Technische Universität München
Department of Mathematics

Kernel Based Image Reconstruction from Spherical Radon Data

Error Estimates and Fast Implementation

Master's Thesis by Thomas Wiatowski

supervisor : Prof. Dr. Rupert Lasser
academic advisor: Dr. Frank Filbir, Dipl.-Math. Jürgen Frikel
submission date : November 14, 2012

"Geld genug, um neun Wochen davon zu leben. Es ist nur ein kleiner Betrag, aber eine lange Zeit. Wir versagen uns jeden Luxus außer dem kostbarsten: Muße."

Nicolas Bouvier

I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

.....

(Munich, 14th November 2012) (signature)

Acknowledgements

I would like to thank all people who have helped and inspired me during my master study.

Especially I want to thank my academic advisors, Dr. Frank Filbir and Dipl.-Math. Jürgen Frikel, for their excellent guidance during my master thesis at the Technische Universität München. They were always accessible and willing to help me with the work on my thesis. As a result, research and writing became smooth and rewarding for me. I am very thankful to Dipl.-Math. Ruben Seyfried and Prof. Dr. Stefan Kunis for their wonderful ideas that improved the numerical results substantially. In addition, thank you to Prof. Dr. Rupert Lasser for the supervision of my master thesis.

I am indebted to my friends Tobias Böttger, Özgür Boztas, Friedrich Prade and Felix Rucker. Their perpetual energy and enthusiasm for life and for science have motivated and helped me throughout my studies.

Thank you to Laetitia Vionnet for all the support, affection and patience; it has been essential.

My deepest gratitude goes to my parents, Anna and Norbert, for their enduring love and support throughout my life. Thank you.

Abstract

The reconstruction of images from data modeled by the spherical Radon transform plays an important role in photoacoustic tomography - a rapidly developing modality for *in vivo* imaging. We provide two novel kernel based reconstruction algorithms adapted to this type of data: optimal recovery and algebraic reconstruction technique. The thesis details the algorithms' derivation, L^2-error estimates in terms of the data density, convergence results and a fast implementation. Several numerical examples for *real* and artificially generated data are included.

Plan of the Thesis

In the first chapter we present the image reconstruction problem from spherical Radon data in the setting of photoacoustic tomography. We start by giving a brief overview on the underlying imaging physics and on the mathematical modeling behind. Then we formulate the inverse problem respectively the semi-discrete inverse problem of photoacoustic tomography.
The following chapter introduces the mathematical fundamentals for the kernel based reconstruction process. We recall important notions of reproducing kernel Hilbert spaces and examine the construction of fractional Sobolev spaces by a given positive definite radial kernel function.
The new results are presented in the third chapter. We give a detailed description of two solution methods for the semi-discrete inverse problem of photoacoustic tomography: optimal recovery and algebraic reconstruction technique. For both methods we provide L^2-error estimates in terms of the data density and a proof of convergence.
Finally, we present a fast algorithm for the algebraic reconstruction technique and give numerical results for *real* and artificially generated data. We study the behavior of the method with respect to its image reconstruction capability and its computational efficiency.

Contents

1 Introduction

The theory and practice of recovering a function from spherical Radon data has been developed especially in the setting of photoacoustic tomography. In this chapter, we give a brief overview on the underlying imaging physics and on the mathematical modeling behind. Further, we formulate the reconstruction problem from continuous and discrete spherical Radon data.

1.1 Scheme of Photoacoustic Tomography

Photoacoustic tomography (PAT) - a rapidly emerging imaging technique - proves to have potential for biomedical *in vivo* imaging. The method is based on the generation of an acoustic wave due to absorption of energy. A very short laser pulse triggers a thermoelastic expansion and contraction of the radiated tissue by absorption. The generated pressure wave - the acoustic signal - propagates through the tissue and beyond and is recorded by transducers distributed around the object. PAT aims to produce an image which represents a map of the electromagnetic absorption properties of the tissue from knowledge of the measured acoustic signals. Because the absorption properties of tissues are highly related to their molecular constitution, PAT images reveal the pathological condition of tissues and facilitate a wide-range of diagnostic tasks.

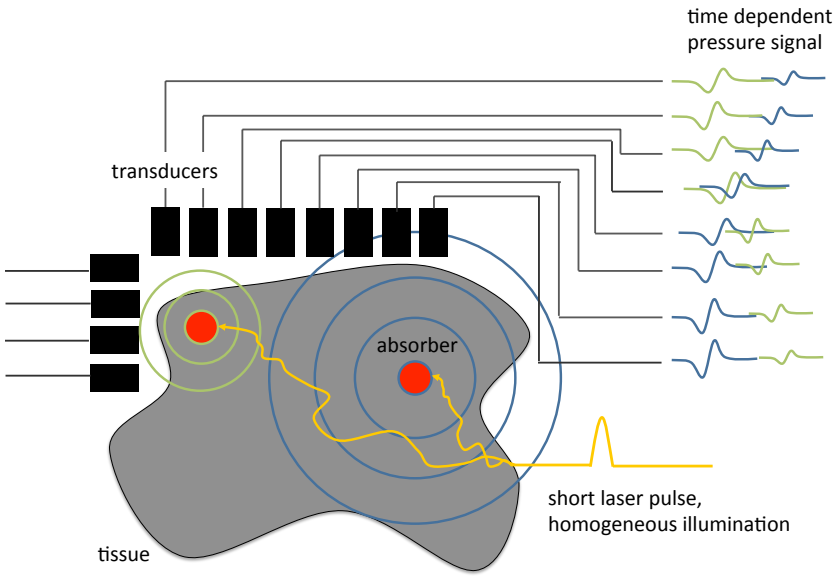

Figure 1: Photoacoustic procedure with a partially surrounding acquisition surface.

1.2 The Inverse Problem of Photoacoustic Tomography

In the following we describe the commonly accepted reconstruction problem of photoacoustic tomography from continuous spherical Radon data [3, 45]. The photoacoustic effect is modeled in terms of the inhomogeneous wave equation

$$\partial_t^2 p(x,t) - \nu_s^2 \Delta p(x,t) = \partial_t H(x,t), \tag{1}$$

where $p(x,t)$ is the pressure, $\nu_s(x)$ is the speed of sound within the medium and $H(x,t)$ is a function that models the heating of the material. Neglecting the effect of finite speed propagation of the in-falling laser pulses, we assume that this function separates in its variables, i.e. $H(x,t) = f(x)\iota(t)$. The function f models the spatial absorption of energy and is assumed to be compactly supported. Due to the very short duration of the laser pulses (in the nanosecond range [55]), we assume the function ι to be a delta function, i.e. $\iota(t) = \delta_t$. Under these assumptions we can show by approximating δ_t through regular distributions [46, Theorem 1.1] that (1) is equivalent to the following Cauchy problem

$$\begin{aligned} \partial_t^2 p(x,t) - \nu_s^2 \Delta p(x,t) &= 0 \\ p(x,0) = f(x), \quad \partial_t p(x,0) &= 0. \end{aligned} \tag{2}$$

Let the hypersurface $\partial\Omega$ - the boundary of an open bounded set $\Omega \subseteq \mathbb{R}^d$ - be the location of the transducers enclosing the support of f. Then, the inverse problem of PAT consists in the determination of the absorption density $f(x)$ for $x \in \Omega$ if $p(x,t)$ is known for $(x,t) \in \partial\Omega \times \mathbb{R}^+$.

In general, the speed of sound depends on the space variable. In this situation, in order to recover f, we need to work with the wave equation directly [45, 48]. However, experimental set ups very often allow to work with constant speed of sound. By rescaling the time variable, we assume throughout this thesis $\nu_s = 1$. As a result, the solution of (2) for odd dimensions $d \geq 3$ and a sufficiently well behaved function f is given as

$$p(x,t) = \frac{1}{1 \cdot 3 \cdots (d-2)} \partial_t (t^{-1}\partial_t)^{\frac{d-3}{2}} [t^{d-2} \mathcal{M}(f)(x,t)], \quad x \in \mathbb{R}^d, t \geq 0, \tag{3}$$

where

$$\mathcal{M}(f)(x,t) := \frac{1}{w_{d-1}} \int_{S^{d-1}} f(x+tu)\, \mathrm{d}\sigma(u)$$

is the spherical mean operator [13]. $S^{d-1} := \{x \in \mathbb{R}^d \mid |x| = 1\}$ denotes the $(d-1)$-dimensional unit sphere embedded in the Euclidean space $\mathbb{R}^d$ and σ denotes the surface measure on S^{d-1} with $\sigma(S^{d-1}) = 2\pi^{\frac{d}{2}}/\Gamma(\frac{d}{2}) =: w_{d-1}$. We derive from (3) that knowledge of the pressure signal $p(x,t)$ for $(x,t) \in \partial\Omega \times \mathbb{R}^+$ is equivalent to knowing the spherical mean $\mathcal{M}(f)(x,t)$ of the function f for any $(x,t) \in \partial\Omega \times \mathbb{R}^+$. Therefore, the inverse problem of PAT essentially amounts to the inversion of the spherical mean operator restricted to the hypersurface $\partial\Omega$. In the literature, the restriction of $\mathcal{M}$ to $\partial\Omega$ is known as spherical Radon transform and defined by

$$\mathcal{M}_{\partial\Omega}(f)(x,t) := \frac{1}{w_{d-1}} \int_{S^{d-1}} f(x+tu)\, \mathrm{d}\sigma(u), \quad x \in \partial\Omega,\ t \geq 0. \tag{4}$$

The inverse problem of photoacoustic tomography as well as various aspects of PAT have attracted considerable attention in the last decade with several results appearing in various articles. Especially work on the spherical mean transform respectively the spherical Radon transform has been done in great detail. We refer the interested reader to [3, 17, 38] for exact inversion formulas, to [6] for a description of its range and to [5] for a characterization of so-called injectivity sets. Further, a variety of image reconstruction algorithms have been developed for three-dimensional PAT data assuming point-like transducers [16, 30].

1.3 Facing Discrete Data: the Semi-Discrete Inverse Problem

The inverse problem of PAT aims to solve the operator equation

$$\mathcal{M}_{\partial\Omega}(f) = g, \tag{5}$$

where $\mathcal{M}_{\partial\Omega} : H(\Omega) \to H(\Omega')$ is the spherical Radon transform (4) defined on Hilbert spaces $H(\Omega)$ and $H(\Omega')$ of real valued functions $f : \Omega \to \mathbb{R}$ respectively $f : \Omega' \to \mathbb{R}$. However, in much of experimental science, the given information (here, the right hand side of the equation) is rather discrete than continuous. In PAT, the data are obtained by measuring the pressure signal $p(x, t)$ using acoustic transducers located at discrete locations $\Xi = \{\xi_k\}_{k=1}^K$ on the acquisition surface $\partial\Omega$ at fixed measurement times $\mathcal{T} = \{t_l\}_{l=1}^L \subseteq (0\ T]$. Hence, the final output of a photoacoustic experiment is a finite vector of real numbers $\bar{g} \in \mathbb{R}^{K \cdot L}$, where the elements of the Euclidean space $\mathbb{R}^{K \cdot L}$ have the representation as a vector (not as a matrix)

$$\bar{g} = (g_{11}, g_{21}, \ldots, g_{(K-1)1}, g_{K1}, g_{12}, g_{22}, \ldots, g_{(K-1)L}, g_{KL})^T.$$

Throughout this thesis, we assume the output vector $\bar{g}$ to be a sampled version of the spherical mean $g(x, t)$ for an unknown absorption function f, i.e.

$$g_{kl} := g(\xi_k, t_l) := \frac{1}{w_{d-1}} \int_{S^{d-1}} f(\xi_k + t_l u)\, \mathrm{d}\sigma(u) \tag{6}$$

for $\xi_k \in \Xi$ and $t_l \in \mathcal{T}$. By considering the experimental information, we replace the continuous operator equation (5) by its semi-discrete analogue $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$ with collocation operator

$$\mathcal{M}_{\mathcal{X}} : H(\Omega) \to \mathbb{R}^{K \cdot L}, \quad (\mathcal{M}_{\mathcal{X}}(f))_{kl} = \mathcal{M}_{\partial\Omega}(f)(\xi_k, t_l), \quad (\xi_k, t_l) \in \mathcal{X} := \Xi \times \mathcal{T}. \tag{7}$$

Then, the semi-discrete inverse problem of PAT can be formulated as follows: given spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$, find a solution $f^+ \in H(\Omega)$ of the equation

$$\mathcal{M}_{\mathcal{X}}(f) = \bar{g}. \tag{8}$$

1.4 Objective of the Thesis

This thesis presents two innovative algorithms for the reconstruction of f from spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$ by solving the semi-discrete inverse problem of PAT (8). In particular, we determine solutions f^+ of (8) in finite-dimensional subspaces X (*recovery spaces*) of a reproducing kernel Hilbert space $H(\Omega)$ with radial kernel Φ. The advantages of using such an approach are

- an analytical description of the action of the spherical mean operator (respectively spherical Radon transform) on a great class of radial kernels Φ, which provides explicit and manageable ways for the computation of f^+,
- a-priori smoothness information of f^+ due to the identification of reproducing kernel Hilbert spaces with Sobolev spaces of fractional order,
- "meshless" recovery of f by exploiting the geometry of the problem (determined by the transducer locations on the acquisition surface $\partial\Omega$) by using kernel related basis functions for the recovery spaces.

Clearly, the explicit choice of a recovery space determines the solution of (8). In this thesis, our focus is on two recovery spaces leading to different reconstruction algorithms:

$$X_{K\cdot L} := \operatorname{span}\{\, \mathcal{M}^y_{\partial\Omega}(\Phi(\cdot, y))(\xi_k, t_l) \mid k = 1, \dots, K \text{ and } l = 1, \dots, L\},$$

where the superscript y at the operator indicates the argument to which the operator is applied, and

$$X_{\mathcal{Y}} := \operatorname{span}\{\, \Phi(\cdot, y_i) \mid i = 1, \dots, N\},$$

where $\mathcal{Y} := \{y_1, \dots, y_N\} \subseteq \Omega$ is an arbitrary, but fixed, set. According to the well-known Golomb and Weinberger theory [24], we refer to the situation $X = X_{K\cdot L}$ as optimal recovery problem. For $X = X_{\mathcal{Y}}$, we call the recovery problem algebraic reconstruction technique.

The present thesis gives sufficient answers to the following questions:

1. How do the solutions of (8) for $X = X_{K\cdot L}$ and $X = X_{\mathcal{Y}}$ look like? How does the kernel Φ of the reproducing kernel Hilbert space $H(\Omega)$ influence the smoothness of the resulting approximation f^+?
2. Under which assumptions can we determine a function which solves the equation (8) exactly? When does a function satisfy the equation (8) just approximatively?
3. Can we guarantee the approximation f^+ to be uniquely determined? And if so, in which sense do we have to understand the uniqueness?
4. Are there requirements on the transducer locations Ξ, on the measurement times $\mathcal{T}$ or on the set $\mathcal{Y}$ in order to calculate f^+?

5. What is the approximation order of f^+ with respect to the unknown absorption density f?

6. Can we compute f^+ in a fast and efficient way?

1.5 Notes and Comments

The spherical Radon transform $\mathcal{M}_{\partial\Omega}$ resembles the classical Radon transform: the common tool of computerized tomography [31], which integrates functions over planes rather than spheres. This analogy with the classical Radon transform - although often purely ideological rather than technical - provides important intuition and frequently points in reasonable directions of study. However, if the speed of sound ν_s cannot be assumed to be constant, the relation between photoacoustic tomography and integral transforms breaks down [45, chapter 19].

With the problem formulation in (8), we regularize the continuous operator equation (5) by the method of *collocation in reproducing kernel Hilbert spaces* [27, chapter 3]. Roughly speaking, the collocation scheme consists in seeking an approximate solution by requiring that the continuous equation (5) is satisfied only at a finite number of collocation points. Hence, by formulating the recovery of f as a semi-discrete inverse problem, we stabilize the inversion of the spherical Radon transform. However, additional regularization methods for the semi-discrete equation (8) should be considered, such as truncated singular value decomposition, the Tikhonov-Phillips method or iterative procedures such as the Landweber and conjugate gradient method. We refer the reader to [9, 10, 28] for a comprehensive discussion.

2 Mathematical Fundamentals

In this chapter, we introduce the fundamentals for the solution methods of the semi-discrete inverse problem of PAT (8). We review the definition and basic properties of reproducing kernel Hilbert spaces and examine the construction of fractional Sobolev spaces by a given positive definite radial kernel function.

2.1 Preliminaries

We require fundamental knowledge about functional analysis, integral and measure theory, operator theory and inverse problems. The topics can be studied in the textbooks of Halmos [25], Rudin [40], Werner [53] and Rieder [36].

The following notation is used throughout this thesis. The Euclidean inner product of two vectors $x, y \in \mathbb{R}^d$ is denoted by $\langle x, y\rangle := \sum_{i=1}^d x_i y_i$. For the associated Euclidean norm of the vector $x \in \mathbb{R}^d$ we use the notation $|x| := \sqrt{\langle x, x\rangle}$. We denote with V' the topological dual space of a real topological vector space V. $\mathcal{S}$ denotes the space of rapidly decreasing functions. The Fourier transform of $\phi \in \mathcal{S}$ is defined by

$$\hat{\phi}(\xi) := (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \phi(x) e^{-i\langle x,\xi\rangle} \mathrm{d}x.$$

For any real number $s \in \mathbb{R}$,

$$H^s(\mathbb{R}^d) := \{f \in \mathcal{S}' \mid \hat{f} \in L^1_{loc}(\mathbb{R}^d), \quad \|f\|_{H^s(\mathbb{R}^d)} := \|(1 + |\cdot|^2)^{\frac{s}{2}} \hat{f}\|_{L^2(\mathbb{R}^d)} < \infty\}$$

is the Sobolev space of fractional order. $H^s(\mathbb{R}^d)$ for $s \in \mathbb{N}_0$ consists of functions whose (generalized) derivatives up to the order s are square integrable [44, Theorem 2.55]. Hence, the larger s the smoother are the elements in $H^s(\mathbb{R}^d)$. For more details on fractional Sobolev spaces and their definition on "sufficiently regular" sets $\Omega \subseteq \mathbb{R}^d$, see Appendix 6.2.

2.2 Reproducing Kernel Hilbert Spaces

We follow the textbook of Wendland [52, chapter 10], as it provides a very structured and comprehensive introduction into the topic of reproducing kernel Hilbert spaces. Additional information is taken from [28].

Definition 1.
Let $\Omega \subseteq \mathbb{R}^d$ be a given domain and let $H(\Omega)$ be a real Hilbert space of functions $f : \Omega \to \mathbb{R}$. $H(\Omega)$ is called reproducing kernel Hilbert space (RKHS) with reproducing kernel Φ, if a function $\Phi : \Omega \times \Omega \to \mathbb{R}$ exists with

1. *$\Phi(\cdot, y) \in H(\Omega)$ for all $y \in \Omega$,*
2. *$f(y) = \langle f, \Phi(\cdot, y)\rangle_H$ for all $y \in \Omega$ and for all $f \in H(\Omega)$.*

It is easy to see that the reproducing kernel of a Hilbert space is uniquely determined. Suppose there are two kernels Φ_1 and Φ_2. Then the second property gives $\langle f, \Phi_1(\cdot, y) - \Phi_2(\cdot, y)\rangle_H = 0$ for all $f \in H(\Omega)$ and all $y \in \Omega$. Setting $f = \Phi_1(\cdot, y) - \Phi_2(\cdot, y)$ for a fixed $y \in \Omega$ shows the uniqueness.

Let us give a first characterization of a RKHS which makes use of point evaluation functionals $\delta_y(f) := f(y)$, $y \in \Omega$, $f \in H(\Omega)$.

Lemma 1. *[52, Theorem 12.3]*
Let $H(\Omega)$ be a Hilbert function space. Then the following statements are equivalent:

1. *Point evaluation functionals are continuous, i.e. $\delta_y \in H(\Omega)'$ for all $y \in \Omega$;*
2. *$H(\Omega)$ has a reproducing kernel.*

Proof. Suppose that the point evaluation functionals are continuous. By the Riesz representation theorem [Appendix 6.1] we find for every $y \in \Omega$ a unique $\Phi_y \in H(\Omega)$ such that $\delta_y(f) = \langle f, \Phi_y\rangle_H$ for all $f \in H(\Omega)$. Thus, $\Phi(x, y) := \Phi_y(x)$ is the reproducing kernel. Now suppose that $H(\Omega)$ has a reproducing kernel Φ. This means that $\delta_y = \langle \cdot, \Phi(\cdot, y)\rangle_H$ for $y \in \Omega$. Since the inner product is continuous, so is δ_y. □

Interesting features of a RKHS are collected in the following lemma.

Lemma 2. *[52, Theorem 10.3]*
Suppose $H(\Omega)$ is a Hilbert space with reproducing kernel Φ. Then

1. $\Phi(x, y) = \langle \Phi(\cdot, x), \Phi(\cdot, y)\rangle_H = \langle \delta_x, \delta_y\rangle_{H'}$,
2. $\Phi(x, y) = \Phi(y, x)$

for every $x, y \in \Omega$.

Proof. By the reproducing property of the kernel [Definition 1] and the uniqueness of the Riesz representers [Appendix 6.1] it follows that $\Phi(\cdot, y)$ is the Riesz representer of δ_y, where $y \in \Omega$. The Riesz representation theorem [Appendix 6.1] yields $\langle \delta_x, \delta_y\rangle_{H'} = \langle \Phi(\cdot, x), \Phi(\cdot, y)\rangle_H$ and therefore

$$\Phi(x, y) = \delta_x(\Phi(\cdot, y)) = \langle \Phi(\cdot, y), \Phi(\cdot, x)\rangle_H = \langle \Phi(\cdot, x), \Phi(\cdot, y)\rangle_H.$$

Hence, the first property is proven and the second follows immediately from it. □

The next result focuses on the action of linear continuous functionals on elements of a reproducing kernel Hilbert space.

Proposition 3. *[52, Theorem 16.7]*
Suppose $H(\Omega)$ is a real Hilbert space with reproducing kernel Φ. Let $\lambda, \mu \in H(\Omega)'$. Then

1. $\lambda^y \Phi(\cdot, y) \in H(\Omega)$,

2. $\lambda(f) = \langle f, \lambda^y \Phi(\cdot, y)\rangle_H \quad \forall f \in H(\Omega)$,

3. $\langle \lambda, \mu\rangle_{H'} = \lambda^x \mu^y \Phi(x, y)$,

where the superscripts x and y at the functionals indicate the arguments to which the functionals are applied.

Proof. Let $\lambda \in H(\Omega)'$. The Riesz representation theorem [Appendix 6.1] guarantees the existence of a unique h_λ such that $\lambda(f) = \langle f, h_\lambda\rangle_H$ for all $f \in H(\Omega)$. Let $f_x := \Phi(\cdot, x) \in H(\Omega)$ for $x \in \Omega$, then

$$\lambda(f_x) = \langle f_x, h_\lambda\rangle_H = \langle h_\lambda, f_x\rangle_H = \langle h_\lambda, \Phi(\cdot, x)\rangle_H = h_\lambda(x)$$

by the reproducing property of the kernel. Thus $\lambda^y \Phi(y, x) = h_\lambda(x)$ for all $x \in \Omega$. Since the kernel of a RKHS is symmetric [Lemma 2], we have $\lambda^y \Phi(\cdot, y) = h_\lambda \in H(\Omega)$. Further,

$$\langle \lambda, \mu\rangle_{H'} = \langle h_\lambda, h_\mu\rangle_H = \langle h_\mu, h_\lambda\rangle_H = \lambda(h_\mu) = \lambda(\mu^y \Phi(\cdot, y)) = \lambda^x \mu^y \Phi(x, y). \qquad \square$$

One of the most important examples of reproducing kernel Hilbert spaces are Sobolev spaces of fractional order.

Theorem 4. *[44, Example 2.59]*
The fractional Sobolev space $H^s(\mathbb{R}^d)$ is a reproducing kernel Hilbert space if $s > \frac{d}{2}$.

