Regularization of Linear Ill-posed Problems in Two Steps:

Combination of Data Smoothing and Reconstruction Methods

von Esther Klann

Dissertation
zur Erlangung des Grades einer Doktorin der Naturwissenschaften
– Dr. rer. nat. –

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Oktober 2005

Regularization of Linear Ill-posed Problems in Two Steps:

Combination of Data Smoothing and Reconstruction Methods

von Esther Klann

Dissertation
zur Erlangung des Grades einer Doktorin der Naturwissenschaften
– Dr. rer. nat. –

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Oktober 2005

Datum des Promotionskolloquiums: 11.01.2006

Gutachter: Prof. Dr. Peter Maaß, Universität Bremen
PD Dr. habil. Ronny Ramlau, Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz

Zusammenfassung

Diese Arbeit ist ein Beitrag zum Themengebiet *schlechtgestellte inverse Probleme*. Dieses Gebiet ist seit etwa vier Jahrzehnten Gegenstand mathematischer Forschung. In den letzten zehn Jahren gab es folgende Entwicklung: Neben operatorangepassten Lösungsverfahren werden zunehmend solche Verfahren betrachtet, die insbesondere Glattheitseigenschaften von Funktionen berücksichtigen.

Die vorliegende Arbeit schließt an diese Entwicklung an. Es werden *Zwei-Schritt-Verfahren* zur Lösung von linearen schlechtgestellten inversen Problemen eingeführt und untersucht. Die generelle Idee von Zwei-Schritt-Verfahren besteht darin zunächst eine Datenschätzung von möglicherweise verrauschten Messdaten vorzunehmen und dann ausgehend von den geschätzten Daten das inverse Problem zu lösen.

Neben der allgemeinen Darstellung von Zwei-Schritt-Verfahren werden zwei konkrete Realisierungen analysiert. Zum einen werden klassische Regularisierungsverfahren wie die von Tikhonov oder Landweber eingeführten Methoden als Zwei-Schritt-Verfahren interpretiert. Von diesen klassischen Methoden ist bekannt, dass sie in der Regel zu stark glätten. Die Interpretation als Zwei-Schritt-Verfahren erklärt diese Eigenschaft. Darüber hinaus werden die klassischen Methoden modifiziert, so dass durch einen Parameter direkt Einfluss auf die Stärke der Glättungseigenschaften des Verfahrens genommen werden kann. Die Ordnungsoptimalität bleibt dabei für einen bestimmten Parameterbereich erhalten.

Zum anderen wird die Kombination von Wavelet Shrinkage und klassischen Regularisierungsverfahren untersucht. Es ergibt sich dadurch ein ordnungsoptimales Verfahren, das sowohl an Glattheitseigenschaften von Funktionen in Sobolev- und Besovräumen angepasst ist, als auch das singuläre System des Operators berücksichtigt. Weiter wird gezeigt, dass auch die Kombination von Wavelet Shrinkage und den modifizierten klassischen Verfahren – mit vergrößertem Parameterbereich – ordnungsoptimal ist.

Die Kombination von Wavelet Shrinkage und Tikhonov Regularisierung wird auf ein Problem aus der medizinischen Bildverarbeitung angewendet.

Abstract

This thesis is a contribution to the field of *ill-posed inverse problems*. This field has been in the focus of mathematical research for the past four decades. During the last ten years a new development has taken place: Besides operator-adapted methods for the solution of inverse problems also methods adjusted to smoothness properties of functions are studied.
The thesis at hand is linked to this development. Its intention is to present and analyze *two-step methods* for the solution of linear ill-posed problems. It is the fundamental idea of a two-step method to perform first a data estimation step of probably noisy data and then to perform a reconstruction step to solve the inverse problem using the data estimate.
Besides the general description of two-step methods two realizations are analyzed. On the one hand classical regularization methods like the ones proposed by Tikhonov or Landweber are interpreted as two-step methods. It is well-known that solving inverse problems by classical regularization methods generally results in an oversmoothing. This effect is explained by the two-step approach. Furthermore the classical methods are modified such that a parameter allows to control the amount of smoothing. For a certain range of this parameter the modified method is order optimal.
On the other hand the combination of wavelet shrinkage and classical regularization methods is analyzed. This yields an order optimal method which is, by the use of wavelet shrinkage, adapted to smoothness properties of functions in Sobolev and Besov spaces and, by the use of the singular system, adapted to the operator under consideration. In addition it is shown that the combination of wavelet shrinkage and the modified classical methods – with a greater parameter range – is order optimal.
The combination of wavelet shrinkage and Tikhonov regularization is applied to a problem from medical imaging.

