Handbook of Mathematical Geodesy
Geosystems Mathematics, 605–639

Parameter Choices for Fast Harmonic Spline Approximation

Martin Gutting

Abstract. The approximation by harmonic trial functions allows the construction of the solution of boundary value problems in geoscience where the boundary is often the known surface of the Earth itself. Using harmonic splines such a solution can be approximated from discrete data on the surface. Due to their localizing properties regional modeling or the improvement of a global model in a part of the Earth's surface is possible with splines.

Fast multipole methods have been developed for some cases of the occurring kernels to obtain a fast matrix-vector multiplication. The main idea of the fast multipole algorithm consists of a hierarchical decomposition of the computational domain into cubes and a kernel approximation for the more distant points. This reduces the numerical effort of the matrix-vector multiplication from quadratic to linear in reference to the number of points for a prescribed accuracy of the kernel approximation. In combination with an iterative solver this provides a fast computation of the spline coefficients.

The application of the fast multipole method to spline approximation which also allows the treatment of noisy data requires the choice of a smoothing parameter. We summarize several methods to (ideally automatically) choose this parameter with and without prior knowledge of the noise level.

Keywords. Spline approximation, fast multipole methods.

1. Introduction

Spherical splines have been developed by Freeden [38] and independently by Wahba [134] for interpolation and smoothing/approximation problems on the sphere and have been generalized to harmonic splines by Freeden [37, 39, 40]. These harmonic splines can be used for interpolation/approximation on regular surfaces, but in particular for the solution of boundary value problems where the boundary is a regular surface. Convergence theorems exist for both spherical splines and harmonic splines (cf. [41, 42, 43, 44]), and the different types of spline spaces and their

reproducing kernels have been investigated (cf. [49] and the references therein). In geosciences they have found many applications (see, e.g., [45, 46, 48, 107] and the references therein).

Splines lead to a system of linear equations which in case of harmonic splines has to be densely populated. This makes the solution expensive in terms of the numerical effort. Iterative solvers require fast summation methods corresponding to the reproducing kernels of these splines to be truly efficient. On the sphere there are several possible ways to achieve a fast summation (the problem that is considered determines which method should be preferred): spherical panel clustering (cf. [45, 52] and the references therein), spherical FFT for gridded data points or spherical NFFT for non-equispaced data (cf. [116, 82]).

[61, 62, 63, 124] have introduced the fast multipole method (FMM) in two and three dimensions for fast evaluation of potentials corresponding to the Laplace operator (generalizations to further operators have been introduced later). Since such potentials are closely related to certain reproducing kernels of our splines, the FMM allows fast summation of harmonic splines as well as spherical splines. Such a combination is used in [57] to solve problems of satellite geodesy with harmonic splines corresponding to the singularity kernel. We have extended this to the Abel–Poisson kernel and use the accelerated version of the FMM that was first introduced in [24, 64]. This approach has also been applied to the oblique boundary value problem of potential theory in [67, 68].

We consider the following (generalized) interpolation problems:

Problem 1.1 (Interpolation on a regular surface). Let Σ be a $C^{(0)}$-regular surface (see Definition 2.1 below for details). Let a finite set of points $\{x_1, \dots, x_N\} \subset \Sigma$ on the surface and data F_i, $i = 1, \dots, N$ corresponding to these points be given. The aim is to find a function F in a function space of choice such that $F(x_i) = F_i$, $i = 1, \dots, N$.

If the data F_i are error-affected and strict interpolation is no longer desirable and the interpolation conditions are reduced to $F(x_i) \approx F_i$, $i = 1, \dots, N$, and F has to minimize a functional that balances closeness to the data and smoothness of F, usually with one (or several) parameters.

Closely related to this interpolation/approximation problem is the discrete version of the Dirichlet boundary value problem which requires only the values of the boundary function in a finite set of points on the surface. This is also called a generalized interpolation problem.

Problem 1.2. Let Σ be a $C^{(k)}$-regular surface with $k \geq 2$ (see again Definition 2.1 below for details). Let $\{x_1, \dots, x_N\} \subset \Sigma$ be a discrete set of N points on the surface. For each point x_i let $F_i = U(x_i)$ be given, where $i = 1, \dots, N$.
The task is to determine the potential $U \in C^{(0)}\left(\overline{\Sigma_{\text{ext}}}\right) \cap C^{(2)}\left(\Sigma_{\text{ext}}\right)$ which is harmonic in Σ_{ext}, i.e., the exterior of the surface Σ, and regular at infinity, i.e., for

$|x| \to \infty$,

$$|U(x)| = \mathcal{O}\left(|x|^{-1}\right), \tag{1}$$

$$|\nabla U(x)| = \mathcal{O}\left(|x|^{-2}\right), \tag{2}$$

or an approximation U_N to it which fits the data, i.e., for $i = 1, \ldots, N$,

$$U_N(x_i) = F_i = U(x_i). \tag{3}$$

As before the interpolation conditions (3) are relaxed in case of error-affected data/measurements.

The outline of this paper is as follows: Section 2 summarizes the theory of harmonic splines and spline approximation. In Section 3 we establish the connection between harmonic splines and the sums that can be computed by the fast multipole method, we introduce the adaptive construction of the decomposition of the computational domain and provide our version of the fast multipole algorithm for harmonic splines. Section 4 gives an overview of suitable methods to choose the smoothing parameter of the approximating splines if the data are afflicted with (stochastic) noise.

2. Preliminaries

Spherical harmonics, which we denote by $Y_{n,m}$ (with degree $n \in \mathbb{N}_0$, order $m = -n, \ldots, n$), are known to form a complete orthonormal basis of the space $L^2(\mathbb{S}^2)$ of square integrable functions on the unit sphere $\mathbb{S}^2$ (see, e.g., [30, 47, 130]). The spherical harmonics $\{Y_{n,m}\}_{n\in\mathbb{N}_0, m=-n,\ldots,n}$ also form a closed system in $C(\mathbb{S}^2)$ and are closed and complete in $L^2(\mathbb{S}^2)$. This allows the representation of square-integrable functions on any sphere $\mathbb{S}^2_R$ of radius $R > 0$ by their Fourier series, where the Fourier coefficients of $F \in L^2(\mathbb{S}^2_R)$ are denoted by

$$F^\wedge(n,m) = \int_{\mathbb{S}^2_R} F(x)\, \frac{1}{R} Y_{n,m}\left(\tfrac{x}{R}\right)\, dS_R(x). \tag{4}$$

2.1. Regular surfaces and Runge spheres

Due to the Runge–Walsh approximation theorem, we can use functions which possess a larger domain of harmonicity to approximate the solution of a problem which requires harmonicity only outside the Earth's surface (see [45, 48] for an extensive introduction of this technique). Harmonic splines as introduced in [37, 39, 40, 125] are constructed in such a way that they are subspaces of the space of harmonic functions on a sphere situated inside the Earth, the so-called Runge (or Krarup) sphere (see [109]).

The Earth's surface is considered to be regular as by the following definition.

Definition 2.1. A $C^{(k)}$-regular surface $\Sigma \subset \mathbb{R}^3$ is a surface in $\mathbb{R}^3$ which has to fulfill the following properties:

(i) Σ divides $\mathbb{R}^3$ into the interior Σ_{int} and the exterior Σ_{ext}, where Σ_{int} is a bounded region and Σ_{ext} is an unbounded region.

(ii) The origin is contained in Σ_{int}.
(iii) Σ is closed (and therefore compact) and free of double points.
(iv) Σ is a $C^{(k)}$-surface, i.e., for each $x \in \Sigma$ there exists a neighborhood $U \subset \mathbb{R}^3$ of x such that $\Sigma \cap U$ possesses a $C^{(k)}$-parametrization.

We can also define $C^{(k,\lambda)}$-regular surfaces $\Sigma \subset \mathbb{R}^3$ with $\lambda \in (0, 1)$ as a $C^{(k)}$-regular surface where every point $x \in \Sigma$ possesses a neighborhood U such that $\Sigma \cap U$ can locally be parameterized by a k-times λ-Hölder continuously differentiable parametrization. Such surfaces are required for oblique derivative boundary value problems as in [68], but not here. Note that any sphere $\mathbb{S}_R^2$ of radius $R > 0$ is obviously a $C^{(\infty)}$-regular surface.

For regular surfaces Σ we can define Runge spheres (cf. [45, 46, 48], see also [109] where it is called Krarup sphere).

Definition 2.2. The *Runge sphere* $\mathbb{S}_R^2$ is a sphere of radius R around the origin such that the exterior of the Runge sphere, i.e., $\mathbb{S}_{R,\text{ext}}^2$, contains the exterior of the regular surface Σ, i.e., $\overline{\Sigma_{\text{ext}}} \subset \mathbb{S}_{R,\text{ext}}^2$.

2.2. Sobolev spaces

Now we briefly introduce the Sobolev spaces of the form $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ using the Runge sphere $\mathbb{S}_R^2$. For more details, the reader is referred to [45, 48, 107] and the references in these books.

Definition 2.3. Let $\{A_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}$ be a sequence which satisfies the summability condition

$$\sum_{n=0}^{\infty} \frac{2n+1}{4\pi A_n^2} < \infty. \tag{5}$$

The Sobolev space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ is defined by

$$\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right) = \overline{\mathcal{E}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)}^{\|\cdot\|_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)}}, \tag{6}$$

where $\mathcal{E}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right) \subset C^{(\infty)}\left(\overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ is the set of all functions that are harmonic in $\mathbb{S}_{R,\text{ext}}^2$, infinitely often differentiable on the Runge sphere $\mathbb{S}_R^2$ and regular at infinity (i.e., (1) and (2) hold) and whose Fourier coefficients $F^\wedge(n,m)$ with respect to $L^2(\mathbb{S}_R^2)$ (as defined in (4)) fulfill

$$\|F\|_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^2 \left(F^\wedge(n,m)\right)^2 < \infty\,. \tag{7}$$

$\mathcal{H}$ is a Hilbert space with the inner product defined by

$$\langle F, G\rangle_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^2\, F^\wedge(n,m) G^\wedge(n,m) \tag{8}$$

for $F, G \in \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$.

It is well known (cf. [45, 107] and the references therein) that such a space possesses a so-called reproducing kernel (see [3] for an overview on reproducing kernels in general).

Definition 2.4. Let U be a non-empty set and $(X, \langle\cdot,\cdot\rangle_X)$ be a separable Hilbert space of real-valued functions on U. Let $\{B_n\}_{n\in\mathbb{N}_0}$ be a complete orthonormal system in $(X, \langle\cdot,\cdot\rangle_X)$. Any function $K : U \times U \longrightarrow \mathbb{R}$ of the form

$$K(x,y) = \sum_{n=0}^{\infty} K^{\wedge}(n) B_n(x)\, B_n(y) \tag{9}$$

with $x, y \in U$ and $K^{\wedge}(n) \in \mathbb{R}$ for $n \in \mathbb{N}_0$ is called an *X-product kernel* (briefly an X-kernel).

An X-kernel $K(\cdot,\cdot) : U \times U \longrightarrow \mathbb{R}$ is called a *reproducing kernel* (or shortly *repro-kernel*) for $(X, \langle\cdot,\cdot\rangle_X)$ if:

(i) $K(x,\cdot) \in X$ for all $x \in U$.
(ii) $\langle K(x,\cdot), F\rangle_X = F(x)$ for all $x \in U$ and all $F \in X$.

