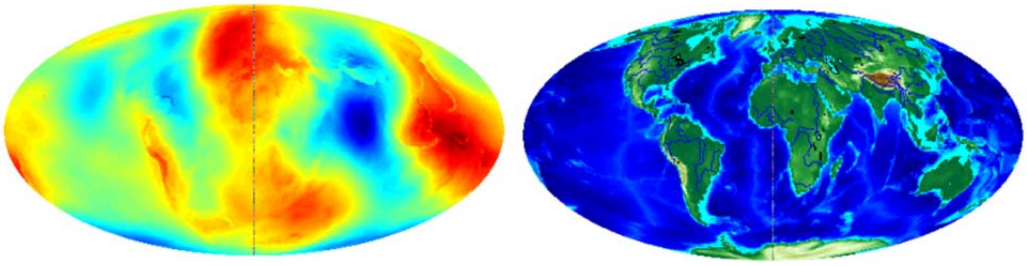

Siegen Preprints on Geomathematics

A Non-linear Approximation Method on the Sphere

V. Michel and R. Telschow

Geomathematics Group
Department of Mathematics
University of Siegen, Germany

www.geomathematics-siegen.de

10

UNIVERSITÄT SIEGEN

A Non-linear Approximation Method on the Sphere

Volker Michel* Roger Telschow†

Abstract

We show the applicability of a modified version of the recently developed Regularized Functional Matching Pursuit (RFMP) to the approximation of functions on the sphere from grid-based data. We elaborate the mathematical details of the choice of trial functions and the specifics of the algorithm. Moreover, we show numerical examples for some benchmark functions. The dictionary of trial functions contains orthogonal polynomials (spherical harmonics) as well as spherical scaling functions and wavelets. It turns out that the greedy algorithm RFMP yields sparse approximations by combining different types of trial functions in a (particular) optimal way, where the sparsity can essentially be increased by a-priori choosing the dictionary appropriately. Moreover, the result of the RFMP can be used for a multiresolution analysis of the investigated function.

Key Words: Greedy algorithm, matching pursuit, multiresolution analysis, non-linear approximation, regularization, reproducing kernel, scaling function, scattered data, sphere, spherical harmonic, wavelet.

MSC2010: primary: 65D15; secondary: 33C50, 46E22, 65D07, 86-08.

dedicated to Willi Freeden's 65th Birthday

*Geomathematics Group, Department of Mathematics, University of Siegen, Emmy-Noether-Campus, Walter-Flex-Str. 3, 57068 Siegen, Germany; Email: michel@mathematik.uni-siegen.de

†Geomathematics Group, Department of Mathematics, University of Siegen, Emmy-Noether-Campus, Walter-Flex-Str. 3, 57068 Siegen, Germany; Email: telschow@mathematik.uni-siegen.de

1 Introduction

The approximation of functions on the sphere, in particular the 2-sphere, plays an important role in many applications, with the geosciences leading the way. Many mathematical techniques have been developed to find 'good' approximations to unknown functions based on given samples. Several new challenges have occurred in the last decades due to the coming of novel technologies like airborne or satellite-based measurements. This implied the need for tools which include two important features: first, high-resolution models should be obtained from huge amounts of data. Second, heterogeneities in the data sets have to be handled, i.e. data can be locally perturbed or data grids could be extremely scattered on the sphere.
The use of a polynomial approximation, which is in the case of the sphere an expansion in spherical harmonics, was considered as a suitable tool in numerical analysis and in the geosciences, as long as very small and regular data sets had to be handled. It is well-known that modern numerical requirements can no longer be appropriately fulfilled by using a basis which merely consists of global trial functions such as orthogonal polynomials. It is an essential progress due to scientists like Willi Freeden that the benefits from the use of localized trial functions have been explored for more than the last three decades. Some of the first works in this respect are the papers [13, 14], where a spherical spline approximation method was developed by Willi Freeden. One major advantage of this spline method was the fact that hat-like basis functions were used, i.e. basis functions which were concentrated around a certain centre. These centres coincided with the data points. As a consequence, the resolution of the resulting approximant was locally adapted to the (maybe spatially heterogeneous) density of the data grid. Moreover, in the case of an exact interpolation, the interpolant minimizes a kind of spherical bending energy, which is why this interpolant is called a (spherical) spline due to the obvious analogy to the predominant property of the natural cubic spline.
In the 1990s, wavelet methods became very popular, also on the sphere. One of the major novelties of this type of approach was the ability to produce a multiresolution analysis of an unknown function, i.e. one can decompose the function into components corresponding to different spatial scales. Moreover, the analysis of the differences between consecutive scales, the so-called scale steps, yielded hidden aspects of particular details in the investigated function. Nowadays, many wavelet methods on the sphere are known, including methods developed by Willi Freeden and his Geomathematics Group (see e.g. [24] and later publications such as [16, 20, 22, 21]). For an overview of the wide range of numerical methods based on localized trial functions which were developed by Willi Freeden and his group for approximation as well as inverse problems, see the books [15, 17, 18, 19, 23].

Of course, there also exist many other approximation and interpolation methods on the sphere. As an incomplete list of examples (without order of priority), we would like to mention the wavelet methods by Schröder and Sweldens (see [43]), by Antoine, Vandergheynst et al. (see [1, 2]) and by Holschneider et al. (see [6, 26, 27, 28]). This list should also include other methods, which are not necessarily wavelet methods, where a stronger focus was on the localization of the trial functions. Examples are the optimally localized wavelets by Laín Fernández and Prestin in [29, 30], the optimally localized approximate identities by one of the authors in [34] and the spherical Slepian functions by Simons et al. in [7, 44, 45, 46, 49, 50] (note also the vec-

torial version by Plattner and Simons in [39, 40]), where the localization is particularly tailored with respect to a predefined region.

In this paper, we will show that, as a continuation in the line of Freeden's research, the use of localized trial functions can be further improved. We demonstrate a non-linear approximation algorithm which iteratively constructs an approximation by choosing trial functions out of a large and redundant set, which is called a dictionary (motivated by a concept known for Euclidean domains). This dictionary contains, for example, a mixture of global trial functions (such as spherical harmonics) and localized trial functions (such as spline basis functions, scaling functions and wavelets). In our numerical experiments, we show, in particular, that we are able to combine the features of Freeden's splines and wavelets. From the splines, we preserve the numerical stability even in the case of irregular data grids and the ability to locally adapt the resolution of results. From the wavelets, we preserve the technique of a multiresolution analysis, i.e. we are able to look at the approximation at different scales. In other words, we can separate features of different spatial size or resolution. Moreover, most of our experiments reveal that the algorithm is able to achieve the same approximation accuracy like a spline but with essentially less trial functions, i.e. we additionally gain a sparsity feature.
In Section 2, we introduce some basic notations. In Section 3, the algorithm, which is based on the Regularized Functional Matching Pursuit (RFMP) in [3, 9, 10] and which we will call the 'RFMP for the Approximation on the Sphere' (briefly RFMP_AoS), is derived and explained. Since it turns out that the theoretical foundations of the algorithm require the use of a reproducing kernel Hilbert space, the theory of such spaces is explained in Section 4, where we also extend an existing concept for the construction of such spaces from the case of a 2-sphere to the general case of a q-sphere. In Section 5, we discuss possible choices of trial functions for the dictionary which is used by the algorithm. In Section 6, we summarize the theoretical results about the RFMP_AoS based on known properties of the RFMP from [36]. After this, we show the results of some numerical experiments with respect to known benchmark functions and the example of gravity field modelling in Section 7. Finally, in Section 8, we summarize the obtained results and discuss the pros and cons of the new method.

2 Preliminaries

The set of positive integers is denoted by $\mathbb{N}$ such that $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Furthermore, we denote the q-dimensional unit sphere in the $(q+1)$-dimensional Euclidean space $\mathbb{R}^{q+1}$ by $\mathbb{S}^q$. The usual dot product in $\mathbb{R}^n$ will be denoted by $\langle x, y\rangle_{\mathbb{R}^n} := \sum_{j=1}^n x_j y_j$ with the corresponding norm $\|x\|_{\mathbb{R}^n} := \sqrt{\langle x, x\rangle_{\mathbb{R}^n}}$. Alternatively, in particular for $n = q+1$, we will also write $x \cdot y := \langle x, y\rangle_{\mathbb{R}^n}$ and $|x| := \|x\|_{\mathbb{R}^n}$. Moreover, for a measurable subset $D \subset \mathbb{R}^n$, the Hilbert space $\mathrm{L}^2(D)$ is the well-known space of (equivalence classes of almost everywhere identical) square-integrable functions $F : D \to \mathbb{R}$.

3 The Algorithm

In the following, we will briefly summarize the RFMP, which was introduced in [3, 9, 10]. The RFMP has been designed for and applied to several inverse problems in geodesy and geophysics, see the latter publications and [11, 12]. For a survey on the current state-of-the-art of the RFMP, including the proofs of the theorems mentioned in Section 6, see [36]. We will see here that the algorithm also performs well for the approximation of functions on the sphere from grid-based data, if we slightly modify it.

The problem to be solved by the RFMP is as follows: We are given a continuous and linear operator $\mathcal{F} : \mathrm{L}^2(D) \to \mathbb{R}^l$, a set of trial functions $\mathcal{D} \subset \mathrm{L}^2(D)$ (which we call the dictionary and which should satisfy some reasonable conditions like $\overline{\operatorname{span} \mathcal{D}}^{\|\cdot\|_{\mathrm{L}^2(D)}} = \mathrm{L}^2(D)$) and a vector $\mathbf{y} = (y_j)_{j=1,\dots,l} \in \mathbb{R}^l$. Find a function $F \in \mathrm{L}^2(D)$ of the form $F = \sum_{k=1}^{\infty} \alpha_k d_k$ (with $\alpha_k \in \mathbb{R}$, $d_k \in \mathcal{D}$ for all $k \in \mathbb{N}$) such that $\mathcal{F}F = \mathbf{y}$. Note that we use bold letters for vectors in $\mathbb{R}^l$.

This is achieved by starting with $F_0 := 0$ and the corresponding residual $\mathbf{R}^0 := \mathbf{y} - \mathcal{F}F_0 = \mathbf{y}$ and choosing the summands $\alpha_n d_n$ iteratively in the following manner: if $F_n = \sum_{k=1}^{n} \alpha_k d_k$ and $\mathbf{R}^n$ are given, then $\alpha_{n+1} \in \mathbb{R}$ and $d_{n+1} \in \mathcal{D}$ are selected such that

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

is minimized by $\alpha = \alpha_{n+1}$ and $d = d_{n+1}$, where $\lambda \in \mathbb{R}_0^+$ is a regularization parameter. In other words, the data misfit is iteratively minimized, where the minimization has to be understood per iteration step. To handle ill-posed problems and numerical instabilities, a Tikhonov-type regularization term is added to the data misfit. Note that the objective of this term is not to obtain sparsity (though the algorithm yields some sense of sparsity in the end) but to stabilize the approximation process.

It is easy to verify that the dictionary element d_{n+1} and its corresponding coefficient α_{n+1} can be obtained as follows:

$$\begin{aligned} d_{n+1} &:= \arg\max_{d \in \mathcal{D}} \frac{\left(\langle \mathbf{R}^n, \mathcal{F}d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(D)} \right)^2}{\|\mathcal{F}d\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathrm{L}^2(D)}^2}, \\ \alpha_{n+1} &:= \frac{\langle \mathbf{R}^n, \mathcal{F}d_{n+1} \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(D)}}{\|\mathcal{F}d_{n+1}\|_{\mathbb{R}^l}^2 + \lambda \|d_{n+1}\|_{\mathrm{L}^2(D)}^2}. \end{aligned}$$

Within this paper, we investigate the following particular case: We are given a vector $\mathbf{y} = (y_j)_{j=1,\dots,l} \in \mathbb{R}^l$ and a point grid $\{\eta^j\}_{j=1,\dots,l} \subset \mathbb{S}^q$, find a function $F \in \mathrm{L}^2(\mathbb{S}^q)$ of the form $F = \sum_{k=1}^{\infty} \alpha_k d_k$ such that $F(\eta^j) \approx y^j$ for all $j = 1, \dots, l$, where $\alpha_k \in \mathbb{R}$ and $d_k \in \mathcal{D} \subset \mathrm{L}^2(\mathbb{S}^q)$ for all $k \in \mathbb{N}$. We skip here the requirement of exactness, i.e. we choose an approximation over an interpolation, because typical data sets in the geosciences are contaminated with noise and have a size which makes a stabilization of the numerical calculations more important than a completely accurate interpolation. This aspect is reflected by the regularization term and the choice of the parameter $\lambda > 0$. Note that the reduction of the RFMP to a pure interpolation problem without a regularization (i.e. $\lambda = 0$) would yield an algorithm which is similar to the Matching Pursuit (MP) introduced in [33, 48]. The RFMP, however, is a generalization of the MP, since it allows to handle also some other kinds of problems (like the inverse problems

mentioned above) and it allows the inclusion of a regularization in the numerical algorithm. Within this paper, we profit from the latter aspect.
Before we proceed, we have to be aware of the following difficulty: In our case, the operator $\mathcal{F}$ would map a function F to the vector of its samples on the given point grid. This operator is, however, not continuous on $\mathrm{L}^2(\mathbb{S}^q)$. For this reason, we have to change the underlying Hilbert space of the algorithm to a space $\mathcal{H}(\mathbb{S}^q) \subsetneqq \mathrm{L}^2(\mathbb{S}^q)$ where the evaluation functional $\mathcal{H}(\mathbb{S}^q) \ni F \mapsto F(\eta)$ is continuous for every fixed $\eta \in \mathbb{S}^q$. According to Aronszajn's Theorem (see, e.g., [8, p. 317]), this is valid if and only if $\mathcal{H}(\mathbb{S}^q)$ is a reproducing kernel Hilbert space. We will elaborate this topic in Section 4 in further detail.
We now formulate the RFMP for our particular kind of problem that we want to solve.