I would like to thank:

Prof. Dr. Peter Maaß and PD Dr. Ronny Ramlau for supervision, support and motivation in writing this thesis.

All the people from the "Zentrum für Technomathematik" for mathematical and technical support, especially the whole "AG Technomathematik" for a very good working atmoshpere and the nice time I had.

My colleagues Jenny Niebsch, Caroline Böß, Mascha Zhariy, Mathias Lindemann, Lutz Justen, Kristian Bredies, Dirk Lorenz, Henning Thielemann and Gerd Teschke for support, advices and discussions - both mathematically and personally.

Esther Klann
Zentrum für Technomathematik
Universität Bremen

Zentrum für
Technomathematik

Contents

Chapter 1

Introduction

The mathematical field of inverse problems has been in the focus of mathematical literature for many years. This is so because inverse problems occure in a wide variety of applications such as medical imaging, heat conduction, atmospheric imaging, image reconstruction and parameter estimation.
Very frequently ill-posedness is associated with inverse problems. That is, there are instabilities caused by inexact or noisy data and instabilities caused by numerical solutions. Since A. Tikhonov initiated the research on stable methods for the numerical solution of ill-posed problems, there is considerable work on the development, improvement and generalization of such methods.
This thesis is a contribution to the field of solution methods for ill-posed problems. Its intention is to present two-step regularization methods for linear ill-posed problems. The two steps are a data estimation step and a reconstruction step. The motivation for this approach is given in a natural way to model the noisy data. We assume that the noise is caused by a measurement process. We model the noisy data as application of the operator to a given function followed by a noise process. A two-step method aims at constructing a solution by tracing back the way the data were generated.
Besides the general idea of two-step methods, two realizations of this idea are studied in detail:

1. Classical methods like the ones proposed by Tikhonov and Landweber are embedded in the two-step context. This gives rise to a class of *reduced methods* based on a modification of the classical methods.

2. Wavelet shrinkage and classical regularization methods are combined.

The first realization aims at a deeper understanding of classical regulariza-

tion methods. The modification of classical methods is given by a parameter which controls, in addition to the regularizing parameter, the amount of damping.
The second realization combines in particular operator-adapted methods (based on the singular system) and data-oriented nonlinear methods (approximation spaces based on wavelets). With this we tie up to a recent development in the field of inverse problems.
Classical regularization methods like the ones proposed by Tikhonov and Landweber or the truncated singular value decomposition define *linear* estimation operators, i.e., the estimate for the exact solution is linear in the given data. Moreover these methods are defined by means of the operator or its singular system. Hence they are well adapted to properties of the operator. However, special properties of the data are neglected. This may not always be the best approach. Standard regularization methods concentrate on the operator-relevant singular functions. If such a method is applied to exact data with local singularities or discontinuities the regularized solution is a smoothed version of the exact solution and sharp features are lost.
The problems with the standard regularization methods described above are well-known and several approaches have been proposed to treat them. One way to overcome these difficulties is to choose a basis better adapted to data properties. Wavelets provide orthonormal bases of the space $L_2(\mathbb{R}^d)$ with localization in space and time. This makes them suitable for an efficient representation of functions that have space-varying smoothness properties. For a particular family of smoothness spaces, the Besov spaces, wavelet expansions provide an equivalent norm.
In this thesis we use wavelets to construct a non-linear data estimate from the noisy data, i.e., we denoise the data. Since the paper of Donoho and Johnstone [DJ94], the use of nonlinear thresholding techniques based on wavelet expansions of noisy signals in order to remove the noise has been discussed in the literature. We recall those articles which consider the use of wavelet methods for solving inverse problems.
Donoho introduces in [Don95] the so-called wavelet-vaguelette decomposition (WVD); the similar vaguelette-wavelet decomposition (VWD) is studied by Abramovich and Silverman [AS98]. The authors construct wavelet or wavelet-inspired bases that are in some sense adapted to the operator to be inverted. These bases are used instead of the singular function basis.
Other articles explore the application of Galerkin-type methods to inverse problems [DM96, LMR97, CHR04]. In [CHR04] Cohen et al. introduce the combination of wavelet-Galerkin projection and wavelet shrinkage. In the context of two-step methods as studied in this thesis both steps of [CHR04],