If there exists such a repro-kernel in X, then X is called a reproducing kernel Hilbert space and the repro-kernel is unique (cf. [3]). In the space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}^2_{R,\mathrm{ext}}}\right)$ with a summable sequence $\{A_n\}$ the repro-kernel (9) can be represented by its expansion in Legendre polynomials due to the well-known addition theorem for spherical harmonics:

$$K_{\mathcal{H}}(x,y) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi A_n^2} \frac{1}{|x||y|} \left(\frac{R^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right). \tag{10}$$

2.3. Harmonic splines

We use the reproducing kernels of Section 2.2 to define harmonic splines.

Definition 2.5. Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\} \subset \mathcal{H}^*$ be a set of N linearly independent bounded linear functionals on the reproducing kernel Hilbert space $\mathcal{H}$. Then any function S of the form

$$S = \sum_{i=1}^{N} a_i \mathcal{L}_i K_{\mathcal{H}}(\cdot,\cdot) \tag{11}$$

with a set of so-called *spline coefficients* $\{a_1, \ldots, a_N\} \subset \mathbb{R}$ is called an *$\mathcal{H}$-spline* relative to $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$. The function space of all $\mathcal{H}$-splines relative to $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ is denoted by $\mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \ldots, \mathcal{L}_N)$.

By construction any $\mathcal{H}$-spline is a harmonic function. The interpolating spline S^F for the function $F \in \mathcal{H}$ has to fulfill the interpolation conditions

$$\mathcal{L}_i S^F = \mathcal{L}_i F \quad \text{for } i = 1, \ldots, N. \tag{12}$$

The interpolation conditions (12) can be rewritten as a system of linear equations for the spline coefficients a_i:

$$\sum_{i=1}^{N} a_i \mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot,\cdot) = \mathcal{L}_j F, \quad j = 1, \dots, N, \tag{13}$$

whose corresponding matrix possesses the entries $\mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot,\cdot)$ and is symmetric and positive definite (for linear functionals $\mathcal{L}_1, \dots, \mathcal{L}_N \in \mathcal{H}^*$ which are linearly independent).

In this paper, we consider only evaluation functionals $\mathcal{L}_x$, i.e., $\mathcal{L}_x F = F(x)$ where $x \in \overline{\Sigma_{\text{ext}}}$. Furthermore, $\mathcal{L}_1, \dots, \mathcal{L}_N$ are given by $\mathcal{L}_i F = F(x_i)$ where $x_i \in \Sigma$. For other types of functionals see [43, 44] or [67, 68] (for the case of oblique derivatives). In the following theorem we summarize the properties of $\mathcal{H}$-splines.

Theorem 2.6. *Let $F \in \mathcal{H}$ and let $\{\mathcal{L}_1, \dots, \mathcal{L}_N\} \subset \mathcal{H}^*$. Then the $\mathcal{H}$-spline interpolation problem with the interpolation conditions* (12) *is uniquely solvable and its solution $S^F \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$ possesses the following properties:*

(i) *S^F is the $\mathcal{H}$-orthogonal projection of F onto $\mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$.*

(ii) $\left\|S^F\right\|_{\mathcal{H}} \leq \|F\|_{\mathcal{H}}$.

(iii) *If $G \in \mathcal{H}$ also satisfies the interpolation conditions* (12)*, then the first minimum property holds:*

$$\|G\|_{\mathcal{H}}^2 = \left\|S^F\right\|_{\mathcal{H}}^2 + \left\|G - S^F\right\|_{\mathcal{H}}^2, \tag{14}$$

i.e., S^F is the interpolating function of F in $\mathcal{H}$ with minimal norm.

(iv) *If $S \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$ and $G \in \mathcal{H}$ also satisfies the interpolation conditions* (12)*, then the second minimum property holds:*

$$\|S - G\|_{\mathcal{H}}^2 = \left\|S^F - G\right\|_{\mathcal{H}}^2 + \left\|S - S^F\right\|_{\mathcal{H}}^2. \tag{15}$$

For the proof and for further details on splines, the reader is referred to [37, 43, 45, 51, 107] and the references therein.

Example 2.7. The choice $A_n = h^{-\frac{n}{2}}$, $h \in (0,1)$, fulfills (5) and provides us with the reproducing kernel called *Abel–Poisson kernel* which is given by

$$K_{\mathcal{H}}(x,y) = \frac{1}{4\pi} \frac{|x|^2|y|^2 - h^2R^4}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x \cdot y)^{\frac{3}{2}}}. \tag{16}$$

Example 2.8. The sequence $A_n = (n + \frac{1}{2})^{\frac{1}{2}} h^{-\frac{n}{2}}$, $h \in (0,1)$, also satisfies (5) and leads to the *singularity kernel* given by

$$K_{\mathcal{H}}(x,y) = \frac{1}{2\pi} \frac{1}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x \cdot y)^{\frac{1}{2}}}. \tag{17}$$

In [48] the existence of approximations fulfilling interpolation conditions is shown by the Runge–Walsh approximation theorem and an extension of Helly's theorem (cf. [138]). Convergence results for harmonic splines (cf. [43, 44]) can be

derived that show the convergence to the solution of the Dirichlet boundary value problem for an increasing density of data points, i.e., if the largest data gap goes to zero (cf. [43]).

We consider two specific types of splines (using Abel–Poisson and singularity kernels) and propose the fast multipole method to quickly compute the sums $\sum a_i K_{\mathcal{H}}(x_i, y_j)$ for many points in Section 3. This can be used to solve the systems of linear equations (13) that occur in the solution of the interpolation problems using harmonic splines.

2.4. Spline approximation

For noisy data, i.e., $F_i = U(x_i) + \delta_i$, $i = 1, \ldots, N$, where the noise δ_i is modeled by some stochastic process, e.g., white noise (see Section 4.1 for details), in Problem 1.2, it makes no sense to compute an interpolation problem. We look for an approximation to U which can be interpreted as a smoothing of the data (see [38, 51, 50, 135] for the spherical spline approximation, [37, 45] for the case of harmonic spline approximation). Minimizing the following functional

$$\mu(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathcal{L}_i S - F_i) C_{ij} (\mathcal{L}_j S - F_j) + \beta \|S\|_{\mathcal{H}} \tag{18}$$

in the reproducing kernel Hilbert space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}^2_{R,\text{ext}}}\right)$ yields the desired smoothed approximation of the data. $C = (C_{ik}) \in \mathbb{R}^{N \times N}$ denotes a positive definite matrix which allows us to include covariance information on the data if available. $\beta > 0$ is a constant smoothing parameter which balances closeness to the data and smoothing. The following theorem of [45, 107] (see also the references therein) summarizes the existence and uniqueness of a spline approximation.

Theorem 2.9. *Let F_i, $i = 1, \ldots, N$, correspond to a set of linearly independent bounded linear functionals $\mathcal{L}_1, \ldots, \mathcal{L}_N \in \mathcal{H}^*$.*
Then there exists a unique element $S \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \ldots, \mathcal{L}_N)$ such that

$$\mu(S) \leq \mu(F) \quad \text{for all } F \in \mathcal{H} \tag{19}$$

and $\mu(S) = \mu(F)$ if and only if $S = F$. This element is called the smoothing spline *or* approximating spline. *Its spline coefficients a_i, $i = 1, \ldots, N$, are uniquely determined by the system of linear equations*

$$\sum_{i=1}^{N} a_i \left(\mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot, \cdot) + \beta (C^{-1})_{ij} \right) = \mathcal{L}_j F, \quad j = 1, \ldots, N. \tag{20}$$

The matrix in (20) corresponds to the sum of the matrix in (13) and βC^{-1}. It is still positive definite. If C is the identity matrix, there is only the one smoothing parameter β. Using a diagonal matrix as C it is possible to introduce weights for the data F_i and include additional information on the noise of the data. The choice of the smoothing parameter(s) can be interpreted as the application of a parameter choice method in the regularization theory of ill-posed problems (see Section 4).

3. The fast multipole method for splines

The interpolation conditions (12) as well as the minimization of the smoothing functional (18) lead to a system of linear equations (13), or (20) respectively, with a dense matrix whose size is the number of data points. This matrix can be large and the solution of the corresponding system of linear equations becomes difficult.

Reproducing kernels of a reproducing kernel Hilbert space $\mathcal{H}$ defined by the summable sequence $\{A_n\}$ can be expanded in terms of Legendre polynomials as in (10). The singularity kernel (17) and the Abel–Poisson kernel (16) possess a representation as an elementary function and both kernels are closely related to the single pole $\frac{1}{|x-y|}$ by the Kelvin transform. Therefore, we can use the fast multipole method (FMM), which has been introduced by Greengard [61], Greengard and Rokhlin [62, 63], and Rokhlin [124], for the fast summation of harmonic splines, i.e., of the sum $\sum a_i K_{\mathcal{H}}(x_i, \cdot)$. This corresponds to the matrix-vector products occurring in an iterative solver for (13) or (20).

The FMM creates a hierarchical subdivision of the computational domain into nested cubes that are organized in an octtree data structure. Instead of single point interaction the cubes summarize the part of the kernel sum corresponding to the points they contain and interact with other cubes via the coefficients of truncated inner/outer harmonics expansions. This kernel approximation is applied as often as possible and on the coarsest possible level of the tree data structure. Direct evaluation is used only for the closest cubes where the approximation is not accurate enough. The algorithm has been improved several times to increase its efficiency (cf., e.g., [24, 64, 137]). We summarize our implementation and show the application of the FMM to harmonic splines (see [67, 69] for a more detailed analysis).

3.1. Kelvin transform of reproducing kernels

The Kelvin transform yields the connection between the kernels (17) and (16) and the fundamental solution of the Laplace equation, i.e., the single pole. The transform can be seen as a reflection on a sphere around the origin and we choose the Runge sphere $\mathbb{S}^2_R$ of Definition 2.2 for this (see also text books on potential theory, e.g., [46, 83]).

Definition 3.1. Let $\Gamma \subseteq \mathbb{R}^3$ be a domain, $W : \Gamma \longrightarrow \mathbb{R}$ a function. Let the reflection of Γ on the sphere $\mathbb{S}^2_R$ be given by

$$\Gamma^{\mathrm{KT}} = \left\{ x^{\mathrm{KT}} \in \mathbb{R}^3 : \frac{R^2}{|x^{\mathrm{KT}}|^2} x^{\mathrm{KT}} = x \in \Gamma \right\}. \tag{21}$$

The function

$$W^{\mathrm{KT}} : \Gamma^{\mathrm{KT}} \longrightarrow \mathbb{R},$$

$$x^{\mathrm{KT}} \mapsto W^{\mathrm{KT}}(x^{\mathrm{KT}}) = \frac{R}{|x^{\mathrm{KT}}|} W\left(\frac{R^2}{|x^{\mathrm{KT}}|^2} x^{\mathrm{KT}} \right) = \frac{R}{|x^{\mathrm{KT}}|} W(x), \tag{22}$$

is called the *Kelvin transform* of W with respect to the sphere of radius R.

