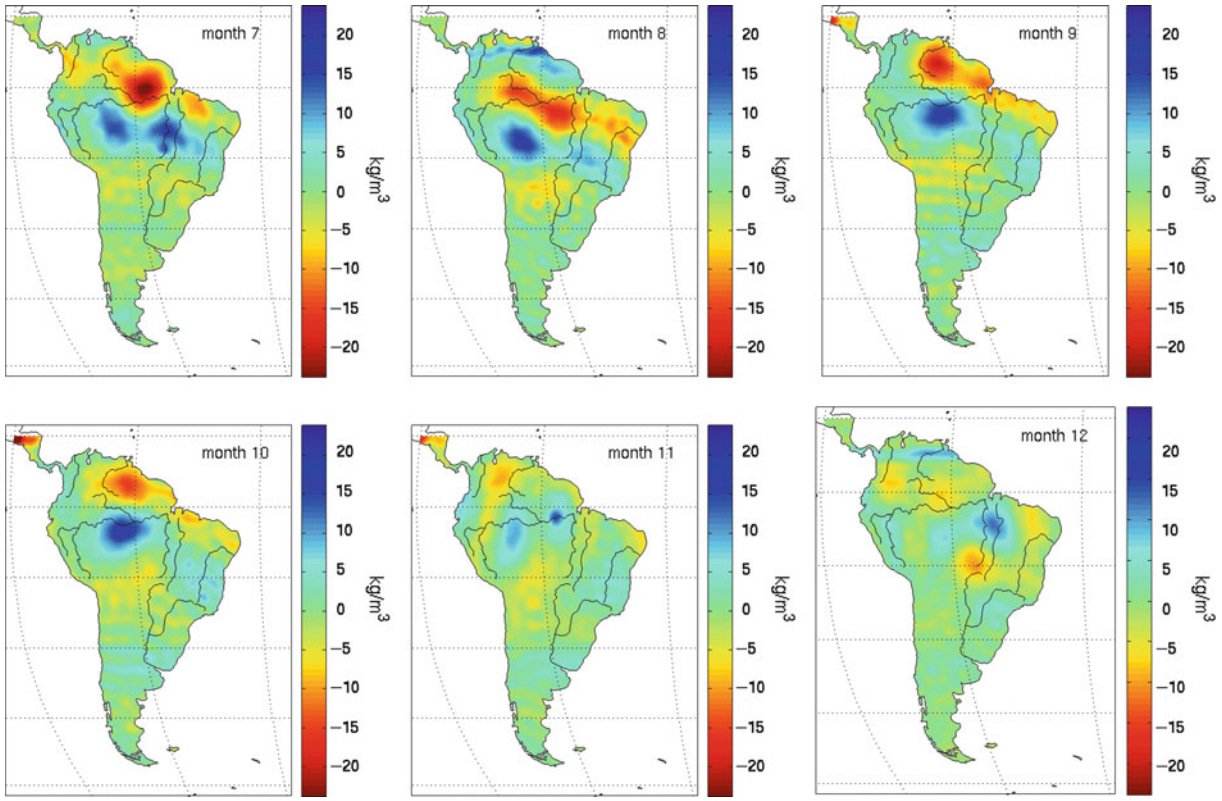

Fig. 6 Here, the remaining months corresponding to Fig. 5 are shown from July 2008 (*top left*) via September 2008 (*top right*) until December 2008 (*bottom right*). Note, in particular, the mass surplus south of the equator in September and October 2008; from Fischer and Michel (2012)

the solution has a high detail density. This appears to be reasonable, since locally complicated structures have higher degrees of freedom and, consequently, require more basis functions for their representation. Furthermore, we also demonstrated that the algorithm is well appropriate for the handling of ill-posed problems, also in the presence of noise.

We consider the RFMP to be a promising approach. It is worth investigating its properties further in the future and to look for additional applications. Several improvements of the algorithm are also possible. For example, different norms than the L^2-norm could be tested for the regularizing penalty term, where Sobolev norms are one possible alternative. Note that the choice of the L^1-norm or anything similar is probably not reasonable, as first investigations in our group suggest, though such a norm is commonly used for sparsity methods. The reason why this is probably not appropriate is the fact that the purpose of the penalty term in our approach is not to obtain sparsity in the solution but to regularize an ill-posed problem.