Algorithm 3.1 (RFMP for the Approximation on the Sphere (RFMP_AoS))
Let $\mathbf{y} \in \mathbb{R}^l$ be a given vector and $X := \{\eta^j\}_{j=1,\dots,l} \subset \mathbb{S}^q$ be a given point grid.

1. ***Initialization:*** *Choose a reproducing kernel Hilbert space $(\mathcal{H}(\mathbb{S}^q), \langle\cdot,\cdot\rangle_{\mathcal{H}(\mathbb{S}^q)})$ and a dictionary $\mathcal{D} \subset \mathcal{H}(\mathbb{S}^q)$. Furthermore, choose an initial approximation $F_0 \in \mathcal{H}(\mathbb{S}^q)$ (e.g. $F_0 := 0$) and calculate its samples $\mathbf{F}_0 := (F_0(\eta^j))_{j=1,\dots,l} \in \mathbb{R}^l$. Set $n := 0$ and $\mathbf{R}^0 := \mathbf{y} - \mathbf{F}_0$. Moreover, choose a regularization parameter $\lambda \in \mathbb{R}_0^+$ and a stopping criterion (e.g. require that $\|\mathbf{R}^{n+1}\|_{\mathbb{R}^l} < \varepsilon$ for a given $\varepsilon > 0$ or require $n+1 \le N$ for a given $N \in \mathbb{N}$).*

2. ***Preprocessing:*** *Provided that enough memory space is available, compute and store the samples $\mathbf{d} = (d(\eta^j))_{j=1,\dots,l} \in \mathbb{R}^l$ as well as the norms $\|\mathbf{d}\|_{\mathbb{R}^l}^2$ and $\|d\|_{\mathcal{H}(\mathbb{S}^q)}^2$ for all $d \in \mathcal{D}$.*

3. ***Core part of the algorithm:*** *Determine*

$$d_{n+1} := \operatorname*{arg\,max}_{d\in\mathcal{D}} \frac{\left(\langle \mathbf{R}^n, \mathbf{d}\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d\rangle_{\mathcal{H}(\mathbb{S}^q)}\right)^2}{\|\mathbf{d}\|_{\mathbb{R}^l}^2 + \lambda \|d\|_{\mathcal{H}(\mathbb{S}^q)}^2}, \tag{1}$$

$$\alpha_{n+1} := \frac{\langle \mathbf{R}^n, \mathbf{d}_{n+1}\rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1}\rangle_{\mathcal{H}(\mathbb{S}^q)}}{\|\mathbf{d}_{n+1}\|_{\mathbb{R}^l}^2 + \lambda \|d_{n+1}\|_{\mathcal{H}(\mathbb{S}^q)}^2}, \tag{2}$$

   where $\mathbf{d}_{n+1} := (d_{n+1}(\eta^j))_{j=1,\dots,l} \in \mathbb{R}^l$.

4. ***Update:*** *Set $F_{n+1} := F_n + \alpha_{n+1} d_{n+1}$ and $\mathbf{R}^{n+1} := \mathbf{R}^n - \alpha_{n+1}\mathbf{d}_{n+1}$.*

5. ***Stopping Criterion:*** *If the stopping criterion is fulfilled, return F_{n+1} as the approximating function and $\mathbf{R}^{n+1}$ as the approximation error on the point grid X. Otherwise, increase n by 1 and go to step 3.*

Note that the dictionary certainly has to be finite in practice such that the requirement $\mathcal{H}(\mathbb{S}^q) = \overline{\operatorname{span}\mathcal{D}}^{\|\cdot\|_{\mathcal{H}(\mathbb{S}^q)}}$ plays a role in the theoretical investigations only (unless $\mathcal{H}(\mathbb{S}^q)$ is finite-dimensional as well).
The most expensive task in the core part of the algorithm is the determination of d_{n+1} in (1), because (in each iteration step n) all dictionary elements $d \in \mathcal{D}$ have to be tested here to find the maximizer. For this reason, the algorithm is essentially accelerated, if as much work as

possible is transferred to the preprocessing in step 2. Moreover, one can additionally compute and store the inner products $\langle d, \tilde{d} \rangle_{\mathcal{H}(\mathbb{S}^q)}$ and $\langle \mathbf{d}, \tilde{\mathbf{d}} \rangle_{\mathbb{R}^l}$ for all $d, \tilde{d} \in \mathcal{D}$ in the preprocessing. This also reduces the effort to compute the inner products

$$\begin{aligned} \langle \mathbf{R}^n, \mathbf{d} \rangle_{\mathbb{R}^l} &= \left\langle \mathbf{R}^{n-1}, \mathbf{d} \right\rangle_{\mathbb{R}^l} - \alpha_n \left\langle \mathbf{d}_n, \mathbf{d} \right\rangle_{\mathbb{R}^l} , \\ \langle F_n, d \rangle_{\mathcal{H}(\mathbb{S}^q)} &= \langle F_{n-1}, d \rangle_{\mathcal{H}(\mathbb{S}^q)} + \alpha_n \langle d_n, d \rangle_{\mathcal{H}(\mathbb{S}^q)} \end{aligned}$$

in (1).

4 Reproducing Kernel Hilbert Spaces on the Sphere

If we have a Hilbert space $\mathcal{H}(\mathbb{S}^q)$ of functions on $\mathbb{S}^q$, then a kernel $K : \mathbb{S}^q \times \mathbb{S}^q \to \mathbb{R}$ is called a reproducing kernel, if

$$\begin{aligned} K(\xi, \cdot) &\in \mathcal{H}(\mathbb{S}^q) \text{ for all } \xi \in \mathbb{S}^q , \\ \langle K(\xi, \cdot), F \rangle_{\mathcal{H}(\mathbb{S}^q)} &= F(\xi) \text{ for all } F \in \mathcal{H}(\mathbb{S}^q) \text{ and all } \xi \in \mathbb{S}^q . \end{aligned}$$

For further details on the theory of reproducing kernels, see, for example, [8, Section 12.6]. Aronszajn's Theorem states, as we mentioned above, that a Hilbert space $\mathcal{H}(\mathbb{S}^q)$ of functions on $\mathbb{S}^q$ is a reproducing kernel Hilbert space (RKHS), i.e. it possesses a reproducing kernel, if and only if the evaluation functional $L_\eta : \mathcal{H}(\mathbb{S}^q) \to \mathbb{R}$ given by $L_\eta F := F(\eta)$, $F \in \mathcal{H}(\mathbb{S}^q)$, is continuous for every fixed $\eta \in \mathbb{S}^q$. Furthermore, it is well-known that a reproducing kernel is always unique and symmetric (i.e. $K(\xi, \eta) = K(\eta, \xi)$ for all $\xi, \eta \in \mathbb{S}^q$).
Since linear operators acting between finite-dimensional spaces are always continuous, every finite-dimensional subspace of $\mathrm{L}^2(\mathbb{S}^q)$ is automatically a RKHS. Since all norms are equivalent in a finite-dimensional space, the choice of the norm (respectively the inner product) is arbitrary such that, for example, also the $\mathrm{L}^2(\mathbb{S}^q)$-norm may be chosen in this case. The representation of the reproducing kernel, however, depends on the choice of the inner product.
Certainly, there also exist infinite-dimensional RKHSs on the sphere. These can be constructed in analogy to the Sobolev spaces on the 2-sphere $\mathbb{S}^2$ which are known from the works by W. Freeden, see, for example, [13, 14, 18]. These spaces were constructed by inserting weights into the Parseval identity of the $\mathrm{L}^2(\mathbb{S}^2)$-inner product. Particular cases $\mathcal{H}_s(\mathbb{S}^2)$ were obtained by choosing weights which are related to the eigenvalues of the Laplace-Beltrami operator on $\mathbb{S}^2$. This idea was used to construct Sobolev spaces $\mathcal{H}_s(\mathbb{S}^q)$ in [5] and, in a slightly modified way, in [4, 32]. We will follow here Freeden's approach to construct equally general Sobolev spaces $\mathcal{H}(\mathbb{S}^q)$ like his spaces on $\mathbb{S}^2$.
For this purpose, we use the well-known system of spherical harmonics in $\mathrm{L}^2(\mathbb{S}^q)$, see also [37]. A spherical harmonic of degree n is the restriction of a homogeneous harmonic polynomial on $\mathbb{R}^{q+1}$ to the unit sphere $\mathbb{S}^q$. The space of all such functions is denoted by $\mathrm{Harm}_n(\mathbb{S}^q)$ and its dimension is denoted by $N(q+1, n)$. One can prove that

$$\frac{1+x}{(1-x)^q} = \sum_{n=0}^{\infty} N(q+1, n) x^n$$

for all x in the neighbourhood of 0. For instance, the geometric series yields that

$$\begin{aligned}\frac{1+x}{(1-x)^2} &= (1+x)\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{1-x}\\ &= (1+x)\sum_{n=1}^{\infty} n\,x^{n-1}\\ &= \sum_{n=1}^{\infty} n\,x^{n-1}+\sum_{n=1}^{\infty} n\,x^{n}\\ &= \sum_{n=0}^{\infty}(2n+1)x^{n}\end{aligned}$$

such that $N(3,n) = 2n+1$ for all $n \in \mathbb{N}_0$. We assume that $\{Y_{n,j}\}_{j=1,\dots,N(q+1,n)}$ denotes an orthonormal basis in $\mathrm{Harm}_n(\mathbb{S}^q)$ with respect to $\langle\cdot,\cdot\rangle_{\mathrm{L}^2(\mathbb{S}^q)}$. Two fundamental results are valid for all such systems (see [37]):

(a) The system $\{Y_{n,j}\}_{n\in\mathbb{N}_0;\ j=1,\dots,N(q+1,n)}$ is complete and orthonormal in $\mathrm{L}^2(\mathbb{S}^q)$.

(b) For all $\xi,\eta \in \mathbb{S}^q$, the addition theorem

$$\sum_{j=1}^{N(q+1,n)} Y_{n,j}(\xi)Y_{n,j}(\eta) = \frac{N(q+1,n)}{\omega_q}\,P_n^{(q)}(\xi\cdot\eta)$$

holds, where $\omega_q = \int_{\mathbb{S}^q} 1\,\mathrm{d}\omega$ is the surface area of the unit sphere and the system of orthogonal polynomials $\{P_n^{(q)}\}_{n\in\mathbb{N}_0}$ is uniquely determined by

(i) $P_n^{(q)}$ is a polynomial of degree n,

(ii) $\int_{-1}^{1} P_n^{(q)}(t)\,P_m^{(q)}(t)\,(1-t^2)^{(q-2)/2}\,\mathrm{d}t = 0$, if $n \neq m$,

(iii) $P_n^{(q)}(1) = 1$.

The $P_n^{(q)}$ are particular examples of Jacobi polynomials, more precisely Gegenbauer polynomials (in particular, $q=2$ yields the Legendre polynomials). One of their properties is that $|P_n^{(q)}(t)| \leq 1$ for all $t \in [-1,1]$.
We are now in the position to define the Sobolev spaces.