Proof. From Lemma 1 we know that $H^s(\mathbb{R}^d)$ is a reproducing kernel Hilbert space if and only if point evaluation functionals are continuous, i.e. $\delta_\xi \in (H^s(\mathbb{R}^d))'$ for every $\xi \in \mathbb{R}^d$. As $(H^s(\mathbb{R}^d))'$ is isometrically isomorphic to $H^{-s}(\mathbb{R}^d)$ [Appendix 6.2, Remark 7 c)] it is sufficient to show that $\delta_\xi \in H^{-s}(\mathbb{R}^d)$ for every $\xi \in \mathbb{R}^d$. Further, $\delta_\xi \in \mathcal{S}'$ [Appendix 6.2, Example 9] and $\widehat{\delta_\xi} = (2\pi)^{-\frac{d}{2}} e^{-i\langle \cdot, \xi\rangle}$ as

$$\widehat{\delta_\xi}(\phi) = \delta_\xi(\hat{\phi}) = \hat{\phi}(\xi) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \phi(x) e^{-i\langle x, \xi\rangle} \mathrm{d}x = ((2\pi)^{-\frac{d}{2}} e^{-i\langle \cdot, \xi\rangle})(\phi)$$

for every $\phi \in \mathcal{S}$. Therefore, $\widehat{\delta_\xi} \in L^1_{loc}(\mathbb{R}^d)$ and

$$\begin{aligned} \|\delta_\xi\|^2_{H^{-s}(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} |1 + |x|^2|^{-s} |\widehat{\delta_\xi}(x)|^2 \mathrm{d}x = (2\pi)^{-d} \int_{\mathbb{R}^d} (1 + |x|^2)^{-s} \mathrm{d}x \\ &= (2\pi)^{-d} w_{d-1} \int_0^\infty (1 + r^2)^{-s} r^{d-1} \mathrm{d}r < \infty, \end{aligned}$$

if and only if $-2s + d - 1 < -1$ [18, chapter 9]. Thus $s > \frac{d}{2}$. $\square$

Note that the reproducing kernel of a fractional Sobolev space for $s > \frac{d}{2}$ is difficult to determine. The next section helps to handle this problem by constructing a fractional Sobolev space with reproducing property by a given kernel function.

2.3 Construction of a Fractional Sobolev Space by a Kernel Function

Given a kernel function Φ, we aim to construct a reproducing kernel Hilbert space respectively a Sobolev space of fractional order. For the subsequent construction process, the notion of positive (semi)-definiteness is crucial.

Definition 2.
A continuous function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called positive (semi)-definite if for all $M \in \mathbb{N}$ and for all sets of pairwise distinct centers $\mathcal{C} = \{x_1, \dots, x_M\} \subseteq \mathbb{R}^d$, the quadratic matrix

$$A_{\Phi,\mathcal{C}} := (\Phi(x_i, x_j))_{1 \le i,j \le M}$$

is positive (semi)-definite.

We call a function positive definite if the associated matrices $A_{\Phi,\mathcal{C}}$ are positive definite and positive semi-definite if $A_{\Phi,\mathcal{C}}$ are positive semi-definite. Although this seems to be natural, for historical reasons an alternative terminology exists in the literature: some authors call a function positive definite if the associated matrices are positive semi-definite and strictly positive definite if the matrices are positive definite. We do not follow this historical approach here.

Our next result shows that a kernel of a RKHS is positive semi-definite.

Corollary 5. *[52, Theorem 10.4]*
Suppose $H(\Omega)$ is a RKHS with kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$. Then Φ is positive semi-definite. Further, Φ is positive definite if and only if the point evaluation functionals are linearly independent in $H(\Omega)'$.

Proof. For pairwise distinct $x_1, \dots, x_N$ and $\alpha \in \mathbb{R}^N \backslash \{0\}$ we have

$$\langle \alpha, A_{\Phi,\mathcal{C}}\alpha \rangle = \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_j \alpha_k \Phi(x_j, x_k) = \langle \sum_{j=1}^{N} \alpha_j \delta_{x_j}, \sum_{k=1}^{N} \alpha_k \delta_{x_k} \rangle_{H'} = \| \sum_{j=1}^{N} \alpha_j \delta_{x_j} \|^2_{H'} \ge 0.$$

The last expression will only be zero if point evaluation functionals are linearly dependent. □

The reproducing kernel of a function space $H(\Omega)$ is positive semi-definite and symmetric [Corollary 5, Lemma 2]. It is therefore not too surprising that the usage of a symmetric positive definite kernel function Φ is already sufficient for the construction of a RKHS with Φ as its reproducing kernel.

Proposition 6. *[52, Theorem 10.7]*
Let $\Phi : \Omega \times \Omega \to \mathbb{R}$ be continuous, symmetric and positive definite. Define the $\mathbb{R}$-linear space

$$H_\Phi(\Omega) := \operatorname{span}\{ \Phi(\cdot, y) \mid y \in \Omega \}$$

with the bilinear form

$$\langle \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j), \sum_{k=1}^{M} \beta_k \Phi(\cdot, x_k) \rangle_\Phi := \sum_{j=1}^{N} \sum_{k=1}^{M} \alpha_j \beta_k \Phi(x_j, x_k).$$

Then $H_\Phi(\Omega)$ is a pre-Hilbert space with inner product $\langle \cdot, \cdot \rangle_\Phi$ and reproducing kernel Φ.

Proof. It is strainghtforward to see that $\langle \cdot, \cdot \rangle_\Phi$ is bilinear and symmetric due to the symmetry of Φ. Further, if we choose an arbitrary function $\tilde{f} = \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j) \in H_\Phi(\Omega) \backslash \{0\}$ it follows that

$$\langle \tilde{f}, \tilde{f} \rangle_\Phi = \sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_j \alpha_k \Phi(x_j, x_k) = \langle \alpha, A_{\Phi, \mathcal{C}} \alpha \rangle > 0$$

due to the positive definiteness of Φ. Finally, we obtain for $\tilde{f}$

$$\langle \tilde{f}, \Phi(\cdot, y) \rangle_\Phi = \sum_{j=1}^{N} \alpha_j \Phi(x_j, y) = \tilde{f}(y),$$

which establishes the reproducing property of the kernel. □

The completion of $H_\Phi(\Omega)$ with respect to the $\| \cdot \|_\Phi$ norm is the suitable candidate for a Hilbert space with reproducing kernel Φ.

Lemma 7. *[52, Theorem 10.10]*
Suppose $\Phi : \Omega \times \Omega \to \mathbb{R}$ is a continuous, symmetric and positive definite kernel. Then

$$N_\Phi(\Omega) := \overline{H_\Phi(\Omega)}^{\| \cdot \|_\Phi}$$

is a Hilbert space with reproducing kernel Φ. $N_\Phi(\Omega)$ is called the native space.

Remark 1.

a) The functions in $N_\Phi(\Omega)$ are continuous, i.e. $N_\Phi(\Omega) \subseteq C(\Omega)$. This can be shown by using the reproducing property of the kernel Φ, the Cauchy Schwartz inequality and the continuity of the kernel Φ.

b) The native space is unique in the following sense: if $G(\Omega)$ is a Hilbert space with positive definite kernel Φ, then $N_\Phi(\Omega) = G$ and the inner products are the same [52, Theorem 10.11].

The upcoming result forges the bridge between native spaces and fractional Sobolev spaces. The theorem makes use of the Fourier transform and certain invariance properties of the kernel Φ. Especially the invariance properties seem quite restrictive, but we refer the reader to [52, chapter 10] for a discussion about invariances of function spaces which are inherited by their reproducing kernels. In particular for $\Omega = \mathbb{R}^d$ those invariance properties seem to arise naturally.

Definition 3.
Let $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a given kernel function. We call the kernel Φ translation invariant, if there exists a function $\tilde{\Phi} : \mathbb{R}^d \to \mathbb{R}$ such that $\Phi(x, y) = \tilde{\Phi}(x - y)$ for every $x, y \in \mathbb{R}^d$. Further, we call Φ radial, if there exists a continuous function $\phi : [0\ \infty) \to \mathbb{R}$ such that $\Phi(x, y) = \phi(|x - y|)$ for every $x, y \in \mathbb{R}^d$.

Theorem 8. *[52, Corollary 10.13, Theorem 10.45, Corollary 10.48]*
Suppose that $\tilde{\Phi} \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ and let the kernel $\Phi(x, y) := \tilde{\Phi}(x - y)$ be positive definite. If

$$C_1(1 + |\xi|^2)^{-s} \leq \widehat{\tilde{\Phi}}(\xi) \leq C_2(1 + |\xi|^2)^{-s}, \quad \xi \in \mathbb{R}^d, \tag{9}$$

with $s > k + \frac{d}{2}$ for $k \in \mathbb{N}_0$ and constants $C_1, C_2 > 0$, then

$$N_\Phi(\mathbb{R}^d) = H^s(\mathbb{R}^d) \subseteq C^k(\mathbb{R}^d), \quad \tilde{\Phi} \in C^{2k}(\mathbb{R}^d),$$

where the native space inner product

$$\langle f, g \rangle_{N_\Phi(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \frac{\hat{f}(\xi)\hat{g}(\xi)}{\widehat{\tilde{\Phi}}(\xi)} \mathrm{d}\xi$$

and the usual inner product on $H^s(\mathbb{R}^d)$ are equivalent. If in addition $\Omega \subseteq \mathbb{R}^d$ is a Lipschitz domain and $s \in \mathbb{N}$, then $N_\Phi(\Omega) = H^s(\Omega)$ again with equivalent inner products, where

$$N_\Phi(\Omega) = \{f|_\Omega \mid f \in N_\Phi(\mathbb{R}^d)\}.$$

Given a continuous integrable function whose Fourier transform decays like (9), we can construct a native space for a translation invariant kernel function which is in particular a fractional Sobolev space (up to the equivalence of the norms). In the upcoming subsection we present well-known radial kernels satisfying the conditions stated in Theorem 8. Even though most kernels which generate fractional Sobolev spaces are radial, there exist also kernels which are not even translation invariant [33].

2.4 Positive Definite Radial Kernel Functions

In this thesis we turn our attention to positive definite radial kernels which generate fractional Sobolev spaces of particular order. Radial kernels have the property of being invariant under all Euclidean transformations, i.e. under translations, rotations and reflections. This is due to fact that Euclidean transformations are characterized by orthogonal transformation matrices and are therefore Euclidean norm invariant. Further, the usage of radial kernels benefits from the fact that the problem becomes insensitive to the dimension $d \in \mathbb{N}$ of the space in which the problem is embedded. Instead of having to deal with a multivariate function Φ (whose complexity will increase with increasing space dimension $d \in \mathbb{N}$) we work with the same univariate function ϕ for all choices of d. When dealing with a radial kernel function

$$\Phi(x, y) = \tilde{\Phi}(x - y) = \phi(|x - y|),$$

it is convenient to also refer to the univariate function ϕ as a positive definite radial function. While this presents a slight abuse of the terminology for positive definite functions, this is what is commonly done in the literature. With this notational convention we can give the subsequent integral characterization proved by Wendland [52], which is nothing else than Bochner's characterization [11] for the radial kernel case.

Theorem 9. *[52, Theorem 6.18]*
Suppose that $\phi \in C([0\ \infty))$ satisfies $r \mapsto r^{d-1}\phi(r) \in L^1([0\ \infty))$. Then ϕ is positive definite on $\mathbb{R}^d$ if and only if ϕ is bounded and the d-dimensional Fourier transform

$$\mathcal{F}_d(\phi)(r) := r^{-\frac{d-2}{2}} \int_0^\infty \phi(t) t^{\frac{d}{2}} J_{\frac{d-2}{2}}(rt) \mathrm{d}t$$

is non-negative and non-vanishing. Here, $J_{\frac{d-2}{2}}$ is the classical Bessel function of the first kind of oder $\frac{d-2}{2}$.

The operator $\mathcal{F}_d$ defined by $\mathcal{F}_d(\phi) = \widehat{\tilde{\Phi}}$ acts on univariate functions. Hence, if working with radial functions we are in a situation where we can do most of the analysis in a univariate setting, which often makes things easier.

We now present a number of functions that are covered by the theory presented thus far and give as a reference the textbook of Fasshauer [14]. In order to influence the shape and localization of a given kernel, we make use of a shape parameter $\varepsilon > 0$ by rescaling $x \in \mathbb{R}^d$ to εx. Our use of the shape parameter as a factor applied directly to x has the advantage of providing a unified treatment in which a decrease of the shape parameter always has the effect of producing a "flat" kernel, while increasing ε leads to a more "peaked" (or localized) kernel. The correct choice of the shape parameter in a particular reconstruction problem is crucial.

Example 1. *Gaussian kernel*
Let $\phi : [0\ \infty) \to \mathbb{R}$, $\phi(r) := e^{-r^2}$. The Gaussian kernel

$$\Phi_\varepsilon(x, y) = \phi(\varepsilon|x - y|) = e^{-\epsilon^2|x-y|^2}, \quad x, y \in \mathbb{R}^d$$

is positive definite and radial on $\mathbb{R}^d$ for any dimension $d \in \mathbb{N}$ [52, Theorem 6.10], since

$$\widehat{\tilde{\Phi}}_\varepsilon(\xi) = 2^{-\frac{d}{2}} \varepsilon^{-d} e^{-\frac{|\xi|^2}{4\varepsilon^2}} > 0, \quad \xi \in \mathbb{R}^d.$$

The native space of a Gaussian is contained in the fractional Sobolev space $H^s(\mathbb{R}^d)$ for any $s \in \mathbb{R}^+$ [41] , i.e.

$$N_{\Phi_\varepsilon}(\mathbb{R}^d) \subseteq \bigcap_{s \in \mathbb{R}^+} H^s(\mathbb{R}^d) \subseteq C^\infty(\mathbb{R}^d).$$

For a bounded Lipschitz set $\Omega \subseteq \mathbb{R}^d$, $N_{\Phi_\varepsilon}(\Omega) \subseteq \bigcap_{s \in \mathbb{N}} H^s(\Omega)$ holds [41]. Even though the native space of Gaussians is small, it contains the important class of so-called band-limited functions, i.e. functions whose Fourier transform is compactly supported.

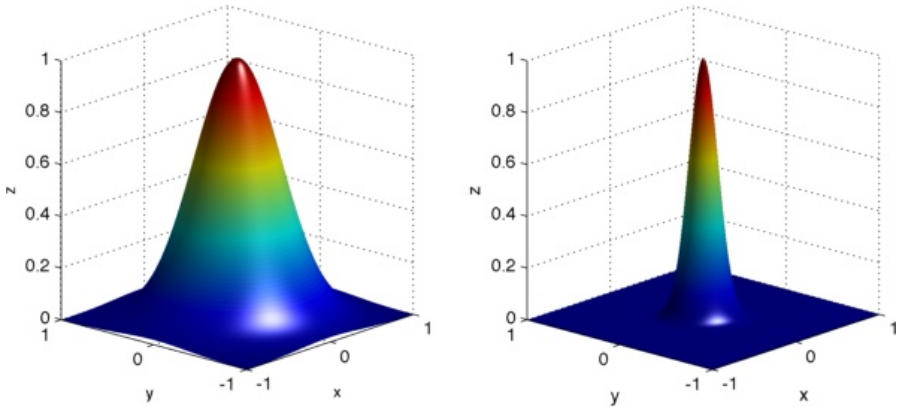

Figure 2: Gaussian kernel $\tilde{\Phi}_\varepsilon$ with shape parameter $\varepsilon = 2$ (left) and $\varepsilon = 6$ (right) in $\mathbb{R}^2$.

Example 2. *Wendland's compactly supported functions*
Probably the most popular family of compactly supported radial functions presently in use was constructed by Wendland in [51]. To this end, let the cut-off function be defined as

$$(x)_+ := \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} .$$

Wendland starts with the truncated power function $\phi_{\varepsilon,l}(r) = (1 - \varepsilon r)_+^l$ and then walks through dimensions by repeatedly applying an integral operator to generate further positive radial functions of higher smoothness [15,51,52]. Functions constructed in this way are positive definite, radial and have compact support. On their support they can be represented by univariate polynomials. An overview about the most important Wendland functions is given in Table 1. For the solution of problem (8), it is mainly important that they satisfy the condition of Theorem 8 with $s = \tau + \frac{d+1}{2}$, where $\tau \in \mathbb{N}_0$ is a given smoothness index. Thus, $N_{\Phi_\varepsilon}(\mathbb{R}^d) = H^{\tau+\frac{d+1}{2}}(\mathbb{R}^d)$ [52, Theorem 10.35]. For a bounded Lipschitz domain $\Omega \subseteq \mathbb{R}^d$ and $\tau + \frac{d+1}{2} \in \mathbb{N}$, $N_{\Phi_\varepsilon}(\Omega) = H^{\tau+\frac{d+1}{2}}(\Omega)$ holds.

dimension	function	smoothness
$d = 1$	$\phi_{\varepsilon,1,0}(r) = (1 - \varepsilon r)_+$	$C^0(\mathbb{R}^d)$
	$\phi_{\varepsilon,1,1}(r) = (1 - \varepsilon r)_+^3(3\varepsilon r + 1)$	$C^2(\mathbb{R}^d)$
	$\phi_{\varepsilon,1,2}(r) = (1 - \varepsilon r)_+^5(8\varepsilon^2 r^2 + 5\varepsilon r + 1)$	$C^4(\mathbb{R}^d)$
$d \leq 3$	$\phi_{\varepsilon,3,0}(r) = (1 - \varepsilon r)_+^2$	$C^0(\mathbb{R}^d)$
	$\phi_{\varepsilon,3,1}(r) = (1 - \varepsilon r)_+^4(4\varepsilon r + 1)$	$C^2(\mathbb{R}^d)$
	$\phi_{\varepsilon,3,2}(r) = (1 - \varepsilon r)_+^6(35\varepsilon^2 r^2 + 18\varepsilon r + 3)$	$C^4(\mathbb{R}^d)$
	$\phi_{\varepsilon,3,3}(r) = (1 - \varepsilon r)_+^8(32\varepsilon^3 r^3 + 25\varepsilon^2 r^2 + 8\varepsilon r + 1)$	$C^6(\mathbb{R}^d)$

Table 1: Wendland functions $\phi_{\varepsilon,d,\tau}$.

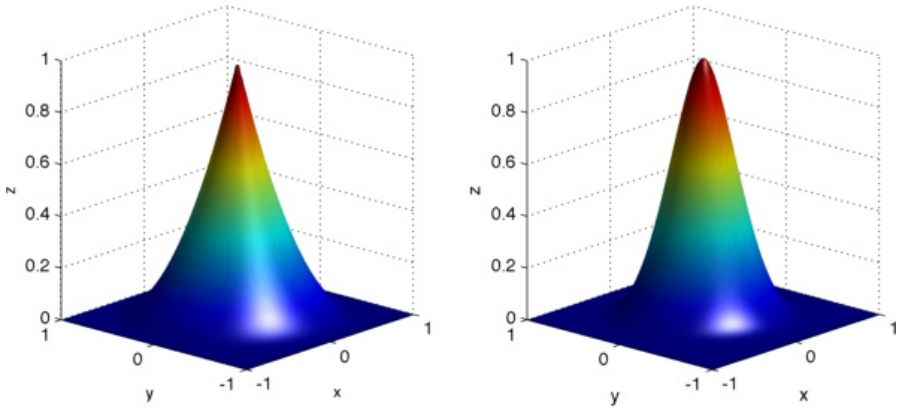

Figure 3: Wendland's compactly supported kernel $\tilde{\Phi}_\varepsilon$ with generating univariate functions $\phi_{1,3,0}$ (left) and $\phi_{1,3,1}$ (right) in $\mathbb{R}^2$.

2.5 Notes and Comments

RKHS are also frequently used in machine learning or learning theory, which considers the synthesis of an unknown function f^* from discrete data $(x_1, y_1), \ldots, (x_N, y_N)$ from a statistical point of view. A reconstruction f should have a good approximation property on training data sites x_i, i.e. $f(x_i) \approx y_i$. Further, the reconstruction should predict function values on previously unknown data. Central questions are how to choose the hypothesis space $H(\Omega)$ in which an approximation is to be reconstructed, and in which way the given data is to be used efficiently [12, 35].

Positive (semi)-definite functions play an important role not only in the construction of a RKHS, but also in other mathematical areas such as scattered data interpolation or probability theory. In probability theory, a positive (semi)-definite function is nothing other than the characteristic function of a probability distribution. But in contrast with probability theory, where semi-definite functions work as well as positive definite ones, scattered data interpolation has to stick with positive definite functions. This is due to the fact that being positive definite is crucial for the inversion of the fundamental interpolation matrix. For a comprehensive textbook on multivariate approximation theory and scattered data interpolation, see [52].

3 Solving the Semi-Discrete Inverse Problem of Photoacoustic Tomography

In this chapter, we give a detailed description of two solution methods for the semi-discrete inverse problem of PAT (8). The basic idea is to formulate (8) as a generalized interpolation problem. To explain the general framework of generalized interpolation, let $\lambda_1, \dots, \lambda_M$ be real-valued linear functionals on a Hilbert function space $H(\Omega)$. Suppose we are given a data vector $(g_1, \dots, g_M)^T \in \mathbb{R}^M$. Then a generalized interpolant is a function $f^+ \in H(\Omega)$ satisfying

$$\lambda_i(f^+) = g_i, \quad i = 1, \dots, M. \tag{10}$$

This general framework covers our reconstruction problem (8) when not only the data vector is given as in (6), but also linear functionals $\lambda_{kl} : H(\Omega) \to \mathbb{R}$ exist such that the collocation operator (7) satisfies

$$\mathcal{M}_{\mathcal{X}} : H(\Omega) \to \mathbb{R}^{K \cdot L}, \qquad (\mathcal{M}_{\mathcal{X}}(f))_{kl} = \lambda_{kl}(f).$$

The existence of these functionals is examined in section 3.1.

In order to solve the generalized interpolation problem (10), we introduce the theory of optimal recovery [section 3.2] and the algebraic reconstruction technique [section 3.3]. Both methods determine a generalized interpolant in a finite-dimensional subspace X (*recovery space*) of a native space $H(\Omega) = N_\Phi(\Omega)$ generated by a radial kernel Φ. First, let us collect assumptions used within this chapter.

Assumption 1.

1. *The acquisition surface $\partial\Omega$ is the boundary of the open ball of radius $r > 0$ centered around the origin, i.e. $\Omega = B_r(0)$ and $\partial\Omega = rS^{d-1}$.*

2. *For a fixed integer $\tau > \frac{d}{2}$, the experimental output vector $\bar{g} \in \mathbb{R}^{K \cdot L}$ is a sampled version of the spherical mean $g = \mathcal{M}_{\partial\Omega}(f^*)$ of an unknown absorption function $f^* \in H_c^\tau(\Omega)$, i.e.*

$$g_{kl} = g(\xi_k, t_l) = \mathcal{M}_{\partial\Omega}(f^*)(\xi_k, t_l), \quad (\xi_k, t_l) \in \Xi \times \mathcal{T}.$$

3. *We are given a radial kernel*

$$\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad \Phi(x, y) = \tilde{\Phi}(x - y) = \phi(|x - y|), \tag{11}$$

where $\tilde{\Phi}$ satisfies the conditions of Theorem 8 with the above specified $\tau > \frac{d}{2}$. Hence, according to Theorem 8, $N_\Phi(\Omega) = H^\tau(\Omega)$ with equivalent inner products.

3.1 Getting Started: Generalized Interpolation Formulation

In order to solve the semi-discrete inverse problem of PAT (8), we formulate it as a generalized interpolation problem (10). To this end, we need to prove the existence of linear functionals $\lambda_{kl} : N_\Phi(\Omega) \to \mathbb{R}$ to write the collocation operator (7) as

$$\mathcal{M}_\mathcal{X} : N_\Phi(\Omega) \to \mathbb{R}^{K\cdot L}, \qquad (\mathcal{M}_\mathcal{X}(f))_{kl} = \lambda_{kl}(f). \tag{12}$$

The two solution methods we propose for the generalized interpolation problem (10) have different requirements in terms of the continuity of the functionals. Continuity is essential when solving (10) with the theory of optimal recovery [section 3.2] (due to the usage of the Riesz representation theorem [Appendix 6.1]). Whereas continuity can be neglected when dealing with the algebraic reconstruction technique [section 3.3]. We therefore present two different sets of *experimental functionals*: one with continuous functionals for the optimal recovery theory and another with not necessarily continuous functionals for the algebraic reconstruction technique.

Let us start with the functionals for the algebraic reconstruction technique. Clearly, the functionals

$$\lambda_{kl} : N_\Phi(\Omega) \to \mathbb{R}, \quad \lambda_{kl} := \delta_{(\xi_k,t_l)} \circ \mathcal{M}_{\partial\Omega}, \quad (\xi_k, t_l) \in \Xi \times \mathcal{T} \tag{13}$$

satisfy (12). However, continuity of the functionals in (13) cannot be guaranteed, since a continuity result is so far missing for the spherical Radon transform $\mathcal{M}_{\partial\Omega}$ as an operator defined on the whole space $N_\Phi(\Omega)$ respectively $H^\tau(\Omega)$. Solely a weaker version exists [Proposition 10].