Another aspect that could be improved refers to the algorithm itself. At present, we keep all chosen dictionary elements and their corresponding coefficients fixed when we go to the next iteration step. The minimization could, however, be improved by reconsidering these previous choices in the next step. The ideal case would be the calculation of a best n-term approximation (see Temlyakov 2003) in the sense that we have to find, in iteration step n, coefficients $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and dictionary elements $d_1, \ldots, d_n \in \mathcal{D}$ such that

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

$$\left\| y - \mathcal{F}\left(\sum_{k=1}^{n} \alpha_k d_k\right)\right\|_{\mathbb{R}^l} = \inf_{\substack{c_1,\ldots,c_n \in \mathbb{R} \\ g_1,\ldots,g_n \in \mathcal{D}}} \left\| y - \mathcal{F}\left(\sum_{k=1}^{n} c_k g_k\right)\right\|_{\mathbb{R}^l}$$

or (taking into account that we want to add a regularization)

$$\begin{aligned} &\left\| y - \mathcal{F}\left(\sum_{k=1}^{n} \alpha_k d_k\right)\right\|_{\mathbb{R}^l}^2 + \lambda \left\| \sum_{k=1}^{n} \alpha_k d_k \right\|_{\mathrm{L}^2(D)}^2 \\ &= \inf_{\substack{c_1,\ldots,c_n \in \mathbb{R} \\ g_1,\ldots,g_n \in \mathcal{D}}} \left(\left\| y - \mathcal{F}\left(\sum_{k=1}^{n} c_k g_k\right)\right\|_{\mathbb{R}^l}^2 + \lambda \left\| \sum_{k=1}^{n} c_k g_k \right\|_{\mathrm{L}^2(D)}^2 \right). \end{aligned}$$

This is, however, very time-consuming such that appropriate trade-offs between the required CPU time on the one hand and the obtained sparsity and accuracy of the solution on the other hand have to be found. In Vincent and Bengio (2002), improvements of the MP in this regard are investigated. However, such techniques have currently not been published for a regularized version which handles ill-posed inverse problems.

These mentioned challenges for future research have partially already been tackled in our group and will be addressed in forthcoming publications.

Acknowledgments The support by the German Research Foundation (project number DFG MI 655/7-1) is gratefully acknowledged. Moreover, the author wishes to thank Roger Telschow for proofreading the manuscript.

References

Amirbekyan A, Michel V (2008) Splines on the three-dimensional ball and their application to seismic body wave tomography. Inverse Probl 24:015022 (25 pp)

Antoine JP, Vandergheynst P (1999) Wavelets on the 2-sphere: a group-theoretic approach. Appl Comput Harmon Anal 7:1–30

Ballani L, Engels J, Grafarend EW (1993) Global base functions for the mass density in the interior of a massive body (Earth). Manuscr Geod 18:99–114

Beckmann J, Mhaskar HN, Prestin J (2012) Quadrature formulas for integration of multivariate trigonometric polynomials on spherical triangles. Int J Geomath 3:119–138

Berkel P (2009) Multiscale methods for the combined inversion of normal mode and gravity variations. PhD thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern, Shaker, Aachen

Berkel P, Michel V (2010) On mathematical aspects of a combined inversion of gravity and normal mode variations by a spline method. Math Geosci 42:795–816

Berkel P, Fischer D, Michel V (2011) Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation. Int J Geomath 1:167–204

Dahlen FA, Simons FJ (2008) Spectral estimation on a sphere in geophysics and cosmology. Geophys J Int 174:774–807

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

Dufour HM (1977) Fonctions orthogonales dans la sphère — résolution théorique du problème du potentiel terrestre. Bull Geod 51:227–237

Engl HW, Kunisch K, Neubauer A (1989) Convergence rates for Tikhonov regularization of nonlinear ill-posed problems. Inverse Probl 5:523–540

Engl HW, Hanke M, Neubauer A (1996) Regularization of inverse problems. Mathematics and its applications, vol 375. Kluwer Academic, Dordrecht

Fengler M, Michel D, Michel V (2006) Harmonic spline-wavelets on the 3-dimensional ball and their application to the reconstruction of the earth's density distribution from gravitational data at arbitrarily shaped satellite orbits. ZAMM-Z Angew Math Me 86:856–873

Fischer D (2011) Sparse regularization of a joint inversion of gravitational data and normal mode anomalies. PhD thesis, Geomathematics Group, Department of Mathematics, University of Siegen, Verlag Dr. Hut, Munich

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

Fischer D, Michel V (2013a) Automatic best-basis selection for geophysical tomographic inverse problems. Geophys J Int 193:1291–1299

Fischer D, Michel V (2013b) Inverting GRACE gravity data for local climate effects. J Geod Sci 3:151–162

Freeden W (1981a) On approximation by harmonic splines. Manuscr Geod 6:193–244

Freeden W (1981b) On spherical spline interpolation and approximation. Math Methods Appl Sci 3:551–575

Freeden W, Schreiner M (1995) Non-orthogonal expansions on the sphere. Math Methods Appl Sci 18:83–120

Freeden W, Schreiner M (1998) Orthogonal and non-orthogonal multiresolution analysis, scale discrete and exact fully discrete wavelet transform on the sphere. Constr Approx 14:493–515