Definition 4.1 (Sobolev space) *Let $(A_n)_{n\in\mathbb{N}_0}$ be a sequence of real numbers. We consider the subspace $\mathcal{E}((A_n);\mathbb{S}^q)$ of $\mathrm{C}^{(\infty)}(\mathbb{S}^q)$ consisting of all F which satisfy*

$$\sum_{n=0}^{\infty}\sum_{j=1}^{N(q+1,n)} A_n^2\,\langle F, Y_{n,j}\rangle^2_{\mathrm{L}^2(\mathbb{S}^q)} < +\infty$$

and $\langle F, Y_{n,j}\rangle_{\mathrm{L}^2(\mathbb{S}^q)} = 0$ if $A_n = 0$. Then the completion of $\mathcal{E}((A_n);\mathbb{S}^q)$ with respect to the inner product

$$\langle F,G\rangle_{\mathcal{H}} := \sum_{n=0}^{\infty}\sum_{j=1}^{N(q+1,n)} A_n^2\,\langle F, Y_{n,j}\rangle_{\mathrm{L}^2(\mathbb{S}^q)}\,\langle G, Y_{n,j}\rangle_{\mathrm{L}^2(\mathbb{S}^q)}$$

is called the Sobolev space $\mathcal{H}((A_n);\mathbb{S}^q)$. *If no confusion is likely to arise, we will simply write* $\mathcal{H}(\mathbb{S}^q)$. *The induced norm is denoted by* $\|\cdot\|_{\mathcal{H}}$.

Definition 4.2 (Summability) *A real sequence* $(A_n)_{n\in\mathbb{N}_0}$ *is called summable if it satisfies the summability condition*

$$\sum_{\substack{n=0\\A_n\neq 0}}^{\infty} A_n^{-2}\, N(q+1,n) < +\infty\,.$$

Note that the summability condition for $q=2$ is the usual condition

$$\sum_{\substack{n=0\\A_n\neq 0}}^{\infty} A_n^{-2}(2n+1) < +\infty\,.$$

Two fundamental properties of the Sobolev spaces $\mathcal{H}(\mathbb{S}^q)$ are available in the case of summable sequences. They provide us with the foundations for the construction of approximating structures.

Theorem 4.3 *Let the Sobolev space* $\mathcal{H}(\mathbb{S}^q)$ *correspond to a summable real sequence* $(A_n)_{n\in\mathbb{N}_0}$. *Then the following holds true:*

1) Every $F\in\mathcal{H}(\mathbb{S}^q)$ *is continuous and has a uniformly convergent Fourier expansion in the basis* $\{Y_{n,j}\}_{n\in\mathbb{N}_0,\,A_n\neq 0;\,j=1,\ldots,N(q+1,n)}$.

2) $\mathcal{H}(\mathbb{S}^q)$ *has a unique reproducing kernel. This is given by*

$$K_{\mathcal{H}}(\xi,\eta) = \sum_{\substack{n=0\\A_n\neq 0}}^{\infty} A_n^{-2}\,\frac{N(q+1,n)}{\omega_q} P_n^{(q)}(\xi\cdot\eta)\,;\quad \xi,\eta\in\mathbb{S}^q\,.$$

Proof. 1) We use the Cauchy-Schwarz inequality, the addition theorem, the definition of the Sobolev space (see Definition 4.1), the summability condition (see Definition 4.2) and the fact that $P_n^{(q)}(1)=1$ for all $n\in\mathbb{N}_0$ and $q\in\mathbb{N}$ to derive that

$$\begin{aligned}
&\left|\sum_{\substack{n=M\\A_n\neq 0}}^{\infty}\sum_{j=1}^{N(q+1,n)} \langle F, Y_{n,j}\rangle_{\mathrm{L}^2(\mathbb{S}^q)}\, Y_{n,j}(\xi)\right| \\
&\quad\leq \left(\sum_{n=M}^{\infty}\sum_{j=1}^{N(q+1,n)} A_n^2\,\langle F, Y_{n,j}\rangle^2_{\mathrm{L}^2(\mathbb{S}^q)}\right)^{1/2}\left(\sum_{\substack{n=M\\A_n\neq 0}}^{\infty} A_n^{-2}\,\frac{N(q+1,n)}{\omega_q}\,P_n^{(q)}(\xi\cdot\xi)\right)^{1/2} \\
&\quad\longrightarrow\quad 0\quad \text{as } M\to\infty
\end{aligned}$$

for all $F\in\mathcal{H}(\mathbb{S}^q)$.
2) For each $\xi\in\mathbb{S}^q$, we define the linear functional $L_\xi:\mathcal{H}(\mathbb{S}^q)\to\mathbb{R}$ by $L_\xi F := F(\xi)$, $F\in\mathcal{H}(\mathbb{S}^q)$.

Then each L_ξ is bounded due to the summability condition, because part 1 of this proof yields

$$\begin{aligned} |L_\xi F| &= \left| \sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} \sum_{j=1}^{N(q+1,n)} \langle F, Y_{n,j} \rangle_{\mathrm{L}^2(\mathbb{S}^q)} \, Y_{n,j}(\xi) \right| \\ &\leq \|F\|_{\mathcal{H}} \left(\sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} A_n^{-2} \, \frac{N(q+1,n)}{\omega_q} \right)^{1/2} . \end{aligned}$$

Hence, Aronszajn's Theorem (see, for example, [8, p. 317]) implies the existence of the reproducing kernel. Furthermore, due to the construction of the inner product in Definition 4.1, the system $\{A_n^{-1} Y_{n,j}\}_{n \in \mathbb{N}_0, \, A_n \neq 0; \, j=1,\ldots,N(q+1,n)}$ is an orthonormal basis in $\mathcal{H}(\mathbb{S}^q)$. Moreover,

$$\sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} \sum_{j=1}^{N(q+1,n)} \left(A_n^{-1} Y_{n,j}(\xi) \right)^2 = \sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} A_n^{-2} \, \frac{N(q+1,n)}{\omega_q} < +\infty$$

due to the addition theorem, the identity $P_n^{(q)}(1) = 1$ and the summability condition. Hence, Theorem 6.2. in [35, p. 147] tells us that the reproducing kernel of $\mathcal{H}(\mathbb{S}^q)$ is represented by

$$\begin{aligned} K_{\mathcal{H}}(\xi, \eta) &= \sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} \sum_{j=1}^{N(q+1,n)} A_n^{-1} Y_{n,j}(\xi) \, A_n^{-1} Y_{n,j}(\eta) \\ &= \sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} A_n^{-2} \, \frac{N(q+1,n)}{\omega_q} \, P_n^{(q)}(\xi \cdot \eta) \end{aligned}$$

for all $\xi, \eta \in \mathbb{S}^q$. ∎

Note that the reproducing kernel is a radial basis function, i.e. it only depends on the inner product $\xi \cdot \eta$ of its two arguments $\xi, \eta \in \mathbb{S}^q$. Furthermore, due to the behaviour

$$N(q+1,n) = \mathcal{O}\left(n^{q-1}\right)$$

as $n \to \infty$ (see [37, p. 3]), the summability condition is equivalent to the requirement that

$$\sum_{\substack{n=0 \\ A_n \neq 0}}^{\infty} A_n^{-2} \, n^{q-1} < +\infty \, . \tag{3}$$

Example 4.4 *We consider here some examples of reproducing kernels.*

(a) In the case $q = 2$, the Abel-Poisson kernel

$$K_h(\xi \cdot \eta) := K_{\mathcal{H}}(\xi, \eta) = \frac{1}{4\pi} \, \frac{1 - h^2}{(1 + h^2 - 2h\xi \cdot \eta)^{3/2}} \, ; \quad \xi, \eta \in \mathbb{S}^2 \, ;$$

is a popular choice. It corresponds to the sequence $A_n = h^{-n/2}$ for a fixed $h \in]0,1[$. We will use the representation $K_h(\xi \cdot \eta)$ in the section on our numerical experiments (Section 7).

(b) The Laplace operator Δ_{q+1} in $q+1$ dimensions can be decomposed, in terms of $(q+1)$-dimensional polar coordinates, into a radial part and an angular part (see [37, p. 38]) by

$$\Delta_{q+1} = \frac{\partial^2}{\partial r^2} + \frac{q}{r}\,\frac{\partial}{\partial r} + \frac{1}{r^2}\,\Delta^*_{q+1}\,,$$

where the operator $\Delta^*_{q+1} : \mathrm{C}^{(2)}(\mathbb{S}^q) \to \mathrm{C}(\mathbb{S}^q)$ is the $(q+1)$-dimensional Laplace-Beltrami operator. Note that we consider only cases here where $q \geq 2$. From [37, Lemma 22], we know that all $Y_{n,j}$ are eigenfunctions of Δ^*_{q+1} corresponding to the eigenvalue $-n(n+q-1)$. With the calculations

$$n(n+q-1) = n^2 + (q-1)n = \left(n + \frac{q-1}{2}\right)^2 - \left(\frac{q-1}{2}\right)^2$$

in mind, we set $A_n := (n + \frac{q-1}{2})^s$ for all $n \in \mathbb{N}_0$ and a fixed $s \in \mathbb{R}^+$. This sequence satisfies the summability condition (3), i.e.

$$\sum_{n=0}^{\infty} \left(n + \frac{q-1}{2}\right)^{-2s} n^{q-1} < +\infty\,,$$

if and only if $-2s + q - 1 < -1$, i.e. $s > \frac{q}{2}$. Provided that this is satisfied, we have the Sobolev spaces

$$\mathcal{H}_s\left(\mathbb{S}^q\right) := \mathcal{H}\left(\left(\left(n + \frac{q-1}{2}\right)^s\right)_{n\in\mathbb{N}_0}; \mathbb{S}^q\right)$$

with the reproducing kernel

$$K_{\mathcal{H}}(\xi,\eta) = \sum_{n=0}^{\infty} \left(n + \frac{q-1}{2}\right)^{-2s} \frac{N(q+1,n)}{\omega_q}\, P_n^{(q)}(\xi\cdot\eta)\,; \quad \xi,\eta \in \mathbb{S}^q\,.$$

Note that the choice $\frac{s}{2} \in \mathbb{N}$ (with $s > \frac{q}{2}$) yields here, for all $F \in \mathcal{H}_s(\mathbb{S}^q)$,

$$\begin{aligned} \|F\|^2_{\mathcal{H}} &= \sum_{n=0}^{\infty} \sum_{j=1}^{N(q+1,n)} \left(n + \frac{q-1}{2}\right)^{2s} \langle F, Y_{n,j}\rangle^2_{\mathrm{L}^2(\mathbb{S}^q)} \\ &= \sum_{n=0}^{\infty} \sum_{j=1}^{N(q+1,n)} \left[n(n+q-1) + \left(\frac{q-1}{2}\right)^2\right]^s \langle F, Y_{n,j}\rangle^2_{\mathrm{L}^2(\mathbb{S}^q)}\,, \end{aligned}$$

where the particular case $s = 2$ corresponds to (provided that Δ^*_{q+1} can be interchanged with the summation of the Fourier expansion of F)

$$\|F\|^2_{\mathcal{H}} = \int_{\mathbb{S}^q} \left[\left(-\Delta^*_{q+1} + \left(\frac{q-1}{2}\right)^2\right) F(\xi)\right]^2 \mathrm{d}\omega(\xi)\,,$$

which is a useful measure for the non-smoothness of a function on the sphere since it takes a second-order derivative into account as it is common for the well-known cubic splines.

Moreover, it does not cancel out $0th$ *and* $1st$ *degree polynomials since the null space of* $-\Delta^*_{q+1} + (\frac{q-1}{2})^2$ *is trivial.*
These spaces $\mathcal{H}_s(\mathbb{S}^q)$*, which were contructed here, can be associated to known RKHSs on the sphere. More precisely,* $\mathcal{H}_s(\mathbb{S}^2)$ *is identical to the homonymous space in [18] and* $\mathcal{H}_s(\mathbb{S}^q)$ *equals the spaces considered in [5], which can, from this point of view, be regarded as the most intuitive generalization of the spaces* $\mathcal{H}_s(\mathbb{S}^2)$*. These analogies also show that the spaces introduced here, indeed, generalize the existing concepts on the sphere.*

Theorem 4.5 *Let* X *be a countable and dense subset of* $\mathbb{S}^q$*. Moreover, let* $(A_n)_{n\in\mathbb{N}_0}$ *be a sequence of non-vanishing real numbers and let* $K_\mathcal{H}$ *be the reproducing kernel in the corresponding Sobolev space* $\mathcal{H}(\mathbb{S}^q) = \mathcal{H}((A_n); \mathbb{S}^q)$*. Then the linear space* $\operatorname{span}\{K_\mathcal{H}(\xi, \cdot) \,|\, \xi \in X\}$ *is dense in*

(a) $\mathcal{H}(\mathbb{S}^q)$ *with respect to* $\|\cdot\|_{\mathcal{H}(\mathbb{S}^q)}$*,*

(b) $\mathrm{C}(\mathbb{S}^q)$ *with respect to* $\|\cdot\|_{\mathrm{C}(\mathbb{S}^q)}$*,*

(c) $\mathrm{L}^2(\mathbb{S}^q)$*with respect to* $\|\cdot\|_{\mathrm{L}^2(\mathbb{S}^q)}$*.*

The proof is analogous to the known proof for the case $q = 2$ (see, for example, [18, pp. 155-156] or [35, pp. 175-177]).