data estimation as well as reconstruction, are based on wavelets. As a variation of this we study the combination of wavelet shrinkage and classical regularization methods. On the data side we use a wavelet-based scheme which is adapted to the unknown smoothness properties of the exact data. For the reconstruction we stick to the operator-adapted methods based on the singular system of the operator.
Wavelet shrinkage is also used by Daubechies et al. [DDD04]. The authors study regularization methods for linear inverse problems with a sparsity constraint. To solve the equation $Kf = g$ for f the authors consider regularization methods as minimization of a functional consisting of the defect, $\|Kf - g\|_X$, and a penalty term $P(f)$. In classical regularization theory the penalty is quadratic, e.g., $P(f) = \|f\|^2_{\mathcal{H}}$ with $\mathcal{H}$ a normed space. The functional to be minimized is then given by

$$\|Kf - g\|^2_X + \alpha \|f\|^2_{\mathcal{H}}.$$

The parameter α weights the influence of both terms. In [DDD04] the authors study regularization where a non-quadratic, ℓ_p-penalty term is used. As a special case the ℓ_p-penalty term is a Besov norm given by wavelet norm equivalences. Wavelet shrinkage is then used to obtain a sparse representation. The corresponding regularized solution is computed by an iterative algorithm which amounts to a Landweber iteration with nonlinear shrinkage applied in each step. We emphasize the difference to the two-step approach as studied in this thesis. We aim at first estimating the data and then reconstructing a solution from the data estimate. The two steps are separated from each other and do not intertwine.

After this review of related works we specify the two-step methods studied in this thesis. For Hilbert spaces X and Y we aim at solving

$$\begin{aligned} K : X \to Y \\ Kf = g \end{aligned}$$

from given data g^δ close to g. We assume the equation to be ill-posed in the sense that the solution does not depend continuously on the data. Therefore small errors in the data can yield arbitrarily large errors in the reconstruction. We construct a solution by a two-step method consisting of a data smoothing step and a reconstruction step.
As has been said, this view reflects a commonly used model of the noisy data g^δ. The exact data g are generated by the operator K applied to some function f. Usually the exact data g are not available. Only a noisy version

g^δ is known which is modelled by adding noise to the exact data g. Hence we assume that

$$g^\delta = Kf + \text{noise}.$$

The index δ indicates the amount of noise and is specified later on together with the noise model. The construction of a solution for $Kf = g$ from noisy data g^δ by a two-step method is depicted in Figure 1.1, which demonstrates that a two-step method takes both component parts of the noisy data into account.

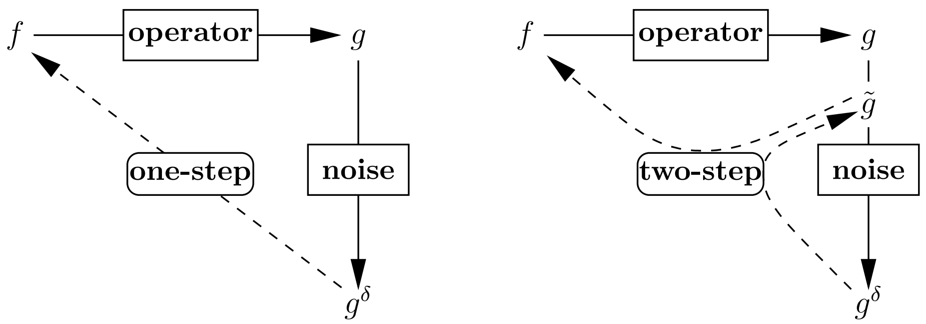
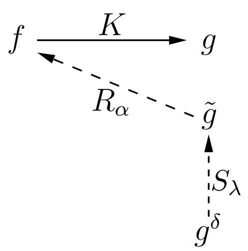


Figure 1.1: Construction of a solution by a one-step method (left) and by a two-step method (right); note the "stopover" at the data estimate $\tilde{g}$.