The Kelvin transform is applied to the reproducing kernels with respect to one argument (the other is kept fixed). The Kelvin transform $K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}})$ of the singularity kernel (17) can be computed, e.g., by its expansion

$$\begin{aligned} K_{\mathcal{H}}(x,y) &= \sum_{n=0}^{\infty} \frac{h^n}{n+\frac{1}{2}} \frac{2n+1}{4\pi|x||y|} \left(\frac{R^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right) \\ &= \frac{1}{2\pi|y|} \sum_{n=0}^{\infty} \frac{(h|y^{\mathrm{KT}}|)^n}{|x|^{n+1}} P_n\left(\frac{x}{|x|}\cdot\frac{y^{\mathrm{KT}}}{|y^{\mathrm{KT}}|}\right) \\ &= \frac{1}{2\pi|y|}\frac{1}{|x-hy^{\mathrm{KT}}|} = \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}), \end{aligned} \tag{23}$$

where $y^{\mathrm{KT}} = \frac{R^2}{|y|^2} y$ and

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R}\frac{1}{|x-hy^{\mathrm{KT}}|}. \tag{24}$$

The Kelvin transform $K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}})$ of the Abel–Poisson kernel (16) is given by

$$\begin{aligned} K_{\mathcal{H}}(x,y) &= \frac{1}{4\pi} \frac{|x|^2|y|^2 - h^2R^4}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x\cdot y)^{\frac{3}{2}}} \\ &= \frac{|y^{\mathrm{KT}}|}{R}\frac{1}{4\pi R}\frac{|x|^2 - h^2|y^{\mathrm{KT}}|^2}{|x-hy^{\mathrm{KT}}|^3} = \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}), \end{aligned} \tag{25}$$

which is related to (24) by

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R}\left(-x\cdot\nabla_x - \tfrac{1}{2}\mathrm{Id}\right)\frac{1}{|x-hy^{\mathrm{KT}}|}. \tag{26}$$

We summarize both (24) and (26) by use of the operator D_x such that

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R} D_x \frac{1}{|x-hy^{\mathrm{KT}}|}, \tag{27}$$

where $D_x = \mathrm{Id}$ (singularity kernel) or $D_x = -x\cdot\nabla_x - \frac{1}{2}\mathrm{Id}$ (Abel–Poisson kernel).

3.2. Adaptive decomposition of the domain

Now we consider the evaluation of the sum

$$\begin{aligned} \sum_{i=1}^{N} a_i K_{\mathcal{H}}(x_i, y) &= \sum_{i=1}^{N} a_i \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x_i, y^{\mathrm{KT}}) \\ &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{i=1}^{N} \frac{a_i}{2\pi R} D_x \left.\frac{1}{|x-hy^{\mathrm{KT}}|}\right|_{x=x_i} \end{aligned} \tag{28}$$

at the points y_j, $j = 1, \ldots, M = \mathcal{O}(N)$, which can be the points x_i as in the systems of linear equations (13) or (20). To better distinguish the points x_i and y_j we call the Kelvin transformed points $hy_j^{\mathrm{KT}} = h\frac{R^2}{|y_j|^2} y_j$ which are used in (28) *targets*.

In the beginning, a bounding cube is determined which is large enough such that it contains all points and all targets. This single cube forms level 0 of the octtree structure and is subdivided into eight equally sized cubes of half its edge length which then form level 1. Each cube is adaptively divided into nested cubes where a cube of level l has half the edge length of a cube of level $l-1$ as proposed, e.g., by Cheng et al. [24]. Points and targets are sorted into the currently available cubes. If a cube contains more than the prescribed maximal number of points or targets m, it is split into eight smaller cubes of the next level. All its points/targets are redistributed into these eight cubes. We summarize the necessary vocabulary in some definitions (see also [24, 64]).

Definition 3.2.

(a) A cube C is called *child* of the cube B if C results from a single subdivision of B which in return is named the *parent* of C.
(b) A cube that is not further subdivided is called *childless* or a *leaf*.
(c) Cubes are said to be *neighbors* if they are of the same size (same level) and share at least one boundary point. Each cube is a neighbor of itself.
(d) If two cubes are at the same level, but are no neighbors, they are called *well-separated*, i.e., between these cubes exists at least one cube of their size.

Each cube carries the relevant information about other cubes, in particular its neighbors, in four lists as suggested by Cheng et al. [24]. Figures 1 and 2 give a two-dimensional illustration.

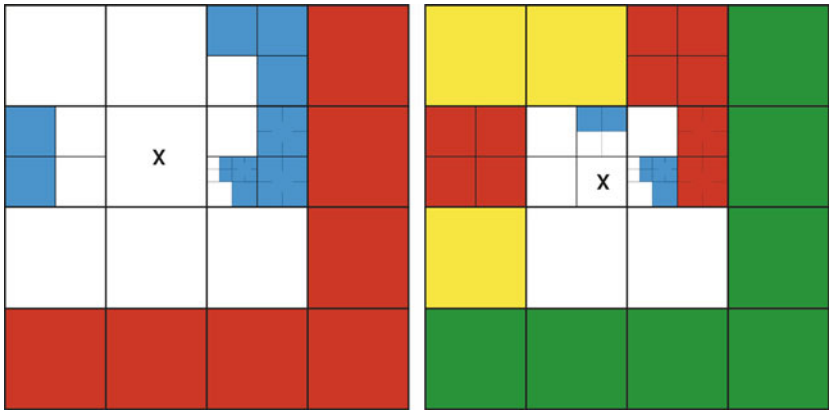

FIGURE 1. Two-dimensional illustration of an adaptive decomposition for levels 2 (left) and 3 (right). If the marked square is a leaf, the white cubes form its list 1 and the blue cubes correspond to list 3. If the marked square is not a leaf, all white and all blue cubes are treated at the next level. List 2 cubes are red and list 4 cubes are yellow. Cubes in green have been handled at coarser levels.

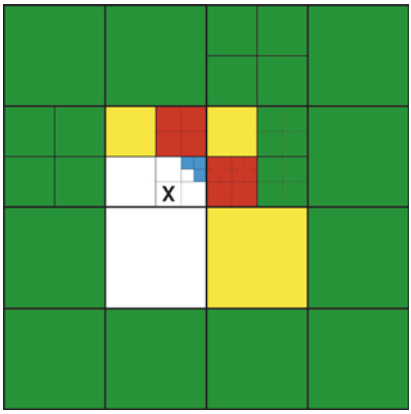

FIGURE 2. Two-dimensional illustration of an adaptive decomposition for level 4. If the marked square is a leaf, the white cubes form its list 1 and the blue cubes correspond to list 3. If the marked square is not a leaf, all white and all blue cubes are treated at the next level. List 2 cubes are red and list 4 cubes are yellow. Cubes in green have been handled at coarser levels.

Definition 3.3.

(a) In *list* 1 of the childless cube X are all childless cubes directly adjacent to X. List 1 only contains any cubes if X is a leaf. In this case it always contains at least X itself.
(b) *List* 2 of a cube X consists of all children of neighbors of the parent cube of X which are well separated from X. The cube X does not need to be childless.
(c) Children of neighbors of the leaf X (or smaller cubes descending from neighbors of X) which do not have any point in common with X form *list* 3. Their parents have to be adjacent to X. If X is not childless, then *list* 3 is empty.
(d) *List* 4 consists of childless cubes which are neighbors of the parent cube of X, but these childless cubes are not adjacent to X.

Notice the following observations:

(i) List 1 is the list of all neighbors.
(ii) All cubes in list 2 of a cube X are of the same size as X and well separated from X.
(iii) The elements of list 3 are all smaller than X and the distance between them and X is at least their side length and at most the side length of X.
(iv) List 4 of a cube X only contains cubes that are larger than X. They are separated from X by a distance that is at least the side length of X and at most their own edge length.
(v) A cube X is in list 3 of a cube Y if and only if Y is in list 4 of X.
(vi) All members of list 1 and list 4 are leaves and list 1 as well as list 3 of a cube X remain empty if X is not childless.

After finishing the adaptive construction of the octtree and sorting all points and targets into cubes, the algorithm removes childless cubes that contain neither points nor targets and are no longer required.

3.3. Single pole expansion

In addition to the decomposition of the domain, the other part of the FMM is the kernel approximation of the single pole by a truncated expansion and the use of translation theorems to shift the expansion center and to change the type of expansion. Similar to (23) we use the generating function of the Legendre polynomials P_n and the addition theorem of spherical harmonics (see, e.g., [1, 47, 102]) to expand the single pole.

$$\begin{aligned}\frac{1}{|x-y|} &= \frac{1}{|y-x_0-(x-x_0)|} \\ &= \sum_{n=0}^{\infty} \frac{|x-x_0|^n}{|y-x_0|^{n+1}} P_n\left(\frac{y-x_0}{|y-x_0|}\cdot\frac{x-x_0}{|x-x_0|}\right) \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} I^*_{n,m}(x-x_0) O_{n,m}(y-x_0), \end{aligned} \tag{29}$$

where $|y-x_0| > |x-x_0|$ for the expansion center $x_0 \in \mathbb{R}^3$. The upper star $*$ in (29) denotes the complex conjugate. Thereby, we use the (complex-valued) outer and inner harmonics for $n \in \mathbb{N}_0$, $m = -n, \ldots, n$:

$$O_{n,m}(x) = \sqrt{\frac{4\pi}{2n+1}} \frac{\sqrt{(n+m)!(n-m)!}}{|x|^{n+1}} Y_{n,m}\left(\tfrac{x}{|x|}\right), \tag{30}$$

$$I_{n,m}(x) = \sqrt{\frac{4\pi}{2n+1}} \frac{|x|^n}{\sqrt{(n+m)!(n-m)!}} Y_{n,m}\left(\tfrac{x}{|x|}\right), \tag{31}$$

where $\vartheta \in [0,\pi]$, $\varphi \in [0, 2\pi)$ are the usual spherical coordinates of $\frac{x}{|x|}$ and $Y_{n,m} : \mathbb{S}^2 \to \mathbb{C}$ with

$$Y_{n,m}(\xi) = (-1)^m \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_{n,m}(\cos(\vartheta)) e^{im\varphi}, \qquad \xi \in \mathbb{S}^2, \tag{32}$$

are complex-valued fully normalized spherical harmonics of degree n and order m. $P_{n,m} : [-1,1] \to \mathbb{R}$ are the *associated Legendre functions* with

$$P_{n,m}(t) = \frac{1}{2^n n!}(1-t^2)^{\frac{m}{2}} \frac{d^{n+m}}{dt^{n+m}}\left((t^2-1)^n\right), \quad m = 0, \ldots, n. \tag{33}$$

The symmetry relation $P_{n,-m}(t) = (-1)^m \frac{(n-m)!}{(n+m)!} P_{n,m}(t)$ extends them for negative orders (cf., e.g., [16, 30]).

Well-known translation theorems for these outer and inner harmonics allow to shift the expansion center (see, e.g., [34] for a detailed derivation).

Theorem 3.4 (Translation Theorem for Outer Harmonics). *Let $x, y \in \mathbb{R}^3$ such that $|x| > |y|$. Then the outer harmonic of degree $n \in \mathbb{N}_0$ and order $m \in \mathbb{Z}$, $-n \leq m \leq n$, at $x - y$ can be expanded in terms of inner and outer harmonics as follows*

$$O_{n,m}(x-y) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} I^*_{n',m'}(y) O_{n+n',m+m'}(x) \tag{34}$$

$$= \sum_{n'=n}^{\infty} \sum_{m'=-n'}^{n'} I^*_{n'-n,m'-m}(y) O_{n',m'}(x). \tag{35}$$

Note that in (35) we make use of the convention that $I_{n,m} = 0$ if $|m| > n$. Obviously, this infinite series as well as the infinite sum in (29) have to be truncated for the algorithm which are sources of (truncation) errors. Error estimates for the truncation errors can be found in [61, 63, 64] and the references therein. Another approach which investigates the combined error of both truncations in (29) and (35) is considered in [67].

Theorem 3.5 (Translation Theorem for Inner Harmonics). *Let $x, y \in \mathbb{R}^3$. Then the inner harmonic of degree $n \in \mathbb{N}_0$ and order $m \in \mathbb{Z}$, $-n \leq m \leq n$, at $x - y$ can be expanded in a finite sum of inner harmonics*

$$I_{n,m}(x-y) = \sum_{n'=0}^{n} \sum_{m'=-n'}^{n'} (-1)^{n'} I_{n',m'}(y) I_{n-n',m-m'}(x). \tag{36}$$

For orders with $|m| > n$ we have again by convention $I_{n,m} = 0$. Note that no truncation is necessary for this translation theorem, i.e., no truncation errors occur.