5 Options to Choose a Dictionary

For our dictionary $\mathcal{D}$, we intend to choose a redundant combination of different types of basis functions. The purpose of this choice is to combine the advantages of different sets of trial functions. For example, strongly localized ansatz functions appear to be more appropriate for the representation of fine, i.e. strongly spatially localized, details, whereas weakly localized trial functions are probably more useful to cover regional to global trends in the signal. Based on the previous considerations, we are able to consider, i.e. combine, the following basis systems:

- *Spherical harmonics:* These trial functions are typical examples of global functions with no space localization but a maximum frequency (i.e. degree) localization. Moreover, they constitute a complete orthonormal system in $\mathrm{L}^2(\mathbb{S}^q)$.

- *Truncated reproducing kernels:* The series expansions of the reproducing kernels above can be truncated. This corresponds to the restriction to a polynomial subspace of $\mathcal{H}(\mathbb{S}^q)$. For the choice of the sequence $(A_n)_{n\in\mathbb{N}_0}$, the generators of bandlimited spherical scaling functions (see [18, pp. 290-296] and the references therein) represent one possibility. In this case, the scale of the scaling function (or the associated wavelet) yields a possibility to vary the localization.

- *Non-bandlimited reproducing kernels:* Reproducing kernels like the Abel-Poisson kernel (see above) usually possess a stronger localization than bandlimited kernels, though their support is still global. In the case of Abel-Poisson kernels, the parameter h provides us with a possibility to control the localization of the kernel.

- *Locally supported kernels:* Locally supported kernels appear to be well appropriate for the approximation of signals which vanish on some regions of the sphere.

Note that, if we have a countable and dense subset X of $\mathbb{S}^q$, then one single reproducing kernel $K_{\mathcal{H}}$ (with $A_n \neq 0$ for all $n \in \mathbb{N}_0$) suffices to construct a basis $K_{\mathcal{H}}(\xi, \cdot)$, $\xi \in X$, of $\mathrm{L}^2(\mathbb{S}^q)$, see Theorem 4.5. Hence, if we use spherical harmonics $\{Y_{n,j}\}_{n\in\mathbb{N}_0;\, j=1,\dots,2n+1}$ and the rotationally symmetric "hat-functions" $\{K_{\mathcal{H}((A_n);\mathbb{S}^q)}(\xi, \cdot)\}_{\xi\in X}$ for different sequences $(A_n)_{n\in\mathbb{N}_0}$ (e.g. for different parameters h in the case of Abel-Poisson kernels), then our dictionary is a redundant union of several basis systems with different localization characteristics. This is what we want to achieve. We will, to keep this paper acceptably brief, only discuss here examples of combinations of the functions of the kind above.

6 Properties of the Algorithm

In the following, we will study some of the properties of our algorithm RFMP_AoS. First of all, we will demonstrate in the following simple example that it is non-linear.

Example 6.1 *We consider here the 2-sphere $\mathbb{S}^2$ and a finite-dimensional space $\mathcal{H}(\mathbb{S}^2)$ which is equipped with the $\mathrm{L}^2(\mathbb{S}^2)$-inner product and is large enough to contain the dictionary chosen below. For our particular problem, we use the simple point grid X consisting of the two points*

$$\eta^1 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \eta^2 := \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

and two different vectors of values on this grid which have to be approximated

$$\mathbf{y} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{z} := \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

i.e. in the first case (data vector $\mathbf{y}$) we seek $F \in \mathrm{L}^2(\mathbb{S}^2)$ such that $F(0,0,1) \approx 1$ and $F(0,0,-1) \approx 0$ and in the second case (data vector $\mathbf{z}$) we seek $F \in \mathrm{L}^2(\mathbb{S}^2)$ such that $F(0,0,1) \approx 0$ and $F(0,0,-1) \approx 1$.
Moreover, our dictionary consists of the three functions $I, N, S \in \mathrm{L}^2(\mathbb{S}^2)$ which are defined by

$$I(\xi) := 1, \quad N(\xi) := \begin{cases} 1, & \xi_3 \geq \frac{1}{2} \\ 0 & else \end{cases}, \qquad S(\xi) := \begin{cases} 1, & \xi_3 \leq -\frac{1}{2} \\ 0 & else \end{cases}$$

for all $\xi \in \mathbb{S}^2$, where ξ_3 is the third cartesian coordinate of ξ. In the preprocessing, we obtain the samples

$$\mathbf{I} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{N} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and the corresponding norms and inner products

$$\|\mathbf{d}\|_{\mathbb{R}^2}^2 + \lambda \|d\|_{\mathrm{L}^2(\mathbb{S}^2)}^2 = \begin{cases} 2 + 4\pi\lambda, & d = I, \\ 1 + \pi\lambda, & d = N, \\ 1 + \pi\lambda, & d = S \end{cases}$$

$$\langle \mathbf{y}, \mathbf{d} \rangle_{\mathbb{R}^2} = \begin{cases} 1, & d = I, \\ 1, & d = N, \\ 0, & d = S \end{cases}, \quad \langle \mathbf{z}, \mathbf{d} \rangle_{\mathbb{R}^2} = \begin{cases} 1, & d = I, \\ 0, & d = N, \\ 1, & d = S \end{cases}, \quad \langle \mathbf{y} + \mathbf{z}, \mathbf{d} \rangle_{\mathbb{R}^2} = \begin{cases} 2, & d = I, \\ 1, & d = N, \\ 1, & d = S \end{cases}.$$

We start the algorithm with $F_0 = 0$. If $\mathbf{y}$ is the data vector, i.e. $\mathbf{R}^0 := \mathbf{y}$, then $d_1 = N$ is the maximizer in (1) and (2) yields $\alpha_1 = \frac{1}{1+\pi\lambda}$, i.e. $F_1 = \frac{1}{1+\pi\lambda} N$. On the other hand, if $\mathbf{z}$ is the data vector, i.e. $\mathbf{R}^0 := \mathbf{z}$, then (analogously) we get $F_1 = \frac{1}{1+\pi\lambda} S$. However, if we set $\mathbf{y} + \mathbf{z}$ as the data vector, then $\mathbf{R}^0 := \mathbf{y} + \mathbf{z}$ yields $d_1 = I$ and $\alpha_1 = \frac{1}{1+2\pi\lambda}$ such that $F_1 = \frac{1}{1+2\pi\lambda} I$.
As a consequence, if we truncate the algorithm already after $n = 1$, then the result F_1 obtained for the data vector $\mathbf{y} + \mathbf{z}$ is not the sum of the results for $\mathbf{y}$ and $\mathbf{z}$. Hence, the approximations depend in a non-linear sense on the given data.

Our further considerations refer to the known theorems about the RFMP, where we simply formulate them for the particular case that $D = \mathbb{S}^q$ and the operator $\mathcal{F} : \mathcal{H}(\mathbb{S}^q) \to \mathbb{R}$ is given by

$$\mathcal{F}F := \left(F\left(\eta^j\right)\right)_{j=1,\dots,l} \in \mathbb{R}^l .$$

Obviously, F is linear and, due to the choice of a RKHS $\mathcal{H}(\mathbb{S}^q)$, also continuous. We omit most of the proofs of the following theorems, since they are analogous to the proofs for the RFMP, see, [36].

Theorem 6.2 *The sequence $(\|\mathbf{R}^n\|^2_{\mathbb{R}^n} + \lambda \|F_n\|^2_{\mathcal{H}(\mathbb{S}^q)})_n$ produced by Algorithm 3.1 is monotonically decreasing and convergent.*

Theorem 6.3 (Convergence Theorem) *Let the dictionary $\mathcal{D}$ satisfy the following conditions:*

1. *'semi-frame condition': Every expansion $H = \sum_{k=1}^{\infty} \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k need not be pairwise distinct but none of them occurs infinitely often, continuously depends on the sequence of coefficients $(\beta_k)_k$, i.e. there exists a constant $\gamma > 0$ such that*

$$\left\| \sum_{k=1}^{\infty} \beta_k d_k \right\|^2_{\mathcal{H}(\mathbb{S}^q)} \leq \gamma \sum_{k=1}^{\infty} \beta_k^2 ,$$

   where γ is independent of H.

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

Moreover, let $\mathbf{y} \in \mathcal{F}(\mathcal{H}(\mathbb{S}^q))$ be a given vector of samples and $(F_n)_n$ be the sequence produced by the RFMP_AoS, where no dictionary element is chosen infinitely often. Then the sequence $(F_n)_n$ converges in $\mathcal{H}(\mathbb{S}^q)$ to $F_\infty := \sum_{k=1}^{\infty} \alpha_k d_k \in \mathcal{H}(\mathbb{S}^q)$ and this limit has the following properties:

(a) If $\overline{\operatorname{span} \mathcal{D}}^{\|\cdot\|_{\mathcal{H}(\mathbb{S}^q)}} = \mathcal{H}(\mathbb{S}^q)$, $C_2 := \sup_{d \in \mathcal{D}} \|d\|_{\mathcal{H}(\mathbb{S}^q)} < +\infty$, and $\lambda \in \mathbb{R}_0^+$, then F_∞ solves the Tikhonov-regularized normal equation

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

where $\mathcal{F}^*$ is the adjoint operator of $\mathcal{F}$ given by

$$\mathcal{F}^*\mathbf{z} = \sum_{j=1}^{l} z_j K_{\mathcal{H}}\left(\eta^j, \cdot\right), \quad \mathbf{z} \in \mathbb{R}^l.$$

This means that F_∞ minimizes the regularized least-square error in the sense that

$$\sum_{j=1}^{l}\left(y_j - F_\infty\left(\eta^j\right)\right)^2 + \lambda \left\|F_\infty\right\|^2_{\mathcal{H}(\mathbb{S}^q)} = \min_{F\in\mathcal{H}(\mathbb{S}^q)}\left[\sum_{j=1}^{l}\left(y_j - F\left(\eta^j\right)\right)^2 + \lambda \left\|F\right\|^2_{\mathcal{H}(\mathbb{S}^q)}\right],$$

where the solution if unique, if $\lambda > 0$.

(b) If $\operatorname{span}\left\{(d(\eta^j))_{j=1,\dots,l} \,|\, d \in \mathcal{D}\right\} = \mathbb{R}^l$ and $\lambda = 0$, then $F_\infty(\eta^j) = y_j$ for all $j = 1, \dots, l$.

Proof. The formula for the adjoint operator $\mathcal{F}^*$ is valid, because

$$\begin{aligned}
\langle \mathbf{z}, \mathcal{F}F\rangle_{\mathbb{R}^l} &= \sum_{j=1}^{l} z_j\, F\left(\eta^j\right) \\
&= \sum_{j=1}^{l} z_j \left\langle K_{\mathcal{H}}\left(\eta^j, \cdot\right), F\right\rangle_{\mathcal{H}(\mathbb{S}^q)} \\
&= \left\langle \sum_{j=1}^{l} z_j\, K_{\mathcal{H}}\left(\eta^j, \cdot\right), F\right\rangle_{\mathcal{H}(\mathbb{S}^q)} \\
&= \langle \mathcal{F}^*\mathbf{z}, F\rangle_{\mathcal{H}(\mathbb{S}^q)}
\end{aligned}$$

for all $\mathbf{z} \in \mathbb{R}^l$ and all $F \in \mathcal{H}(\mathbb{S}^q)$. The rest of the proof is analogous to the proof of Theorem 4 in [36]. ∎

Remark 6.4 *In the implementation, it can be reasonable to normalize all dictionary elements d to $\|d\|_{\mathcal{H}(\mathbb{S}^q)} = 1$. This reduces the computational expenses in (1) and (2). Moreover, in this case, two conditions in Theorem 6.3 become trivial (at least, if $\lambda > 0$) since then*

$$\begin{aligned}
C_1 &= \inf_{d\in\mathcal{D}}(\|\mathcal{F}d\|^2_{\mathbb{R}^l} + \lambda\|d\|^2_{\mathcal{H}(\mathbb{S}^q)}) \geq \lambda\,, \\
C_2 &= \sup_{d\in\mathcal{D}} \|d\|_{\mathcal{H}(\mathbb{S}^q)} = 1\,.
\end{aligned}$$

Theorem 6.3 tells us that the RFMP_AoS yields a smoothed approximation of the unknown function, where the smoothness is controlled by the Sobolev norm. In the case $\lambda = 0$, we obtain an exact interpolant. Note that these results are similar to the properties of a spline interpolation or approximation (see [13, 14, 18]). The difference is, however, the flexibility of the representation of F_∞. We will see below in the numerical experiments (Section 7) that we are able to achieve the approximation quality of a spline but (in several examples) with an essentially sparser representation of the solution.