Definition 4.
Given spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$, the set of experimental functionals for the algebraic reconstruction technique is denoted by $\Lambda_\Omega := \{\lambda_{11}, \dots, \lambda_{KL}\}$, where λ_{kl} is defined in (13).

In order to prove the existence of linear *continuous* functionals admitting a representation of the collocation operator as in (12), the subsequent Sobolev space estimates for the spherical Radon transform are essential.

Proposition 10. *[34, Corollary 8.1]*
Let $\mathcal{K}$ be a compact subset of Ω. The spherical Radon transform $\mathcal{M}_{\partial\Omega} : L^2(\Omega) \to L^2(\partial\Omega \times (0\ T])$ is smoothing of degree $\frac{d-1}{2}$, i.e. for every $s \in \mathbb{R}$ there are positive constants $m_{s,\mathcal{K}}$, $M_{s,\mathcal{K}}$ such that

$$m_{s,\mathcal{K}} \|f\|_{H^s(\Omega)} \leq \|\mathcal{M}_{\partial\Omega}(f)\|_{H^{s+(d-1)/2}(\partial\Omega\times(0\ T])} \leq M_{s,\mathcal{K}} \|f\|_{H^s(\Omega)}$$

for every $f \in H^s(\Omega)$ with $\operatorname{supp}(f) \subseteq \mathcal{K}$.

The previous proposition provides an important continuity result for the spherical Radon transform as an operator defined on fractional Sobolev spaces consisting of functions with compact support. Therewith we can give the following existence theorem.

Theorem 11.
Let $\mathcal{K} \subseteq \Omega$ be an arbitrary, but fixed, compact set and let $(\xi, t) \in \partial\Omega \times (0\ T]$. Suppose there exists a linear continuous extension operator

$$E : H^{\tau+\frac{d-1}{2}}(\partial\Omega \times (0\ T]) \to H^{\tau+\frac{d-1}{2}}(\mathbb{R}^{d+1})$$

such that $E(g)(x) = g(x)$ for almost every $x \in \partial\Omega \times (0\ T]$ and every $g \in H^{\tau+\frac{d-1}{2}}(\partial\Omega \times (0\ T])$. Then there exists a linear continuous functional $\lambda^c_{(\xi,t)} : N_\Phi(\Omega) \to \mathbb{R}$ such that $\lambda^c_{(\xi,t)}(f) = \mathcal{M}_{\partial\Omega}(f)(\xi, t)$ for every $f \in N_\Phi(\Omega)$ with $\operatorname{supp}(f) \subseteq \mathcal{K}$.

Proof. Define $H^\tau_{\mathcal{K}}(\Omega) := \{f \in H^\tau(\Omega) \mid \operatorname{supp}(f) \subseteq \mathcal{K}\}$. From Proposition 10 it follows that $\mathcal{M}_{\partial\Omega}$ is continuous as an operator from $H^\tau_{\mathcal{K}}(\Omega)$ into $H^{\tau+\frac{d-1}{2}}(\partial\Omega \times (0\ T])$. Theorem 4 shows that $H^{\tau+\frac{d-1}{2}}(\mathbb{R}^{d+1})$ is a RKHS, as $\tau + \frac{d-1}{2} > \frac{d+1}{2}$ for any space dimension $d \geq 2$. Hence, point evaluation functionals defined on $H^{\tau+\frac{d-1}{2}}(\mathbb{R}^d)$ are continuous [Lemma 1]. Let E be the linear continuous extension operator, then

$$\widetilde{\lambda}^c_{(\xi,t)} := \delta_{(\xi,t)} \circ E \circ \mathcal{M}_{\partial\Omega} \in H^\tau_{\mathcal{K}}(\Omega)'$$

for every $(\xi, t) \in \partial\Omega \times (0\ T]$. As $H^\tau_{\mathcal{K}}(\Omega)$ is a subspace of $H^\tau(\Omega)$, the Hahn-Banach Theorem [Appendix 6.1] guarantees the existence of a linear continuous extension $\lambda^c_{(\xi,t)} : H^\tau(\Omega) \to \mathbb{R}$ of the functional $\widetilde{\lambda}^c_{(\xi,t)}$ such that

$$\widetilde{\lambda}^c_{(\xi,t)}(f) = \mathcal{M}_{\partial\Omega}(f)(\xi, t) = \lambda^c_{(\xi,t)}(f)$$

for every $f \in H^\tau_{\mathcal{K}}(\Omega)$. Due to the equivalence of the native space norm and the H^τ-norm [Theorem 8], $\lambda^c_{(\xi,t)}$ is continuous as a functional from $N_\Phi(\Omega)$ into $\mathbb{R}$. This proves the assertion. □

Remark 2.

a) *To the present days, results for an extension of $f \in H^s(\Omega')$ to a function in $H^s(\mathbb{R}^{d+1})$ are solely given for Ω' being an open subset of $\mathbb{R}^{d+1}$ [2, 50]. As $\partial\Omega \times (0\ T]$ is not open in $\mathbb{R}^{d+1}$, the requirement in the theorem is needed. However, throughout this thesis we assume the existence of such an extension.*

b) *Note that, given a fixed compact set $\mathcal{K} \subseteq \Omega$, the functional $\lambda^c_{(\xi,t)}$ has the representation $\lambda^c_{(\xi,t)}(f) = \mathcal{M}_{\partial\Omega}(f)(\xi, t)$ only for functions f with support in $\mathcal{K}$. Its representation for functions which are not supported in $\mathcal{K}$, is not explicitly given. This illustrates the strong dependency of the functional $\lambda^c_{(\xi,t)}$ on the compact set $\mathcal{K}$. In order to have an explicit representation for a large class of functions, let us consider the previous theorem for*

$$\mathcal{K}_\delta := \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) \geq \delta\}$$

with an arbitrary small $\delta > 0$.

With the previous result, we note the continuous analogue to Definition 4.

Definition 5.
Let $\mathcal{K}_\delta \subseteq \Omega$ for an arbitrary small $\delta > 0$. Given spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$, the set of experimental functionals for the optimal recovery theory is denoted by $\Lambda_{\mathcal{K}_\delta} := \{\lambda^c_{11}, \dots, \lambda^c_{KL}\} \subseteq N_\Phi(\Omega)'$, where

$$\lambda^c_{kl} : N_\Phi(\Omega) \to \mathbb{R}, \quad \lambda^c_{kl}(f) = \mathcal{M}_{\partial\Omega}(f)(\xi_k, t_l), \quad (\xi_k, t_l) \in \mathcal{X} = \Xi \times \mathcal{T}$$

for every $f \in N_\Phi(\Omega)$ with $\operatorname{supp}(f) \subseteq \mathcal{K}_\delta$.

In comparison to the set Λ_Ω, we gained continuity for the functionals, but we lost an explicit representation for functions not supported in $\mathcal{K}_\delta$. Furthermore, the collocation operator $\mathcal{M}_\mathcal{X}$ (7) satisfies

$$(\mathcal{M}_\mathcal{X}(f))_{kl} = \mathcal{M}_{\partial\Omega}(f)(\xi_k, t_l) = \lambda^c_{kl}(f)$$

solely for functions $f \in N_\Phi(\Omega)$ with $\operatorname{supp}(f) \subseteq \mathcal{K}_\delta$. In order to have a representation of the collocation operator as in (12), let us define, in the presence of $\Lambda_{\mathcal{K}_\delta}$,

$$\mathcal{M}_\mathcal{X} : N_\Phi(\Omega) \to \mathbb{R}^{K \cdot L}, \quad (\mathcal{M}_\mathcal{X}(f))_{kl} := \lambda^c_{kl}(f). \tag{14}$$

3.2 Optimal Recovery

The theory of optimal recovery is a well-known solution method for the generalized interpolation problem (10), which determines the norm-minimal generalized interpolant [24]. As the norm-minimal solution of (10) is an element of

$$X_{K\cdot L} = \operatorname{span}\{\, \lambda_{kl}^{c\ y}\Phi(\cdot, y) \mid \lambda_{kl}^{c} \in \Lambda_{\mathcal{K}_\delta}\},$$

$X_{K\cdot L}$ arises "naturally" as a recovery space. In the following we detail the determination of the norm-minimal generalized interpolant, a L^2-error bound in terms of the data density and a proof of convergence.

3.2.1 Solution Concept

Assume we are given continuous experimental functionals $\Lambda = \Lambda_{\mathcal{K}_\delta}$ [Definition 5] for a sufficiently small $\delta > 0$. For the sake of notational simplicity, we drop the superscript c at the continuous functionals $\lambda_{kl}^c \in \Lambda$.

In order to determine the norm-minimal generalized interpolant, an analysis of the collocation operator $\mathcal{M}_\mathcal{X}$ (14) is expedient. Essential for the analysis is the Riesz representation theorem [Appendix 6.1], which identifies with every functional in Λ its unique Riesz representer. Even though the Riesz representation theorem only guarantees the existence of a unique Riesz representer, in our particular case, we know their explicit form: Proposition 3 yields that the Riesz representer of a functional in a RKHS is simply given by applying the functional to one argument of the kernel.

Definition 6.
The set of Riesz representers of the functionals in Λ is denoted by

$$\mathcal{V} := \{\lambda_{kl}^{y}\Phi(\cdot, y) \mid \lambda_{kl} \in \Lambda\} \subseteq N_\Phi(\Omega).$$

We call the linear span of $\mathcal{V}$ optimal recovery space and denote it by $X_D := \operatorname{span}(\mathcal{V})$, where the index $D \in \mathbb{N}$ describes its dimension.

With the previous considerations, we can give the following characterization of the collocation operator $\mathcal{M}_\mathcal{X}$.

Proposition 12.
The collocation operator $\mathcal{M}_\mathcal{X} : N_\Phi(\Omega) \to \mathbb{R}^{K\cdot L}$ has the following properties:

1. *$\mathcal{M}_\mathcal{X}$ is linear and continuous,*
2. *$(\mathcal{M}_\mathcal{X}(f))_{kl} = \langle f, \lambda_{kl}^{y}\Phi(\cdot, y)\rangle_{N_\Phi}$,*
3. *$\operatorname{Ran}(\mathcal{M}_\mathcal{X}) = \mathbb{R}^{K\cdot L}$ if and only if Λ is linear independent,*
4. *$\operatorname{Ker}(\mathcal{M}_\mathcal{X}) = X_D^\perp$.*

Proof. The linearity and continuity of the collocation operator is due to the definition of $\mathcal{M}_{\mathcal{X}}$ (14) and the fact that $\Lambda \subseteq N_\Phi(\Omega)'$. The second property follows from (14) and Proposition 3. Let us now focus on the range of $\mathcal{M}_{\mathcal{X}}$. First, note that the set of functionals Λ is linear independent if and only if the set $\mathcal{V}$ is linear independent. Suppose $\mathcal{M}_{\mathcal{X}}$ is not surjective, i.e. $\mathcal{M}_{\mathcal{X}}(N_\Phi(\Omega)) \subsetneq \mathbb{R}^{K\cdot L}$. Then, there exists a vector $x \in \mathcal{M}_{\mathcal{X}}(N_\Phi(\Omega))^\perp \backslash \{0\}$ such that $\langle \mathcal{M}_{\mathcal{X}}(f), x\rangle = 0$ for all $f \in N_\Phi(\Omega)$. Thus

$$\langle \mathcal{M}_{\mathcal{X}}(f), x\rangle = \sum_{k=1}^{K}\sum_{l=1}^{L} \langle f, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} x_{kl} = \langle f, \sum_{k=1}^{K}\sum_{l=1}^{L} x_{kl}\lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = 0 \tag{15}$$

for all $f \in N_\Phi(\Omega)$. Inserting $f = \sum_{k=1}^{K}\sum_{l=1}^{L} x_{kl}\lambda_{kl}^y \Phi(\cdot, y) \in N_\Phi(\Omega)$ in (15) yields

$$\| \sum_{k=1}^{K}\sum_{l=1}^{L} x_{kl}\lambda_{kl}^y \Phi(\cdot, y)\|_{N_\Phi} = 0$$

and thus

$$\sum_{k=1}^{K}\sum_{l=1}^{L} x_{kl}\lambda_{kl}^y \Phi(\cdot, y) = 0.$$

As $x \neq 0$, the linear dependency of Λ follows; this is in contradiction to the assumption. The opposite direction is shown by applying the same steps backwards. The assertion about the nullspace $\mathrm{Ker}(\mathcal{M}_{\mathcal{X}})$ is a direct consequence of the definition of $\mathcal{M}_{\mathcal{X}}$ (14) and the definition of the orthogonal complement, as

$$f \in X_D^\perp \Leftrightarrow \langle f, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = 0 \Leftrightarrow (\mathcal{M}_{\mathcal{X}}(f))_{kl} = 0$$

for every $k = 1, \ldots, K$ and $l = 1, \ldots, L$. □

Clearly, we are interested in solutions of the equation $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$ in $N_\Phi(\Omega)$. Existence is guaranteed in the situation of linear independency of Λ.

Corollary 13.
Let the set of functionals $\Lambda \subseteq N_\Phi(\Omega)'$ be linearly independent. Then

$$\{f \in N_\Phi(\Omega) \mid \mathcal{M}_{\mathcal{X}}(f) = \bar{g}\} \neq \emptyset.$$

Proof. As the set Λ is linear independent, the collocation operator $\mathcal{M}_{\mathcal{X}}$ is surjective. □

The previous corollary implies - in the situation of linear independency of Λ - that there exists a function in $N_\Phi(\Omega)$ satisfying $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$. However, the function is not unique. To this end, let Λ be linear independent and let $f_0 \in N_\Phi(\Omega)$ be a function satisfying $\mathcal{M}_{\mathcal{X}}(f_0) = \bar{g}$. Further, $X_{K\cdot L}^\perp$ is of infinite dimension, as $N_\Phi(\Omega) = X_{K\cdot L} \oplus X_{K\cdot L}^\perp$. Since $\mathrm{Ker}(\mathcal{M}_{\mathcal{X}}) = X_{K\cdot L}^\perp$, it follows that for any function $f_1 \in X_{K\cdot L}^\perp$, $f = f_0 + f_1$ solves the semi-discrete equation. Thus, there is a lack of uniqueness. However, we can enforce uniqueness by searching for a minimal norm solution.

Theorem 14. *[52, Theorem 16.1]*
Let the set Λ be linear independent. The solution of the optimal recovery problem

$$\|f^+\|_{N_\Phi} = \min_{\substack{f\in N_\Phi(\Omega)\\ \mathcal{M}_\mathcal{X}(f)=\bar{g}}} \|f\|_{N_\Phi} \tag{16}$$

is unique and is given by

$$f^+ = \sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{kl}\lambda_{kl}^y \Phi(\cdot, y) \in X_{K\cdot L},$$

where the coefficient vector $\alpha \in \mathbb{R}^{K\cdot L}$ is determined by the system of linear equations $M_{\Lambda,\Phi}\alpha = \bar{g}$ with

$$(M_{\Lambda,\Phi})_{kl,k'l'} = \langle \lambda_{kl}^y \Phi(\cdot, y), \lambda_{k'l'}^y \Phi(\cdot, y)\rangle_{N_\Phi} = \langle \lambda_{kl}, \lambda_{k'l'}\rangle_{N_\Phi{}'}.$$

Further, the optimal recovery problem (16) *is well-posed in the sense of Hadamard [Appendix 6.3, Definition 9].*

Proof. By definition, the matrix $M_{\Lambda,\Phi}$ is Gramian and hence positive definite, as the set of functionals Λ is linear independent. Positive definite matrices are invertible and so $f^+ = \sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{kl}\lambda_{kl}^y \Phi(\cdot, y)$ is well-defined. It is straightforward to see that f^+ is a solution of equation (8), as

$$\begin{aligned}
(\mathcal{M}_\mathcal{X}(f^+))_{kl} = \lambda_{kl}(f^+) &= \sum_{k'=1}^{K}\sum_{l'=1}^{L} \alpha_{k'l'}\lambda_{kl}^x \lambda_{k'l'}^y \Phi(x, y)\\
&= \sum_{k'=1}^{K}\sum_{l'=1}^{L} \alpha_{k'l'}\langle \lambda_{k'l'}^y \Phi(\cdot, y), \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi}\\
&= \sum_{k'=1}^{K}\sum_{l'=1}^{L} \alpha_{k'l'}\langle \lambda_{kl}^y \Phi(\cdot, y), \lambda_{k'l'}^y \Phi(\cdot, y)\rangle_{N_\Phi} = \sum_{k'=1}^{K}\sum_{l'=1}^{L} (M_{\Lambda,\Phi})_{kl,k'l'}\alpha_{k'l'} = g_{kl}
\end{aligned}$$

for $k = 1, \dots, K$ and $l = 1, \dots, L$.
Next, we show that f^+ is norm-minimal. To this end, let $\tilde{f} \in N_\Phi(\Omega)$ satisfy $\mathcal{M}_\mathcal{X}(\tilde{f}) = \bar{g}$. Then $f^+ - \tilde{f} \in X_{K\cdot L}^\perp$, as

$$\langle f^+ - \tilde{f}, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = \langle f^+, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} - \langle \tilde{f}, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = 0$$

for $k = 1, \dots, K$ and $l = 1, \dots, L$. Since $f^+ \in X_{K\cdot L}$, this shows that

$$\|f^+\|_{N_\Phi}^2 = \langle f^+, (f^+ - \tilde{f}) + \tilde{f}\rangle_{N_\Phi} = \langle f^+, \tilde{f}\rangle_{N_\Phi} \le \|f^+\|_{N_\Phi}\|\tilde{f}\|_{N_\Phi}.$$

As $f^+ \neq 0$, the minimal norm property follows. Let us now focus on the uniqueness of the solution. Let f' be any solution of the optimal recovery problem, then necessarily

$\langle f', h\rangle_{N_\Phi} = 0$ for all $h \in X_{K\cdot L}^{\perp}\backslash\{0\}$. Otherwise define $s = f' + \beta h$ with $\beta < \frac{-2\langle f', h\rangle_{N_\Phi}}{\|h\|^2_{N_\Phi}}$. Then, $\lambda_{kl}(s) = \lambda_{kl}(f') + \beta\lambda_{kl}(h) = g_{kl}$ and

$$\|s\|^2_{N_\Phi} = \|f'\|^2_{N_\Phi} + \beta^2\|h\|^2_{N_\Phi} + 2\beta\langle f', h\rangle_{N_\Phi} < \|f'\|^2_{N_\Phi},$$

which would make s to a solution of the optimal recovery problem with strictly smaller norm than f'. Thus, if f_1^+ and f_2^+ are both solutions of the optimal recovery problem, then $\langle f_1^+, f_1^+ - f_2^+\rangle_{N_\Phi} = 0$ and $\langle f_2^+, f_1^+ - f_2^+\rangle_{N_\Phi} = 0$ as $f_1^+ - f_2^+ \in X_{K\cdot L}^{\perp}$. This shows

$$\langle f_1^+, f_1^+ - f_2^+\rangle_{N_\Phi} - \langle f_2^+, f_1^+ - f_2^+\rangle_{N_\Phi} = \|f_1^+ - f_2^+\|^2_{N_\Phi} = 0,$$

which proves the uniqueness.
In order to show the well-posedness of the optimal recovery problem, only the stability is left to prove. To this end, note that the Gramian matrix $M_{\Lambda,\Phi}$ is invertible due to the linear independency of Λ. Further, denote with $(M^{-1}_{\Lambda,\Phi})_{kl,k'l'}$ the matrix components of the inverse of $M_{\Lambda,\Phi}$. Then we introduce in $X_{K\cdot L}$ the dual basis

$$w_{kl} = \sum_{k'=1}^{K}\sum_{l'=1}^{L}(M^{-1}_{\Lambda,\Phi})_{kl,k'l'}\lambda^y_{k'l'}\Phi(\cdot, y) \tag{17}$$

for $k = 1, \ldots, K$ and $l = 1, \ldots, L$. It follows that

$$f^+ = \sum_{k=1}^{K}\sum_{l=1}^{L} g_{kl}w_{kl} \tag{18}$$

by inserting (17) into (18) and using the fact that $M^{-1}_{\Lambda,\Phi}\bar{g} = \alpha$. This representation of f^+ clearly shows the continuous dependency: if $\delta\bar{g}$ denotes a variation of $\bar{g}$ and δf^+ the corresponding variation of f^+, then $\|\delta f^+\|_{N_\Phi}$ tends to zero when $|\delta\bar{g}|$ tends to zero. It follows that the computation of f^+ is well-posed in the sense of Hadamard. □

Remark 3.

a) *The latter shows that $X_{K\cdot L}$ arises "naturally" as a recovery space for the semi-discrete inverse problem* (8).

b) *We can interpret the previous theorem as follows: the finite vector $\bar{g} \in \mathbb{R}^{K\cdot L}$ only allows the determination of the projection of the unknown absorption density f^* on the subspace $X_{K\cdot L}$. As a result, the solution $f^+ \in X_{K\cdot L}$ gives no information about the component of f^* orthogonal to $X_{K\cdot L}$. The characterization of the component of f^* orthogonal to $X_{K\cdot L}$ depends not only on the set of functionals Λ, but also on the choice of the reproducing kernel Φ, as $X_{K\cdot L}$ is spanned by the functions $\lambda^y_{kl}\Phi(\cdot, y)$.*

c) *As an effect of the ill-posedness of the continuous problem [Proposition 10 and Appendix 6.3], the collocation matrix $M_{\Lambda,\Phi}$ can be extremely ill-conditioned [9, 42].*

This phenomenon arises when the number of data is large, since in that case the finite-dimensional problem is "closer" to the infinite-dimensional one. By adding more and more spherical Radon data, we add smaller and smaller eigenvalues and therefore we get a collocation matrix with smaller and smaller determinant.

d) It is easy to see that the matrix $M_{\Lambda,\Phi}$ with components

$$(M_{\Lambda,\Phi})_{kl,k'l'} = \langle \lambda_{kl}^y \Phi(\cdot, y), \lambda_{k'l'}^y \Phi(\cdot, y) \rangle_{N_\Phi} = \langle \lambda_{kl}, \lambda_{k'l'} \rangle_{N_\Phi'}$$

is Gramian and hence symmetric positive semi-definite. Positive definiteness (and thus regularity) follows, when the set of functionals Λ is linear independent. Hence, linear independency is crucial for the determination of f^+. Though, the actual number of linearly independent functionals can be much smaller than the number $K \cdot L$ of data we are dealing with. Unfortunately, a handy method for checking linear independency of the set Λ is not developed thus far.

In order to determine the functions in $X_{K \cdot L}$ and the components of the collocation matrix $M_{\Lambda,\Phi}$, we have to examine the expressions $\lambda_{kl}^y \Phi(\cdot, y)$ and

$$\langle \lambda_{kl}^y \Phi(\cdot, y), \lambda_{k'l'}^y \Phi(\cdot, y) \rangle_{N_\Phi} = \lambda_{kl}^x \lambda_{k'l'}^y \Phi(x, y).$$

Following Remark 2, an explicit representation for both expressions is given solely in the situation of $\tilde{\Phi}$ (11) and $\lambda_{kl}^y \Phi(\cdot, y)$ being functions supported in $\mathcal{K}_\delta$ (respectively in Ω). Fortunately in that case, we are able to express $\lambda_{kl}^y \Phi(\cdot, y)$ and $\lambda_{kl}^x \lambda_{k'l'}^y \Phi(x, y)$ with the help of Theorem 26 in Appendix 6.4. Additionally, Sigl showed in [47] that for a large class of radial kernel functions, $\lambda_{kl}^y \Phi(\cdot, y)$ can be given even analytically.

Assumption 2.
The function $\tilde{\Phi}$ and the functions in the set $\mathcal{V} = \{\lambda_{kl}^y \Phi(\cdot, y) \mid \lambda_{kl} \in \Lambda\}$ are compactly supported in $\mathcal{K}_\delta$ (respectively in Ω).

The solution method presented so far is very attractive as there is no use of pre-defined reconstruction points or even a mesh. We fully operate with the given spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$. To solve the optimal recovery problem explicitly, though, not only a couple of assumptions (Assumption 1, 2 and linear independency of the set Λ) have to be fulfilled, but also the generation of the functions $\lambda_{kl}^y \Phi(\cdot, y)$ and the collocation matrix $M_{\Lambda,\Phi}$ is needed. Especially the computation of the matrix components by double integrals with spherical weight functions involved might cause numerical difficulties [Algorithm 1]. Therefore, we present in section 3.3 a second method for the determination of a generalized interpolant by a change of the recovery space. This method avoids calculation of spherical double integrals and implies a relaxation of assumptions.