By applying (35) of Theorem 3.4 we can translate an outer harmonics expansion with expansion center x_0 such as

$$F(x) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_0}(n,m) O_{n,m}(x - x_0) \tag{37}$$

which converges uniformly for $x \in \mathbb{S}^2_{r_0,\mathrm{ext}}(x_0)$ with some $r_0 > 0$. $\mathbb{S}^2_{r_0,\mathrm{ext}}(x_0)$ denotes the exterior of the sphere of radius r_0 around x_0. The outer harmonics series resulting from the translation possesses the expansion center x_1 and the coefficients

$$F^{\wedge,\mathrm{O}}_{x_1}(n',m') = \sum_{n=0}^{n'} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_0}(n,m) I^*_{n'-n,m'-m}(x_0 - x_1). \tag{38}$$

This expansion converges uniformly for $x \in \mathbb{S}^2_{r_1,\mathrm{ext}}(x_1)$ where

$$\mathbb{S}^2_{r_1,\mathrm{ext}}(x_1) \subset \mathbb{S}^2_{r_0,\mathrm{ext}}(x_0).$$

This translation of coefficients is called *multipole to multipole translation* (M2M). Note that one can show that no further errors arise if the series in (37) has already been truncated before the translation.

By using formulation (34) of Theorem 3.4 we also find that the outer harmonics expansion with expansion center x_1 can be translated into an inner harmonics series centered around x_2 which converges uniformly for $x \in \mathbb{S}^2_{r_2,\mathrm{int}}(x_2)$ if the new ball of convergence is situated completely in $\mathbb{S}^2_{r_1,\mathrm{ext}}(x_1)$, i.e., $\overline{\mathbb{S}^2_{r_1,\mathrm{int}}(x_1)} \cap \overline{\mathbb{S}^2_{r_2,\mathrm{int}}(x_2)} = \emptyset$. The resulting coefficients of the inner harmonic expansion are

$$F^{\wedge,\mathrm{I}}_{x_2}(n',m') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_1}(n,m)(-1)^{n'+m} O^*_{n+n',m'-m}(x_2 - x_1) \tag{39}$$

and this translation is named *multipole to local translation* (M2L).

Finally, Theorem 3.5 lets us shift the expansion center of such inner harmonics expansions to the new center x_3 which possesses the coefficients

$$F^{\wedge,\mathrm{I}}_{x_3}(n',m') = \sum_{n=n'}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{I}}_{x_2}(n,m) I_{n-n',m-m'}(x_3 - x_2). \tag{40}$$

and converges uniformly for $x \in \mathbb{S}^2_{r_3,\mathrm{int}}(x_3) \subset \mathbb{S}^2_{r_2,\mathrm{int}}(x_2)$. This translation step is called *local to local translation* (L2L). For further details we refer to [67] and the references therein, in particular [34].

3.4. The fast multipole algorithm

To start the algorithm a first set of multipole expansion coefficients for each cube containing any points has to be computed. We consider only the part of the spline related to a single cube X, i.e., the kernel functions $K_{\mathcal{H}}(x_i, \cdot)$, where $x_i \in X$ and $y \in \overline{\Sigma_{\mathrm{ext}}}$:

$$F(y) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} a_i K_{\mathcal{H}}(x_i, y) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} a_i \left(\frac{|y^{\mathrm{KT}}|}{R} \frac{1}{2\pi R} D_x \frac{1}{|x - hy^{\mathrm{KT}}|} \right) \Bigg|_{x=x_i}. \tag{41}$$

We find the following expansion for $|hy^{\mathrm{KT}} - x_0| > |x_i - x_0|$, $x_i \in X$, i.e., if x_0 is the center of the cube X, the targets hy^{KT} and the cube X need to fulfill a distance requirement, i.e., targets must be contained in a well-separated cube.

$$\begin{aligned} F(y) &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{\substack{i=1 \\ x_i \in X}}^{N} \frac{a_i}{2\pi R} \left(D_x \sum_{n=0}^{\infty} \sum_{m=-n}^{n} I^*_{n,m}(x - x_0) O_{n,m}(hy^{\mathrm{KT}} - x_0) \right) \Bigg|_{x=x_i} \\ &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,O}_{x_0}(n,m) O_{n,m}(hy^{\mathrm{KT}} - x_0) \end{aligned} \tag{42}$$

where the multipole coefficients $F^{\wedge,O}_{x_0}(n,m)$ of the cube X are given by

$$F^{\wedge,O}_{x_0}(n,m) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} \frac{a_i}{2\pi R} \left(D_x I^*_{n,m}(x - x_0) \right) \Bigg|_{x=x_i}. \tag{43}$$

This first step is called *point to multipole* (P2M) step where the infinite sum in (42) has to be truncated at degree p. The degree p essentially determines the accuracy of the algorithm. The coefficients $F_{x_0}^{\wedge,O}(n,m)$ can be translated to other cubes via relations (38), (39) as well as (40) as long as the distance requirements are fulfilled by the construction of the decomposition of the domain into nested cubes.

At the end of the fast multipole cycle, i.e., after several M2M-, M2L-, L2L-translations, each cube Y possesses an inner harmonics expansion centered around the center of the cube. This expansion has to be evaluated at the targets contained by Y. This evaluation is called the *local to targets* (L2T) *step*:

$$\mathcal{L}_j F = F(y_j) = \left(\frac{|y^{\mathrm{KT}}|}{R} \sum_{n=0}^{p} \sum_{m=-n}^{n} F_{x_0}^{\wedge,I}(n,m) I_{n,m}(hy^{\mathrm{KT}} - x_0) \right) \Bigg|_{y=y_j}, \qquad (44)$$

where the variable y is hidden by $y^{\mathrm{KT}} = \frac{R^2}{|y|^2} y$.

Now we briefly summarize the fast multipole algorithm (see, e.g., [20, 24] or [67, 69] for our specific implementation).

Algorithm 3.6 (Fast Multipole Algorithm).
Input:

- A set of points $x_i \in \overline{\Sigma_{\mathrm{ext}}}$ (often $x_i \in \Sigma$), $i = 1, \ldots, N$,
- a set of coefficients a_i, $i = 1, \ldots, N$,
- the choice of the type of the reproducing kernel $K_{\mathcal{H}}$ (singularity or Abel–Poisson with the parameter h and the radius of the Runge sphere R),
- a set of evaluation points $y_j \in \overline{\Sigma_{\mathrm{ext}}}$, $j = 1, \ldots, M$, where $M = \mathcal{O}(N)$,
- the degree of the multipole expansion p,
- the maximal number of points per cube m.

Aim: compute the sum

$$F(y_j) = \sum_{i=1}^{N} a_i K_{\mathcal{H}}(x_i, y_j) \quad \text{for each } j = 1, \ldots, M. \qquad (45)$$

Initialization:

- Compute the targets $hy_j^{\mathrm{KT}} = h\frac{R^2}{|y_j|^2} y_j$, $j = 1, \ldots, M$.
- Create a bounding box that contains all points and all targets, build the adaptive octtree and sort in all points and targets. Set L as the maximum level, eliminate all empty cubes.
- Determine list 1 to list 4 of Definition 3.2. Create a list of all cubes of level l for each level $l = 0, \ldots, L$. Collect all leaves in a list.
- Allocate memory for the different expansion coefficients of each cube X: multipole expansion (coefficient vector M_X), local expansion (coefficient vector L_X).

Fast multipole cycle:

1. Generation of the multipole coefficients:
For all leaves X: P2M, i.e., compute the multipole coefficients M_X of the multipole

expansion up to degree p around the center of X from the points in X as in (43). For level $l = L-1, \ldots, 2$: M2M, i.e., translate the multipole coefficients of the children of X to X itself for all cubes X of level l via (38).

2. Interaction phase for list 4:
For level $l = 2, \ldots, L$: for all cubes X of level l: compute the expansion coefficients of an inner harmonics expansion around the center of X from the points in Y for all cubes Y of list 4 of X and add them to L_X – or use direct evaluation of the kernel sum corresponding to the points in Y to obtain the result at the targets in X if the number of targets in $X \leq p^2$ and X is a leaf.

3. Multipole to local translation:
For level $l = 2, \ldots, L$: for all cubes X of level l: use (39) to translate M_X to L_Y for all cubes Y in list 2 of X.

4. Translation of the inner harmonics expansions:
For level $l = 2, \ldots, L-1$: L2L, i.e., translate the local coefficients L_X to the children of X (if there are any) via (40) and add the resulting coefficients to L_Z where Z denotes the corresponding child of X for all cubes X of level l.

5. Evaluation of the expansions and direct interaction:
For all leaves X: L2T, i.e., evaluate the inner harmonics expansion of X at all targets in X as in (44). Store the result in F.
For all cubes Y in list 1 of X: P2T, i.e., add the direct evaluation of the kernel sum corresponding to the points in Y at the targets in X to F.
For all cubes Y in list 3 of X: evaluate the multipole expansion around the center of Y (coefficients M_Y) at the targets in X and add the results to F – or use direct evaluation of the kernel sum corresponding to the points in Y to add the result at the targets in X to F if the number of points in $Y \leq p^2$ and Y is a leaf.

6. Reverse the effects of the Kelvin transformation:
$\widetilde{F}_j = \frac{|y_j^{\mathrm{KT}}|}{R} F_j$ for $j = 1, \ldots, M$.
Return the result $\widetilde{F}$.

For the computation of the spline coefficients of the spline approximation of Section 2.4 we consider the system of linear equations (20) instead of (13). This means that we have to add $\beta \sum_{i=1}^{N} a_i (C^{-1})_{ij}$ to the matrix-vector product that is computed by the FMM. In order to keep a fast algorithm the matrix C^{-1} has to allow a fast summation method or C has to be a sparse matrix. The trivial cases where C is a diagonal matrix can also be included in the direct evaluation step of the fast multipole algorithm.

3.5. Acceleration of the translations

Newer iterations of the FMM include several ideas to reduce the numerical effort of the translations from the original $\mathcal{O}(p^4)$ to $\mathcal{O}(p^3)$ or even $\mathcal{O}(p^2)$ per translation operation. This includes the ideas of [137] (see also [24, 64]) for the multipole to multipole (M2M) and the local to local (L2L) steps using Wigner rotation

matrices (cf., e.g., [16, 25, 30, 130]). The main point is to rotate the coordinate system such that the shift direction becomes the ε^3-axis, shift there and rotate back. This reduces the numerical costs from $\mathcal{O}(p^4)$ in the M2M- and L2L-steps to $\mathcal{O}(p^3)$, since each rotation as well as the shift along the ε^3-axis requires an effort of $\mathcal{O}(p^3)$. For a detailed description we refer to [137] or [67] with all technical details of our implementation.