Theorem 6.5 (Stability of the Solution) *Let the dictionary satisfy conditions 1, 2, and (a) of Theorem 6.3 and let $\lambda > 0$. For a given convergent sequence $(\mathbf{y}^k)_k \subset \mathbb{R}^l$ with limit $\mathbf{y} \in \mathbb{R}^l$, let $(F_{\infty,\mathbf{y}^k})_k$ and $F_{\infty,\mathbf{y}}$ be the corresponding solutions produced by the RFMP_AoS (according to Theorem 6.3). Then*

$$\lim_{k\to\infty} \left\| F_{\infty,\mathbf{y}^k} - F_{\infty,\mathbf{y}} \right\|_{\mathcal{H}(\mathbb{S}^q)} = 0\,.$$

Theorem 6.5 means that the application of the RFMP_AoS to a given data vector $\mathbf{y}$ is stable. Small perturbations in the data only cause small variations of the approximating function F_∞. This is clearly the effect of the Tikhonov regularization.

Theorem 6.6 (Convergence of the Regularization) *Let the dictionary satisfy conditions 1, 2, and (a) of Theorem 6.3. For an unknown function $F \in \mathcal{H}(\mathbb{S}^q)$, let $\mathbf{y} = (F(\eta^j))_{j=1,\dots,l} \in \mathbb{R}^l$ be the corresponding vector of samples and let $(\mathbf{y}^\varepsilon)_{\varepsilon>0} \subset \mathcal{F}(\mathcal{H}(\mathbb{S}^q))$ be a family of given perturbed data vectors with $\|\mathbf{y} - \mathbf{y}^\varepsilon\|_{\mathbb{R}^l} \le \varepsilon$. Moreover, let F^+ be the minimum-norm solution of the interpolation problem, i.e.,*

$$\left\| F^+ \right\|_{\mathcal{H}(\mathbb{S}^q)} = \min \left\{ \|F\|_{\mathcal{H}(\mathbb{S}^q)} \middle|\ F \in \mathcal{H}(\mathbb{S}^q) \text{ and } F\left(\eta^j\right) = y_j \text{ for all } j = 1,\dots,l \right\}\,.$$

Furthermore, let the regularization parameter $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ be parameterized with respect to the noise level ε such that

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

Moreover, for each data vector $\mathbf{y}^\varepsilon$ ($\varepsilon > 0$) and the corresponding regularization parameter $\lambda(\varepsilon)$, let $F_{\infty,\varepsilon}$ denote the approximation produced by the RFMP_AoS (due to Theorem 6.3). Then

$$\lim_{\varepsilon\to 0+} \left\| F_{\infty,\varepsilon} - F^+ \right\|_{\mathcal{H}(\mathbb{S}^q)} = 0\,.$$

Note that the minimum-norm solution F^+ is known to be the interpolating spline in $\mathcal{H}(\mathbb{S}^q)$ due to the first minimum property of spline interpolation. In this respect and in view of the fact that the Sobolev norm $\|\cdot\|_{\mathcal{H}(\mathbb{S}^q)}$ may be regarded as a non-smoothness measure, we can interpret Theorem 6.6 as follows: the more accurate the available data of the unknown function are, the better we can approximate the unknown function by the smoothest function which fits to the data.
Note that both Theorems 6.5 and 6.6 are the basic ingredients of a regularization.

7 Numerical Experiments

In this section, we present results of some numerical experiments on the unit sphere in $\mathbb{R}^3$. That is, in what follows, the underlying domain is $\mathbb{S}^2$.
First of all, we will consider some benchmark functions, which were originally proposed by various authors (see [31] and the references therein). These functions are, for $\xi = (\xi_1, \xi_2, \xi_3)^{\mathrm{T}} \in \mathbb{S}^2$, given

by

$$
\begin{aligned}
g_1(\xi) &= (\xi_1 - 0.9)_+^{3/4} + (\xi_3 - 0.9)_+^{3/4}\,, \\
g_2(\xi) &= \left[0.01 - \left(\xi_1^2 + \xi_2^2 + (\xi_3 - 1)^2\right)\right]_+ + \exp(\xi_1 + \xi_2 + \xi_3)\,, \\
g_3(\xi) &= \frac{1}{101 - 100\xi_3}\,, \\
g_4(\xi) &= \frac{1}{|\xi_1| + |\xi_2| + |\xi_3|}\,, \\
g_5(\xi) &= \begin{cases} \cos^2\left(\frac{3\pi}{2}|\xi - \eta|\right)\,, & \text{if } |\xi - \eta| < 1/3\,, \\ 0\,, & \text{if } |\xi - \eta| \geq 1/3\,, \end{cases}
\end{aligned}
\tag{4}
$$

where $\eta = (-1/2, -1/2, 1/\sqrt{2})^{\mathrm{T}}$ and $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. The benchmark functions are plotted in Figure 1.

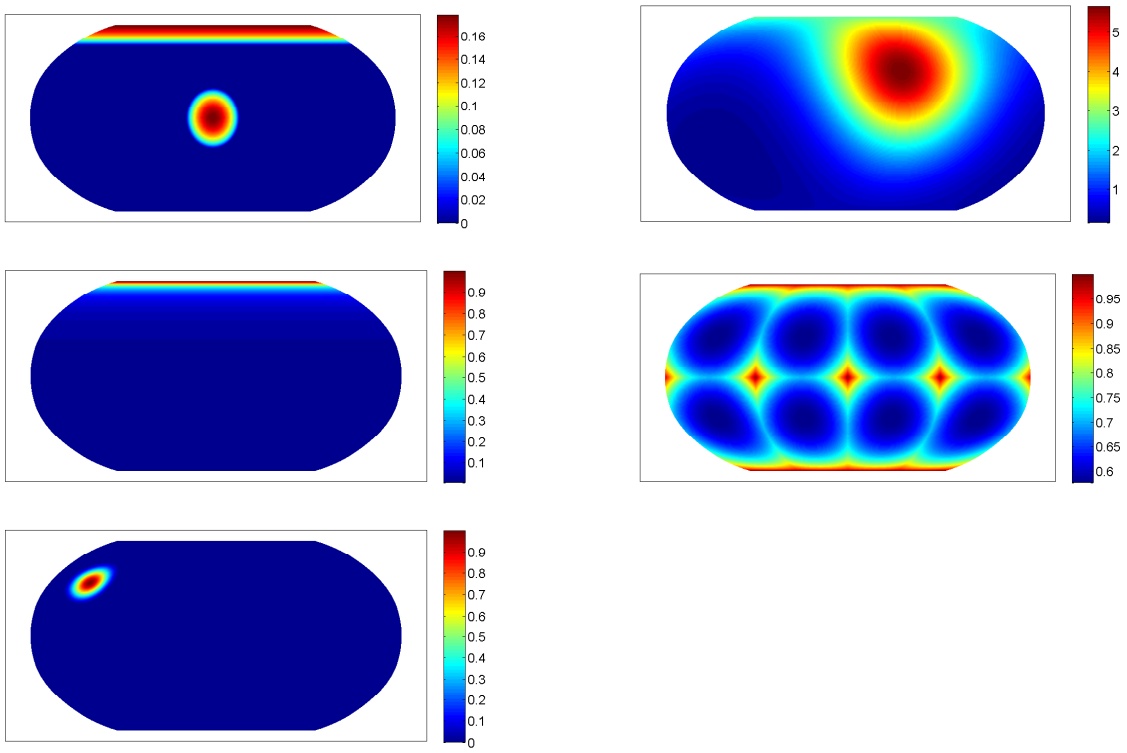

Figure 1: Benchmark functions g_1 (top left), g_2 (top right), g_3 (middle left), g_4 (middle right), and g_5 (bottom).

In a second experiment, we will consider the Earth's gravitational potential as an example of an application of the algorithm in the geosciences. In order to be able to use data of the gravitational potential on a point grid as the input, we sample the Earth's Gravitational Model

2008 (EGM2008, see [38]) on a given set of irregularly distributed points, which is displayed in Figure 4.
We will compare different dictionaries. In the first one, we want to combine spherical harmonics as well as Abel-Poisson kernels. Let now $N \in \mathbb{N}$ be a prescribed number and let us denote by

$$\mathcal{Y}^N := \{Y_{n,j} \,|\, n = 0, \dots, N;\, j = 1, \dots, 2n+1\}$$

the set of spherical harmonics up to degree N and order $2N+1$. Furthermore, let $X_M := \{\xi^1, \dots, \xi^M\} \subset \mathbb{S}^2$ be a set of $M \in \mathbb{N}$ pairwise distinct points which are regularly distributed over the entire unit sphere. These points will be used for the locations of the kernels which are contained in our dictionary. In fact, we use a so-called Reuter grid as proposed in [41] with a parameter $\gamma = 80$ which results in a total of 8102 pairwise distinct points. Moreover, let

$$\mathfrak{h} := \{0.75, 0.80, 0.85, 0.89, 0.91, 0.93, 0.94, 0.95, 0.96, 0.97\}$$

be the parameter set for the localization of the Abel-Poisson kernels (see Example 4.4) in order to ensure a large variety of localized functions in our dictionary. Note that the localized trial functions $K_h(\xi\cdot)$ correspond to spline basis functions (or to scaling functions, see also the construction of dictionary $\mathcal{D}_2$ below).
The first dictionary which we are going to use for the numerical experiments then reads as follows,

$$\mathcal{D}_1 := \mathcal{Y}^{25} \cup \{K_h(\xi\cdot) \,|\, h \in \mathfrak{h},\, \xi \in X_M\}\,.$$

Furthermore, we want to take spherical wavelets into account as well. For an overview on this topic, the reader is referred to [18] and the references therein. In the present case, we construct the wavelets with the Abel-Poisson scaling function, which is generated by

$$\varphi_0(x) := \exp(-Rx) \quad \text{for all } \; x \in \mathbb{R}^+\,. \tag{5}$$

Consequently, the Legendre coefficients of the scaling function are, for all $n \in \mathbb{N}_0$ and all scales $J \in \mathbb{N}_0$, given by $\Phi_J^\wedge(n) = \varphi_J(n) := \varphi_0(2^{-J}n)$ such that, for all $t \in [-1,1]$,

$$\Phi_J(t) = \sum_{n=0}^{\infty} \exp\left(-R2^{-J}\right)^n \frac{2n+1}{4\pi} P_n(t)$$

which is a reproducing kernel due to Theorem 4.3. More specifically, we know from Example 4.4 that

$$\Phi_J(t) = \frac{1}{4\pi} \frac{1 - \exp(-R2^{1-J})}{(1 + \exp(-R2^{1-J}) - 2\exp(-R2^{-J})t)^{3/2}}\,, \quad t \in [-1,1]\,,$$

which means that the Abel-Poisson scaling function is an Abel-Poisson kernel with the parameter $h = \exp(-R2^{-J})$, where $t = \xi \cdot \eta$ for $\xi, \eta \in \mathbb{S}^2$. We choose $R = 0.1791$ as the parameter for the generator (see (5)) such that the iterated scaling function $\Phi_0^{(2)} := \Phi_0 * \Phi_0$ corresponds to an Abel-Poisson kernel with parameter $h = 0.6989$, where the convolution of spherical kernels $K, L \in \mathrm{L}^2[-1,1]$ is (as usual) defined by

$$(K * L)(\xi \cdot \eta) := \int_{\mathbb{S}^2} F(\xi \cdot \zeta) L(\eta \cdot \zeta)\, \mathrm{d}\omega(\zeta)\,, \quad \xi, \eta \in \mathbb{S}^2\,.$$

Moreover, the corresponding P-wavelets are, for every $n \in \mathbb{N}_0$, defined by the Legendre coefficients

$$\Psi_J^\wedge(n) := \tilde{\Psi}_J^\wedge(n) := \sqrt{\exp\left(-R2^{-J}n\right) - \exp\left(-R2^{1-J}n\right)}\,.$$

It is well-known that the Legendre coefficients of the associated spherical convolution are given by

$$\left(\tilde{\Psi}_J * \Psi_J\right)^\wedge(n) = \Psi_J^\wedge(n) \cdot \tilde{\Psi}_J^\wedge(n)\,, \quad n \in \mathbb{N}_0\,,$$

such that we are able to conclude that, for every scale $J \in \mathbb{N}_0$ and every $t \in [-1,1]$, the Abel-Poisson P-wavelets are given by

$$\left(\tilde{\Psi}_J * \Psi_J\right)(t) = \sum_{n=0}^{\infty} \exp\left(-R2^{-J}\right)^n \frac{2n+1}{4\pi} P_n(t) - \sum_{n=0}^{\infty} \exp\left(-R2^{1-J}\right)^n \frac{2n+1}{4\pi} P_n(t)\,.$$

Again, by virtue of Theorem 4.3, the wavelets above are simply a difference of two reproducing kernels which are Abel-Poisson kernels with $h = \exp(-R2^{-J})$ and $h = \exp(-R2^{1-J})$, respectively. Hence, we have a closed representation for these kernels as we already saw above. We now define another dictionary by

$$\mathcal{D}_2 := \mathcal{Y}^{25} \cup \left\{ \Phi_0^{(2)}(\xi\cdot) \,\middle|\, \xi \in X_M \right\} \cup \left\{ \left(\tilde{\Psi}_J * \Psi_J\right)(\xi\cdot) \,\middle|\, \xi \in X_M,\, J = 0,1,2 \right\}\,,$$

which combines spherical harmonics and scaling functions as well as wavelets for different scales, where the grid for the centres of the localized trial functions is the same as above. The latter dictionary additionally gives us the opportunity to obtain a multiresolution of the signal where we do not have to solve any integration problem as it would be the case for the application of the corresponding wavelet method.