We introduce two-step methods as the composition of two operators,

$$T_{\alpha,\lambda} = R_\alpha S_\lambda.$$



Here S_λ is a smoothing operator defined on the data side,

$$S_\lambda : Y \to Y.$$

The data estimate is defined as $\tilde{g} := S_\lambda g^\delta$. The operator R_α is used to construct a solution f from the estimate $\tilde{g}$,

$$R_\alpha : Y \to X.$$

An approximate or regularized solution is then given by $f_{\text{reg}} = R_\alpha S_\lambda g^\delta$. The work of Jonas and Louis [JL01] fits into this two-step context. The authors interpret linear regularization methods as a combination of the pseudoinverse (or generalized inverse) and a smoothing operator in either order. In the terminology used in [JL01] the intended two-step method of first smoothing the data and then constructing a solution is a *prewhitening method* or a *range mollification*. The version where first the pseudoinverse is

applied and then the result is smoothed in suitable Sobolev norms is called *domain mollification*, see [HA98].
As a particular realization of two-step methods we study classical regularization methods like the ones introduced by Tikhonov and Landweber. For this we consider compact operators with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$. By classical regularization methods we have in mind operators R_α given as filtered versions of the generalized inverse,

$$R_\alpha g = \sum_{\sigma_n > 0} F_\alpha(\sigma_n)\sigma_n^{-1}\langle g, v_n\rangle u_n.$$

For the interpretation as a two-step method we split the filter function $F_\alpha(\sigma)$ according to

$$F_\alpha(\sigma) = F_\alpha(\sigma)^\gamma F_\alpha(\sigma)^{1-\gamma} \quad \text{with} \quad \gamma \in [0,1].$$

One part of the filter, for instance $F_\alpha(\sigma)^{1-\gamma}$, is assigned to the data side and results in the smoothing operator. The remaining part of the filter, i.e., $F_\alpha(\sigma)^\gamma$ defines the reconstruction operator. For the extreme cases where all of the filter is used on the data side (or for the reconstruction, which is the classical way) we get the results of [JL01]. One interesting result of filter splitting is an explanation of the well-known effect of oversmoothing when using the Tikhonov or the Landweber filter. It turns out that the filter functions $F_\alpha(\sigma)^\gamma$ induce order optimal regularization methods as long as the exponent γ varies within $(1/2, 1]$, see Propositions 4.2 and 4.4. We call these methods the *reduced Tikhonov* or the *reduced Landweber* method.
As another particular realization of a two-step method $T_{\alpha,\lambda} = R_\alpha S_\lambda$ we consider the combination of wavelet shrinkage S_λ and classical regularization methods R_α. This is related to the work of Cohen et al. [CHR04] as introduced above. For the nonlinear method of wavelet shrinkage and classical regularization we prove order optimality up to a logarithmic factor, see Theorem 4.12.
As a generalization we enlarge the class of filter induced reconstruction operators $R_\alpha : Y \to X$ which achieve, in combination with wavelet shrinkage, the optimal convergence rate. For this we consider again the reduced regularization methods where an exponent is applied to the filter function. We show that wavelet shrinkage allows to use an exponent that is equal or smaller than $1/2$ as long as the wavelet basis complies with a smoothness condition, see Corollary 4.14.
There are many other possibilities to realize the two steps given by the operators S_λ and R_α. Especially for the data estimation step S_λ every

method for data denoising is of interest. In this context we make a few remarks on the Sobolev embedding operator as studied by Ramlau and Teschke in [RT04a]. However, we do not give general conditions which would assure that the combined method $T_{\alpha,\lambda} = R_\alpha S_\lambda$ is working. We introduce the concept of regularization as a two-step method and study the details for the mentioned examples.
For instance, the adaptation of wavelet shrinkage to classical regularization methods requires a translation of the mutual additional information of the exact solution: For wavelet shrinkage, smoothness conditions in Sobolev or Besov spaces are used. For classical regularization methods, smoothness conditions are given as *source conditions* in spaces defined by the operator K. The operator-defined spaces as well as Sobolev spaces define a Hilbert scale. In this thesis Hilbert scales are used as common theoretical framework for studying ill-posed inverse problems.

To summarize, this thesis is on linear operators between Hilbert spaces with error in the data. Many related works exist and references are given throughout the thesis. Moreover numerous works treat non-linear operators [Neu92, Ram03, EN05], or a non-Hilbert space approach [Res05], or the case that not only the data but also the operator is not known exactly [HR04, CH05].