For the M2L translation, [64, 24] have replaced this step with *exponential translations* which are based on the numerical integration of the integral representation

$$\begin{aligned}\frac{1}{|x-y|} &= \frac{1}{2\pi}\int_0^\infty e^{-\lambda(x_3-y_3)}\int_0^{2\pi} e^{i\lambda((x_1-y_1)\cos\alpha)+(x_2-y_2)\sin\alpha)}\,d\alpha\,d\lambda \\ &= \sum_{k=1}^{s(\varepsilon)}\frac{w_k}{M_k}\sum_{j=1}^{M_k} e^{-\lambda_k(x_3-y_3)}e^{i\lambda_k((x_1-y_1)\cos\alpha_{j,k})+(x_2-y_2)\sin\alpha_{j,k})} + \mathcal{O}(\varepsilon) \quad (46)\end{aligned}$$

for points x, y whose Cartesian coordinates satisfy $1 \le x_3 - y_3 \le 4$ as well as $0 \le \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \le 4\sqrt{2}$. Details as well as integration points λ_k, weights w_k and numbers of points M_k for the trapezoidal rule applied to the inner integral can be found in [24, 64, 139]. The accuracy ε of the numerical integration is determined by the discretization parameter $s(\varepsilon) = \mathcal{O}(p)$ of the outer integral in (46). The total number of numerical integration points, i.e., the number of exponential functions and coefficients, is supposed to be $\mathcal{O}(p^2)$.

By Hobson's formula (cf. [76]) a multipole expansion of F is transformed by (46) into a series of exponentials (multipole to exponential step, briefly M2X), these exponentials can be translated efficiently by the exponential to exponential shift (X2X). Afterwards the local coefficients are computed from the exponential coefficients (X2L). The restrictions on the positions of x and y mean that the exponential translations are applicable for cubes in list 2 (see Definition 3.3) that are situated above the current cube with another cube in between.

However, by combining rotations of the multipole expansion using again the Wigner rotation matrices, the exponential translation can substitute the M2L translation for all cubes in list 2. Therefore, the list of all well-separated cubes (list 2) is split into 6 directional lists (up, down, North, South, East and West) and instead of M2L the following sequence of transformations is used: (rotation), M2X, X2X, X2L, (inverse rotation).

Each exponential shift requires numerical costs of $\mathcal{O}(p^2)$ and the rotations can be applied using $\mathcal{O}(p^3)$ operations (as do the M2X and X2L steps). Thus, this improves the performance compared to the M2L step's $\mathcal{O}(p^4)$ effort. Moreover, we can save translations by recombination (see [24, 64, 67, 69] for more on the technical details). It should also be noted that there are several symmetries in the coefficients of the exponential expansion since we are dealing with a real-valued function F. These symmetries can be used to further reduce the constant of the numerical costs (cf. [24, 64]).

3.6. Parameters and results of our FMM implementation

Here we present just a few of the parameters that we use in our implementation of the FMM with exponential translations. For more detailed investigations of our version (in particular of the recombination of exponential translations mentioned at the end of Section 3.5) we refer to [67, 68, 69].

The truncation degree p is investigated for different accuracies of the exponential translation $s(\varepsilon)$. We increase p while $s(\varepsilon)$ is kept fixed and determine when the integration error of the numerical integration in the exponential translation (46) dominates the truncation error. This leads to the choices of p for different levels of $s(\varepsilon)$ given by Table 1. Note that the kernels of Abel–Poisson type require a slightly higher degree and therefore a bit more numerical effort.

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	4	5
17	12	13
26	23	25

TABLE 1. Resulting truncation degrees p for different $s(\varepsilon)$ for the two types of kernels.

The maximal number of points or targets per cube m has a strong influence on the adaptive octtree construction and the performance of the FMM. If m is too small, there are many cubes each containing only very few points. Thus, the kernel expansion coefficients no longer combine the information of enough points to be efficient. If m is too large, there are only few cubes each with a large number of points. This means that far too often instead of kernel expansion direct interaction is used. Therefore, m can be used to balance the effort of the direct interaction and the kernel approximation. It has been suggested to choose $m = \mathcal{O}(p^{3/2})$ (cf. [24, 64]) which may serve as a guideline. Many empirical tests (cf. [67]) led us to the conclusion that the choices for m given by Table 2 provide a good performance

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	85	75
17	130	140
26	380	240

TABLE 2. Chosen maximal numbers of points m per cube for the singularity kernel and the Abel–Poisson kernel and the different error levels.

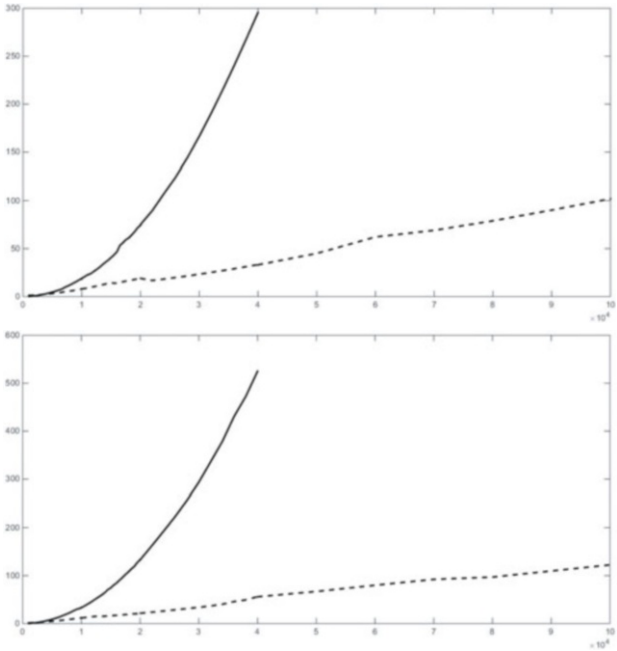

FIGURE 3. Break-even points by comparison of computation times for direct (solid line) and FMM accelerated (dashed line) computation (top: Singularity kernel, bottom: Abel–Poisson kernel), the number of points forms the abscissae.

in our implementation. Note again that there are remarkable differences between the two different types of kernels under investigation. Using these optimizations of the parameters of the FMM we compare its performance with direct computation and find the break-even points of our implementation, i.e., the minimal number of points that is necessary for our algorithm to be faster than the direct approach (see Table 3).

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	530	360
17	1,160	960
26	2,670	2,250

TABLE 3. Break-even points for the singularity kernel and the Abel–Poisson kernel.

Note that such results are always very dependent on the implementation. Our implementation turns out to be efficient even for rather small problem sizes. In general, the Abel–Poisson kernel requires some more computational time since it leads to a more difficult P2M-step. Finally, we show the linear asymptotic behavior which we expect from the FMM in Figure 3 compared to the quadratic behavior of the direct approach.

4. Parameter choice methods for spline approximation

In this section we summarize several parameter choice methods for the determination of the smoothing parameter in (18) and (20) that are known from the theory of regularization of inverse problems. For this context the reader is referred to [9, 14] where also many numerical tests for a wide range of inverse problems are presented.

We consider the problem of solving $Ax = y$ where $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator. In our case $\mathcal{X} = \mathcal{Y} = \mathbb{R}^N$ and A is the matrix of the linear system (13) (in [9, 14] this is called case C2). This is in fact not an ill-posed problem, but can be severely ill-conditioned (depending on the distribution of the data points) and as seen in Section 2.4 noisy data lead to the use of a form of Tikhonov regularization, i.e., the linear system (20) with the smoothing parameter $\beta > 0$. It should be noted that we want to use the FMM for a fast matrix-vector-multiplication, i.e., A is never given as an actual matrix. This excludes some of the parameter choice methods in order to keep reasonably low computational costs. Moreover, it should be pointed out that the use of an iterative solver (e.g., cg-method or GMRES) requires a stopping criterion and the number of iterations can be seen as another regularization parameter which needs to be chosen in some way.

The smoothing parameter $\beta > 0$ is discretized exponentially, i.e., $\beta_n = \beta_0 q_\beta^n$ with $q_\beta \in (0, 1)$ and $n \in \mathbb{N}$ (actually only $n = 1, \ldots, n_{\max}$). The use of a discrete set of regularization parameters with a fine enough resolution does not alter the behavior of most parameter choice methods. For the efficient implementation of these methods, it is useful to have a bound on the value of the optimal parameter (i.e., a maximal regularization parameter), especially if the method minimizes some function (see also [9, 14] and Section 4.3). We denote the vector of noisy data by y^δ (see Section 4.1) and $x_n^\delta \in \mathcal{X}$ is the vector of spline coefficients resulting from the minimization of the smoothing functional (18) using β_n and y^δ, i.e., x_n^δ solves the system of linear equations (20):

$$x_n^\delta = (A + \beta_n I)^{-1} y^\delta = A_n^{-1} y^\delta, \tag{47}$$

where $A_n = A + \beta_n I$. Furthermore, $x_n^0 = A_n^{-1} y$ with noise-free data $y \in \mathcal{Y}$.

4.1. Noise models

We investigate additive noise models, i.e., $y^\delta = y + \delta\xi$, where ξ is a normalized noise element and $\delta > 0$ is the noise level. The most common noise model in the classical inverse problems literature is *deterministic noise* (cf. [33]), where $\xi \in \mathcal{Y}$

with $\|\xi\| \leq 1$, so $\|y^\delta - y\| \leq \delta$. This models discretization errors, but only poorly represents random measurement errors arising in practice.

A *stochastic noise* model for a discrete data vector $y^\delta \in \mathbb{R}^N$ (see [135]) uses ξ, where the components ξ_i are i.i.d. random variables with mean $\mathbb{E}\xi_i = 0$ and variance $\mathbb{E}\xi_i^2 = 1$. δ is the standard deviation of each error component $\delta\xi_i$ and $\mathbb{E}\|y^\delta - y\|^2 = \delta^2\mathbb{E}\|\xi\|^2 = N\delta^2$. Note that for the number of data $N \to \infty$ this is unbounded. For correlated errors, $\delta\xi_i$ possesses the covariance matrix C. If known, this matrix can be used in (20). We restrict ourselves to diagonal C here. If $y^\delta = y + \delta\xi$ with $\xi \sim N(0, I)$, the noise model corresponds to Gaussian white noise. For colored noise, if the entries C_{kk} are increasing, it is called blue noise, and, if they are decreasing, it is called red noise (see [9] and the references therein for more details).

4.2. Parameter choice methods

A parameter choice method is a rule that assigns a value for the regularization/smoothing parameter. In case of a discrete set of parameters, the method selects a value for the index, which will be denoted by n_*. Parameter choice methods can be classified as three basic types by their input (see, e.g., [8, 33]):

- a priori method, i.e., n_* is a function of δ and information about x which is not known in practice. Thus, we do not discuss such methods here.
- a posteriori method, i.e., $n_* = n_*(\delta, y^\delta)$ requiring the noise level δ. If δ is not known, an estimate is used.
- data-driven method, i.e., $n_* = n_*(y^\delta)$ requiring only the data y^δ as input (sometimes called "heuristic method").

If y^δ contains stochastic noise, then n_* is a random variable. n_* is defined as either the point at which a function $F(n)$ falls below a threshold (Type 1) or the minimizer of F (Type 2). Some methods need (sensitive) tuning parameters (in particular methods of type 1).

It should be pointed out that [4] states that, for an ill-posed problem, a parameter choice rule that does not explicitly use the noise level (e.g., data-driven methods) cannot yield a regularization method such that the worst case error converges to 0 as $\delta \to 0$. This Bakushinskii veto is important for deterministic noise, but it is not really appropriate for stochastic noise (cf. [8, 15]). There are data-driven rules yielding regularization methods that converge and perform very well in practice (see also [9]).

Since $x = A^{-1}y$ is unknown (noise-free data y is not available), a practical parameter choice method must use some other known or easily computed/estimated quantities such as, e.g., the norm of the residual defined as $\|y^\delta - Ax_n^\delta\|$. In our case this is a Euclidean norm which is easily computed. Splitting the error $\|x - x_n^\delta\|$ such that

$$\|x - x_n^\delta\| \leq \|x - x_n^0\| + \|x_n^0 - x_n^\delta\|, \tag{48}$$

the first term (regularization error) is usually bounded by a decreasing function $\varphi(n)$ reflecting smoothness properties and the so-called qualification of the regularization method (see [9, 33] and the references therein). The second term (propagated noise error) on the right-hand side of (48) can often be bounded for regularization methods as

$$\|x_n^0 - x_n^\delta\| \leq \delta\varrho(n), \tag{49}$$

where ϱ is a known increasing function of n, indicating that, with less smoothing, there is more influence of the data noise (cf. [33]).