We start with the approximation of data which is generated by the various benchmark functions given in (4) and compare our results with the associated results obtained with the spherical spline method based on Abel-Poisson kernels. For more information on this method see, for example, [14, 18] and the references therein. To do so, let now $\mathcal{H}(\mathbb{S}^2)$ be a certain Sobolev space with corresponding reproducing kernel $K_{\mathcal{H}}$. Moreover, let $l \in \mathbb{N}$ be a prescribed number and let $y = (y_1, \ldots, y_l)^{\mathrm{T}} \in \mathbb{R}^l$ be the underlying data which is given at distinct points $\eta^i \in \mathbb{S}^2$ for $i = 1, \ldots, l$. Then, for every $\xi \in \mathbb{S}^2$, a function S defined by

$$S(\xi) = \sum_{j=1}^{l} a_j K_{\mathcal{H}}(\eta^j, \xi)$$

is a spherical spline in $\mathcal{H}(\mathbb{S}^2)$ relative to $\{\eta^1, \ldots, \eta^l\} \subset \mathbb{S}^2$. If we require the spline to interpolate the data, i.e. $S(\eta^i) = y_i$ for $i = 1, \ldots, l$, then the vector of the coefficients $a = (a_1, \ldots, a_l)^{\mathrm{T}} \in \mathbb{R}^l$ of the corresponding interpolating spline is given by the solution of the system

$$\left(K_{\mathcal{H}}(\eta^j, \eta^i)\right)_{i,j=1,\ldots,l} a = y\,.$$

As the input data for our first experiments, we sample the different benchmark functions at locations which are regularly distributed over the sphere. Therefore, we use another Reuter grid

with $\gamma = 100$, which results in a set $\{\eta^1, \dots, \eta^l\} \subset \mathbb{S}^2$ with a total of $l = 12\,684$ pairwise distinct points. That is, we consider the samples

$$g_k(\eta^i) = y_i\,, \quad i = 1, \dots, 12\,684$$

for every function g_k, $k = 1, \dots, 5$ and use the RFMP_AoS as well as the spherical spline method to reconstruct these data. To construct the spline, we choose Abel-Poisson kernels with $h = 0.93$, i.e. $\mathcal{H}(\mathbb{S}^2) = \mathcal{H}((h^{-n/2}); \mathbb{S}^2)$ is the corresponding Sobolev space. The associated system of equations to obtain the vector of spline coefficients has a condition number of $6.6077 \cdot 10^3$ such that there is no need for a regularization to stabilize the solution in the case of the spline method.
In the case of the RFMP_AoS, we will at least use a very slight regularization to improve the results of the algorithm. That means, we use a very low parameter $\lambda > 0$ while the penalizing norm will be the $\mathcal{H}_2(\mathbb{S}^2)$-norm (see Example 4.4). Also note that the dictionaries which we defined do not contain localized functions which are located at the data points and the locations of the kernels and wavelets are totally independent of the data. In general, due to the construction of the spline, since the reproducing kernels are located at the data points, the data misfit of the spline is expected to be lower than the one obtained with the RFMP_AoS. In fact, for all benchmark functions, the data misfit of the spline is smaller than $2 \cdot 10^{-15}$ which we are not able to obtain with our algorithm. However, if we denote by F the underlying signal which produces the data and G is an approximation of the given data, then the relative approximation error which is defined by

$$\epsilon_{\mathrm{r}} := \frac{\|F - G\|_{\mathrm{L}^2(\mathbb{S}^2)}}{\|F\|_{\mathrm{L}^2(\mathbb{S}^2)}} = \left(\frac{\int_{\mathbb{S}^2} (F(\xi) - G(\xi))^2 \,\mathrm{d}\omega(\xi)}{\int_{\mathbb{S}^2} (F(\xi))^2 \,\mathrm{d}\omega(\xi)} \right)^{1/2}$$

is more significant for the quality of the approximation than the data misfit, in particular in the case of irregularly distributed data. Let now

$$\{\zeta_i\}_{i=1,\dots,45\,000}$$

be an equispaced grid of the longitude and the latitude with $300 \times 150 = 45\,000$ grid points which we will use to display our results below, then we can approximately compute the relative approximation error by

$$e_{\mathrm{r}} := \left(\frac{\sum_i (F(\zeta_i) - G(\zeta_i))^2}{\sum_i (F(\zeta_i))^2} \right)^{1/2}.$$

In the case of the benchmark functions g_1, g_2 and g_5 we run the RFMP_AoS with a regularization parameter $\lambda = 1 \cdot 10^{-5}$. Note that for these functions the approximation is not very sensitive to the regularization parameter since the signals are very smooth. Moreover, we will terminate each approximation process after 2000 iterations which already leads to satisfactory results. We also have to keep in mind that the iteration process itself has a regularizing effect on the solution such that too many iterations might result in a too severe regularization. In contrast to the spline, which is constructed with 12 684 reproducing kernels, the RFMP_AoS needs essentially less functions to attain slightly smaller approximation errors. In the case of g_5, the ROFMP

yields an expansion with only 793 different trial functions which even produces an essentially lower approximation error in comparison to the spline. The function g_4, however, contains much more detailed structures which demand a higher amount of trial functions to reconstruct such that we, in that case, run the algorithm for 20 000 iterations and use a smaller regularization parameter of $\lambda = 1 \cdot 10^{-8}$ to support the reconstruction of the details. The solution is an expansion with about half of the number of functions which are used for the spline and which still results in an approximation error which is slightly lower than the one obtained with the spline method. Note that the algorithm chooses dictionary elements several times. For this reason, we eventually count the number of pairwise distinct dictionary elements used for the approximation.

Benchmark	Rel. Approximation Error		Number of Functions
	Spline	RFMP_AoS	RFMP_AoS
g_1	0.066493	0.061319	1325
g_2	0.016497	0.012493	1106
g_3	0.075570	0.094869	750
g_4	0.008980	0.008941	6315
g_5	0.132759	0.093599	793

Table 1: Relative approximation errors corresponding to the solutions with a spherical spline and with the RFMP_AoS for data given by the various benchmark functions g_k, $k = 1, \ldots, 5$, as well as the number of pairwise distinct trial functions chosen by the RFMP_AoS algorithm. Note that the spline always uses $l = 12\,864$ basis functions.

The only exception to the former results is the benchmark function g_3 where our algorithm attains a higher approximation error even after 10 000 iterations with a parameter $\lambda = 1 \cdot 10^{-8}$. However, the number of chosen trial functions is extremely low in comparison to the number of kernels in the spline and in comparison to the number of iterations. For this reason, the result might also be improved with the use of a different dictionary. All the other approximation errors are comparable and all important numbers can be found in Table 1.
If we now use the other dictionary $\mathcal{D}_2$, we are also able to obtain a multiresolution of the underlying signal. Therefore, let us again consider benchmark function g_4 (which has a relatively complicated structure) and approximate the given samples with the RFMP_AoS where we now provide $\mathcal{D}_2$ for the algorithm to pick functions from. As it was the case with the dictionary $\mathcal{D}_1$, the algorithm reconstructs the major features of the signal in terms of spherical harmonics in the beginning of the approximation process. However, since the dictionary now contains spherical scaling functions as well as wavelets, we can take a look at the details which are added by these functions. Particularly interesting in that regard are the wavelets of the highest scale, i.e. wavelets of scale $J = 2$. We choose again the regularization parameter $\lambda = 1 \cdot 10^{-8}$ and run the algorithm for 10 000 iterations. The result is displayed in Figure 2 and the associated approximation error is with $e_{\mathrm{r}} = 0.008960$ between the one of the spherical spline method and the one obtained with the RFMP_AoS with dictionary $\mathcal{D}_1$ after 20 000 iterations, which is double the amount. We see that the wavelet scales contribute to a 'sharper' image of the smoothed

version of g_4 which is obtained from the chosen spherical harmonics of the solution. In total, the RFMP_AoS chooses 3924 pairwise distinct trial functions (56 spherical harmonics and 3868 localized functions) from $\mathcal{D}_2$ to obtain an approximation which is comparable to a spline (with 6968 basis functions) but is also equipped with a multiresolution. Hence, in the case of function g_4, we can say that the use of the wavelet dictionary is advantageous.

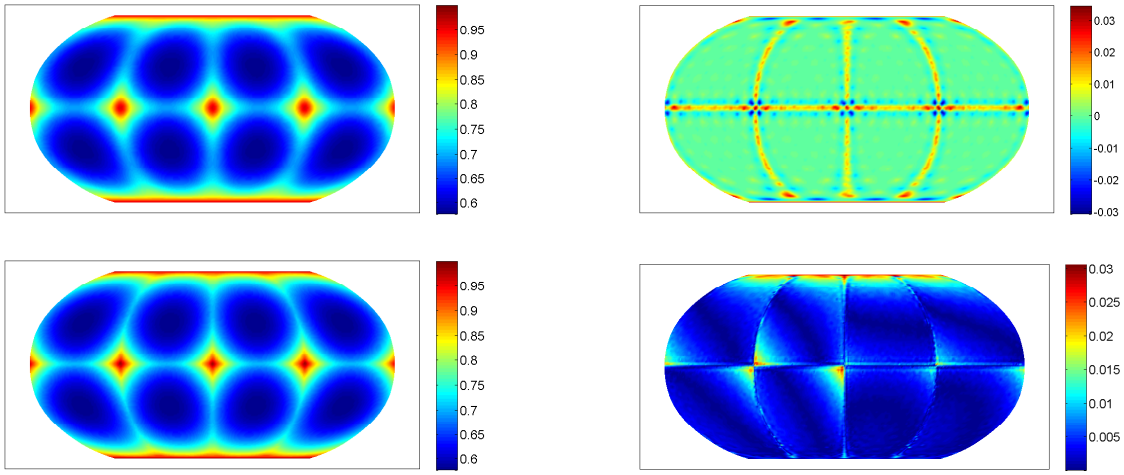

Figure 2: Approximation of benchmark function g_4 with RFMP_AoS and $\mathcal{D}_2$ after 10 000 iterations. The upper row contains the chosen spherical harmonics (top left) as well as the picked wavelets of scale $J = 2$ (top right) while the entire expansion is displayed in the bottom row (bottom left) with the corresponding absolute approximation error (bottom right).

In that regard, it is also interesting to have a look at the centres for the localized basis functions which were chosen by the algorithm, that is, the subset of X_M which is used to locate the trial functions in our expansion of the signal. These points are displayed in Figure 3 and one can easily recognize the structure of the function in the chosen centres, which means that the outcome of the RFMP_AoS is, in fact, adapted to the detail structure of the underlying signal.

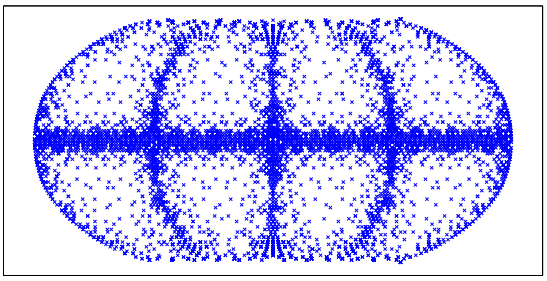

Figure 3: The centres of the 3868 localized trial functions which were chosen by the algorithm for benchmark function g_4.

As a second experiment (which is more appropriate to test the multiresolution feature), let us consider a geoscientific application and approximate the gravitational potential on the Earth's surface. Therefore, we sample the EGM2008 expansion on an irregular data grid consisting of 6968 points (see Figure 4). As we can see, the point grid contains essentially more points on the continents than in other areas which is reasonable for the case of terrestrial measurements, although this grid is certainly contrived.

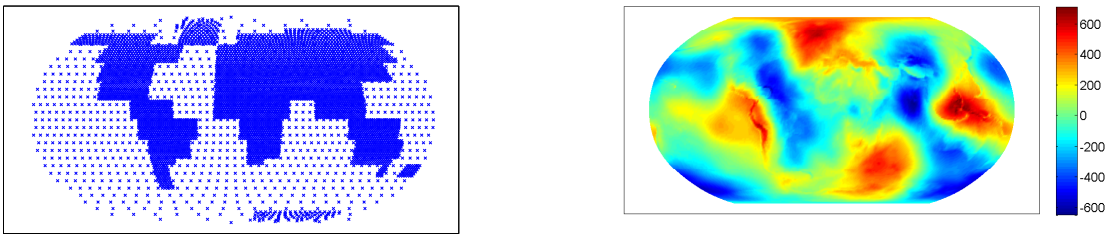

Figure 4: Irregular point grid for the approximation of terrestrial gravitational data (6968 points, left) and the EGM2008 signal (right) in $\mathrm{m}^2/\mathrm{s}^2$.