Algorithm 1 Optimal recovery solution in $X_{K\cdot L}$

Assume linear independency of Λ and the conditions stated in Assumption 1, 2. Let f^* be the solution of the continuous operator equation $\mathcal{M}_{\partial\Omega}(f) = g$.

Input: Spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$, where

$\Xi = \{\xi_k\}_{k=1}^K \subseteq \partial\Omega$,
$\mathcal{T} = \{t_l\}_{l=1}^L \subseteq (0\ T]$,
$\bar{g} \in \mathbb{R}^{K\cdot L}$.

1: Generate the collocation matrix

$$(M_{\Lambda,\Phi})_{kl,k'l'} = \frac{w_{d-2}^2}{w_{d-1}^2} \int_{-1}^{1}\int_{-1}^{1} \phi\Big(\Psi_{k'l'}^{kl}(\tau_1,\tau_2)\Big)(1-\tau_1^2)^{\frac{d-3}{2}}(1-\tau_2^2)^{\frac{d-3}{2}}\mathrm{d}\tau_2\mathrm{d}\tau_1,$$

with

$$\Psi_{k'l'}^{kl}(x,y) = \Big(|\xi_k-\xi_{k'}|^2+t_l^2+t_{l'}^2+2t_l|\xi_k-\xi_{k'}|x+2t_{l'}y\,(|\xi_k-\xi_{k'}|^2+t_l^2+2t_l|\xi_k-\xi_{k'}|x)^{\frac{1}{2}}\Big)^{\frac{1}{2}}$$

for $k, k' = 1, \ldots, K$ and $l, l' = 1, \ldots, L$.

2: Generate the basis functions

$$v_{kl} = \frac{w_{d-2}}{w_{d-1}} \int_{-1}^{1} \phi\Big(\sqrt{|\xi_k - \cdot|^2 + t_l^2 + 2t_l|\xi_k - \cdot|\tau}\Big)(1-\tau^2)^{\frac{d-3}{2}}\,\mathrm{d}\tau$$

for k=1,. . . ,K and $l = 1, \ldots, L$.

3: Solve the linear system

$$M_{\Lambda,\Phi}\alpha = \bar{g}.$$

Output:

$$f^+ = \sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{kl} v_{kl} \in X_{K\cdot L},$$

which is the unique norm-minimal solution of $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$ and satisfies

$$\mathcal{M}_{\partial\Omega}(f^+)(\xi_k, t_l) = \mathcal{M}_{\partial\Omega}(f^*)(\xi_k, t_l)$$

for every $k = 1 \ldots, K$ and $l = 1, \ldots, L$.

3.2.2 L^2-Error Bound and Convergence

In the following we derive a L^2-error bound for the residual $f^+ - f^*$, where $f^* \in H_c^\tau(\Omega)$ is the solution of the continuous problem $\mathcal{M}_{\partial\Omega}(f) = g$. We start by calculating a bound on the residual in the RKHS norm. We proceed with a L^2-error bound by the usage of a sampling inequality. As before, we consider Assumption 1, 2 and assume throughout this paragraph linear independency of the set of functionals Λ.

Proposition 15. *[52, chapter 16]*
Assume linear independency of Λ and the conditions stated in Assumption 1, 2. The optimal recovery solution f^+ is stable in terms of the native space norm, i.e.

$$\|f^+\|_{N_\Phi} \leq \|f^*\|_{N_\Phi}, \quad \|f^* - f^+\|_{N_\Phi} \leq \|f^*\|_{N_\Phi}.$$

Proof. Clearly, f^* satisfies $\mathcal{M}_\mathcal{X}(f^*) = \bar{g}$. As $f^* \in H_c^\tau(\Omega)$, it follows that $f^* \in N_\Phi(\Omega)$. The optimality of f^+ yields $\|f^+\|_{N_\Phi} \leq \|f^*\|_{N_\Phi}$. Further, $f^+ - f^* \in X_{K\cdot L}^\perp$, as

$$\langle f^+ - f^*, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = \langle f^+, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} - \langle f^*, \lambda_{kl}^y \Phi(\cdot, y)\rangle_{N_\Phi} = 0$$

for $k = 1, \ldots, K$ and $l = 1, \ldots, L$. The Pythagorean theorem yields

$$\|f^* - f^+\|_{N_\Phi}^2 + \|f^+\|_{N_\Phi}^2 = \|(f^* - f^+) + f^+\|_{N_\Phi}^2 = \|f^*\|_{N_\Phi}^2. \qquad \square$$

To derive a L^2-error bound, we introduce the concept of data density. We measure the data density of $\mathcal{X} = \Xi \times \mathcal{T}$ in $\Delta = \partial\Omega \times (0\ T]$ by the fill distance

$$h_{\mathcal{X},\Delta} := \sup_{(\xi,t)\in\Delta} \min_{(\xi_k,t_l)\in\mathcal{X}} \sqrt{|\xi - \xi_k|^2 + |t - t_l|^2}.$$

The fill distance $h_{\mathcal{X},\Delta}$ can be interpreted in the following way: $h_{\mathcal{X},\Delta}$ is the radius of the largest ball which is completely contained in Δ and which does not contain any point of $\mathcal{X}$. In this sense, the fill distance describes the largest data-site-free hole in Δ.

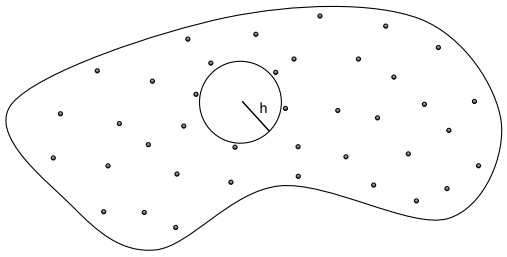

Figure 4: Scheme of the fill distance $h_{\mathcal{X},\Delta}$.

The following sampling inequality is essential in order to incorporate the concept of fill distance into the optimal recovery theory.

Proposition 16. *[28, Theorem 4.8]*
Let $\Omega \subseteq \mathbb{R}^d$ be a Lipschitz domain and let $\theta > \frac{d}{2}$. Then for any arbitrary discrete set $\mathcal{X} \subseteq \Omega$ with sufficiently small data density $h = h_{\mathcal{X},\Omega}$, there is a constant $C > 0$ such that for all $f \in H^\theta(\Omega)$

$$\|f\|_{H^\sigma(\Omega)} \leq C\Big(h^{\theta-\sigma}\|f\|_{H^\theta(\Omega)} + h^{\frac{d}{2}-\sigma}\Big(\sum_{x_i\in\mathcal{X}} |f(x_i)|^2\Big)^{\frac{1}{2}}\Big) \tag{19}$$

for $\sigma \in [0, \lfloor\theta\rfloor]$.

Theorem 17. *L^2-error bound and convergence*
Assume linear independency of Λ and the conditions stated in Assumption 1, 2. Let $f^* \in H_c^\tau(\Omega)$ be the solution of the continuous operator equation $\mathcal{M}_{\partial\Omega}(f) = g$. Then the following error estimate holds for sufficiently small data density $h = h_{\mathcal{X},\Delta}$:

$$\|f^* - f^+\|_{L^2(\Omega)} \leq Ch^\tau \|f^*\|_{N_\Phi(\Omega)}$$

with a constant $C > 0$. Thus, the solution f^+ of the optimal recovery problem (16) is convergent of order τ.

Proof. Due to Assumption 1, 2, $f^* - f^+ \in H^\tau(\Omega)$ with $\operatorname{supp}(f^* - f^+) \subseteq \mathcal{K}_\delta$ for a sufficiently small $\delta > 0$. Hence, the Sobolev space estimates in Proposition 10 can be applied to $f^* - f^+$ with $\mathcal{K} = \mathcal{K}_\delta$. Further, Proposition 10 and the sampling inequality (19) for $\theta = \tau + \frac{d-1}{2}$ and $\sigma = \frac{d-1}{2}$ yield

$$\begin{aligned}
\|f^* - f^+\|_{L^2(\Omega)} &\leq \frac{1}{m_{0,\mathcal{K}}}\|\mathcal{M}_{\partial\Omega}(f^* - f^+)\|_{H^{(d-1)/2}(\Delta)} \\
&\leq \frac{C}{m_{0,\mathcal{K}}}\Big(h^\tau\|\mathcal{M}_{\partial\Omega}(f^* - f^+)\|_{H^\theta(\Delta)} + \Big(h\sum_{k=1}^{K}\sum_{l=1}^{L}|\mathcal{M}_{\partial\Omega}(f^* - f^+)(\xi_k, t_l)|^2\Big)^{\frac{1}{2}}\Big) \\
&= \frac{C}{m_{0,\mathcal{K}}}\Big(h^\tau\|\mathcal{M}_{\partial\Omega}(f^* - f^+)\|_{H^\theta(\Delta)} + h^{1/2}|\mathcal{M}_{\mathcal{X}}(f^* - f^+)|\Big) \\
&= \frac{C}{m_{0,\mathcal{K}}}h^\tau\|\mathcal{M}_{\partial\Omega}(f^* - f^+)\|_{H^\theta(\Delta)} \leq \frac{CM_{\tau,\mathcal{K}}}{m_{0,\mathcal{K}}}h^\tau\|f^* - f^+\|_{H^\tau(\Omega)} \\
&\leq \tilde{C}h^\tau\|f^* - f^+\|_{N_\Phi(\Omega)} \leq \tilde{C}h^\tau\|f^*\|_{N_\Phi(\Omega)},
\end{aligned}$$

by using the equivalence of the native space norm and the corresponding fractional Sobolev norm. □

The error bound in the previous theorem can be generalized to H^s-error estimates easily when applying the sampling inequality (19) with $\theta = \tau + \frac{d-1}{2} + s$ and $\sigma = \frac{d-1}{2} + s$.

3.3 Algebraic Reconstruction Technique

The algebraic reconstruction technique (ART) is a solution method for the generalized interpolation problem (10), which determines a generalized interpolant in the recovery space

$$X_{\mathcal{Y}} = \operatorname{span}\{\, \Phi(\cdot, y_i) \mid i = 1, \dots, N\}.$$

In the following we detail the determination of the ART-solution, a L^2-error bound in terms of the data density and a proof of convergence.

3.3.1 Solution Concept

Assume we are given experimental functionals $\Lambda = \Lambda_\Omega$ [Definition 4]. Let us denote with $\mathcal{Y} := \{y_1, \dots, y_N\} \subseteq \Omega$ an arbitrary, but fixed, set of reconstruction points.

In order to find a solution of the semi-discrete problem $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$ in $X_{\mathcal{Y}}$, we can restrict ourselves to functions of the form

$$f_{\mathcal{Y}} = \sum_{i=1}^{N} \alpha_i \Phi(\cdot, y_i) \tag{20}$$

with coefficients α_i to be determined. Inserting the expression (20) into the equation $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$ yields

$$(\mathcal{M}_{\mathcal{X}}(f_{\mathcal{Y}}))_{kl} = \sum_{i=1}^{N} \alpha_i \lambda_{kl}^x \Phi(x, y_i) = g_{kl} \tag{21}$$

for $k = 1, \dots, K$ and $l = 1, \dots, L$. (21) shows that the determining equation for the coefficient vector $\alpha \in \mathbb{R}^N$ is given by

$$N_{\Lambda,\Phi}^{\mathcal{Y}} \alpha = \bar{g}, \tag{22}$$

with asymmetric collocation matrix

$$N_{\Lambda,\Phi}^{\mathcal{Y}} \in \mathbb{R}^{K \cdot L \times N}, \qquad (N_{\Lambda,\Phi}^{\mathcal{Y}})_{kl,i} := \lambda_{kl}^x \Phi(x, y_i).$$

By the ansatz of searching for a generalized interpolant in the recovery space $X_{\mathcal{Y}}$, we thus deal with a collocation matrix generated by only one spherical integral

$$(N_{\Lambda,\Phi}^{\mathcal{Y}})_{kl,i} = \lambda_{kl}^x \Phi(x, y_i) = \frac{w_{d-2}}{w_{d-1}} \int_{-1}^{1} \phi\Big(\sqrt{|\xi_k - y_i|^2 + t_l^2 + 2t_l|\xi_k - y_i|\tau}\Big)(1 - \tau^2)^{\frac{d-3}{2}} \mathrm{d}\tau,$$

see Theorem 26 in Appendix 6.4. Unless the size $K \cdot L$ of the data vector $\bar{g}$ coincides with the number N of reconstruction points, the linear system (22) is either overdetermined or underdetermined. Note that in either case, existence of a solution for (22) is not guaranteed. This is due to the fact that the range of $N_{\Lambda,\Phi}^{\mathcal{Y}}$ depends not only on the set of functionals Λ, but also on the reconstruction points $\mathcal{Y} \subseteq \Omega$.

We focus in the following on a solution concept for (22), which produces a "solution" in any case. In particular, we determine a coefficient vector $\alpha \in \mathbb{R}^N$ (*generalized solution*) which minimizes the defect functional

$$J^{\mathcal{Y},\bar{g}}_{\Lambda,\Phi} : \mathbb{R}^N \to \mathbb{R}, \quad J^{\mathcal{Y},\bar{g}}_{\Lambda,\Phi}(\alpha) = |N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha - \bar{g}|.$$

To this end, we apply the theory of generalized solutions stated in [36, chapter 2] to our particular finite-dimensional setting.

Lemma 18.
Let $N^{\mathcal{Y}}_{\Lambda,\Phi} \in \mathbb{R}^{K\cdot L\times N}$ and $\bar{g} \in \mathbb{R}^{K\cdot L}$. There is at least one solution of the minimization problem

$$\min_{\alpha\in\mathbb{R}^N} |N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha - \bar{g}|. \tag{23}$$

Among all solutions of (23), there exists exactly one with minimal Euclidean norm. This element is called Moore-Penrose generalized solution of (22) and is denoted by α^+. Further, the set of solutions of (23) is equal to the set of solutions of the matrix normal equation

$$(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha = (N^{\mathcal{Y}}_{\Lambda,\Phi})^T \bar{g}.$$

The previous lemma shows that a unique generalized solution for (22) exists. Though, the usage of $\alpha^+ \in \mathbb{R}^N$ as a coefficient vector in (20) does not guarantee that the corresponding function

$$f^+_{\mathcal{Y}} := \sum_{i=1}^{N} \alpha^+_i \Phi(\cdot, y_i) \in X_{\mathcal{Y}}$$

satisfies the equation $\mathcal{M}_{\mathcal{X}}(f) = \bar{g}$. However, we can give the following result.

Theorem 19.
Let $\alpha^+ \in \mathbb{R}^N$ be the Moore-Penrose generalized solution of (22). Then $f^+_{\mathcal{Y}}$ is a solution of the minimization problem

$$\min_{f\in X_{\mathcal{Y}}} |\mathcal{M}_{\mathcal{X}}(f) - \bar{g}|. \tag{24}$$

If $\tilde{f} \in X_{\mathcal{Y}}$ is another solution of (24), then

$$\|f^+_{\mathcal{Y}}\|_{N_\Phi} \leq \sqrt{\kappa(A_{\Phi,\mathcal{Y}})}\|\tilde{f}\|_{N_\Phi}, \tag{25}$$

where $\kappa(A_{\Phi,\mathcal{Y}})$ is the condition number of the matrix $A_{\Phi,\mathcal{Y}} = (\Phi(y_i, y_j))_{1\leq i,j\leq N}$.

Proof. Consider the defect functional

$$J^{\mathcal{Y},\bar{g}}_{\mathcal{X}} : X_{\mathcal{Y}} \to \mathbb{R}, \quad J^{\mathcal{Y},\bar{g}}_{\mathcal{X}}(f) = |\mathcal{M}_{\mathcal{X}}(f) - \bar{g}|.$$

Let $f = \sum_{i=1}^{N} \alpha_i \Phi(\cdot, y_i) \in X_{\mathcal{Y}}$, then

$$(J_{\mathcal{X}}^{\mathcal{Y},\bar{g}}(f))^2 = |\mathcal{M}_{\mathcal{X}}(f) - \bar{g}|^2 = \sum_{k=1}^{K}\sum_{l=1}^{L}(\lambda_{kl}(f) - g_{kl})^2 = \sum_{k=1}^{K}\sum_{l=1}^{L}(\sum_{i=1}^{N}\alpha_i \lambda_{kl}^x \Phi(x, y_i) - g_{kl})^2$$
$$= \sum_{k=1}^{K}\sum_{l=1}^{L}((N_{\Lambda,\Phi}^{\mathcal{Y}}\alpha)_{kl} - g_{kl})^2 = |N_{\Lambda,\Phi}^{\mathcal{Y}}\alpha - \bar{g}|^2 = (J_{\Lambda,\Phi}^{\mathcal{Y},\bar{g}}(\alpha))^2.$$

Hence, the minimization problem (24) is equivalent to the minimization problem (23). Consequently, $f_{\mathcal{Y}}^{+}$ is a solution of (24). The bound (25) is left to show. To this end, let $\tilde{f} = \sum_{i=1}^{N} \beta_i \Phi(\cdot, y_i) \in X_{\mathcal{Y}}$ be another solution of (24). Hence, the coefficient vector $\beta \in \mathbb{R}^N$ is a solution of (23). Lemma 18 yields thus

$$|\alpha^+|^2 = \langle \alpha^+, \alpha^+ \rangle < \langle \beta, \beta \rangle = |\beta|^2.$$

Further, the native space norms satisfy

$$\|f_{\mathcal{Y}}^{+}\|_{N_\Phi}^2 = \langle \sum_{i=1}^{N} \alpha_i^+ \Phi(\cdot, y_i), \sum_{j=1}^{N} \alpha_j^+ \Phi(\cdot, y_j) \rangle_{N_\Phi} = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^+ \alpha_j^+ \langle \Phi(\cdot, y_i), \Phi(\cdot, y_j) \rangle_{N_\Phi}$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^+ \alpha_j^+ \Phi(y_i, y_j) = \langle \alpha^+, A_{\Phi,\mathcal{Y}} \alpha^+ \rangle$$

and similarly $\|\tilde{f}\|_{N_\Phi}^2 = \langle \beta, A_{\Phi,\mathcal{Y}} \beta \rangle$ with positive definite matrix $A_{\Phi,\mathcal{Y}} = (\Phi(y_i, y_j))_{1 \le i,j \le N}$. Since $A_{\Phi,\mathcal{Y}}$ is symmetric, the Courant-Fischer minimax theorem [22, Theorem 8.1.2] yields

$$\mu_{min} \langle x, x \rangle \le \langle x, A_{\Phi,\mathcal{Y}} x \rangle \le \mu_{max} \langle x, x \rangle, \qquad x \in \mathbb{R}^N,$$

where μ_{min} and μ_{max} are the smallest and largest eigenvalues of $A_{\Phi,\mathcal{Y}}$. Therefore

$$\|f_{\mathcal{Y}}^{+}\|_{N_\Phi}^2 = \langle \alpha^+, A_{\Phi,\mathcal{Y}} \alpha^+ \rangle \le \mu_{max} \langle \alpha^+, \alpha^+ \rangle$$
$$< \mu_{max} \langle \beta, \beta \rangle \le \frac{\mu_{max}}{\mu_{min}} \|\tilde{f}\|_{N_\Phi}^2 = \kappa(A_{\Phi,\mathcal{Y}}) \|\tilde{f}\|_{N_\Phi}^2.$$

□

Remark 4.

a) *We refer to the solution $f_{\mathcal{Y}}^{+}$ of the minimization problem (24) as ART-solution.*

b) *So far, no restrictions on the reconstruction grid $\mathcal{Y} = \{y_1, \ldots, y_N\} \subseteq \Omega$ have been formulated. Any arbitrary, but fixed, set $\mathcal{Y}$ can be used. However, the choice of $\mathcal{Y}$ dramatically influences the size and form of the recovery space $X_{\mathcal{Y}} = \operatorname{span}\{ \Phi(\cdot, y_i) \mid y_i \in \mathcal{Y}\}$. Further, as we see in section 4, specific choices of $\mathcal{Y}$ can lead to a fast and efficient algorithm to determine the coefficient vector α^+.*

Algorithm 2 ART-solution in $X_{\mathcal{Y}}$

Assume the conditions stated in Assumption 1. Let f^* be the solution of the continuous operator equation $\mathcal{M}_{\partial\Omega}(f) = g$.

Input: Spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$, where

$\Xi = \{\xi_k\}_{k=1}^K \subseteq \partial\Omega$,
$\mathcal{T} = \{t_l\}_{l=1}^L \subseteq (0\ T]$,
$\bar{g} \in \mathbb{R}^{K \cdot L}$.

1: Choose reconstruction points $\mathcal{Y} = \{y_1, \ldots, y_N\} \subseteq \Omega$.
2: Generate the asymmetric collocation matrix

$$(N_{\Lambda,\Phi}^{\mathcal{Y}})_{kl,i} = \frac{w_{d-2}}{w_{d-1}} \int_{-1}^{1} \phi\Big(\sqrt{|\xi_k - y_i|^2 + t_l^2 + 2t_l|\xi_k - y_i|\tau}\Big)(1-\tau^2)^{\frac{d-3}{2}} \mathrm{d}\tau$$

for $k = 1, \ldots, K$, $l = 1, \ldots, L$ and $i = 1, \ldots, N$.
3: Determine the solution α^+ of

$$(N_{\Lambda,\Phi}^{\mathcal{Y}})^T N_{\Lambda,\Phi}^{\mathcal{Y}} \alpha = (N_{\Lambda,\Phi}^{\mathcal{Y}})^T \bar{g}$$

with minimal Euclidean norm.

Output:

$$f_{\mathcal{Y}}^+ = \sum_{i=1}^{N} \alpha_i^+ \Phi(\cdot, y_i) \in X_{\mathcal{Y}},$$

which is a solution of the problem

$$\min_{f \in X_{\mathcal{Y}}} |\mathcal{M}_{\mathcal{X}}(f) - \bar{g}|$$

with

$$\mathcal{M}_{\partial\Omega}(f_{\mathcal{Y}}^+)(\xi_k, t_l) \approx \mathcal{M}_{\partial\Omega}(f^*)(\xi_k, t_l)$$

for every $k = 1 \ldots, K$ and $l = 1, \ldots, L$.

3.3.2 L^2-Error Bound and Convergence

In the following we derive a L^2-error bound for the residual $f_{\mathcal{Y}}^+ - f^*$, where $f^* \in H_c^\tau(\Omega)$ is the solution of the continuous problem $\mathcal{M}_{\partial\Omega}(f) = g$. Similar to section 3.2.2, we use the sampling inequality (19) and the Sobolev space estimates in Proposition 10. To this end, we give restrictions on the reconstruction grid $\mathcal{Y} = \{y_1, \dots, y_N\} \subseteq \Omega$ and on the support of the kernel $\tilde{\Phi}$.

The set of reconstruction points $\mathcal{Y}$ plays an important role in the solution theory of ART, as it influences the size and form of the recovery space $X_{\mathcal{Y}}$ dramatically. For the derivation of the error bound, we assume $\mathcal{Y}$ to be "large enough" such that $f^* \in X_{\mathcal{Y}}$. This assumption can be interpreted as a denseness condition for $X_{\mathcal{Y}} = \operatorname{span}\{\Phi(\cdot, y_i) \mid y_i \in \mathcal{Y}\}$ in $H^\tau(\Omega)$, as

$$H^\tau(\Omega) = N_\Phi(\Omega) = \overline{\operatorname{span}\{\, \Phi(\cdot, y) \mid y \in \Omega\}}^{\|\cdot\|_\Phi},$$

see Lemma 7.