In the case of stochastic noise, the risk, i.e., the expected squared error $\mathbb{E}\|x - x_n^\delta\|^2$, is considered. For noise with zero mean, instead of (48), the risk can be decomposed exactly into a sum of squared bias $\|x - x_n^0\|^2$ and variance terms $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, i.e.

$$\mathbb{E}\|x - x_n^\delta\|^2 = \|x - x_n^0\|^2 + \mathbb{E}\|x_n^0 - x_n^\delta\|^2. \tag{50}$$

The squared bias can be bounded as before and, under suitable assumptions, the variance can be expressed as $\delta^2\varrho^2(n)$ for some increasing function $\varrho(n)$. For white noise, our Tikhonov regularized solution (47) has variance

$$\delta^2\varrho^2(n) = \delta^2\mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2\sum_k \frac{1}{(\sigma_k + \beta_n)^2}, \tag{51}$$

where σ_k are the singular values of A (in our case they are eigenvalues of A). A much more detailed discussion of the above errors (including, e.g., minimax results) in various situations can be found in [5, 9, 23, 26, 33, 78, 95, 103, 105]. For some methods, there are stronger results involving oracle inequalities (see [12, 19, 21, 22]), which provide, for any noise level, a bound on the risk $\mathbb{E}\|x - x_n^\delta\|^2$ relative to the smallest possible value of the risk, and allow the classification of methods as asymptotically optimal.

4.3. Maximal regularization parameter

Some parameter choice methods, e.g., the balancing principle (Section 4.4.2), require a maximal index $n_{\max}$ as essential input in the algorithm itself. [9] suggest to define the *maximal index* as

$$n_{\max} = \max\{n|\varrho(n) < 0.5\varrho(\infty)\}, \tag{52}$$

where $\mathbb{E}\|x_n^0 - x_n^\delta\|^2 = \delta^2\varrho^2(n)$ and $\delta^2\varrho^2(\infty)$ is the supremum of the variance. This allows us to expect that the optimal parameter index is smaller than $n_{\max}$. Also note that $n_{\max}$ should not be too large for an efficient computation.

We obtain $n_{\max}$ either by an analytic expression for $\delta^2\varrho^2(n)$, as in (51) for white noise, or by a good estimate of it. If there are several independent data sets available, then a good estimate (for any noise color) is

$$\delta^2\varrho^2(n) \approx 2^{-1}\text{mean}\{\|x_{n,i}^\delta - x_{n,j}^\delta\|^2, i \neq j\}. \tag{53}$$

Often two sets of data are sufficient (see [5] for further details). If only a single data set is available, then it may not be possible to estimate $\delta^2\varrho^2(n)$ if the noise is correlated with unknown covariance. Then one can define a maximal index $\tilde{n}_{\max}$ by

$\beta_{\tilde{n}_{\max}} \approx \sigma_{\tilde{n}_{\max}}$ for our Tikhonov regularization if there are at least good estimates of the eigenvalues σ_k of A available. However, methods that perform much worse without the use of the maximal index $n_{\max}$, may yield different results for $\tilde{n}_{\max}$ instead of $n_{\max}$ (see [9, 14]).

4.4. Description and evaluation of methods

In this section, we describe the origin and idea of the method, list the input of the method and the algorithm that we use. We also give a brief discussion of known theoretical and practical issues about the method, including the most relevant references. Several of the methods use a tuning parameter or some other parameter that must be chosen (see [9, 14] and the references therein for further details).

4.4.1. Discrepancy principle. The discrepancy principle of [115, 110, 111] is one of the oldest and most widely used parameter choice procedures (cf. [33] and references therein). Its idea is that for a good regularized solution, the norm of the residual should match the noise level δ of the data. The method needs the following input:

- Norms of residuals $\{Ax_n^\delta - y^\delta\}_{n \le n_{\max}}$ until a certain bound is satisfied.
- Noise level δ.
- Tuning parameter $\tau \ge 1$.

In a deterministic setting with $\|y^\delta - y\| \le \delta$, the parameter choice n_* is the first n such that $\|Ax_n^\delta - y^\delta\| \le \tau\delta$. In a stochastic setting, with the error in each element of $y^\delta \in \mathbb{R}^N$ having standard deviation δ, the choice n_* is the first n such that

$$\|Ax_n^\delta - y^\delta\| \le \tau\delta\sqrt{N}. \tag{54}$$

Originating from a deterministic setting, the discrepancy principle has also been studied in stochastic settings (see, e.g., [17, 29, 97, 132]) and for many regularization methods and many inverse problems. There are many results on convergence properties of this method for both settings (see, e.g., [33, 65, 79, 104, 111, 112, 114] for the deterministic case and [29, 97, 98, 132] for the stochastic case).

The discrepancy principle is one of the fastest methods available, since one only needs to compute the residuals until the bound (54) is satisfied which allows the use of the FMM of Section 3. However, its drawback is the requirement of an accurate estimate of the noise level. Estimations that are just slightly off can lead to very poor solutions (see [73, Chap. 7]).

There are also many variants of the method such as the transformed discrepancy principle (cf. [119, 120, 71]), the modified discrepancy principle (MD rule) (cf. [32, 53, 117, 118]), or the varying discrepancy principle (cf. [17, 94]). Their main drawback is that they are no longer easily compatible with the FMM. For comparative studies in the context of inverse problems with stochastic noise we refer, e.g., to [9, 14].

4.4.2. Balancing principle. The balancing principle of [90] was originally derived for statistical estimation from direct observations in a white noise model. Since then it has been developed further for regularization of linear and nonlinear inverse problems (see, e.g., [6, 13, 11, 58, 129, 105, 106]) in deterministic and stochastic settings. The idea is to balance the known propagated noise error bound $\delta\varrho(n)$ in (49) with the unknown regularization error (48) by an adaptive procedure that employs a collection of differences of regularized solutions. As input the balancing principle needs:

- Maximal index $n_{\max}$, e.g., as defined in (52).
- All regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$ up to the index $n_{\max}$.
- An upper bound $\delta\varrho(n)$ for the propagated noise error $\|x_n^0 - x_n^\delta\|$ or a bound or estimate $\delta^2\varrho^2(n)$ of the variance $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$.
- Noise level δ (and the covariance in the stochastic setting if known). Then one can use known expressions for $\delta\varrho(n)$. Alternatively, if one has two or more independent sets of data y_i^δ, then $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$ can be estimated by (53).
- Tuning constant κ, typically $\kappa \in [0.5, 1.5]$ (cf. [9] and the references therein).

The balancing functional is defined by

$$b(n) = \max_{n<k\le n_{\max}} \left\{4^{-1}\|x_n - x_k\|/(\delta\varrho(k))\right\}. \tag{55}$$

The smoothed balancing functional (which is monotonously decreasing) is defined as $B(n) = \max_{n\le k\le n_{\max}}\{b(k)\}$. Then the parameter choice n_* is the first n such that $B(n) \le \kappa$.

The balancing principle is one of the few parameter choice methods for which oracle inequalities for the error are known (cf. [121, 13]), i.e., there are stronger results than rates of convergence alone. For variants we refer to [10, 105] and for comparative studies we mention [9, 14].

One variant should be noted: The hardened balancing principlewhich is a modified version of the balancing principle in the stochastic setting first proposed in [5]. The input is the same as before, but without the tuning parameter and the noise level. Furthermore, an expression or approximation of the scaled variance $\varrho^2(n) = \delta^{-2}\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, or any scalar multiple of this (so δ can be unknown), is required. The balancing functional $b(n)$ is defined as in (55) and the smoothed balancing functional $B(n) = \max_{n\le k\le n_{\max}}\{b(k)\}$. The parameter choice is

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{B(n)\sqrt{\varrho(n)}\right\}, \tag{56}$$

where any scalar multiple of $\varrho(n)$ gives the same choice. The method has the advantage that it does not require a tuning parameter. Numerical experiments in [5, 12, 9, 14] indicate that the method is very stable even for colored noise.

4.4.3. Quasi-optimality criterion. The quasi-optimality criterion by Tikhonov and Arsenin [127], Tikhonov and Glasko [128] (see also [77]) is one of the oldest and simplest available parameter choice methods. An overview of the method and its

history can be found, e.g., in [8]. As input for the minimization the following is required:

- Maximal index $n_{\max}$, e.g., as defined in (52).
- All regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$ up to $n_{\max}$.

The noise level does not need to be known, and there is no tuning parameter. The parameter choice n_* is defined simply as

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \|x_n^\delta - x_{n+1}^\delta\| \right\}. \tag{57}$$

The well-known continuous version for Tikhonov regularization defines the parameter choice by $\beta_* = \operatorname{argmin} \left\| \beta \frac{d}{d\beta} x_\beta^\delta \right\|$. Using a difference quotient in place of the derivative for the discrete parameters $\beta_n = \beta_0 q_\beta^n$ we obtain (57).

For a discrete set of regularization parameters, the use of a suitable maximal index $n_{\max}$ is essential, because the method is based on a discrete evaluation of a differential. Hence is very sensitive to a situation where the regularization operators A_n^{-1} are formally different, but are practically the same (cf. [9] and the references therein). Convergence results for the Tikhonov regularization with the quasi-optimality criterion for different settings can be found in [89, 56], for further convergence properties see [7, 84, 85, 113].

4.4.4. L-curve method. The L-curve method, proposed by Hansen [72, 73] and Hansen and O'Leary [74], is based on the long-known fact that a log-log parametric plot of $(\|Ax_n^\delta - y^\delta\|, \|x_n^\delta\|)$ often has a distinct L-shape (cf. [88]). Points on the vertical part correspond to large n (under-smoothed solutions) and those on the horizontal part correspond to small n (over-smoothed solutions), which suggests that the "corner point" of the L-curve should define a good value of the parameter n. It is usually applied manually and can provide good results then whereas finding the L-curve corner is hard to automate. As input to minimize a certain function the following is used:

- Norms of all residuals $\{Ax_n^\delta - y^\delta\}_{n\le n_{\max}}$.
- Norms of the regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$.

The noise level does not need to be known. The parameter choice can be defined by the product of the norms of the residual and regularized solution, i.e.,

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \|Ax_n^\delta - y^\delta\| \cdot \|x_n^\delta\| \right\}. \tag{58}$$

Here the "corner point" is defined by the slope of its "tangent" being -1 as in [122] (see also [33]). The generalizations minimize $\|Ax_n^\delta - y^\delta\| \|x_n^\delta\|^\tau$ (see [122, 93]), where τ is a tuning constant.

Since "corner point" is not a well-defined notion, several algorithms have evolved with different definitions (see [74, 18, 75, 122]). [93] derived first rigorous optimality results for the L-curve criterion. In many (but not all) problems, variants of the L-curve method has been observed to give a reasonably good parameter choice which can deal with correlated errors. See [9] for an overview of references where the method works or runs into severe limitations.