First of all, for a better comparison, we again compute a spherical spline consisting of Abel-Poisson kernels with $h = 0.9$ such that again $\mathcal{H}(\mathbb{S}^2) = \mathcal{H}((h^{-n/2}); \mathbb{S}^2)$ and we stabilize the corresponding system of equations with

$$\lambda_{\mathrm{S}} = 0.159513$$

which appears to approximately minimize the relative approximation error of the result. That means, we now have to solve

$$\left[\left(K_{\mathcal{H}}(\eta^j, \eta^i)\right)_{i,j=1,\dots,l} + \lambda_{\mathrm{S}} I\right] a = y$$

where I is the 6968×6968 identity matrix.
If we let the algorithm run for 50 000 iterations while using dictionary $\mathcal{D}_1$ and a regularization parameter $\lambda = 2 \cdot 10^{-5}$, the method picks a total of 10 162 different functions. These functions consist of 148 spherical harmonics which are chosen up to 4 times and 10 014 Abel-Poisson kernels with various centres and different localizations which are picked up to 187 times by the algorithm. The relative approximation error of the spline is already attained after 18 500 iterations with 7374 functions (140 spherical harmonics and 7234 Abel-Poisson kernels). The corresponding errors can be found in Table 2 while the results are displayed in Figure 5.
As we can see, the data fidelity of the spline is again higher which is due to the fact that the reproducing kernels for the spline are located at the data points but these points are not contained in the grid X_M which we use to contruct the dictionary. That means, the spline can again reconstruct the data more accurately. However, the RFMP_AoS is still able to attain the approximation error of the spline after 18500 iterations and even falls below it in the iterations following thereafter although the approximation error decreases very slowly. Note that, in particular in the case of irregularly distributed data, the approximation error is more important than the data misfit for the quality of the approximation.

Method	Spline	RFMP_AoS	
Iterations		18500	50000
Data misfit	0.032974	0.043043	0.041157
Approximation error	0.063583	0.063583	0.062449
Number of functions	6968	7374	10162

Table 2: Errors corresponding to the approximations of the EGM2008 data with a spherical spline as well as with the RFMP_AoS.

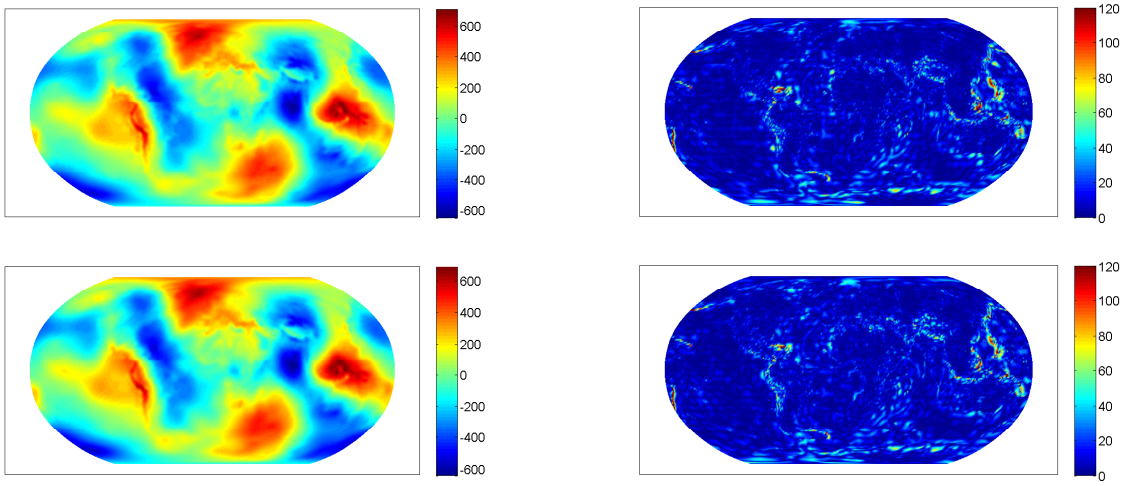

Figure 5: Approximation of EGM2008 samples with spherical spline (top left) and with RFMP_AoS (bottom left) after 50000 iterations with dictionary $\mathcal{D}_1$ as well as the corresponding absolute approximation errors with regard to EGM2008 (right hand). All values in $\mathrm{m}^2/\mathrm{s}^2$.

Let us again use dictionary $\mathcal{D}_2$ and consider a multiscale analysis of the signal. Again, we use the parameter $\lambda = 2 \cdot 10^{-5}$ for the regularization and, this time, we abort the process after 30 000 iterations. The relative approximation error is with

$$e_{\mathrm{r}} = 0.063824 \tag{6}$$

only negligibly larger than the approximation error obtained above with the spline method for the same data. Within the 30 000 iterations, the RFMP chooses 329 different spherical harmonics as well as 26 different scaling functions and 3544 wavelets which is a total of 3899 functions. Apparently, this means that, with the dictionary $\mathcal{D}_2$, the algorithm needs less functions to achieve comparable results than before. This is due to the construction of the wavelets, which is a difference of two scaling functions, such that the algorithm is able to add details more effectively. With the dictionary $\mathcal{D}_1$ which we used at first, the RFMP had to pick at least two functions to add to the expansion what it is now able to add with one single wavelet. This explains the smaller number of functions which are necessary for an appropriate representation of the signal.

In Figure 6 and Figure 7, we see the approximation of the potential data at different scales in the left-hand column and the details which are added to each scale in the right-hand column. That is, in the top left image of 6 the chosen spherical harmonics are displayed and the top right image shows the chosen scaling functions. The combination of both of the latter functions is shown in the left image of the second row. Obviously, the spherical harmonics and the scaling functions alone, which is a total of 371 functions (296 spherical harmonics as well as 75 scaling functions), do not yield a proper approximation of the signal since certain structures, in particular around Greenland near the North pole, are missing. The other image in the right-hand column features the wavelets of scale $J = 0$ while the combination of the chosen spherical harmonics as well as scaling functions and wavelets of scale $J = 0$ is displayed in the bottom row. With this first wavelet scale ($J = 0$), we are now actually able to get a rough image of the signal and the major structures of the signal are contained, although the resolution is not very fine yet. However, this changes as soon as we add the finer wavelet scales, which exactly conforms with the intention of this multiresolution approach.

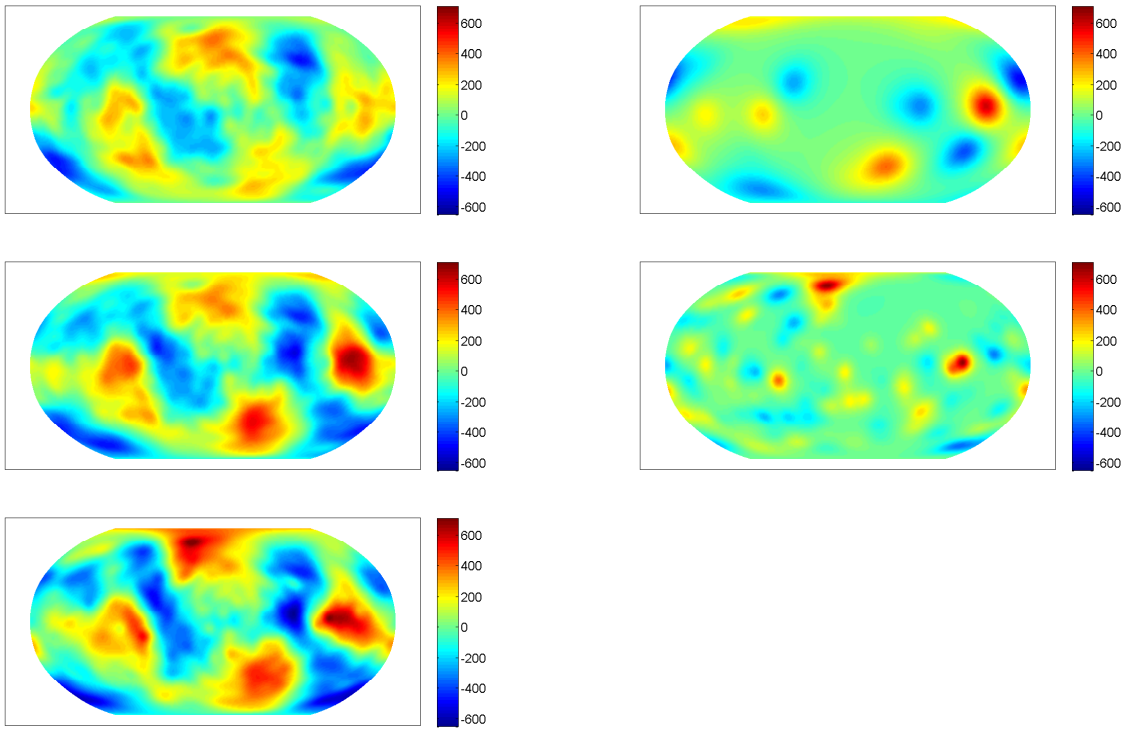

Figure 6: Multiresolution of EGM2008 samples with ROFMP_AoS and Dictionary $\mathcal{D}_2$. The chosen spherical harmonics are shown in the top left image. The right-hand column features the chosen scaling functions (top) as well as the chosen wavelets of scale $J = 0$ (second row). The corresponding approximations are shown in the left-hand column, i.e. the combination of spherical harmonics and scaling functions (second row) and all chosen functions up to wavelet scale $J = 0$ in the bottom row. All values in $\mathrm{m}^2/\mathrm{s}^2$.

In Figure 7, we see again the latter image (top left) as well as the approximations with all of the chosen functions up to the finer wavelet scales $J = 1$ (middle left) and $J = 2$ (bottom) while the right-hand column again features the added details, i.e. the chosen wavelets of scale $J = 1$ (top right) and scale $J = 2$ (middle right). In these details, certain structures of the signal such as the Andes or the Himalayas can be easily recognized such that the details themselves contain a lot of information about the signal.

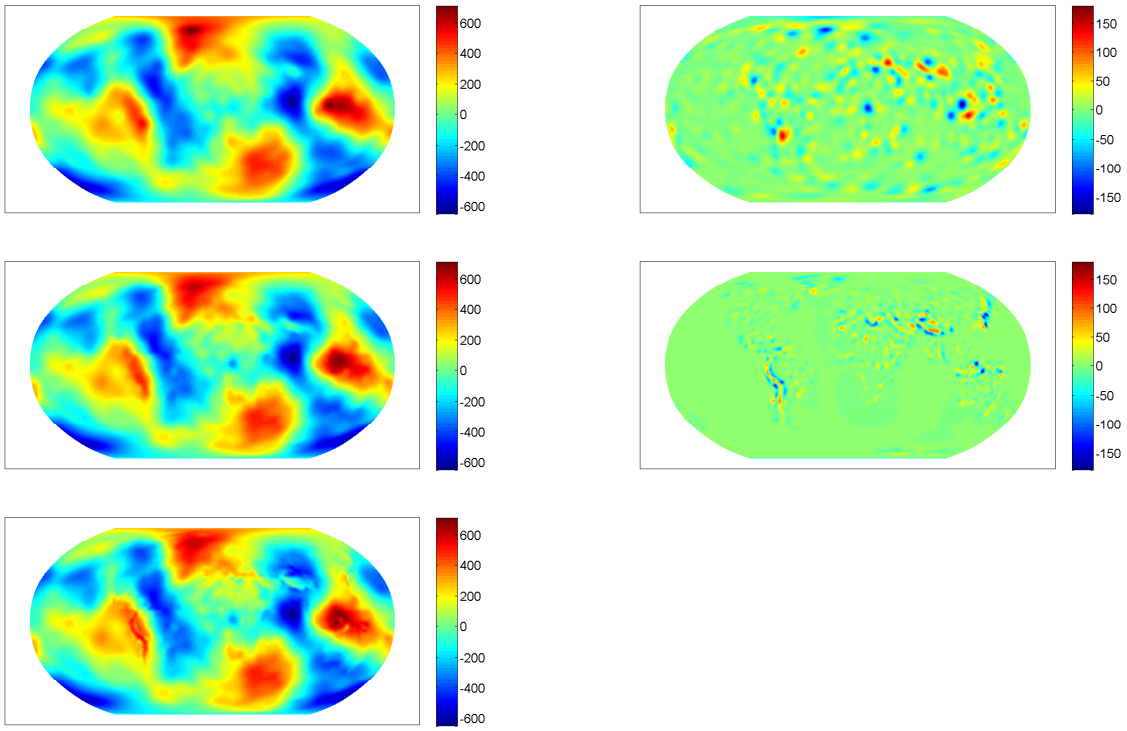

Figure 7: Multiresolution of EGM2008 samples with ROFMP_AoS and Dictionary $\mathcal{D}_2$. The top left iage is again the combination of chosen spherical harmonics, scaling functions as well as wavelets of scale $J = 0$. The right-hand column features the chosen wavelets of scale $J = 1$ and $J = 2$. The corresponding approximations are shown in the left-hand column, i.e. the combination of spherical harmonics and scaling functions (second row) and all chosen functions up to wavelet scale $J = 1$ (middle) and $J = 2$ (bottom). All values in $\mathrm{m}^2/\mathrm{s}^2$.