Proposition 20.
Assume the conditions stated in Assumption 1. Let $\mathcal{Y} \subseteq \Omega$ such that $f^* \in X_{\mathcal{Y}}$. Then the ART-solution $f_{\mathcal{Y}}^+$ satisfies

$$\mathcal{M}_{\mathcal{X}}(f_{\mathcal{Y}}^+) = \bar{g}, \quad \|f_{\mathcal{Y}}^+\|_{N_\Phi} \leq \sqrt{\kappa(A_{\Phi,\mathcal{Y}})}\|f^*\|_{N_\Phi}.$$

Proof. Clearly, f^* satisfies $\mathcal{M}_{\mathcal{X}}(f^*) = \bar{g}$. As $f^* \in X_{\mathcal{Y}}$, it follows that $J_{\mathcal{X}}^{\mathcal{Y},\bar{g}}(f^*) = |\mathcal{M}_{\mathcal{X}}(f^*) - \bar{g}| = 0$. Hence, f^* and $f_{\mathcal{Y}}^+$ are both solutions of the minimization problem (24). Therefore, $J_{\mathcal{X}}^{\mathcal{Y},\bar{g}}(f^*) = 0$ yields $J_{\mathcal{X}}^{\mathcal{Y},\bar{g}}(f_{\mathcal{Y}}^+) = 0$ and thus $\mathcal{M}_{\mathcal{X}}(f_{\mathcal{Y}}^+) = \bar{g}$. The assertion about the bound follows from Theorem 19. □

We now turn our attention to restrictions on the kernel function. According to Proposition 10, the Sobolev space estimates for the spherical Radon transform hold solely for functions compactly supported in Ω. Hence, the estimates can be applied to $f_{\mathcal{Y}}^+ - f^*$ only if the residual is supported in Ω. We can realize this as follows: assume the kernel $\tilde{\Phi}$ to have compact support in Ω and rescale the variables of the kernel with a shape parameter $\varepsilon > 0$,

$$\Phi_\varepsilon(x, y) := \Phi(\varepsilon x, \varepsilon y) = \tilde{\Phi}(\varepsilon(x - y)).$$

A decrease of the shape parameter ε has the effect of producing a "flat" function, while an increase of ε leads to a more "peaked" (or localized) function. Thus, the shape parameter influences the size of the support of a compactly supported kernel.

Proposition 21.
Assume the conditions stated in Assumption 1. Let $\mathcal{Y} \subseteq \Omega$ be arbitrary, but fixed. Suppose the kernel $\tilde{\Phi}$ has compact support in Ω. Then there exists $\delta > 0$ such that for all sufficiently large shape parameters $\varepsilon > 0$, the functions in $\{\Phi_\varepsilon(\cdot, y_i) \mid y_i \in \mathcal{Y}\}$ are supported in $\mathcal{K}_\delta = \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) \geq \delta\}$.

Proof. As Ω is open, there exists $\delta > 0$ such that $\operatorname{supp}(\tilde{\Phi}) \subseteq \mathcal{K}_\delta$ and $\mathcal{Y} \subseteq \operatorname{int}(\mathcal{K}_\delta) := \mathcal{K}_\delta \backslash \partial\mathcal{K}_\delta$. Then

$$\begin{aligned}\operatorname{supp}(\Phi_\varepsilon(\cdot, y_i)) \subseteq \mathcal{K}_\delta \text{ for every } y_i \in \mathcal{Y} &\Leftrightarrow \operatorname{supp}(\tilde{\Phi}(\varepsilon(\cdot - y_i))) \subseteq \mathcal{K}_\delta \text{ for every } y_i \in \mathcal{Y} \\ &\Leftrightarrow \operatorname{supp}(\tilde{\Phi}(\cdot - y_i)) \subseteq \varepsilon\mathcal{K}_\delta \text{ for every } y_i \in \mathcal{Y} \\ &\Leftrightarrow \operatorname{supp}(\tilde{\Phi}) \subseteq \{\varepsilon\mathcal{K}_\delta + y_i\} \text{ for every } y_i \in \mathcal{Y}.\end{aligned}$$

As $0 \in K_\delta$, the assertion follows. □

Theorem 22. *L^2-error bound and convergence*
Assume the conditions stated in Assumption 1. Let $\mathcal{Y} \subseteq \Omega$ be arbitrary, but fixed. Suppose the kernel $\tilde{\Phi}$ has compact support in Ω. Consider the algebraic reconstruction technique for the recovery space

$$X_{\mathcal{Y},\Phi_\varepsilon} := \operatorname{span}\{\Phi_\varepsilon(\cdot, y_i) \mid y_i \in \mathcal{Y}\}$$

for a sufficiently large shape parameter ε. Let $f^* \in H_c^\tau(\Omega)$, the solution of the continuous operator equation $\mathcal{M}_{\partial\Omega}(f) = g$, be an element of the recovery space $X_{\mathcal{Y},\Phi_\varepsilon}$. Then the following error estimate holds for sufficiently small data density $h = h_{\mathcal{X},\Delta}$:

$$\|f^* - f_{\mathcal{Y}}^+\|_{L^2(\Omega)} \leq Ch^\tau \left(1 + \sqrt{\kappa(A_{\Phi,\mathcal{Y}})}\right) \|f^*\|_{N_\Phi(\Omega)}$$

with a constant $C > 0$. Thus, the ART-solution $f_{\mathcal{Y}}^+ \in X_{\mathcal{Y},\Phi_\varepsilon}$ is convergent of order τ.

Proof. Due to Assumption 1 and Proposition 21, $f^* - f_{\mathcal{Y}}^+ \in H^\tau(\Omega)$ with $\operatorname{supp}(f^* - f_{\mathcal{Y}}^+) \subseteq \mathcal{K}_\delta$ for a sufficiently small $\delta > 0$. Hence, the Sobolev space estimates in Proposition 10 can be applied to $f^* - f_{\mathcal{Y}}^+$ with $\mathcal{K} = \mathcal{K}_\delta$. Further, Proposition 10 and the sampling inequality (19) for $\theta = \tau + \frac{d-1}{2}$ and $\sigma = \frac{d-1}{2}$ yield

$$\begin{aligned}\|f^* - f_{\mathcal{Y}}^+\|_{L^2(\Omega)} &\leq \frac{1}{m_{0,\mathcal{K}}} \|\mathcal{M}_{\partial\Omega}(f^* - f_{\mathcal{Y}}^+)\|_{H^{(d-1)/2}(\Delta)} \\ &\leq \frac{C}{m_{0,\mathcal{K}}} \left(h^\tau \|\mathcal{M}_{\partial\Omega}(f^* - f_{\mathcal{Y}}^+)\|_{H^\theta(\Delta)} + \left(h \sum_{k=1}^{K} \sum_{l=1}^{L} |\mathcal{M}_{\partial\Omega}(f^* - f_{\mathcal{Y}}^+)(\xi_k, t_l)|^2\right)^{\frac{1}{2}}\right) \\ &= \frac{C}{m_{0,\mathcal{K}}} \left(h^\tau \|\mathcal{M}_{\partial\Omega}(f^* - f_{\mathcal{Y}}^+)\|_{H^\theta(\Delta)} + h^{\frac{1}{2}} |\mathcal{M}_{\mathcal{X}}(f^* - f_{\mathcal{Y}}^+)|\right) \\ &= \frac{C}{m_{0,\mathcal{K}}} h^\tau \|\mathcal{M}_{\partial\Omega}(f^* - f_{\mathcal{Y}}^+)\|_{H^\theta(\Delta)} \leq \frac{CM_{\tau,\mathcal{K}}}{m_{0,\mathcal{K}}} h^\tau \|f^* - f_{\mathcal{Y}}^+\|_{H^\tau(\Omega)} \\ &\leq \tilde{C}h^\tau \|f^* - f_{\mathcal{Y}}^+\|_{N_\Phi(\Omega)} \leq \tilde{C}h^\tau \left(\|f^*\|_{N_\Phi(\Omega)} + \|f_{\mathcal{Y}}^+\|_{N_\Phi(\Omega)}\right) \\ &\leq \tilde{C}h^\tau \left(1 + \sqrt{\kappa(A_{\Phi,\mathcal{Y}})}\right) \|f^*\|_{N_\Phi(\Omega)}\end{aligned}$$

by using the equivalence of the native space norm and the corresponding fractional Sobolev norm. □

Again, the error bound in the previous theorem can be generalized to H^s-error estimates easily when applying the sampling inequality (19) with $\theta = \tau + \frac{d-1}{2} + s$ and $\sigma = \frac{d-1}{2} + s$.

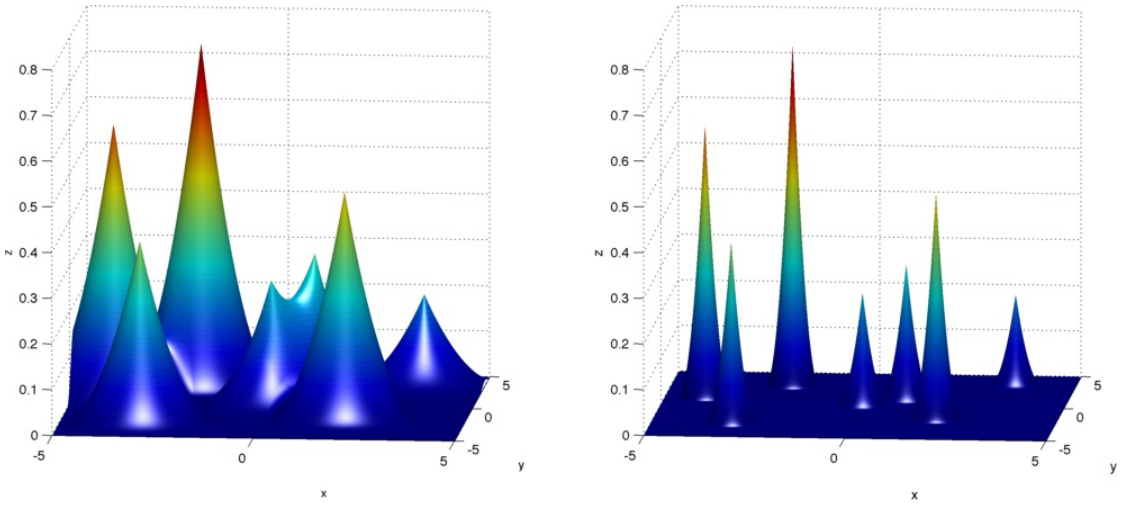

Figure 5: Influence of the shape parameter on a function in $X_{\mathcal{Y},\Phi_\varepsilon}$ with $\varepsilon = 0.5$ (left) and $\varepsilon = 2$ (right) for Wendland's compactly supported kernel with generating univariate function $\phi_{\varepsilon,3,0}$ in $\mathbb{R}^2$ [Example 2]. The set $\mathcal{Y}$ consists of 6 non-uniform reconstruction points lying in the open ball $\Omega = B_{6.5}(0)$.

3.4 Notes and Comments

Firstly described in the late 1950s by Michael Golomb and Hans Weinberger [24], optimal recovery has attracted much attention in the last years for solving various linear inverse problems with discrete data. For an interesting discussion on the solution of partial differential equations using the optimal recovery theory, see [52, chapter 16].

Continuity of the functionals $\lambda_1, \ldots, \lambda_N$ is essential for the optimal recovery theory. To derive the set of continuous functionals $\Lambda_{\mathcal{K}_\delta}$, we used Sobolev space estimates for the spherical Radon transform [Proposition 10]. Even though Proposition 10 has been proven only for the case $\Omega = B_r(0)$, we strongly assume that a generalization to a case, where Ω has just a sufficiently regular boundary, is possible. This would relax the assumption on the acquisition surface in Assumption 1.

Frank Natterer introduced the term algebraic reconstruction technique as a method to derive reconstruction algorithms in the setting of computerized tomography [31]. Note that Natterer's understanding of an algebraic reconstruction technique differs from ours: he means the application of Kaczmarz's method (an iterative method for solving linear systems of equations [31]) to Radon's integral equation. Several algorithms were derived by Kaczmarz's method (ART, SIRT, SART) and we refer the interested reader to [29, chapter 7] and [4, 8, 21] for further information.

This sampling inequality (19) can be used to derive convergence orders for very general

recovery processes when assuming that the unknown function f is of finite smoothness (see [43] for machine learning algorithms and [49] for PDE solvers). We refer the reader to [41] for sampling inequalities for infinitely smooth functions, where the convergence order turns out to depend exponentially on the fill distance. Those inequalities are of great interest when analyzing the convergence order for e.g. a Gaussian or a generalized multiquadrics kernel function.

4 Fast Implementation and Numerical Results for the Algebraic Reconstruction Technique

In this chapter, we present a fast algorithm for the algebraic reconstruction technique and give numerical results to study the image reconstruction capability and the computational efficiency of the method.

4.1 A Fast Algorithm

In order to determine the ART-solution $f_{\mathcal{Y}}^{+}$ by given spherical Radon data $(\Xi, \mathcal{T}, \bar{g})$ and arbitrary, but fixed, reconstruction points $\mathcal{Y} = \{y_1, \ldots, y_N\}$, we need to compute the norm-minimal solution α^{+} of the normal equation

$$(N_{\Lambda,\Phi}^{\mathcal{Y}})^T N_{\Lambda,\Phi}^{\mathcal{Y}} \alpha = (N_{\Lambda,\Phi}^{\mathcal{Y}})^T \bar{g}. \tag{26}$$

In general, the size $K \cdot L$ of a *real experimental* data vector $\bar{g}$ is extremely large and thus, $N_{\Lambda,\Phi}^{\mathcal{Y}} \in \mathbb{R}^{K \cdot L \times N}$ is of large dimension. Hence, direct methods for the solution of (26) are computing time and storage intensive. For instance, the Gaussian elimination applied to (26) has arithmetic complexity of $\mathcal{O}(N^2(KL+N))$ and storage complexity of $\mathcal{O}(NKL)$, if the components of $N_{\Lambda,\Phi}^{\mathcal{Y}}$ are assumed to be pre-computed.

With the application of the conjugate gradient method [section 4.1.1] to (26) and the usage of reconstruction points $\mathcal{Y}$ adapted to spherical transducer locations Ξ [section 4.1.2], we can enhance the computing time and storage complexity dramatically. This is due to the fact that only a few auxiliary vectors have to be stored and circulant block structures in $N_{\Lambda,\Phi}^{\mathcal{Y}}$ enable a fast implementation by the fast Fourier transform (FFT).

4.1.1 Conjugate Gradient Method

The conjugate gradient method (CGM) was proposed by Hestens and Stiefel [26] in the 1950s as an iterative method for solving linear systems with positive definite matrices. It is an alternative to the direct method of Gaussian elimination that is very well suited for solving large problems. In the following we make use of the CGM to determine an approximation to the coefficient vector α^{+}, where we follow the textbook of Nocedal and Wright [32].
Briefly, the conjugate gradient method works as follows: suppose $(N_{\Lambda,\Phi}^{\mathcal{Y}})^T N_{\Lambda,\Phi}^{\mathcal{Y}}$ is positive definite, then, given a starting vector $\alpha^{(0)} \in \mathbb{R}^N$ and a set of nonzero vectors (*conjugate directions*) $\{p^{(0)}, p^{(1)}, \ldots, p^{(n-1)}\} \subseteq \mathbb{R}^N$ satisfying

$$\langle N_{\Lambda,\Phi}^{\mathcal{Y}} p^{(i)}, N_{\Lambda,\Phi}^{\mathcal{Y}} p^{(j)} \rangle = 0, \quad i \neq j,$$

the conjugate gradient method generates the sequence $\{\alpha^{(k)}\}_{k \in \mathbb{N}}$ defined by

$$\alpha^{(k+1)} = \alpha^{(k)} - \frac{\langle r^{(k)}, p^{(k)} \rangle}{|N_{\Lambda,\Phi}^{\mathcal{Y}} p^{(k)}|^2} p^{(k)}, \tag{27}$$

where $r^{(k)} = (N^{\mathcal{Y}}_{\Lambda,\Phi})^T(N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha^{(k)} - \bar{g})$.

Proposition 23. *[32, Theorem 5.1]*
Let $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$ be positive definite. For any $\alpha^{(0)} \in \mathbb{R}^N$, the sequence $\{\alpha^{(k)}\}_{k\in\mathbb{N}}$ generated by the conjugate gradient algorithm (27) converges to the norm-minimal solution α^+ of the linear system (26) in at most N steps.

Remark 5.

a) The positive definiteness of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$ is crucial for the convergence. Though, a sufficient criteria to check the positive definiteness has not been developed so far.

b) There are many ways to choose the set of conjugate directions. One possibility is the set of eigenvectors of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$. However, for large scale problems, the determination of the eigenvectors is not efficient. Another way of choosing the conjugate directions is presented in the subsequent algorithm: a new $p^{(k)}$ can be computed using only the previous vector $p^{(k-1)}$. It does not need to know all the previous elements $p^{(0)}, p^{(1)}, \ldots, p^{(k-2)}$ of the conjugate set. This remarkable property implies that the method below requires little storage and computation. We give [32, Algorithm 5.2] as a reference.

Algorithm 3 Conjugate gradient method

Assume positive definiteness of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$.

Input: $\alpha^{(0)} \in \mathbb{R}^N$, $\bar{g} \in \mathbb{R}^{K\cdot L}$ and $N^{\mathcal{Y}}_{\Lambda,\Phi} \in \mathbb{R}^{K\cdot L\times N}$.

1: $r^{(0)} = (N^{\mathcal{Y}}_{\Lambda,\Phi})^T(N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha^{(0)} - \bar{g})$; $p^{(0)} = -r^{(0)}$; $k = 0$;
2: **while** $r^{(k)} \neq 0$ **do**
3: $\quad \gamma^{(k)} = \frac{|r^{(k)}|^2}{|N^{\mathcal{Y}}_{\Lambda,\Phi}p^{(k)}|^2}$;
4: $\quad \alpha^{(k+1)} = \alpha^{(k)} + \gamma^{(k)}p^{(k)}$;
5: $\quad r^{(k+1)} = r^{(k)} + \gamma^{(k)}(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}p^{(k)}$;
6: $\quad \beta^{(k+1)} = \frac{|r^{(k+1)}|^2}{|r^{(k)}|^2}$;
7: $\quad p^{(k+1)} = -r^{(k+1)} + \beta^{(k+1)}p^{(k)}$;
8: $\quad k = k + 1$;
9: **end while**
10: $\alpha^+ = \alpha^{(k+1)}$;

Output: α^+, which is the norm-minimal solution of the normal equation

$$(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}\alpha = (N^{\mathcal{Y}}_{\Lambda,\Phi})^T\bar{g}.$$

Arithmetic complexity: $\mathcal{O}(NKL)$.
Storage complexity: $\mathcal{O}(NKL)$.

Let us focus on the arithmetic and storage complexity. The major computational tasks to be performed at each step are the computation of the matrix-vector products

with $N_{\Lambda,\Phi}^{\mathcal{Y}}$ and $(N_{\Lambda,\Phi}^{\mathcal{Y}})^T$, calculation of the norms $|r^{(k)}|^2$ and $|N_{\Lambda,\Phi}^{\mathcal{Y}} p^{(k)}|^2$ and calculation of three vector sums. The norms and vector sums can be performed in $\mathcal{O}(N + KL)$ floating-point operations, while the cost of the matrix-vector products is $\mathcal{O}(NKL)$, if the components of $N_{\Lambda,\Phi}^{\mathcal{Y}}$ are assumed to be pre-computed.
The storage complexity arises by storing the entire collocation matrix $N_{\Lambda,\Phi}^{\mathcal{Y}}$ throughout the whole algorithm. Other vectors in the algorithm are not as storage demanding as $N_{\Lambda,\Phi}^{\mathcal{Y}}$. This is due to the fact that, at any given point of the algorithm, we never need to know the vectors α, r and p for more than the last two iterations.

In order to decrease the arithmetic and storage complexity of the conjugate gradient method, we focus on reconstruction points $\mathcal{Y}$ generating block circulant matrix structures in $N_{\Lambda,\Phi}^{\mathcal{Y}}$. This enables not only the usage of the FFT for the matrix-vector products with $N_{\Lambda,\Phi}^{\mathcal{Y}}$ and $(N_{\Lambda,\Phi}^{\mathcal{Y}})^T$, but also the possibility to store these matrices via a few auxiliary vectors.

4.1.2 Circulant Block Matrices

The basic idea to enhance arithmetic and storage complexity of the CGM is the efficient storage of the matrix $N_{\Lambda,\Phi}^{\mathcal{Y}}$ and a fast implementation of the matrix vector products with $N_{\Lambda,\Phi}^{\mathcal{Y}}$ and $(N_{\Lambda,\Phi}^{\mathcal{Y}})^T$. We realize this by choosing a reconstruction grid $\mathcal{Y} = \{y_1, \dots, y_{K \cdot J}\} \subseteq \Omega$, $J \in \mathbb{N}$, which generates circulant block structures in the collocation matrix, i.e.

$$N_{\Lambda,\Phi}^{\mathcal{Y}} = \begin{pmatrix} N_{11} & \cdots & N_{1J} \\ \vdots & & \vdots \\ N_{L1} & \cdots & N_{LJ} \end{pmatrix} \in \mathbb{R}^{K \cdot L \times K \cdot J} \tag{28}$$

where a blockmatrix $N_{lj} \in \mathbb{R}^{K \times K}$ takes the form

$$N_{lj} := \begin{pmatrix} n_1^{lj} & n_K^{lj} & \cdots & n_3^{lj} & n_2^{lj} \\ n_2^{lj} & n_1^{lj} & n_K^{lj} & & n_3^{lj} \\ \vdots & n_2^{lj} & n_1^{lj} & \ddots & \vdots \\ n_{K-1}^{lj} & & \ddots & \ddots & n_K^{lj} \\ n_K^{lj} & n_{K-1}^{lj} & \cdots & n_2^{lj} & n_1^{lj} \end{pmatrix}$$

and is fully specified by its first column

$$c_{lj} := (n_1^{lj}, \dots, n_K^{lj})^T \in \mathbb{R}^K.$$

The remaining columns of N_{lj} are each cyclic permutations of the first column. The following proposition illustrates the particular efficiency of circulant block structures for the conjugate gradient method.

Proposition 24. *[22, chapter 4.7]*
Let $N \in \mathbb{R}^{K\times K}$ be a circulant matrix with first column $c \in \mathbb{R}^K$. Further, let $F_K \in \mathbb{C}^{K\times K}$ be the Fourier matrix with entries $(F_K)_{jk} := K^{-\frac{1}{2}} e^{-\frac{(j-1)(k-1)2\pi i}{K}}$. Then the identity

$$Nv = F_K^*(F_K c \odot F_K v), \quad v \in \mathbb{R}^K$$

holds. Thus, the matrix-vector product Nv can be performed in $\mathcal{O}(K\log(K))$ floating point operations by the usage of the fast Fourier transform. Here, the point-wise multiplication operator $\odot$ is defined by

$$x \odot y := (x_1 y_1, \ldots, x_K y_K)^T, \quad x, y \in \mathbb{R}^K.$$

Algorithm 4 Fast matrix vector product (FMVP)

Assume $N^{\mathcal{Y}}_{\Lambda,\Phi} \in \mathbb{R}^{K\cdot L\times K\cdot J}$ has a circulant block structure (28).

Input: $\{c_{lj} \mid l = 1,\ldots,L \text{ and } j = 1,\ldots,J\} \subseteq \mathbb{R}^K$, $v \in \mathbb{R}^{K\cdot J}$.

1: $w = 0 \in \mathbb{R}^{K\cdot L}$;
2: **for** $l = 1,\ldots,L$ **do**
3: **for** $j = 1,\ldots,J$ **do**
4: $(w_k)_{1+(l-1)K\le k\le lK} = (w_k)_{1+(l-1)K\le k\le lK} + F_K c_{lj} \odot F_K (v_k)_{1+(j-1)K\le k\le jK}$;
5: **end for**
6: $(w_k)_{1+(l-1)K\le k\le lK} = F_K^*(w_k)_{1+(l-1)K\le k\le lK}$;
7: **end for**

Output: $w = N^{\mathcal{Y}}_{\Lambda,\Phi} v$.
Arithmetic complexity: $\mathcal{O}(JK\log(K)L)$.
Storage complexity: $\mathcal{O}(JKL)$.

Remark 6.

a) *Clearly, an analogue algorithm can be given for the matrix-vector product with the transposed collocation matrix $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T$. We refer to it as "fast transposed matrix vector product" (FTMVP).*

b) *It has been shown in [47] that for a large class of radial kernel functions Φ, the components of $N^{\mathcal{Y}}_{\Lambda,\Phi}$ can be given analytically. Hence, the input set $\{c_{lj} \mid l = 1,\ldots,L \text{ and } j = 1,\ldots,J\} \subseteq \mathbb{R}^K$ is not needed. The actual storage complexity is thus $\mathcal{O}(K(J+L))$.*

Before we provide the final algorithm for solving (26), let us focus on examples of $\mathcal{Y}$ which generate block circulant entries in $N^{\mathcal{Y}}_{\Lambda,\Phi}$.