4.4.5. Generalized cross-validation. Generalized cross-validation (GCV), due to [133], is a popular method for problems with discrete data and stochastic noise as (20). It goes back to ordinary cross-validation, where the idea is to consider all the "leave-one-out" regularized solutions and choose the parameter that minimizes the average of the squared prediction errors using each solution to predict the missing data value. These calculations do not require the computation of all the regularized solutions. Weighting the prediction errors, [27, 60, 133, 135] derived the GCV method, which has is invariant under orthogonal transformations of the data. Some other parameter choice methods proposed in the literature have been shown to be closely related to GCV, in particular the Akaike information criterion (AIC) of [2, 35]. As input to minimize a certain function we need:

- Sums of squares of all the residuals $\{Ax_n^\delta - y^\delta\}_{n\le n_{\max}}$ where $y^\delta \in \mathbb{R}^N$.
- The trace of the influence matrix AA_n^{-1} mapping y^δ to Ax_n^δ.

The noise level does not need to be known. The GCV parameter estimate is defined by

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \right\}. \tag{59}$$

GCV is closely related to and behaves like the unbiased prediction risk method (also known as Mallows C_p or C_L; see [31, 92, 135]). It is asymptotically optimal with respect to the prediction risk as $N \to \infty$ for stochastic white noise and the Tikhonov regularization (cf. [66, 91, 96, 132, 136]). The GCV method has been used widely and performs very well for reasonably large data sets with uncorrelated errors (white noise). However, it is known (see, e.g., [31, 86, 87, 99, 101, 126, 135]) that for smaller data sets or correlated errors of red noise type, the method is rather unstable, often resulting in under-smoothing.

The term $\operatorname{tr}(AA_n^{-1})$ in the GCV function is a measure of the degrees of freedom in the regularized solution. For its fast computation making use of the FMM trace estimation methods are needed that use stochastic (Monte-Carlo) algorithms (cf. [54, 55, 59, 80, 81]).

In order to overcome the instability of GCV, several variants have evolved. The robust GCV (RGCV) method has been developed and investigated in [99, 100, 123]. It needs the same input as for GCV and additionally:

- The trace of the square of the influence matrix $(AA_n^{-1})^2$.
- A robustness parameter $\gamma \in (0, 1)$. Note that with $\gamma = 1$ the RGCV method is just GCV.

The RGCV parameter estimate is defined by minimizing a certain function:

$$n_* = \operatorname*{argmin}_{n\le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma)N^{-1}\operatorname{tr}((AA_n^{-1})^2)\right) \right\}. \tag{60}$$

The family of robust GCV methods developed in [100] also includes the strong robust GCV method, denoted $\mathrm{R_1GCV}$. As input one needs the same as for GCV as well as

- The trace of $A_n^{-1\,*}A_n^{-1}$.
- A robustness parameter $\gamma \in (0,1)$. Note that for $\gamma = 1$ the R_1GCV method is just GCV.

The R_1GCV parameter estimate is defined by minimizing a certain function:

$$n_* = \underset{n \le n_{\max}}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma)N^{-1}\operatorname{tr}((A_n^{-1\,*}A_n^{-1})\right) \right\}. \quad (61)$$

The modified GCV method involves a simple modification of the GCV function that is designed to stabilize the method (cf. [28, 131]). The inputs are the same as for GCV plus:

- A stabilization parameter $c > 1$. For $c = 1$ the method reduces to GCV.

The noise level does not need to be known. The modified GCV estimate is defined by

$$n_* = \operatorname{argmin}_{n \le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - cAA_n^{-1}))^2} \right\}. \quad (62)$$

For comparative studies of these variants and further details we refer to [9, 14] and the references therein.

5. Conclusion

Using the FMM in an iterative algorithm like, e.g., conjugate gradients or GMRES is an efficient solution strategy that can treat interpolation problems and Dirichlet boundary value problems with many data points on regular surfaces (e.g., the actual topography of the Earth) (see [67, 68, 69]). It should be pointed out that this spline approach is not restricted to a global treatment, but also applies to regional domains (cf. [67, 68]). This can lead to a local improvement of the gravitational field in areas of particular interest. The approach can be extended to spline approximation (in particular for diagonal covariance matrices) as seen in Section 2.4 and the end of Section 3.4).

The smoothing parameter(s) plays a crucial role in this approach and must be chosen very carefully or a lot of information is lost to oversmoothing, in particular the high-frequent details of the signal. We have presented several parameter choice methods the can be used without losing the advantages of the FMM. Their performance for the regularization of inverse problems has been investigated in several studies with different solution techniques (see, e.g., [9, 14, 70] and the references therein). Tests of the combination of the FMM with these parameter choice methods (cf. [9, 14, 70] and the references therein) are an interesting challenge for the future. In particular the interaction with stopping criteria for iterative solvers needs further investigation. Note that often the solution (even using the FMM) requires much more computational effort than the evaluation of the parameter choice. It can be advisable to apply several parameter choice methods to find the best choice of the parameter.

For highly irregular distributions of data points, the spline approach reaches its limits due to large data gaps which result in severe ill-conditioning. Even smoothing splines cannot completely bridge this gap so far though further investigation is required. However, functional matching pursuit methods (RFMP or ROFMP) can result in better approximations (see [36, 70, 108] and the references therein), but so far these algorithms require high numerical costs. These methods are also iterative regularizations and the combination of stopping criteria and regularization parameters has been investigated for a class of ill-posed problems in [70].

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, 1972.

[2] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor,* 1971), pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[3] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[4] A.B. Bakushinskii. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *U.S.S.R. Comput. Math. Math. Phys.*, 24:181–182, 1984.

[5] F. Bauer. Some considerations concerning regularization and parameter choice algorithms. *Inverse Prob.*, 23(2):837–858, 2007.

[6] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Prob.*, 21:1975–1991, 2005.

[7] F. Bauer and S. Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Prob.*, 24:035002, 20 pp., 2008.

[8] F. Bauer and S. Kindermann. Recent results on the quasi-optimality principle. *J. Inverse Ill-Posed Prob.*, 17(1):5–18, 2009.

[9] F. Bauer and M.A. Lukas. Comparing parameter choice methods for regularization of ill-posed problems. *Math. Comput. Simul.*, 81(9):1795–1841, 2011.

[10] F. Bauer and A. Munk. Optimal regularization for ill-posed problems in metric spaces. *J. Inverse Ill-Posed Probl.*, 15(2):137–148, 2007.

[11] F. Bauer and S. Pereverzev. Regularization without preliminary knowledge of smoothness and error behavior. *European J. Appl. Math.*, 16(3):303–317, 2005.

[12] F. Bauer and M. Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Prob.*, 24:055009, 16 pp., 2008.

[13] F. Bauer, T. Hohage, and A. Munk. Iteratively regularized Gauss–Newton method for nonlinear inverse problems with random noise. *SIAM J. Numer. Anal.*, 47(3): 1827–1846, 2009.

[14] F. Bauer, M. Gutting, and M.A. Lukas. Evaluation of parameter choice methods for regularization of ill-posed problems in geomathematics. In W. Freeden, M.Z.

Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 2nd ed.*, pages 1713–1774. Springer, Heidelberg, 2015.

[15] S.M.A. Becker. Regularization of statistical inverse problems and the Bakushinkii veto. *Inverse Prob.*, 27:115010, 22 pp., 2011.

[16] L.C. Biedenharn and J.D. Louck. *Angular Momentum in Quantum Physics* (*Theory and Application*). Encyclopedia of Mathematics and its Applications. Addison-Wesley, Reading, Massachusetts, 1981.

[17] G. Blanchard and P. Mathé. Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Prob.*, 28:115011, 23 pp., 2012.

[18] D. Calvetti, P.C. Hansen, and L. Reichel. L-curve curvature bounds via Lanczos bidiagonalization. *Electron. Trans. Numer. Anal.*, 14:20–35, 2002.

[19] E.J. Candès. Modern statistical estimation via oracle inequalities. *Acta Numerica*, 15:257–325, 2006.

[20] J. Carrier, L. Greengard, and V. Rokhlin. A fast adaptive multipole algorithm for particle simulations. *SIAM J. Sci. Stat. Comput.*, 9(4):669–686, 1988.

[21] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Prob.*, 24(3): 034004, 19 pp., 2008.

[22] L. Cavalier, G.K. Golubev, D. Picard, and A.B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002.

[23] L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory Probab. Appl.*, 48(3):426–446, 2004.

[24] H. Cheng, L. Greengard, and V. Rokhlin. A fast adaptive multipole algorithm in three dimensions. *J. Comput. Phys.*, 155:468–498, 1999.

[25] C.H. Choi, J. Ivanic, M.S. Gordon, and K. Ruedenberg. Rapid and staple determination of rotation matrices between spherical harmonics by direct recursion. *J. Chem. Phys.*, 111(19):8825–8831, 1999.

[26] D.D. Cox. Approximation of method of regularization estimators. *Ann. Statist.*, 16 (2):694–712, 1988.

[27] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numer. Math.*, 31:377–403, 1979.

[28] D.J. Cummins, T.G. Filloon, and D. Nychka. Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.*, 96(453):233–246, 2001.

[29] A.R. Davies and R.S. Anderssen. Improved estimates of statistical regularization parameters in Fourier differentiation and smoothing. *Numer. Math.*, 48:671–697, 1986.

[30] A.R. Edmonds. *Drehimpulse in der Quantenmechanik*. Bibliographisches Institut, Mannheim, 1964.

[31] B. Efron. Selection criteria for scatterplot smoothers. *Ann. Statist.*, 29(2):470–504, 2001.

[32] H.W. Engl and H. Gfrerer. A posteriori parameter choice for general regularization methods for solving linear ill-posed problems. *Appl. Numer. Math.*, 4(5):395–417, 1988.

[33] H.W. Engl, H. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[34] M.A. Epton and B. Dembart. Multipole translation theory for the three-dimensional Laplace and Helmholtz equations. *SIAM J. Sci. Comput.*, 16 (4): 865–897, 1995.

[35] R.L. Eubank. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York, 1988.

[36] D. Fischer and V. Michel. Sparse regularization of inverse gravimetry – case study: Spatial and temporal mass variation in South America. *Inverse Prob.*, 28:065012, 34 pp., 2012.

[37] W. Freeden. On approximation by harmonic splines. *Manuscripta Geod.*, 6:193–244, 1981a.

[38] W. Freeden. On spherical spline interpolation and approximation. *Math. Method. Appl. Sci.*, 3:551–575, 1981b.

[39] W. Freeden. Interpolation and best approximation by harmonic spline functions. *Boll. Geod. Sci. Aff.*, 1:105–120, 1982a.

[40] W. Freeden. On spline methods in geodetic approximation problems. *Math. Method. Appl. Sci.*, 4:382–396, 1982b.

[41] W. Freeden. Spherical spline interpolation: Basic theory and computational aspects. *J. Comput. Appl. Math.*, 11:367–375, 1984a.

[42] W. Freeden. Ein Konvergenzsatz in sphärischer Spline-Interpolation. *Z. f. Vermessungswes.* (*ZfV*), 109:569–576, 1984b.

[43] W. Freeden. A spline interpolation method for solving boundary value problems of potential theory from discretely given data. *Numer. Methods Partial Differential Equations*, 3:375–398, 1987a.

[44] W. Freeden. Harmonic splines for solving boundary value problems of potential theory. In J.C. Mason and M.G. Cox, editors, *Algorithms for Approximation*, pages 507–529. The Institute of Mathematics and its Applications, Conference Series, Vol. 10, Clarendon Press, Oxford, 1987b.

[45] W. Freeden. *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Stuttgart, Leipzig, 1999.

[46] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Chapman & Hall/CRC, Boca Raton, 2013.

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

[48] W. Freeden and V. Michel. *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, Boston, Basel, Berlin, 2004.

[49] W. Freeden and M. Schreiner. Special functions in mathematical geosciences: An attempt at a categorization. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics.* 1*st ed.*, pages 925–948. Springer, Heidelberg, 2010.