Let us once more take a look at the chosen centres of the localized trial functions, in particular, at the centres of the chosen wavelets. In Figure 8, we display those centres where the color of the points corresponds to the associated coefficent of the chosen function. We can see that the functions are chosen according to the data density (locations) as well as to the detail structure of the underlying signal (value of coefficients). The centres of all of the chosen wavelets are again shown in Figure 9.

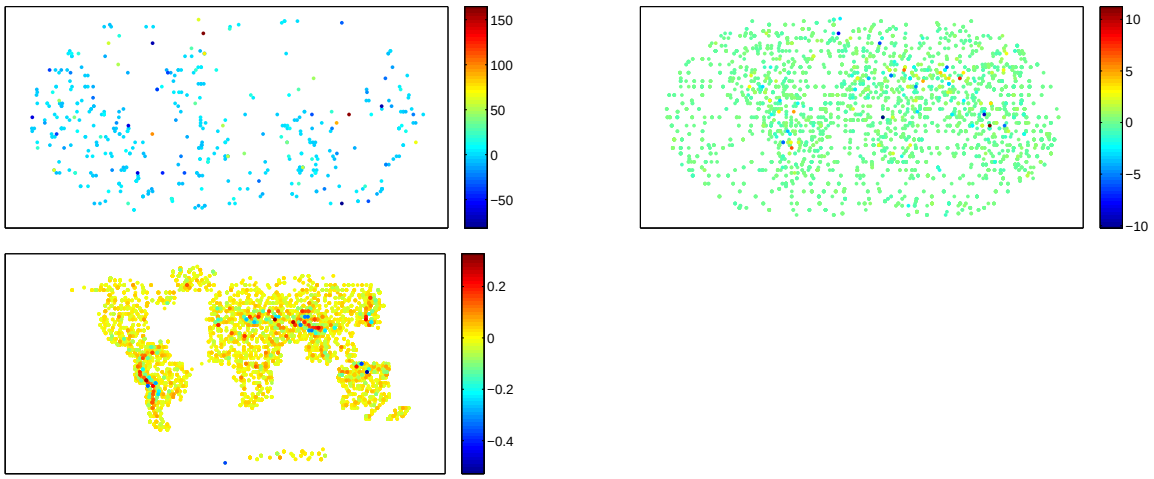

Figure 8: The centres of the wavelets chosen by the RFMP_AoS at scale $J = 0$ (top left), $J = 1$ (top right) and scale $J = 2$ (bottom) where the color indicates the value of the corresponding coefficient of the wavelet within the expansion. Note the different colorbars.

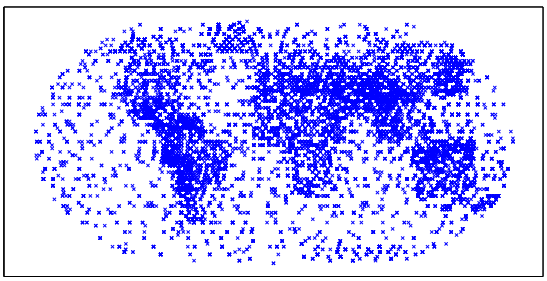

Figure 9: The centres of the 3544 wavelets of scales $J = 0, 1, 2$ which were chosen by the algorithm.

8 Conclusions and Outlook

The RFMP_AoS is able to combine trial functions with different characteristics for a sparse approximation of spherical signals. In particular, a multi-scale analysis of the signal is possible if the obtained approximation is split into the parts corresponding to the different kinds of trial functions. For example, the contribution by the spherical harmonics yields a low-scale, coarse approximation of the signal. More localized parts of the signal are represented by scaling functions (or spline basis functions). Furthermore, the wavelet functions yield strongly localized details in the signal. Another feature of the new method is the fact that strongly irregular data grids do not represent a numerical difficulty — this is a feature which is also known for spherical spline interpolation. In this respect, the new algorithm combines the advantages of previously rather competitive approaches such as spherical harmonics expansion, spline interpolation/approximation and wavelet analysis. Moreover, since the algorithm is iterative and

avoids the resolution of a system of linear equations (as it is necessary for the other methods — either directly for the interpolation or indirectly due to the necessity of a quadrature rule), numerical instabilities in the case of very large data sets are not likely.
Further improvements of the algorithm are possible. For example, the numerical tests revealed that trial functions are often chosen several times. This can be avoided if the coefficients of the chosen trial functions are appropriately adapted. We will address this issue in a forthcoming publication. Moreover, further numerical tests with much larger data sets in the dimension of the degrees of freedom of current gravity models like EGM2008 (approximately $4 \cdot 10^6$ degrees of freedom) are necessary to verify the potential of the method for practical applications in the geosciences and to investigate the possible sparsity effect in such cases. In this context, an efficient parallelization on a high-performance computing cluster appears to be useful.

9 Acknowledgments

The authors gratefully acknowledge the financial support by the German Research Foundation (DFG), projects MI 655/2-2 and MI 655/7-1, as well as the Faculty for Science and Engineering of the University of Siegen.

References

[1] Antoine JP, Demanet L, Jacques L, Vandergheynst P (2002) Wavelets on the sphere: implementations and approximations. Appl. Comput. Harmon. Anal. 13:177-200.

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

[3] 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. 1:167-20.

[4] Brauchart JS, Dick J (2013) A characterization of Sobolev spaces on the sphere and an extension of Stolarsky's invariance principle to arbitrary smoothness. Constr. Approx. 38:397-445.

[5] Brauchart JS, Hesse K (2007) Numerical integration over spheres of arbitrary dimension. Constr. Approx. 25:41-71.

[6] Chambodut A, Panet I, Mandea M, Diament M, Holschneider M, Jamet O (2005) Wavelet frames: An alternative to spherical harmonic representation of potential fields. Geophys. J. Int. 163:875-899.

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

[8] Davis P (1963) Interpolation and approximation. Blaisdell Publishing Company, Waltham.

[9] 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.

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

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

[12] Fischer D, Michel V (2013) Inverting GRACE gravity data for local climate effects. Journal of Geodetic Science, 3:151-162.

[13] Freeden W (1981) On approximation by harmonic splines. Manuscr. Geod. 6:193-244.

[14] Freeden W (1981) On spherical spline interpolation and approximation. Math. Methods Appl. Sci. 3:551-575.

[15] Freeden W (1999) Multiscale modeling of spaceborne geodata. BG Teubner, Stuttgart, Leipzig.

[16] Freeden W, Gerhards C (2010) Poloidal and toroidal field modeling in terms of locally supported vector wavelets. Math. Geosci. 42:817-838.

[17] Freeden W, Gerhards C (2013) Geomathematically oriented potential theory. CRC Press, Boca Raton.

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

[19] Freeden W, Gutting M (2013) Special functions of mathematical (geo-)physics. Birkhäuser, Basel

[20] Freeden W, Michel V (2004) Orthogonal zonal, tesseral and sectorial wavelets on the sphere for the analysis of satellite data. Adv. Comp. Math. 21:181-217.

[21] Freeden W, Schreiner M (2007) Biorthogonal locally supported wavelets on the sphere based on zonal kernel functions. J. Fourier Anal. Appl. 13:693-709.

[22] Freeden W, Schreiner M (2007) Orthogonal and non-orthogonal multiresolution analysis, scale discrete and exact fully discrete wavelet transform on the sphere. Constr. Appr. 14:493-515.

[23] Freeden W, Schreiner M (2009) Spherical functions of mathematical geosciences, a scalar, vectorial, and tensorial setup. Springer, Berlin.

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

[25] Gerhards C (2012) Locally supported wavelets for the separation of spherical vector fields with respect to their sources. Int. J. Wavelets Multiresolut. Inf. Process. 10:1250034 (26 pp.).

[26] Holschneider M (1996) Continuous wavelet transforms on the sphere. J. Math. Phys. 37:4156-4165.

[27] Holschneider M, Chambodut A, Mandea M (2003) From global to regional analysis of the magnetic field on the sphere using wavelet frames. Phys. Earth Planet In. 135:107-124.

[28] Holschneider M, Iglewska-Nowak I (2007) Poisson wavelets on the sphere. J. Fourier Anal. Appl. 13:405-419.

[29] Laín Fernández N (2007) Optimally space-localized band-limited wavelets on $\mathbb{S}^{q-1}$. J. Comput. Appl. Math. 199:68-79.

[30] Laín Fernández N, Prestin J (2006) Interpolatory band-limited wavelet bases on the sphere. Constr. Appr. 23:79-101.

[31] Le Gia QT, Mhasar HN (2009) Localized linear polynomial operators and quadrature formulas on the sphere. SIAM J. Num. Anal. 47:440-466.

[32] Le Gia QT, Sloan IH, Wendland H (2012) Multiscale approximation for functions in arbitrary Sobolev spaces by scaled radial basis functions on the unit sphere. Appl. Comput. Harmon. Anal. 32:401-412.

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

[34] Michel V (2005) Optimally localized approximate identities on the 2-sphere. Numer. Func. Anal. Opt. 32:877-903.

[35] Michel V (2013) Lectures on constructive approximation — Fourier, spline, and wavelet methods on the real line, the sphere, and the ball. Birkhäuser, New York.

[36] Michel V (2013) RFMP — An iterative best basis algorithm for inverse problems in the geosciences. In: Handbook of geomathematics (W. Freeden, M.Z. Nashed and T. Sonar, eds.), 2nd edition, accepted, 2013.

[37] Müller C (1966) Spherical harmonics. Springer, Berlin.

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

[39] Plattner A, Simons FJ (2013) Potential-field estimation using scalar and vector Slepian functions at satellite altitude. Preprint.

[40] Plattner A, Simons FJ (2014) Spatiospectral concentration of vector fields on a sphere. Appl. Comput. Harmon. Anal. 36:1-22.

[41] Reuter R (1982) Integralformeln der Einheitssphäre und harmonische Splinefunktionen. PhD Thesis, RWTH Aachen.

[42] Schreiner M (1997) Locally supported kernels for spherical spline interpolation. J. Approx. Theory 89:172-194.

[43] Schröder P, Sweldens W (1995) Spherical wavelets: efficiently representing functions on the sphere. In: SIGGRAPH'95 proceedings of the 22nd annual conference on computer graphics and interactive techniques, pp. 161-172. ACM, New York.

[44] Simons FJ (2010) Slepian functions and their use in signal estimation and spectral analysis. In: Handbook of geomathematics (W. Freeden, M.Z. Nashed and T. Sonar, eds.), pp. 891-923. Springer, Heidelberg.

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

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

[47] Svensson SL (1984) Finite elements on the sphere. J. Approx. Theory 40:246-260.

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

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

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

Siegen Preprints on Geomathematics

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

1. P. Berkel, D. Fischer, V. Michel: *Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation*, 2010.
2. M. Akram, V. Michel: *Regularisation of the Helmholtz decomposition and its application to geomagnetic field modelling*, 2010.
3. V. Michel: *Optimally Localized Approximate Identities on the 2-Sphere*, 2010.
4. N. Akhtar, V. Michel: *Reproducing Kernel Based Splines for the Regularization of the Inverse Spheroidal Gravimetric Problem*, 2011.
5. D. Fischer, V. Michel: *Sparse Regularization of Inverse Gravimetry - Case Study: Spatial and Temporal Mass Variations in South America*, 2011.
6. A.S. Fokas, O. Hauk, V. Michel: *Electro-Magneto-Encephalography for the three-Shell Model: Numerical Implementation for Distributed Current in Spherical Geometry,* 2011.
7. M. Akram, I. Amina, V. Michel: *A Study of Differential Operators for Complete Orthonormal Systems on a 3D Ball*, 2011.
8. D. Fischer, V. Michel: *How to combine spherical harmonics and localized bases for regional gravity modelling and inversion*, 2012.
9. D. Fischer, V. Michel: *Inverting GRACE gravity data for local climate effects,* 2012.
10. V. Michel, R. Telschow: *A Non-linear Approximation Method on the Sphere, 2014.*

Geomathematics Group Siegen
Prof. Dr. Volker Michel

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

UNIVERSITÄT SIEGEN