Example 3. *Two-dimensional data*
Consider we are given two-dimensional experimental information

$$(\Xi, \mathcal{T}, \bar{g}) \subseteq rS^1 \times (0\ T] \times \mathbb{R}^{K\cdot L}.$$

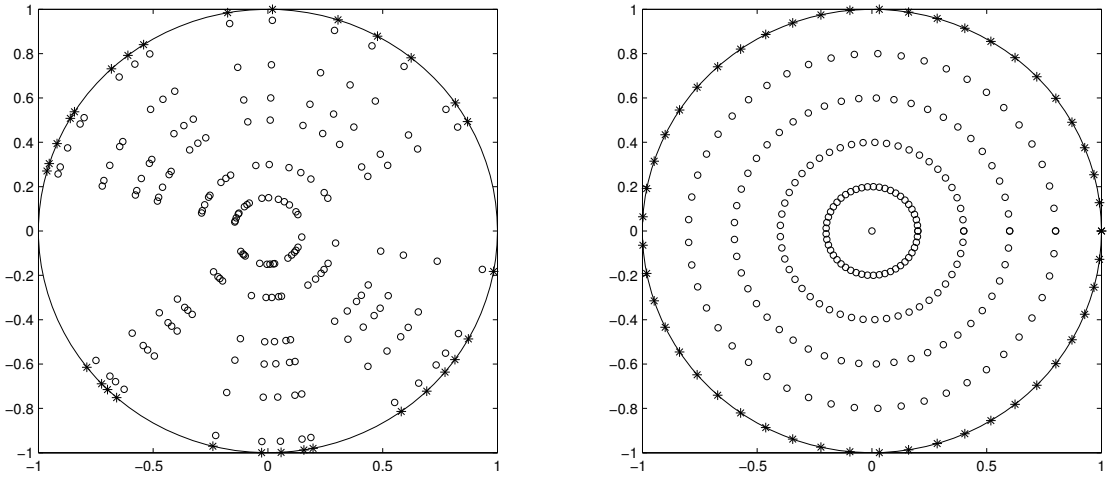

Figure 6: Polar reconstruction grid $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ in $B_1(0) \subseteq \mathbb{R}^2$. Non-uniform transducer locations $\Xi \subseteq S^1$ (left, asterisks) with reconstruction radii $\mathcal{R}_6 = \{0.15, 0.3, 0.5, 0.6, 0.75, 0.95\}$. Uniform transducer locations $\Xi \subseteq S^1$ (right, asterisks) with radii $\mathcal{R}_5 = \{0, 0.2, 0.4, 0.6, 0.8\}$.

Define for an arbitrary, but fixed, $J \in \mathbb{N}$ the polar reconstruction grid

$$\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J} := \{y_{kj} = r_j(\cos(\theta_k), \sin(\theta_k))^T \mid (\theta_k, r_j) \in \Theta_\Xi \times \mathcal{R}_J\}, \tag{29}$$

where

$$\Theta_\Xi := \{\theta_k \in [0\ 2\pi) \mid \xi_k = r(\cos(\theta_k), \sin(\theta_k))^T \in \Xi\}$$

and

$$\mathcal{R}_J := \{r_j \in [0\ r) \mid r_j < r_{j+1},\ j = 1, \ldots, J-1\}.$$

The reconstruction points lie therefore on concentric circles around the origin. For a fixed measurement time $t_l \in \mathcal{T}$ and a fixed radius $r_j \in \mathcal{R}_J$,

$$N_{lj} = ((N^{\mathcal{Y}}_{\Lambda,\Phi})_{kl,k'j})_{1\le k,k'\le K}$$

is a circulant matrix. To this end,

$$\begin{aligned}
|\xi_{(k+c)} - y_{(k'+c)j}| &= |r(\cos(\theta_{(k+c)}), \sin(\theta_{(k+c)}))^T - r_j(\cos(\theta_{(k'+c)}), \sin(\theta_{(k'+c)}))^T| \\
&= |r(\cos(\theta_k), \sin(\theta_k))^T - r_j(\cos(\theta_{k'}), \sin(\theta_{k'}))^T| \\
&= |\xi_k - y_{k'j}|
\end{aligned}$$

and thus

$$\begin{aligned}(N^{\mathcal{Y}}_{\Lambda,\Phi})_{kl,k'j} &= \frac{1}{\pi}\int_{-1}^{1}\phi\Big(\sqrt{|\xi_k - y_{k'j}|^2 + t_l^2 + 2t_l|\xi_k - y_{k'j}|\tau}\Big)(1-\tau^2)^{-\frac{1}{2}}\,d\tau \\ &= \frac{1}{\pi}\int_{-1}^{1}\phi\Big(\sqrt{|\xi_{(k+c)} - y_{(k'+c)j}|^2 + t_l^2 + 2t_l|\xi_{(k+c)} - y_{(k'+c)j}|\tau}\Big)(1-\tau^2)^{-\frac{1}{2}}\,d\tau \\ &= (N^{\mathcal{Y}}_{\Lambda,\Phi})_{(k+c)l,(k'+c)j}\end{aligned}$$

for $k, k' = 1, \ldots, K$, $l = 1, \ldots, L$, $j = 1, \ldots, J$ and any $c \in \mathbb{N}_0$ with $1 \leq k+c,\ k'+c \leq K$.

Example 4. *Three-dimensional data*
Consider we are given three-dimensional experimental information

$$(\Xi, \mathcal{T}, \bar{g}) \subseteq rS^2 \times (0\ T] \times \mathbb{R}^{K\cdot L}.$$

Define for an arbitrary, but fixed, $J \in \mathbb{N}$ the spherical reconstruction cloud

$$\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J} := \left\{ y_{kj} = r_j \begin{pmatrix}\sin(\theta_k)\cos(\phi_k)\\ \sin(\theta_k)\sin(\phi_k)\\ \cos(\theta_k)\end{pmatrix} \mid (\theta_k, \phi_k, r_j) \in \Theta_\Xi \times \mathcal{R}_J \right\}$$

where

$$\Theta_\Xi := \left\{ (\theta_k, \phi_k) \in [0\ \pi] \times [0\ 2\pi) \mid \xi_k = r \begin{pmatrix}\sin(\theta_k)\cos(\phi_k)\\ \sin(\theta_k)\sin(\phi_k)\\ \cos(\theta_k)\end{pmatrix} \in \Xi \right\}$$

and

$$\mathcal{R}_J := \{r_j \in [0\ r) \mid r_j < r_{j+1},\ j = 1, \ldots, J-1\}.$$

The reconstruction points lie therefore on concentric spheres around the origin. For a fixed measurement time $t_l \in \mathcal{T}$ and a fixed radius $r_j \in \mathcal{R}_J$,

$$N_{lj} = ((N^{\mathcal{Y}}_{\Lambda,\Phi})_{kl,k'j})_{1\leq k,k'\leq K}$$

is a circulant matrix. The proof is similar as in Example 3, since

$$|\xi_{(k+c)} - y_{(k'+c)j}| = |\xi_k - y_{k'j}|$$

and thus

$$\begin{aligned}(N^{\mathcal{Y}}_{\Lambda,\Phi})_{kl,k'j} &= \frac{1}{2}\int_{-1}^{1}\phi\Big(\sqrt{|\xi_k - y_{k'j}|^2 + t_l^2 + 2t_l|\xi_k - y_{k'j}|\tau}\Big)\,d\tau \\ &= \frac{1}{2}\int_{-1}^{1}\phi\Big(\sqrt{|\xi_{(k+c)} - y_{(k'+c)j}|^2 + t_l^2 + 2t_l|\xi_{(k+c)} - y_{(k'+c)j}|\tau}\Big)\,d\tau \\ &= (N^{\mathcal{Y}}_{\Lambda,\Phi})_{(k+c)l,(k'+c)j}\end{aligned}$$

for $k, k' = 1, \ldots, K$, $l = 1, \ldots, L$, $j = 1, \ldots, J$ and any $c \in \mathbb{N}_0$ with $1 \leq k+c,\ k'+c \leq K$.

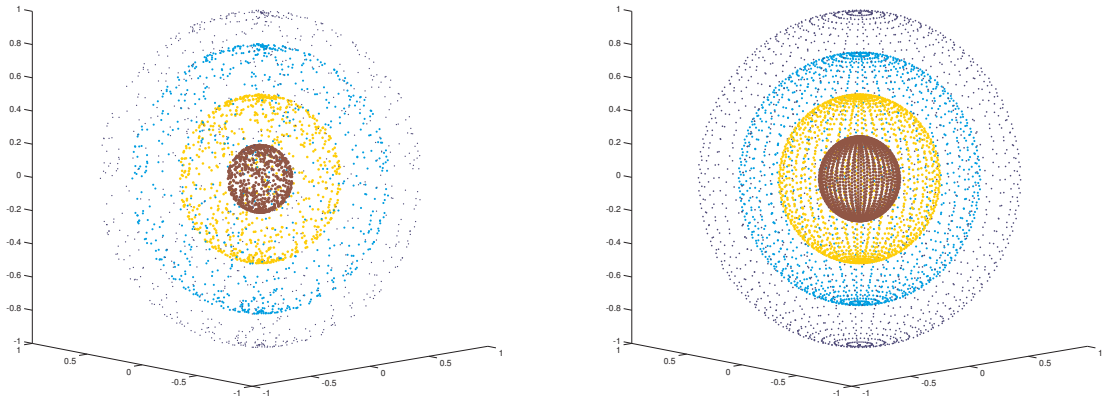

Figure 7: Spherical reconstruction cloud $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ in $B_1(0) \subseteq \mathbb{R}^3$. Non-uniform transducer locations $\Xi \subseteq S^2$ (left, outermost spherical points) with reconstruction radii $\mathcal{R}_3 = \{0.2, 0.5, 0.8\}$. Uniform transducer locations $\Xi \subseteq S^2$ (right, outermost spherical points) with radii $\mathcal{R}_3 = \{0.25, 0.5, 0.75\}$.

Algorithm 5 Fast conjugate gradient method

Let $\mathcal{Y} = \{y_1, \dots, y_{K\cdot J}\}$ be a reconstruction grid such that $N^{\mathcal{Y}}_{\Lambda,\Phi}$ has circulant block structure (28). Assume positive definiteness of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$. Further, assume Φ to be a radial kernel such that the components of $N^{\mathcal{Y}}_{\Lambda,\Phi}$ are analytically given [47].

Input: $\alpha^{(0)} \in \mathbb{R}^{K\cdot J}$ and $\bar{g} \in \mathbb{R}^{K\cdot L}$.

$r^{(0)} = FTMVP(FMVP(\alpha^{(0)}) - \bar{g})$;
$p^{(0)} = -r^{(0)}$;
$k = 0$;
while $r^{(k)} \neq 0$ **do**
  $\widetilde{p^{(k)}} = FMVP(p^{(k)})$;
  $\gamma^{(k)} = \frac{|r^{(k)}|^2}{|\widetilde{p^{(k)}}|^2}$;
  $\alpha^{(k+1)} = \alpha^{(k)} + \gamma^{(k)} p^{(k)}$;
  $r^{(k+1)} = r^{(k)} + \gamma^{(k)} FTMVP(\widetilde{p^{(k)}})$;
  $\beta^{(k+1)} = \frac{|r^{(k+1)}|^2}{|r^{(k)}|^2}$;
  $p^{(k+1)} = -r^{(k+1)} + \beta^{(k+1)} p^{(k)}$;
  $k = k + 1$;
end while
$\alpha^+ = \alpha^{(k+1)}$;

Output: α^+, which is the norm-minimal solution of the normal equation

$$(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi} \alpha = (N^{\mathcal{Y}}_{\Lambda,\Phi})^T \bar{g}.$$

Arithmetic complexity: $\mathcal{O}(JK\log(K)L)$.
Storage complexity: $\mathcal{O}(K(J+L))$.

4.2 Numerical Results

We give numerical results to study the behavior of the algebraic reconstruction technique with respect to its image reconstruction capability and its computational efficiency. The reconstructions are considered in dimension two, which is a case of practical interest.

We provide reconstructions for *artificial* phantoms and for *real* absorption densities. Both, the phantoms and the absorption densities, are assumed to have support in the unit disc $B_1(0) \subseteq \mathbb{R}^2$. The artificial data has been generated by the efficient and accurate algorithm by Görner, Hielscher and Kunis [20]. For the real experimental data, we give thanks to the Institute for Biological and Medical Imaging at Helmholtz Zentrum München.

In order to apply the fast conjugate gradient method [Algorithm 5] for the reconstruction process, we use a polar reconstruction grid $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ (29) to generate block circulant structures in $N^{\mathcal{Y}}_{\Lambda,\Phi}$ and a kernel function for which the components of the collocation matrix are analytically given. In particular, we use the Gaussian kernel with shape parameter $\varepsilon > 0$

$$\Phi_\varepsilon(x,y) = e^{-\varepsilon^2|x-y|^2}, \qquad x,y \in \mathbb{R}^2.$$

With the usage of the Gaussian kernel, Example 10 in Appendix 6.4 yields that the matrix components are given by

$$(N^{\mathcal{Y}}_{\Lambda,\Phi})_{kl,i} = \pi^{-\frac{1}{2}} e^{-\varepsilon^2(|\xi_k - y_i|^2 + t_l^2)} I_0(2t_l\varepsilon^2|\xi_k - y_i|), \tag{30}$$

where I_0 is the modified Bessel function of the first kind of order zero.

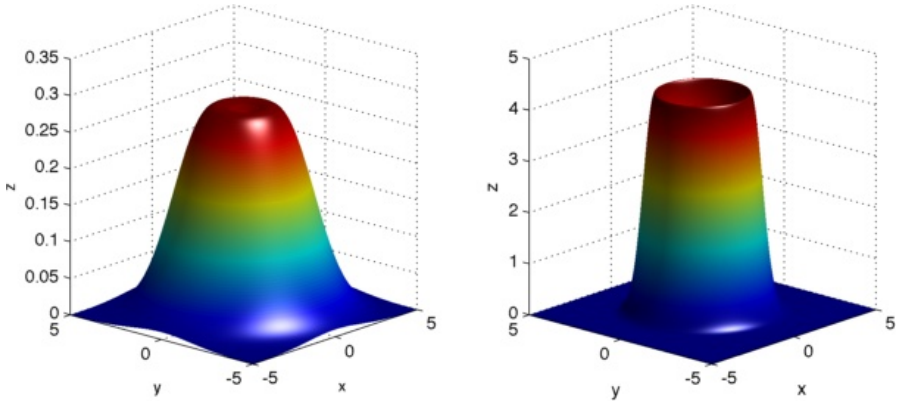

Figure 8: Matrix components of the collocation matrix $N^{\mathcal{Y}}_{\Lambda,\Phi}$ with Gaussian kernel function. Displayed is the expression (30) as a function of y_i for $\xi_k = (0,0)^T$, $t_l = 2$ with shape parameter $\varepsilon = 0.5$ (left) and $\varepsilon = 1.5$ (right).

Algorithm 5 requires a couple of parameters which have to be set a-priori: the maximal number k of iterations, the reconstruction radii $\mathcal{R}_J$ for the polar grid $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ and the shape parameter ε. In Example 6, we discuss their influence on the performance and give advice for a parameter selection.

In our implementation of Algorithm 5, we work with the initial vector $\alpha^{(0)} = 0 \in \mathbb{R}^{K\cdot J}$. When having reached the maximal number k of iterations, we proceed with the coefficient vector $\alpha^{(k)}$ by evaluating the function

$$f_{\mathcal{Y}}^{(k)} := \sum_{i=1}^{N} \alpha_i^{(k)} \Phi_\varepsilon(\cdot, y_i)$$

at $M \times M$ uniformly distributed evaluation points x_j in the unit cube $[-1,1]^2$. Fortunately, this can be done by a single matrix vector multiplication with the matrix $(\Phi_\varepsilon(x_j, y_i))_{1\le j\le M^2, 1\le i\le N}$ and the vector $\alpha^{(k)}$.

In order to measure the error between the approximation $f_{\mathcal{Y}}^{(k)}$ and the phantom f^*, we use the standard *root-mean-square error* (RMS-error)

$$\sqrt{\frac{1}{M^2}\sum_{j=1}^{M^2}\left(f^*(x_j) - f_{\mathcal{Y}}^{(k)}(x_j)\right)^2},$$

where $\{x_j\}_{j=1}^{M^2}$ are the uniformly distributed evaluation points from above. The closer the RMS-error to zero, the more the approximation $f_{\mathcal{Y}}^{(k)}$ is considered to be accurate.

For the comparison of computational efficiency, all simulations are run in MATLAB (R2012a) [1] on a computer equipped with a 2.7 GHz Intel Core i5 CPU and 4GByte main memory.

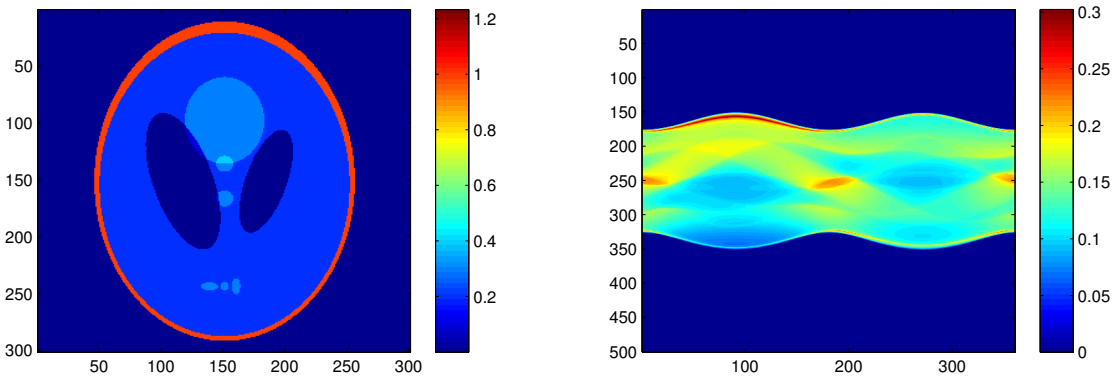

Figure 9: Shepp-Logan head phantom and its spherical mean values for spherical transducer geometry.

[1] MATLAB® is a trademark of The MathWorks, Inc. and is used with a student license.

Example 5.
We start by giving the reconstructed result from the spherical mean values of the Shepp-Logan head phantom [Figure 9]. The experimental setup is as follows: there are $K = 360$ detectors uniformly distributed on the unit circle S^1. Further, the spherical means have been measured $L = 500$ times within the time interval (0 2]. In addition, the image resolution of the phantom is 300×300, i.e. $M = 300$.

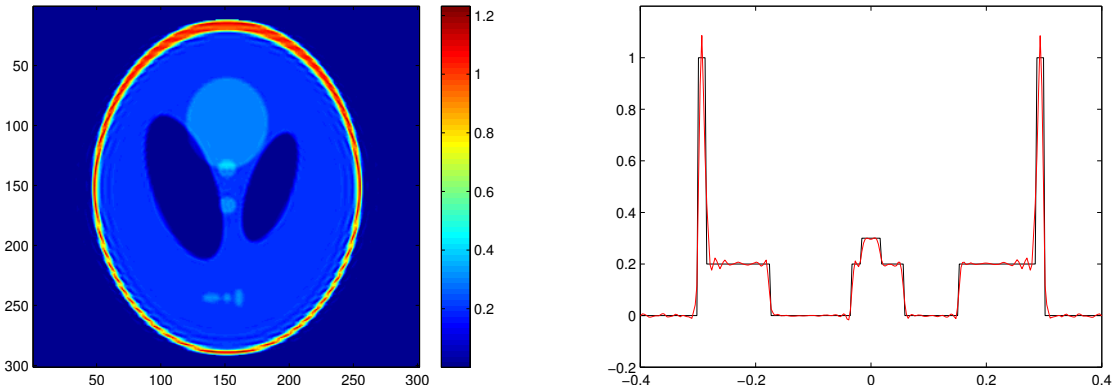

Figure 10: Reconstructed Shepp-Logan head phantom (left) and cross-section plot (right) through horizontal pixel line 140.

On the left in Figure 10, we see the reconstruction for a shape parameter $\varepsilon = 165$ reconstructed on a polar reconstruction grid with radii $\mathcal{R}_{50} = \{r_j = \frac{j}{125} \mid j = 1, \ldots, 50\}$. We stopped the fast conjugate gradient method [Algorithm 5] after $k = 29$ iterations with CPU time 387.4s. The initial residual has been reduced by a factor of 10^{-3}. Comparing the reconstructed result in Figure 10 with the original phantom in Figure 9, we observe that the method performs fairly good with a RMS-error of 0.064.

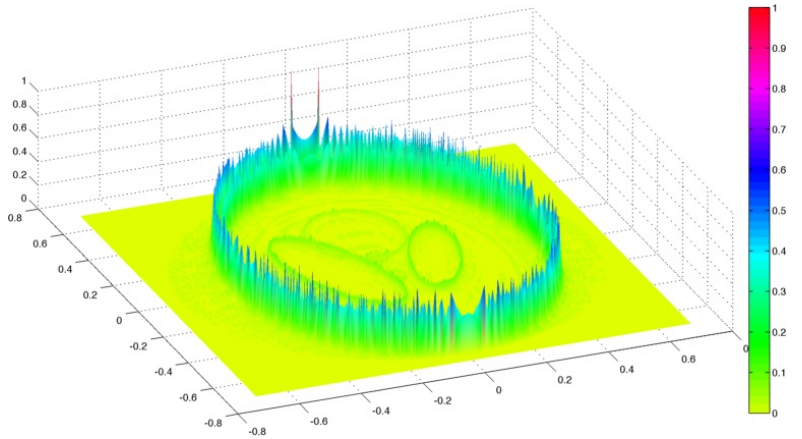

Figure 11: Error plot for the reconstruction of the Shepp-Logan head phantom.

We can see clearly in the cross-section plot on the right in Figure 10 and in the error plot [Figure 11] that most of the error is concentrated near the boundary of the characteristic functions. This is no surprise since the Shepp-Logan phantom (as a superposition of characteristic functions) is not an element of the native space $N_{\Phi_\varepsilon}(B_1(0)) \subseteq \bigcap_{s\in\mathbb{N}} H^s(B_1(0))$ of the Gaussian kernel [Example 1]. Hence, by the theory presented in this thesis, it is not surprising that the reconstructed picture is smoother and thus a bit blurred compared to the original phantom.

Example 6. *Reconstruction parameters*
The fast conjugate gradient method [Algorithm 5] requires three parameters to be set a-priori: the number k of iterations, the reconstruction radii $\mathcal{R}_J$ for the polar grid $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ and the shape parameter ε. In this example we discuss how these parameters influence the reconstruction process. Further, we give advice for a parameter selection.

We discuss the influence of the parameters by reconstructing the smiley phantom from its spherical mean values [Figure 12]. Let us give the experimental setup: there are $K = 360$ detectors uniformly distributed on the unit circle S^1. Further, the spherical means have been measured $L = 500$ times within the time interval $(0\ 2]$. In addition, the image resolution of the phantom is 450×450.

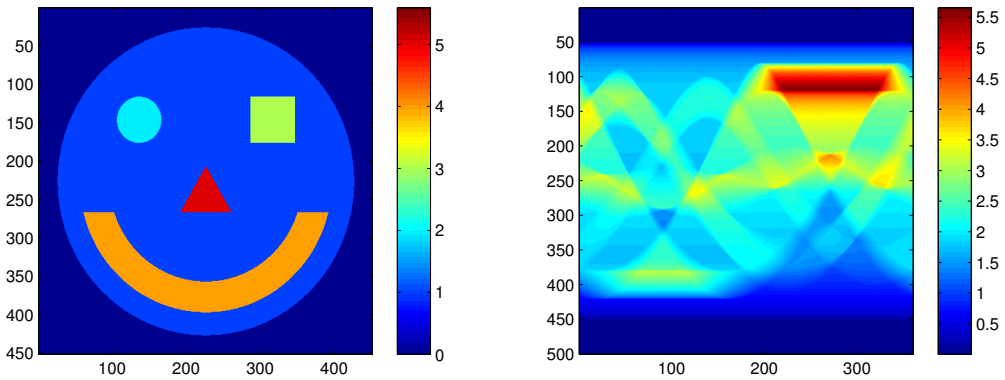

Figure 12: Smiley phantom and its spherical mean values for spherical transducer geometry.