[50] W. Freeden, M. Schreiner, and R. Franke. A survey on spherical spline approximation. *Surv. Math. Ind.*, 7:29–85, 1997.

[51] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere* (*with Applications to Geomathematics*). Oxford Science Publications, Clarendon, 1998a.

[52] W. Freeden, O. Glockner, and M. Schreiner. Spherical panel clustering and its numerical aspects. *J. Geodesy*, 72:586–599, 1998b.

[53] H. Gfrerer. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comput.*, 49(180):507–522, 1987.

[54] D. Girard. A fast "Monte-Carlo cross-validation" procedure for large least squares problems with noisy data. *Numer. Math.*, 56(1):1–23, 1989.

[55] D. Girard. Asymptotic comparison of (partial) cross-validation, gcv and randomized gcv in nonparametric regression. *Ann. Statist.*, 26(1):315–334, 1998.

[56] V. Glasko and Y. Kriksin. On the quasioptimality principle for linear ill-posed problems in Hilbert space. *Vychisl. Mat. Mat. Fiz.*, 24:1603–1613, 1984.

[57] O. Glockner. *On Numerical Aspects of Gravitational Field Modelling From SST and SGG by Harmonic Splines and Wavelets* (*with Application to CHAMP Data*). Ph.D.-thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern. Shaker, Aachen, 2002.

[58] A. Goldenshluger and S. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Prob. Th. Rel. Fields*, 118(2): 169–186, 2000.

[59] G.H. Golub and U. von Matt. Generalized cross-validation for large scale problems. *J. Comput. Graph. Statist.*, 6(1):1–34, 1997.

[60] G.H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[61] L. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems*. MIT Press, Cambridge, MA, 1988.

[62] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73(1):325–348, 1987.

[63] L. Greengard and V. Rokhlin. Rapid evaluation of potential fields in three dimensions. In C. Anderson and L. Greengard, editors, *Vortex Methods*, pages 121–141. Springer, 1988.

[64] L. Greengard and V. Rokhlin. A new version of the fast multipole method for the Laplace equation in three dimensions. *Acta Numerica*, 6:229–269, 1997.

[65] C.W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, Boston, 1984.

[66] C. Gu. *Smoothing Spline ANOVA Models*. Springer, New York, 2002.

[67] M. Gutting. *Fast Multipole Methods for Oblique Derivative Problems*. Ph.D.-thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern. Shaker, Aachen, 2007.

[68] M. Gutting. Fast multipole accelerated solution of the oblique derivative boundary value problem. *GEM – Int. J. Geom.*, 3(2):223–252, 2012.

[69] M. Gutting. Fast spherical/harmonic spline modeling. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 2nd ed.*, pages 2711–2746. Springer, Heidelberg, 2015.

[70] M. Gutting, B. Kretz, V. Michel, and R. Telschow. Study on parameter choice methods for the RFMP with respect to downward continuation. *Front. Appl. Math. Stat.*, 3, doi:10.3389/fams.2017.00010, 17 pp., 2017.

[71] U. Hämarik and T. Raus. On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. *J. Inverse Ill-Posed Probl.*, 14(3):251–266, 2006.

[72] P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[73] P.C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM, Philadelphia, 1998.

[74] P.C. Hansen and D.P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, 1993.

[75] P.C. Hansen, T.K. Jensen, and G. Rodriguez. An adaptive pruning algorithm for the discrete L-curve criterion. *J. Comput. Appl. Math.*, 198(2):483–492, 2007.

[76] E.W. Hobson. *The Theory of Spherical and Ellipsoidal Harmonics* (*Second Reprint*). Chelsea Publishing Company, New York, 1965.

[77] B. Hofmann. *Regularization of Applied Inverse and Ill-Posed Problems.* Teubner, Leipzig, 1986.

[78] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. of Numer. Anal.*, 45(3):1122–1141, 2007.

[79] T. Hohage. Regularization of exponentially ill-posed problems. *Numer. Funct. Anal. Optim.*, 21:439–464, 2000.

[80] M. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Stat. Simulation Comput.*, 18(3):1059–1076, 1989.

[81] M.F. Hutchinson and F.R. de Hoog. Smoothing noisy data with spline functions. *Numer. Math.*, 47:99–106, 1985.

[82] J. Keiner, S. Kunis, and D. Potts. Fast summation of radial functions on the sphere. *Computing*, 78:1–15, 2006.

[83] O.D. Kellogg. *Foundation of Potential Theory.* Springer, Berlin, Heidelberg, New York, 1967.

[84] S. Kindermann. Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. *Electron. Trans. Numer. Anal.*, 38: 233–257, 2011.

[85] S. Kindermann and A. Neubauer. On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. *Inverse Probl. Imaging*, 2(2):291–299, 2008.

[86] R. Kohn, C.F. Ansley, and D. Tharm. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *J. Amer. Statist. Assoc.*, 86:1042–1050, 1991.

[87] S.C. Kou and B. Efron. Smoothers and the C_p, generalized maximum likelihood, and extended exponential criteria: a geometric approach. *J. Amer. Statist. Assoc.*, 97(459):766–782, 2002.

[88] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, N.J., 1974.

[89] A. Leonov. Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria. *U.S.S.R. Comput. Math. Math. Phys.*, 18(6):1–15, 1979.

[90] O. Lepskij. On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35(3):454–466, 1990.

[91] K.-C. Li. Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.*, 14:1101–1112, 1986.

[92] K.-C. Li. Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.*, 15:958–975, 1987.

[93] S. Lu and P. Mathé. Heuristic parameter selection based on functional minimization: optimality and model function approach. *Math. Comput.*, 82(283):1609–1630, 2013.

[94] S. Lu and P. Mathé. Discrepancy based model selection in statistical inverse problems. *J. Complexity*, 30(3):290–308, 2014.

[95] M.A. Lukas. Convergence rates for regularized solutions. *Math. Comput.*, 51(183): 107–131, 1988.

[96] M.A. Lukas. Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. *Numer. Math.*, 66(1):41–66, 1993.

[97] M.A. Lukas. On the discrepancy principle and generalized maximum likelihood for regularization. *Bull. Aust. Math. Soc.*, 52(3):399–424, 1995.

[98] M.A. Lukas. Comparisons of parameter choice methods for regularization with discrete noisy data. *Inverse Prob.*, 14(1):161–184, 1998.

[99] M.A. Lukas. Robust generalized cross-validation for choosing the regularization parameter. *Inverse Prob.*, 22(5):1883–1902, 2006.

[100] M.A. Lukas. Strong robust generalized cross-validation for choosing the regularization parameter. *Inverse Prob.*, 24:034006, 16 pp., 2008.

[101] M.A. Lukas. Robust GCV choice of the regularization parameter for correlated data. *J. Integral Equations Appl.*, 22(3):519–547, 2010.

[102] W. Magnus, F. Oberhettinger, and R.P. Soni. *Formulas and Theorems for the Special Functions of Mathematical Physics*, volume 52 of *Die Grundlehren der mathematischen Wissenschaften*. Springer, New York, 3rd edition, 1966.

[103] B.A. Mair and F.H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[104] P. Mathé. What do we learn from the discrepancy principle? *Z. Anal. Anwend.*, 25 (4):411–420, 2006.

[105] P. Mathé and S.V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert spaces. *Inverse Prob.*, 19(3):789–803, 2003.

[106] P. Mathé and S.V. Pereverzev. Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comput.*, 75(256):1913–1929, 2006.

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

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

[109] H. Moritz. Classical physical geodesy. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 1st ed.*, pages 127–158. Springer, Heidelberg, 2010.

[110] V.A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[111] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag, New York, 1984.

[112] M.T. Nair, E. Schock, and U. Tautenhahn. Morozov's discrepancy principle under general source conditions. *Z. Anal. Anwend.*, 22:199–214, 2003.

[113] A. Neubauer. The convergence of a new heuristic parameter selection criterion for general regularization methods. *Inverse Prob.*, 24(5):055005, 10 pp., 2008.

[114] S. Pereverzev and E. Schock. Morozov's discrepancy principle for tikhonov regularization of severely ill-posed problems in finite dimensional subspaces. *Numer. Funct. Anal. Optim.*, 21:901–916, 2000.

[115] D. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[116] D. Potts and G. Steidl. Fast summation at nonequispaced knots by NFFTs. *SIAM J. Sci. Comput.*, 24(6):2013–2037, 2003.

[117] T. Raus. On the discrepancy principle for the solution of ill-posed problems. *Uch. Zap. Tartu. Gos. Univ.*, 672:16–26, 1984.

[118] T. Raus. The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator. *Tartu Riikl. Ül. Toimetised*, 715:12–20, 1985.

[119] T. Raus. An a posteriori choice of the regularization parameter in case of approximately given error bound of data. In Pedas, A., editor, *Collocation and Projection Methods for Integral Equations and Boundary Value Problems*, pages 73–87. Tartu University, Tartu, 1990.

[120] T. Raus. About regularization parameter choice in case of approximately given error bounds of data. In G. Vainikko, editor, *Methods for solution of integral equations and ill-posed problems*, pages 77–89. Tartu University, Tartu, 1992.

[121] T. Raus and U. Hämarik. On the quasioptimal regularization parameter choices for solving ill-posed problems. *J. Inverse Ill-Posed Probl.*, 15(4):419–439, 2007.

[122] T. Reginska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17(3):740–749, 1996.

[123] T. Robinson and R. Moyeed. Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression. *Commun. Stat. Theory Methods*, 18 (2):523–539, 1989.

[124] V. Rokhlin. Rapid solution of integral equations of classical potential theory. *J. Comput. Phys.*, 60:187–207, 1985.

[125] L. Shure, R.L. Parker, and G.E. Backus. Harmonic splines for geomagnetic modelling. *Phys. Earth Planet. Inter.*, 28:215–229, 1982.

[126] A.M. Thompson, J.W. Kay, and D.M. Titterington. A cautionary note about cross-validatory choice. *J. Statist. Comput. Simul.*, 33:199–216, 1989.

[127] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.

[128] A. Tikhonov and V. Glasko. Use of the regularization method in non-linear problems. *U.S.S.R. Comput. Math. Math. Phys.*, 5(3):93–107, 1965.

[129] A. Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci., Paris, Sér. I, Math.*, 330(9):835–840, 2000.

[130] D.A. Varšalovič, A.N. Moskalev, and V.K. Chersonskij. *Quantum Theory of Angular Momentum*. World Scientific, Singapore, 1988.

[131] R. Vio, P. Ma, W. Zhong, J. Nagy, L. Tenorio, and W. Wamsteker. Estimation of regularization parameters in multiple-image deblurring. *Astron. Astrophys.*, 423: 1179–1186, 2004.

[132] C.R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.

[133] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667, 1977.

[134] G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.*, 2:5–16. Also errata: SIAM J. Sci. Stat. Comput., 3, 385–386, 1981.

[135] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.

[136] G. Wahba and Y.H. Wang. When is the optimal regularization parameter insensitive to the choice of the loss function? *Comm. Statist. Theory Methods*, 19(5):1685–1700, 1990.

[137] C.A. White and M. Head-Gordon. Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J. Chem. Phys.*, 105(12):5061–5067, 1996.

[138] H. Yamabe. On an extension of the Helly's theorem. *Osaka Math. J.*, 2(1):15–17, 1950.

[139] N. Yarvin and V. Rokhlin. Generalized Gaussian quadratures and singular value decomposition of integral equations. *SIAM J. Sci. Comput.*, 20(2):699–718, 1998.

Martin Gutting
Geomathematics Group
University of Siegen
Walter-Flex-Straße 3
D-57068 Siegen, Germany