Freeden W, Windheuser U (1996) Spherical wavelet transform and its discretization. Adv Comput Math 5:51–94

Freeden W, Gervens T, Schreiner M (1998) Constructive approximation on the sphere – with applications to geomathematics. Oxford University Press, Oxford

Gerhards C (2011) Spherical decompositions in a global and local framework: theory and application to geomagnetic modeling. Int J Geomath 1:205–256

Gräf M, Kunis S, Potts D (2009) On the computation of nonnegative quadrature weights on the sphere. Appl Comput Harmon Anal 27:124–132

Gutting M (2012) Fast multipole accelerated solution of the oblique derivative boundary value problem. Int J Geomath 3:223–252

Heiskanen WA, Moritz H (1981) Physical geodesy. Institute of Physical Geodesy, Technical University Graz/Austria (Reprint)

Holschneider M (1996) Continuous wavelet transforms on the sphere. J Math Phys 37:4156–4165

Keiner J, Kunis S, Potts D (2009) Using NFFT 3 – a software library for various non-equispaced fast Fourier transforms. ACM Trans Math Softw 36:Article 19 (30 pp)

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. J Geod 81:733–749

Louis AK (1989) Inverse und schlecht gestellte Probleme. Teubner, Stuttgart

Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41:3397–3415

Mhaskar HN (2004a) Local quadrature formulas on the sphere. J Complex 20:753–772

DOI 10.1007/978-3-642-27793-1_93-1
© Springer-Verlag Berlin Heidelberg 2014

Mhaskar HN (2004b) Local quadrature formulas on the sphere, II. In: Neamtu M, Saff EB (eds) Advances in constructive approximation. Nashboro Press, Brentwood, pp 333–344

Michel V (2002) Scale continuous, scale discretized and scale discrete harmonic wavelets for the outer and the inner space of a sphere and their application to an inverse problem in geomathematics. Appl Comput Harmon Anal 12:77–99

Michel V (2005a) Wavelets on the 3-dimensional ball. Proc Appl Math Mech 5:775–776

Michel V (2005b) Regularized wavelet-based multiresolution recovery of the harmonic mass density distribution from data of the earth's gravitational field at satellite height. Inverse Probl 21:997–1025

Michel V (2013) Lectures on constructive approximation – Fourier, spline, and wavelet methods on the real line, the sphere, and the ball. Birkhäuser, Boston

Michel V, Fokas AS (2008) A unified approach to various techniques for the non-uniqueness of the inverse gravimetric problem and wavelet-based methods. Inverse Probl 24:045019 (25 pp)

Michel V, Telschow R (2014) A non-linear approximation method on the sphere. Int J Geomath, accepted for publication

Müller C (1966) Spherical harmonics. Springer, Berlin

Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2008) An Earth gravitational model to degree 2160: EGM2008. General Assembly of the European Geosciences Union, Vienna

Rieder A (2003) Keine Probleme mit Inversen Problemen. Vieweg, Braunschweig

Schreiner M (1996) A pyramid scheme for spherical wavelets. AGTM report 170, Geomathematics Group, Kaiserslautern

Schröder P, Sweldens W (1995) Spherical wavelets: efficiently representing functions on the sphere. In: Proceedings of the 22nd annual conference on computer graphics and interactive techniques, Los Angeles. ACM, New York, pp 161–172

Seidman TI, Vogel CR (1989) Well posedness and convergence of some regularisation methods for non-linear ill posed problems. Inverse Probl 5:227–238

Simons FJ, Dahlen FA (2006) Spherical Slepian functions and the polar gap in geodesy. Geophys J Int 166:1039–1061

Simons FJ, Dahlen FA, Wieczorek MA (2006) Spatiospectral concentration on a sphere. SIAM Rev 48:504–536

Telschow R (2014) An Orthogonal Matching Pursuit for the Regularization of Spherical Inverse Problems, PhD thesis, Geomathematics Group, Department of Mathematics, University of Siegen

Temlyakov VN (2003) Nonlinear methods of approximation. Found Comput Math 3:33–107

Tscherning CC (1996) Isotropic reproducing kernels for the inner of a sphere or spherical shell and their use as density covariance functions. Math Geol 28:161–168

Vincent P, Bengio Y (2002) Kernel matching pursuit. Mach Learn 48:169–191

Wieczorek MA, Simons FJ (2005) Localized spectral analysis on the sphere. Geophys J Int 162:655–675

Wieczorek MA, Simons FJ (2007) Minimum-variance spectral analysis on the sphere. J Fourier Anal Appl 13:665–692

WWW Center for Space Research, University of Texas, Austin. http://www.csr.utexas.edu/grace/overview.html. Last accessed: 30 July 2013

WWW Jet Propulsion Laboratory, California Institute of Technology, Pasadena. http://podaac.jpl.nasa.gov/GRACE. Last accessed: 30 July 2013