Let us start with investigating the number of iterations. Recall, that under the assumption of positive definiteness of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$, the iterates $\alpha^{(k)}$ in Algorithm 5 tend to produce better and better approximations $f^{(k)}_{\mathcal{Y}}$ to the ART-solution $f^{+}_{\mathcal{Y}}$ (and thus to the phantom f^*). It is further a well-known fact [32, chapter 5.1], that the conjugate gradient method [Algorithm 3] is likely to converge after the first few iterations. In late stages of the iteration, though, the convergence may become very slow. The question thus arises, when to stop the iteration in Algorithm 5 in order to balance the trade-off between the approximation quality (measured by the RMS-error) and the computational efficiency (measured by the number of iterations respectively CPU time).

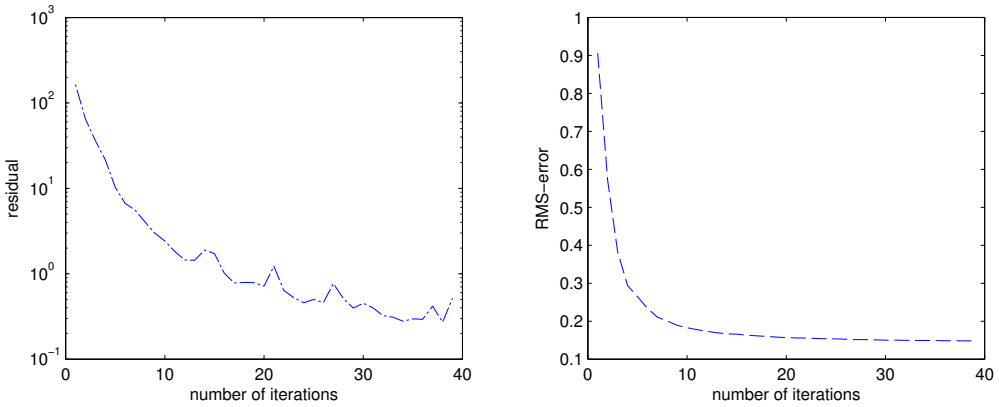

Figure 13: Convergence history for the fast conjugate gradient method applied to the spherical mean values of the smiley phantom [Figure 12].

In Figure 13, we see the convergence history of Algorithm 5, i.e. the residual $r^{(k)}$ respectively the RMS-error as functions of k. The algorithm has been executed for a shape parameter $\varepsilon = 120$ and reconstruction radii $\mathcal{R}_{100} = \{r_j = \frac{2j}{225} \mid j = 1, \dots, 100\}$. Both graphs illustrate fast initial convergence and slow convergence after $k = 20$ iterations, which corresponds to a decrease of the initial residual by a factor of 10^{-3}. Figure 14 shows the reconstructed images for a few iterations. All experiences with Algorithm 5 (even applied to different phantoms) showed a similar behavior: the convergence essentially stagnates after the initial residual decreased by a factor of 10^{-3}. Therefore, we recommend to consider the latter as a stopping condition for the fast conjugate gradient method.

We now turn our attention to the set of reconstruction radii $\mathcal{R}_J$. According to the convergence result for the algebraic reconstruction technique in section 3.3.2, the accuracy of ART improves when more reconstruction points are used. In fact, this corresponds to an increase of the number of reconstruction radii for the polar reconstruction grid $\mathcal{Y}_{\Theta_\Xi, \mathcal{R}_J}$. Due to the trade-off between approximation quality and computational efficiency, we are interested in a reasonable choice for $\mathcal{R}_J$. Clearly, a reasonable choice for $\mathcal{R}_J$ depends on the desired image resolution $M \times M$. In the following we give an advice for $M \approx 500$. In Table 2, we report results of a series of experiments for increasingly larger sets of reconstruction points. We used the fixed shape parameter $\varepsilon = 120$ for all the experiments and stopped the fast conjugate gradient method after a decrease of the initial residual by a factor of 10^{-3}. Further, we reconstructed on equidistant circles and used the a-priori information provided by the sinogram [Figure 12] that the phantom's support is contained in $B_{0.8}(0)$. Table 2 illustrates a stagnation of convergence when reconstructing with more than $J = 100$ reconstruction circles. On the other side, we see an increase of the CPU time. Other experiences with Algorithm 5 (even applied to different phantoms) have provided similar results. Therefore, we recommend to consider 50-100 (depending on a-priori information of the phantom's support) equidistant reconstruction circles for

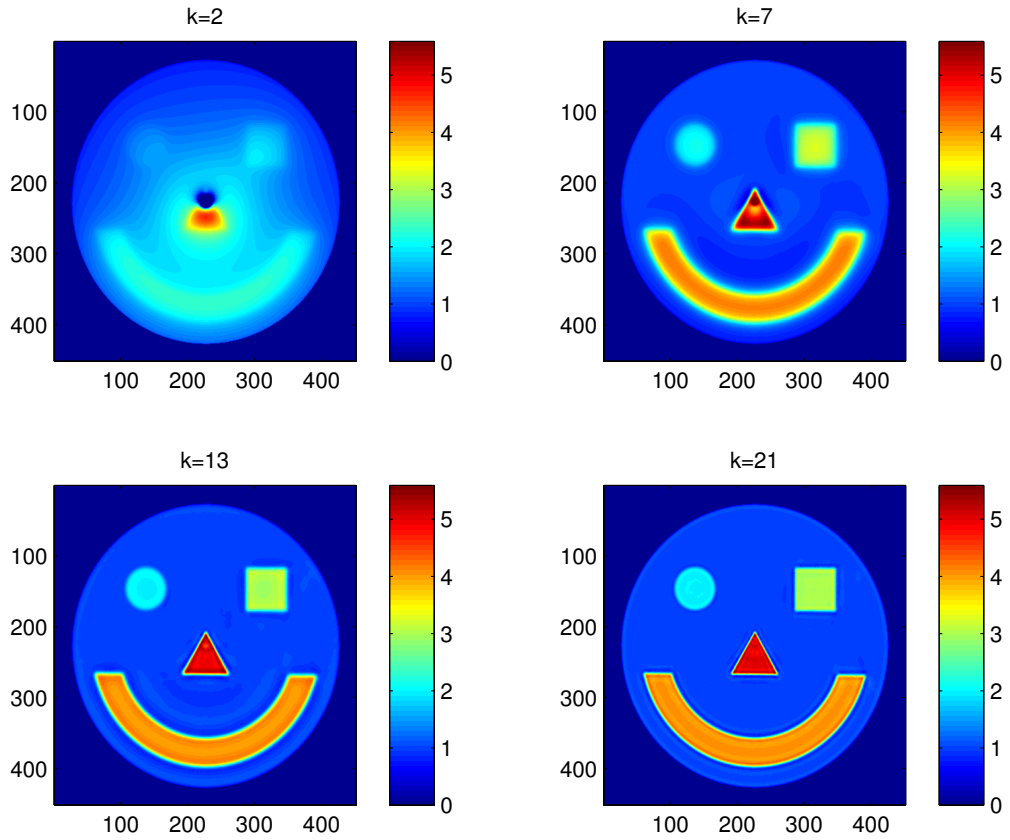

Figure 14: Reconstructed smiley phantom for different iteration steps.

J	# of elements in $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$	RMS-error	CPU time
20	7200	2.116637	58.7
40	14400	0.513901	130.5
60	21600	0.175194	208.9
80	28800	0.159030	277.1
100	36000	0.155130	346.7
120	43200	0.154910	419.7
140	50400	0.154036	493.9
160	57600	0.153489	562.1

Table 2: RMS-error and CPU time for increasingly larger sets of reconstruction points $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ with reconstruction radii $\mathcal{R}_J = \{0.8\frac{i}{J} \mid i = 1,\dots,J\}$.

an image resolution with $M \approx 500$ in order to have a fairly good computational efficiency and approximation quality.

Finally, let us analyze the influence of the the shape parameter ε. It is well-known in the theory of interpolation with radial basis functions [15, 52], that the shape parameter ε governs a trade-off between approximation quality (measured by the RMS-error) and numerical stability (measured by the condition number of the interpolation matrix). There are several strategies to balance this trade-off by an "optimal" choice of ε, and we refer the interested reader to [15, 37, 54] for a comprehensive study. Our strategy used within this chapter is as follows: we scale ε according to the average distance between two neighbouring reconstruction circles of the polar grid $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$. Hence, we end up using a recovery space $X_{\mathcal{Y},\Phi_\varepsilon}$ with "peaked" basis functions $\Phi_\varepsilon(\cdot, y_i)$ for densely spaced reconstruction circles and "flat" basis functions for coarsely spaced circles. Let us focus on an experiment which justifies this strategy for the parameter selection.

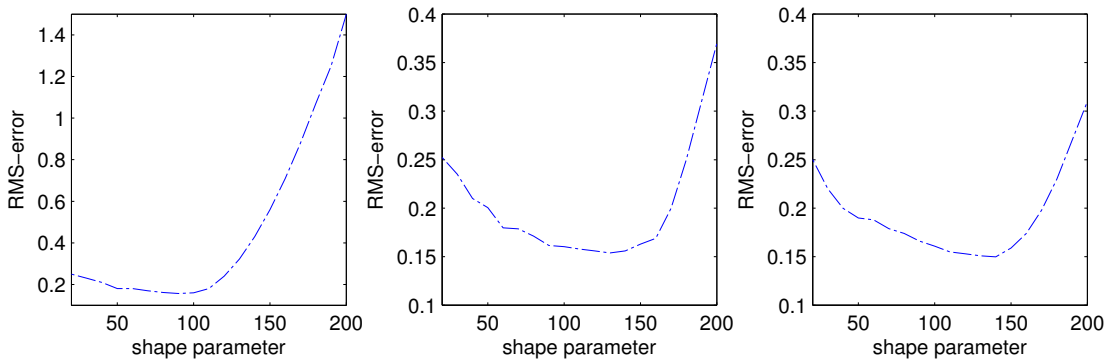

Figure 15: RMS-error as a function of the shape parameter ε for reconstruction sets $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ with equidistant radii $\mathcal{R}_J = \{0.8\frac{i}{J} \mid i = 1, \ldots, J\}$ for J =50 (left), J =100 (middle) and J =150 (right).

In Figure 15, we display the RMS-error as a function of the shape parameter ε for three fixed reconstruction sets $\mathcal{Y}_{\Theta_\Xi,\mathcal{R}_J}$ with equidistant radii $\mathcal{R}_J = \{0.8\frac{i}{J} \mid i = 1, \ldots, J\}$ for $J = 50, 100, 150$. We clearly see that the error curves are not monotonic. Further, we can identify a range for a value of ε for which the RMS-errors are minimal. The interesting fact about the three curves is the following: the bigger the distance between neighbouring reconstruction circles the smaller the value of ε for which the minimum RMS-error is obtained. Therefore, ε has to be chosen according to the distance between two neighbouring reconstruction circles. Of course, if the data we are trying to reconstruct are not generated from a known phantom, then we are not able to choose an "optimal" shape parameter by monitoring the RMS-error as in Figure 15. However, we can give the following heuristically derived rule of thumb for equidistant reconstruction circles, which provided fairly good results in most of the experiments: $\varepsilon \approx -5000\omega + 170$, where ω is the distance between neighbouring reconstruction circles.

Example 7. *Noisy data*
We test Algorithm 5 on data corrupted by white Gaussian noise. To this end, let us consider spherical Radon data as in Example 6, where we added 5% white Gaussian noise to the spherical mean values [Figure 16]. On the left in Figure 17, we see the reconstruction for a shape parameter $\varepsilon = 65$ reconstructed on a polar reconstruction grid with radii $\mathcal{R}_{100} = \{r_j = \frac{2j}{225} \mid j = 1, \ldots, 100\}$. We stopped Algorithm 5 after $k = 11$ iterations with CPU time 180.3s, where the initial residual has been reduced by a factor of 10^{-2}. Further, the reconstruction has a RMS-error of 0.31. It is interesting to note that the reconstruction quality is fairly good even though no explicit regularization strategy has been incorporated.

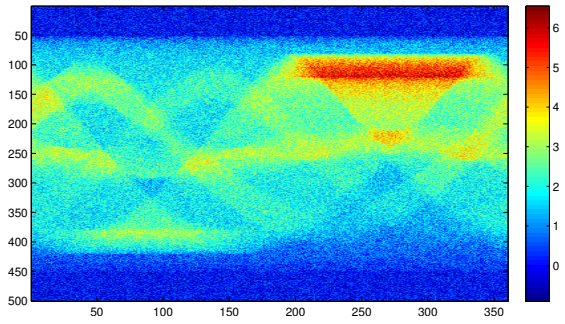

Figure 16: Spherical mean values of the smiley phantom with 5% white Gaussian noise.

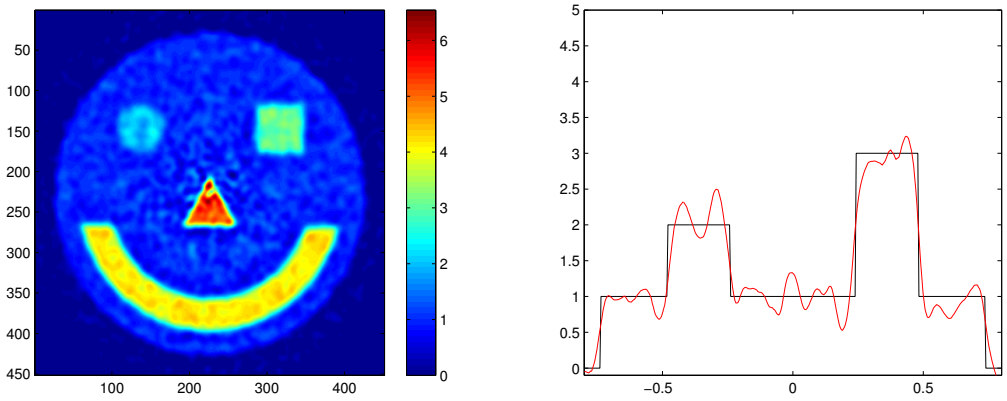

Figure 17: Reconstructed smiley phantom (left) and cross-section plot (right) through horizontal pixel line 150.

Example 8. *Reconstruction of real experimental data*
In the following we reconstruct experimental data (pressure signals) provided by the Institute for Biological and Medical Imaging at Helmholtz Zentrum München. As Algorithm 5 works solely with spherical mean values, we transformed the pressure data into spherical mean values via numerical integration. In what follows, we display the images in grayscale.

On the left in Figure 18, we see pressure signals recorded with a sampling rate of 100MHz. A cylinder containing inked acoustic gel was laterally scanned with $K = 180$ detectors uniformly distributed on a spherical acquisition surface of diameter 0.081m. Figure 19 shows the reconstruction for a shape parameter $\varepsilon = 80$ reconstructed on a grid with radii $\mathcal{R}_{50} = \{r_j = \frac{3j}{225} \mid j = 1, \ldots, 50\}$. We stopped Algorithm 5 after $k = 11$ iterations with CPU time 121.3s, where the initial residual has been reduced by a factor of 10^{-3}.

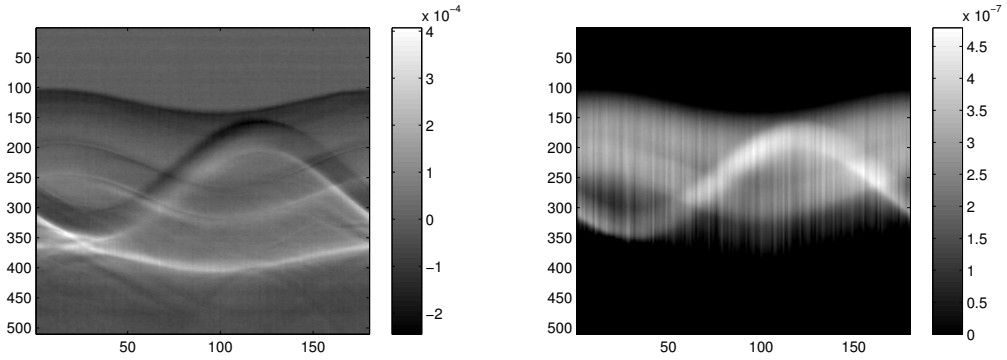

Figure 18: Experimental pressure data for an inked acoustic gel cylinder scan (left) and the corresponding spherical mean values (right).

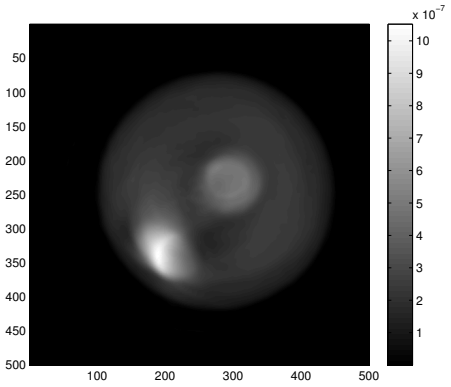

Figure 19: Reconstruction of an inked acoustic gel cylinder phantom.

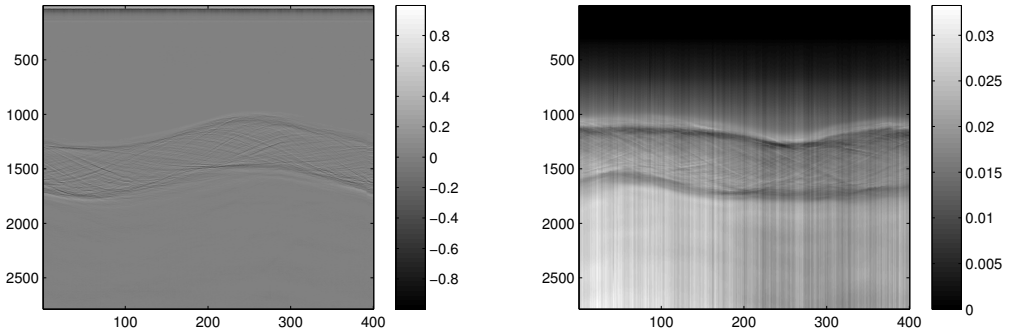

Figure 20: Experimental pressure data (left) for a mouse scan and the corresponding spherical mean values (right).

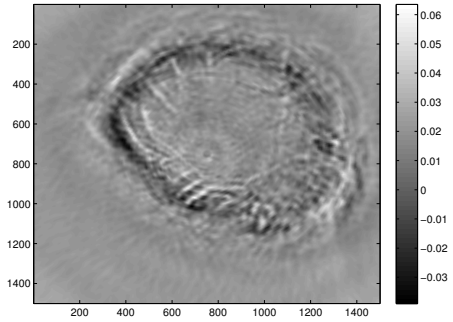

Figure 21: Reconstruction of a mouse scan.

A second data set is displayed on the left in Figure 20. The pressure signals were recorded with a sampling rate of 100MHz. $K = 400$ uniformly distributed detectors laterally scanned a mouse to record the pressure signals. The spherical acquisition surface had a diameter of 0.0418m. In Figure 21, we see the reconstruction for a shape parameter $\varepsilon = 300$ reconstructed on a polar grid with radii $\mathcal{R}_{100} = \{r_j = \frac{2j}{500} \mid j = 1, \ldots, 100\}$. We stopped Algorithm 5 after $k = 25$ iterations with CPU time 3632.2s, where the initial residual has been reduced by a factor of 10^{-3}.

4.3 Notes and Comments

The conjugate gradient method is recommended only for large problems; otherwise, Gaussian elimination or other factorization algorithms for the solution of (26) are to be preferred, since they are less sensitive to rounding errors.

The performance of the conjugate gradient method is tied to the distribution of the eigenvalues of the involved matrix. By *preconditioning* the linear system, it is possible to make the distribution more favorable and improve the convergence of the method significantly [32, chapter 5].

A lot of work has been done in investigating strategies for the selection of an "optimal" shape parameter for radial basis functions. We refer the interested reader to [37] for a method based on cross validation, to [15, chapter 17.1.2] for a method based on the so-called power function, and to [54] for the Contour-Padé algorithm.

5 Conclusion and Open Problems

In this master thesis we proposed innovative image reconstruction techniques for data modeled by the spherical Radon transform in the setting of photoacoustic tomography. In particular, we detailed the derivation of two kernel based solution methods for the semi-discrete inverse problem of PAT: optimal recovery and algebraic reconstruction technique. For both methods, we gave L^2-error estimates in terms of the data density and convergence results. Further, we presented a fast algorithm for the algebraic reconstruction technique. The algorithm's fairly good image reconstruction capability and computational efficiency were proved by several numerical experiments for real and artificially generated data.

Interesting problems arose during the development of this master thesis. Due to the more involved nature of complexity, their elaboration has not been possible.

1. A continuity result for the spherical Radon transform $\mathcal{M}_{\partial\Omega}$ as an operator defined on the whole space $N_\Phi(\Omega)$ respectively $H^\tau(\Omega)$ for a sufficienty regular set $\Omega \subseteq \mathbb{R}^d$ should be investigated. Only a weaker version exists [Proposition 10].

2. An extension of $f \in H^s(\Omega)$ to a function in $H^s(\mathbb{R}^d)$ is given if Ω is an open subset of $\mathbb{R}^d$ [2, 50]. Similar results for not necessarily open "thick" sets are missing.

3. Fractional Sobolev spaces $H^s(\mathbb{R}^d)$ are reproducing kernel Hilbert spaces if $s > \frac{d}{2}$ [Theorem 4]. Similar results for spaces $H^s(\Omega)$, defined for not necessarily open sets Ω, should be investigated.

4. Linear independency of the set of functionals $\Lambda_{\mathcal{K}_\delta}$ [Definition 5] is crucial for the determination of the optimal recovery solution f^+. A handy method for checking linear independency should be developed.

5. An implementation of the optimal recovery algorithm [Algorithm 1] is missing.

6. The positive definiteness of $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$ is important for the convergence of the conjugate gradient method [Algorithm 3]. A sufficient criteria to check the positive definiteness has not been developed so far.

7. As the performance of the conjugate gradient method is tied to the distribution of the eigenvalues of the matrix $(N^{\mathcal{Y}}_{\Lambda,\Phi})^T N^{\mathcal{Y}}_{\Lambda,\Phi}$, preconditioning should be considered as a means to make the distribution of the eigenvalues more favorable for improving the convergence of the method.

8. Strategies to select an "optimal" shape parameter for the algebraic reconstruction technique should be developed by cross validation [37] or the Contour-Padé algorithm [54].

6 Appendix

6.1 Useful Facts from Functional Analysis

Riesz Representation Theorem. *[39, Theorem 6.19]*
Let H be a Hilbert space and H' be its topological dual space. Then for every $\lambda \in H'$ there exists a unique $x_\lambda \in H$ such that

$$\lambda(y) = \langle y, x_\lambda \rangle_H, \quad y \in H.$$

The vector x_λ is called Riesz representer of the functional λ. Further, let $\lambda, \mu \in H'$ and let $x_\lambda, x_\mu \in H$ be their Riesz representers, then

$$\langle \lambda, \mu \rangle_{H'} = \langle x_\lambda, x_\mu \rangle_H.$$

Hahn-Banach Theorem. *[40, Thereom 3.6]*
Let V be a normed vector space with subspace $U \subseteq V$ (not necessarily closed). Let $\psi : U \to \mathbb{R}$ be a linear continuous functional on U, then there exists a linear continuous extension $\Psi : V \to \mathbb{R}$ of ψ such that $\Psi|_U = \psi$.

6.2 Sobolev Spaces of Fractional Order

In the following we introduce fractional Sobolev spaces. The members of fractional Sobolev spaces are tempered distributions whose Fourier transform satisfies certain decay properties. In order to define these function spaces, let us review the Fourier transform of tempered distributions. For more details we refer the interested reader to [2, 40, 44]. The Schwartz space $\mathcal{S}$ is the set of rapidly decreasing functions, i.e. functions $f \in C^\infty(\mathbb{R}^d)$ for which

$$\|f\|_N := \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^d} (1 + |x|^2)^N |(D^\alpha f)(x)| < \infty$$

for every $N \in \mathbb{N}_0$ with multi-index α. The family of semi-norms $\{\|\cdot\|_N\}_{N \in \mathbb{N}}$ induces a topology on $\mathcal{S}$, which makes $\mathcal{S}$ locally convex. Due to the fact that $\mathcal{S}'$ can be identified with a certain subspace of distributions [40, chapter 7] with particular growth restrictions at infinity, one refers to $\mathcal{S}'$ as tempered distributions.

Example 9.
Every $f \in L^p(\mathbb{R}^d)$ with $1 \leq p \leq \infty$ is a tempered distribution. So is every point evaluation functional, every polynomial and, more generally, every measurable function whose absolute value is majorized by some polynomial. Note that in general $f \in L^p_{loc}(\mathbb{R}^d)$ for $1 \leq p \leq \infty$ is not a tempered distribution. The proofs can be found in [40].

Definition 7.
Let $f \in \mathcal{S}'$. Then the Fourier transform of a tempered distribution is defined by

$$\hat{f}(\phi) := f(\hat{\phi}), \quad \phi \in \mathcal{S}.$$

With the extension of the Fourier transform to tempered distribution, we can now define Sobolev spaces of fractional order.

Definition 8.
The Sobolev space of fractional order $s \in \mathbb{R}$ is defined by

$$H^s(\mathbb{R}^d) := \{f \in \mathcal{S}' \mid \hat{f} \in L^1_{loc}(\mathbb{R}^d), \quad \|f\|_{H^s(\mathbb{R}^d)} := \|(1+|\cdot|^2)^{\frac{s}{2}}\hat{f}\|_{L^2(\mathbb{R}^d)} < \infty\}.$$

Remark 7.

a) $H^s(\mathbb{R}^d)$ is a Hilbert space with the inner product

$$\langle f, g\rangle_{H^s(\mathbb{R}^d)} := \int_{\mathbb{R}^d} (1+|\xi|^2)^s \hat{f}(\xi)\hat{g}(\xi)\mathrm{d}\xi.$$

b) The inclusions

$$H^s(\mathbb{R}^d) \subseteq H^{\tilde{s}}(\mathbb{R}^d), \quad \mathcal{S} \subseteq H^s(\mathbb{R}^d), \quad \bigcap_{s\in\mathbb{R}} H^s(\mathbb{R}^d) \subseteq C^\infty(\mathbb{R}^d)$$

hold for every $s, \tilde{s} \in \mathbb{R}$ with $\tilde{s} \leq s$ and can be proved with the subsequent Sobolev embedding theorem [44, Theorem 2.60]:

$$H^s(\mathbb{R}^d) \hookrightarrow C^k(\mathbb{R}^d) \text{ for } s > k + \frac{d}{2}.$$

c) It is interesting to note that the topological dual space $(H^s(\mathbb{R}^d))'$ is isometrically isomorphic to $H^{-s}(\mathbb{R}^d)$ for every $s \in \mathbb{R}$ [44, Theorem 2.58].

d) It is possible to define fractional Sobolev spaces even for Lipschitz sets Ω, i.e. for open bounded connected sets Ω with a boundary $\partial\Omega$ "sufficiently regular" in the sense that it can be thought of as being locally the graph of a Lipschitz-continuous function. We define $H^s(\Omega)$ for $s \in \mathbb{N}$ as the space consisting of all functions $f : \Omega \to \mathbb{R}$ with distributional derivatives $D^\alpha f \in L^2(\Omega)$ for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq s$. Associated with this space is the norm

$$\|f\|^2_{H^s(\Omega)} := \Big(\sum_{|\alpha|\leq k} \Big(\int_\Omega |f(x)|^2 \mathrm{d}x\Big)^2\Big)^{\frac{1}{2}}.$$

One approach for defining $H^s(\Omega)$ for $s = k + \tau$ with $k \in \mathbb{N}_0$ and $0 < \tau < 1$ is via the norm

$$\|f\|_{H^{k+\tau}(\Omega)} := \big(\|f\|^2_{H^k(\Omega)} + |f|^2_{H^{k+\tau}(\Omega)}\big)^{\frac{1}{2}},$$

with involved semi-norm

$$|f|_{H^{k+\tau}(\Omega)} := \Big(\sum_{|\alpha|=k} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^2}{|x-y|^{d+2\tau}} \mathrm{d}x\mathrm{d}y\Big)^{\frac{1}{2}}.$$

For further information we refer the interested reader to [2].

e) In some settings it is essential that the functions involved have compact support. Therefore, we define for $s \in \mathbb{R}$ and a Lipschitz domain $\Omega \subseteq \mathbb{R}^d$

$$H_c^s(\Omega) := \{f \in H^s(\Omega) \mid \operatorname{supp}(f) := \overline{\{x \in \Omega \mid f(x) \neq 0\}} \subseteq \Omega\}.$$

6.3 Characterization of Ill-posedness via Sobolev Space Estimates

In the following we introduce fundamental notions of the theory of ill-posed problems and give a characterization via Sobolev space estimates. Note that we discuss ill-posed problems only in the framework of linear bounded operators between Hilbert function spaces.

Definition 9.
Let $H(\Omega)$ and $H(\Omega')$ be Hilbert function spaces and let $A : H(\Omega) \to H(\Omega')$ be a linear bounded operator. The equation $A(f) = g$ is called well-posed in the sense of Hadamard if the following holds:

1. *Existence: for every $g \in H(\Omega')$ there is (at least one) $f \in H(\Omega)$ such that $A(f) = g$.*
2. *Uniqueness: for every $g \in H(\Omega')$ there is at most one $f \in H(\Omega)$ such that $A(f) = g$.*
3. *Stability: the solution f depends continuously on the data g. That is, for every sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq H(\Omega)$ with $\lim_{n\to\infty} A(f_n) = A(f)$, then $\lim_{n\to\infty} f_n = f$.*

Equations for which at least one of these properties does not hold are called ill-posed.

Remark 8.

a) Existence and uniqueness of solutions for the operator equation $A(f) = g$ depend only on the algebraic nature of the spaces and the operator; that is, whether the operator is surjective or even invertible. Stability, however, depends also on the topologies of the spaces, i.e., whether the inverse operator $A^{-1} : H(\Omega') \to H(\Omega)$ is continuous

b) The practical difficulty with an ill-posed problem is the following. Either $A(f) = g$ is not solvable or not uniquely solvable; this is the case if A is not invertible. If A is invertible with A^{-1} not continuous, then $\|A^{-1}(g') - A^{-1}(g)\|_{H(\Omega)}$ need not to be small even if $\|g - g'\|_{H(\Omega')}$ is. Thus, it is difficult to solve an ill-posed problem in the presence of experimental errors.

In the following we show that a smoothing effect of an operator, described by Sobolev space estimates, is a sufficient criteria for the ill-posedness of an operator equation in a L^2- setting.

Definition 10.
Let $\Omega \subseteq \mathbb{R}^d$, $\Omega' \subseteq \mathbb{R}^q$ be Lipschitz domains and let $s \in \mathbb{R}$ be arbitrary, but fixed. We

call a linear bounded operator $A : L^2(\Omega) \to L^2(\Omega')$ smoothing of degree α, if there is a number $\alpha > 0$ such that there exist positive constants m_s, M_s with

$$m_s \|f\|_{H^s(\Omega)} \leq \|Af\|_{H^{s+\alpha}(\Omega')} \leq M_s \|f\|_{H^s(\Omega)}$$

for all $f \in H_c^s(\Omega)$.

Clearly, a linear bounded operator with smoothing effect is injective and thus invertible by restricting it on its range. Further, the inverse is a continuous operator from $H^{s+\alpha}(\Omega')$ into $H_c^s(\Omega)$. In particular, if $s = -\alpha$, $H_c^s(\Omega)$ is a Sobolev space with negative order whose norm is weaker than the L^2- norm on $L^2(\Omega)$, i.e., there exists a constant $C > 0$ such that $\|f\|_{H^{-\alpha}(\Omega)} \leq C\|f\|_{L^2(\Omega)}$ for every $f \in H_c^{-\alpha}(\Omega)$. This shows that A^{-1} is not continuous as an operator from $L^2(\Omega')$ into $L^2(\Omega)$. Hence, the equation $A(f) = g$ is ill-posed in the sense of Hadamard.

Corollary 25. *[31, chapter 4]*
Let $A : L^2(\Omega) \to L^2(\Omega')$ be a linear bounded α-smoothing operator for Lipschitz domains $\Omega \subseteq \mathbb{R}^d$ and $\Omega' \subseteq \mathbb{R}^q$. Then the equation $A(f) = g$ is ill-posed in the sense of Hadamard.

Proof. The proof follows directly from the considerations above. □

Remark 9.
The smoothing effect of an operator A is nothing else than a continuity result for A and its inverse A^{-1} between fractional Sobolev spaces of certain orders. It is interesting to mention that those continuity results are important for many theoretical and practical purposes in various types of tomography. Several work has been done in these fields. For a good introductory discussion in the setting of X-ray, spherical and classical Radon transform, see [45].

6.4 Action of the Spherical Mean Operator on Radial Kernel Functions

Theorem 26.
Let $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a radial kernel function, i.e. $\Phi(x, y) = \phi(|x - y|)$ for $x, y \in \mathbb{R}^d$. Then

$$\mathcal{M}^x(\Phi(x, y))(\xi, t) = \frac{w_{d-2}}{w_{d-1}} \int_{-1}^{1} \phi(\sqrt{|\xi - y|^2 + t^2 + 2t|\xi - y|\tau})(1 - \tau^2)^{\frac{d-3}{2}} \mathrm{d}\tau$$

for every $t \geq 0$ and $y, \xi \in \mathbb{R}^d$ with $y \neq \xi$. The superscript at the operator indicates the argument to which the operator is applied. Further,

$$\mathcal{M}^x(\mathcal{M}^y(\Phi(x, y))(\nu, s))(\xi, t) = \frac{w_{d-2}^2}{w_{d-1}^2} \int_{-1}^{1} \int_{-1}^{1} \phi(\Psi_{\nu,s}^{\xi,t}(\tau_1, \tau_2))(1 - \tau_1^2)^{\frac{d-3}{2}} (1 - \tau_2^2)^{\frac{d-3}{2}} \mathrm{d}\tau_2 \mathrm{d}\tau_1$$

with

$$\Psi_{\nu,s}^{\xi,t}(x, y) = \sqrt{|\xi - \nu|^2 + s^2 + t^2 + 2t|\xi - \nu|x + 2sy\sqrt{|\xi - \nu|^2 + t^2 + 2t|\xi - \nu|x}}$$

for every $t, s \geq 0$ and $\nu, \xi \in \mathbb{R}^d$ with $|\xi - \nu| \notin \{0, t\}$.

Proof. The essential idea of the proof is to use the Funk-Hecke formula [1, chapter 9.7] in order to transform a spherical integral into an univariate integral: for every function $f \in C([-1,1])$ and every unit vector $\alpha \in S^{d-1}$,

$$\frac{1}{w_{d-2}} \int_{S^{d-1}} f(\langle u, \alpha \rangle) \, \mathrm{d}\sigma(u) = \int_{-1}^{1} f(\tau)(1-\tau^2)^{\frac{d-3}{2}} \, \mathrm{d}\tau \, .$$

Let us prove the first identity. For $y \neq \xi$ and $u \in S^{d-1}$ we have

$$\begin{aligned} |\xi + tu - y|^2 &= |\xi - y|^2 + t^2 + 2\langle tu, \xi - y \rangle \\ &= |\xi - y|^2 + t^2 + 2t|\xi - y| \langle u, (\xi - y)|\xi - y|^{-1} \rangle \end{aligned}$$

and therefore

$$\begin{aligned} \mathcal{M}^x(\Phi(x,y))(\xi,t) &= \frac{1}{w_{d-1}} \int_{S^{d-1}} \Phi(\xi + tu, y) \, \mathrm{d}\sigma(u) \\ &= \frac{1}{w_{d-1}} \int_{S^{d-1}} \phi(\sqrt{|\xi - y|^2 + t^2 + 2t|\xi - y| \langle u, (\xi - y)|\xi - y|^{-1} \rangle}) \, \mathrm{d}\sigma(u) \\ &= \frac{w_{d-2}}{w_{d-1}} \int_{-1}^{1} \phi(\sqrt{|\xi - y|^2 + t^2 + 2t|\xi - y|\tau})(1-\tau^2)^{\frac{d-3}{2}} \, \mathrm{d}\tau \, . \end{aligned}$$

For the second identity we apply the Funk-Hecke formula twice. For $u, w \in S^{d-1}$ and $|\xi - \nu| \notin \{0, t\}$ we have

$$\begin{aligned} |\xi + tu - (\nu + sw)|^2 &= |(\xi - \nu - sw) + tu|^2 = |\xi - \nu - sw|^2 + 2\langle \xi - \nu - sw, tu \rangle + t^2 \\ &= |\xi - \nu|^2 + s^2 + t^2 + 2t\langle \xi - \nu, u \rangle \ - 2s\langle \xi - \nu, w \rangle - 2st\langle w, u \rangle \\ &= |\xi - \nu|^2 + s^2 + t^2 + 2t\langle \xi - \nu, u \rangle - 2s\langle \xi - \nu + tu, w \rangle \\ &= |\xi - \nu|^2 + s^2 + t^2 + 2t\langle \xi - \nu, u \rangle \\ &+ 2s|\xi - \nu + tu| \langle (\nu - \xi - tu)|\nu - \xi - tu|^{-1}, w \rangle \\ &= |\xi - \nu|^2 + s^2 + t^2 + 2t|\xi - \nu| \langle (\xi - \nu)|\xi - \nu|^{-1}, u \rangle \\ &+ 2s\sqrt{|\xi - \nu|^2 + t^2 + 2t|\xi - \nu| \langle (\xi - \nu)|\xi - \nu|^{-1}, u \rangle} \langle \frac{\nu - \xi - tu}{|\nu - \xi - tu|}, w \rangle. \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{M}^x(\mathcal{M}^y(\Phi(x,y))(\nu,s))(\xi,t) &= \frac{1}{w_{d-1}^2} \int_{S^{d-1}} \int_{S^{d-1}} \Phi(\xi + tu, \nu + sw) \, \mathrm{d}\sigma(u) \, \mathrm{d}\sigma(w) \\ &= \frac{1}{w_{d-1}^2} \int_{S^{d-1}} \int_{S^{d-1}} \phi(|\xi + tu - (\nu + sw)|) \, \mathrm{d}\sigma(u) \, \mathrm{d}\sigma(w) \\ &= \frac{w_{d-2}^2}{w_{d-1}^2} \int_{-1}^{1} \int_{-1}^{1} \phi(\Psi_{\nu,s}^{\xi,t}(\tau_1, \tau_2))(1-\tau_1^2)^{\frac{d-3}{2}} (1-\tau_2^2)^{\frac{d-3}{2}} \, \mathrm{d}\tau_2 \, \mathrm{d}\tau_1 \end{aligned}$$

with

$$\Psi_{\nu,s}^{\xi,t}(x,y) = \sqrt{|\xi - \nu|^2 + s^2 + t^2 + 2t|\xi - \nu|x + 2sy\sqrt{|\xi - \nu|^2 + t^2 + 2t|\xi - \nu|x}}. \qquad \square$$

Example 10.
Consider the Gaussian kernel in $\mathbb{R}^2$ with shape parameter $\varepsilon > 0$

$$\Phi_\varepsilon(x,y) = \phi(\varepsilon|x-y|) = e^{-\epsilon^2|x-y|^2}, \quad x,y \in \mathbb{R}^2.$$

Then

$$\begin{aligned}
\mathcal{M}^x(\Phi_\varepsilon(x,y))(\xi,t) &= \frac{w_0}{w_1}\int_{-1}^{1} \phi\Big(\varepsilon\sqrt{|\xi-y|^2+t^2+2t|\xi-y|\tau}\Big)(1-\tau^2)^{-\frac{1}{2}}\,d\tau \\
&= \frac{1}{\pi}e^{-\varepsilon^2(|\xi-y|^2+t^2)}\int_{-1}^{1} e^{-\varepsilon^2 2t|\xi-y|\tau}(1-\tau^2)^{-\frac{1}{2}}\,d\tau \\
&= \pi^{-\frac{1}{2}}e^{-\varepsilon^2(|\xi-y|^2+t^2)}J_0(2ti\varepsilon^2|\xi-y|) \\
&= \pi^{-\frac{1}{2}}e^{-\varepsilon^2(|\xi-y|^2+t^2)}I_0(2t\varepsilon^2|\xi-y|)
\end{aligned}$$

by using the Poisson representation formula for Bessel functions [23, Equation 8.411.10]

$$J_v(z) = \frac{(\frac{z}{2})^v}{\Gamma(v+\frac{1}{2})}\int_{-1}^{1} e^{izs}(1-s^2)^{v-\frac{1}{2}}\,ds, \quad v > -\frac{1}{2}, \quad z \in \mathbb{C}$$

and the definition of modified Bessel functions of the first kind

$$I_\alpha(x) := J_\alpha(ix)i^{-\alpha}, \quad x \geq 0.$$

References

[1] G. E. Andrews, R. Askey and R. Roy, "Special Functions." *Cambridge University Press*, (1999).

[2] R. A. Adams, "Sobolev Spaces." *Academic Press*, New York, (1975).

[3] M. Ansorg, F.Filbir, W. R. Madych and R. Seyfried, "Summability Methods for Reconstruction from Circular Mean Data." submitted, (2012).

[4] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the art algorithm." *Ultrason. Imaging* vol 6, pp. 81-84, (1984).

[5] G. Ambartsoumian and P. Kuchment, "On the injectivity of the circular Radon transform." *Inverse Problems* 21, 473-485, (2005).

[6] M. L. Agranovsky, P. Kuchment and E.T. Quinto, "Range description of the spherical mean Radon transform." *J. Funct. Anal.*, 248 (2), 344-386, (2007).

[7] M. Agranovsky and E. T. Quinto, "Injectivity sets for the Radon transform over circles and complete systems of radial functions." *J. Funct. Anal.*, 139:383-414, (1996).

[8] R. Bender, R. Gordon and G. Herman, "Algebraic reconstruction techniques (ART) for three dimensional electron microscopy and X-ray photography." *Journal of Theoretical Biology* 29(3), 471-481, (1970).

[9] M. Bertero, C. De Mol and E. Pike, "Linear inverse problems with discrete data I: General formulation and singular system analysis." *Inverse Problems* 1(4):301-330, (1985).

[10] M. Bertero, C. De Mol and E. Pike, "Linear inverse problems with discrete data II: Stability and regularization." *Inverse Problems* 4(3):573-594, (1988).

[11] S. Bochner, "Monotone Funktionen, Stieltjes Integrale und harmonische Analyse." *Math. Ann.* 108, pp. 378-410, (1933).

[12] F. Cucker and S. Smale, "On the mathematical foundation on learning." *Bull. Amer. Math. Soc.* 39, 1-49, (2002).

[13] L. C. Evans, "Partial Differential Equations." *American Mathematical Society*, first Indian edition, (2009).

[14] G. E. Fasshauer, "Solving partial differential equations by collocation with radial basis functions." *Surface Fitting and Multiresolution Methods* , Vanderbilt University Press 131-138, Nashville, (1997).

[15] G. E. Fasshauer, "Meshfree Approximation Methods with MATLAB." *World Scientic*, Singapore, (2007).

[16] F. Filbir, R. Hielscher and W. R. Madych, "Reconstruction from circular and spherical mean data." *Appl. Comput. Harmon. Anal.* 29, 111-120, (2009).

[17] D. Finch, M. Haltmeier and Rakesh, "Inversion of the spherical means and wave equation in even dimensions." *SIAM J. Appl. Math.* 86 (2), 392-412, (2007).

[18] O. Forster, "Analysis 3." *Vieweg-Teubner*, 5. Auflage, Wiesbaden, (2009).

[19] C. Franke and R. Schaback, "Solving partial differential equations by collocation using radial basis functions." *Appl. Math. Comput.* 93, 73-82, (1998).

[20] T. Görner, R. Hielscher and S. Kunis, "Efficient and accurate computation of spherical mean values at scattered center points." *Preprint*, http://www.analysis.uos.de/goerner/software/, (2011).

[21] P. Gilbert, "Iterative methods for the reconstruction of three dimensional objects from their projections." *J. Theor. Biol.*, vol. 36, pp. 105-117, (1972).

[22] G. Golub and C. van Loan, "Matrix Computations." *The Johns Hopkins University Press*, third edition, Baltimore, (1996).

[23] I.S. Gradshteyn, I.M Ryzhik, A. Jeffrey, D. Zwinllinger, "Table of Integrals, Series, and Products." *Academic Press*, seventh edition, (2007).

[24] M. Golomb and H. F. Weinberger, "Optimal approximation and error bounds." *On Numerical Approximation*, R.E. Langer(ed), University of Wisconsin Press, pp.117-190, (1959).

[25] P. R. Halmos, "Measure Theory." *Van Nostrand*, New York, (1950).

[26] M. R. Hestens and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems." *J. Res. Nat. Bur. Stand.* 49, 409-36, (1952).

[27] A. Kirsch, "An Introduction to the Mathematical Theory of Inverse Problems." *Springer*, (2011).

[28] J. Krebs, A. K. Louis and H. Wendland, "Sobolev error estimates and a priori parameter selection for semi-discrete Tikhonov regularization." *Journal of Inverse and Ill-Posed Problems* 17(9):845-869, (2009).

[29] A. C. Kak and M. Slaney, "Principles of Computerized and Tomographic Imaging." *SIAM Classics*, (2001).

[30] L. A. Kunyansky, "A series solution and fast algorithm for the inversion of the spherical mean Radon transform." *Inverse Problems*, 23, 11-20, (2007).

[31] F. Natterer, "The Mathematics of Computerized Tomography." *Wiley-Teubner*, Chichester-Stuttgart, (1986).

[32] J. Nocedal and S. J. Wright, "Numerical Optimization." *Springer*, New York, (1999).

[33] R. Opfer, "Multiscale kernels." *Adv. Comput. Math.* 25, 357-380, (2006).

[34] V. Palamodov, "Remarks on the general Funk-Radon transform and thermoacoustic tomography." *Preprint*, math.AP/0701204, (2007).

[35] T. Poggio and S. Smale, "The mathematics of learning: Dealing with the data." *Amer. Math. Soc.* 50, 537-544, (2003).

[36] A. Rieder, "Keine Probleme mit Inversen Problemen: Eine Einführung in ihre stabile Lösung." *Vieweg*, Wiesbaden, (2003).

[37] S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation." *Adv. in Comput. Math.* 11, pp.193-210, (1999).

[38] B. Rubin, "Inversion formulae for the spherical mean in odd dimensions and the Euler-Poisson-Darboux equation." *Inverse Problems* 24, 025021, (2008).

[39] W. Rudin, "Real and Complex Analysis." *McGraw-Hill*, third edition, New York, (1987).

[40] W. Rudin, "Functional Analysis." *McGraw-Hill*, second edition, (2006).

[41] C. Rieger and B. Zwicknagel, "Sampling inequalities for infinitely smooth functions with applications to interpolation and machine learning." *Adv Comput Math*, 32:103–129, (2010).

[42] R. Schaback, "Error estimates and condition numbers for radial basis function interpolation." *Adv. Comput. Math.* 3, 251-264, (1995).

[43] R. Schaback, "Convergence of unsymmetric Kernel-based meshless collocation methods." *SIAM J. Numer. Anal.* 45/1, 333–351, (2007).

[44] B. Schmidt, "Weak Convergence Methods for Nonlinear Partial Differential Equations." *Lecture notes for the summer term 2010*, http://www.math.uni-augsburg.de/prof/ana/arbeitsgruppe/schmidt/skripten/WeakMethods.pdf, (2010).

[45] O. Scherzer, "Handbook of Mathematical Methods in Imaging." *Springer Science+Business Media LLC*, (2011).

[46] B. Sieber. "Spectral Methods for Reconstruction of Functions from Spherical Radon Data." *Masters's Thesis*, München, (2011).

[47] J. Sigl. "Reconstruction of Functions from Spherical Mean Data using Radial Basis Functions." *Bachelor's Thesis*, München, (2011).

[48] P. Stefanov and G. Uhlmann, "Thermoacoustic tomography with variable sound speed." *Inverse Problems*, 25, 075011, (2009).

[49] R. Schaback and H. Wendland, "Kernel techniques: from machine learning to meshless methods." *Acta Numer.* 15, 543–639, (2006).

[50] R. A. DeVore and R. C. Sharpley, "Besov-Spaces on Domains in RD." *Trans. of the Amer. Math. Soc.* vol 335, issue 2, 843-864, (1993).

[51] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree." *Adv. in Comput. Math.* 4, pp. 389-396, (1995).

[52] H. Wendland, "Scattered Data Approximation." *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge, UK, (2005).

[53] D. Werner, "Funktionalanalysis." *Springer*, 6. Auflage, (2007).

[54] G. B. Wright, "Radial Basis Function Interpolation: Numerical and Analytical Developments." *Ph.D. Dissertation*, University of Colorado at Boulder, (2003).

[55] M. Xu and L.H. Wang, "Photoacoustic imaging in biomedicine." *Review of Scientific Instruments*, 77 (4): 041101, (2006).

[56] Q. Ye, "Reproducing kernels of generalized Sobolev spaces via a Green function approach with differential operators." *Preprint*, (2010).