Organization of the thesis

The thesis is organized as follows. Chapter 1 is the introductory chapter – and it is almost over now.
In Chapter 2 we introduce basic notations and facts about ill-posed problems and regularization methods. Special emphasis is given to the *degree of ill-posedness* and optimality for deterministic as well as for stochastic noise.
In Chapter 3 we are concerned with data estimation. In the context of the intended two-step methods $T_{\alpha,\lambda} = R_\alpha S_\lambda$ we consider some possibilities for the data smoothing step S_λ. We present wavelet shrinkage including some results which are needed later. Furthermore we interpret classical regularization as smoothing operation with respect to a Hilbert scale generated by the operator K.
In Chapter 4 we present results on two-step methods. The chapter is divided into two sections: The first one is about optimality results for the reduced Tikhonov and the reduced Landweber method (Propositions 4.2 and 4.4) while the second one is about the order optimality of the combination of classical regularization methods and wavelet shrinkage (Theorem 4.12). At the end of the chapter we show that the use of wavelet shrinkage makes it

possible to enlarge the class of reconstruction operators R_α for which the optimal result is obtained (Corollary 4.14).
In Chapter 5 we apply the combination of Tikhonov regularization and wavelet shrinkage named TikShrink to a problem from medical imaging. We reconstruct activity functions from simulated SPECT (Single Photon Emission Computerized Tomography) data. The results illustrate the performance of our method and confirm the theoretical results.
In Chapter 6 we summarize the obtained results and discuss some starting points for further research.
In Appendix A we give the basic notations necessary to define stochastic processes which are used to model white noise.

Chapter 2

Basics on Inverse Problems and Regularization

In this chapter we present the necessary basic definitions, notations and results for ill-posed inverse problems and filter regularization methods used in this work. We rely on the textbooks [Lou89, EHN96, Rie03]. Those readers familiar with the classical theory of inverse problems might be interested in Section 2.2 and Section 2.4. In Section 2.2 we define Hilbert scales and introduce the theory of ill-posedness and optimality for inverse problems in Hilbert scales. In Section 2.4 we summarize and compare the optimal rates depending on the noise model, whether deterministic or stochastic. Section 2.3 presents the definition of regularization methods as filtered versions of the generalized inverse (or pseudoinverse) as well as optimality conditions. These conditions become significant for the *reduced filter* methods studied later on.

2.1 Ill-posed Inverse Problems

We consider normed spaces X, Y and operators

$$K : X \to Y.$$

The *inverse problem* consists in finding an $f \in X$ which for given $g \in Y$ fulfills

$$Kf = g. \tag{2.1}$$

Depending on the properties of the operator K this task involves some effort. If K is a bijection and K^{-1} is continuous with respect to properly chosen topologies in X and Y we have

i) the existence of a solution of $Kf = g$ for all $g \in Y$,

ii) the uniqueness of the solution and

iii) the continuous dependence of the solution on the data.

Properties i) - iii) are well-known as definition of a *well-posed* problem by Hadamard, see, e.g., [Lou89]. If one condition is violated the problem of solving $Kf = g$ is called *ill-posed* according to Hadamard.
If condition i) is violated we have the situation that not for all $g \in Y$ a solution exists. Let us consider $g = Kf$ as indirect observations of the quantity f searched for. In practice observations are made usually in the presence of noise. We model the noisy data g^δ as

$$g^\delta = Kf + \delta\xi. \tag{2.2}$$

Here δ is a small positive number, used for measuring the noise level and ξ denotes the noise. We assume that the noise ξ is *normalized*, that is, ξ is contained in some space $\tilde{Y}$ and it is $\|\xi\|_{\tilde{Y}} \leq 1$ or $E(\|\xi\|_{\tilde{Y}}) \leq 1$. So far we do not restrict the noise ξ to be of deterministic or stochastic nature. In Section 2.2 we introduce the theory of optimality for both types of noise and compare the results. Independent of the kind of noise, g^δ is usually not contained in the so-called *range* or *image* of the operator K

$$\mathrm{rg}\,(K) := \{g \in Y : \exists f \in X \text{ with } Kf = g\}.$$

The closure of the range is denoted by $\overline{\mathrm{rg}\,(K)}$. For elements $g \in Y \backslash \mathrm{rg}\,(K)$ we do not solve the equation

$$Kf = g$$

but we minimize the distance

$$\|Kf - g\|. \tag{2.3}$$

For X and Y Hilbert spaces and K a linear operator the minimization of the defect (2.3) is possible for $g \in \mathrm{rg}\,(K) \oplus \mathrm{rg}\,(K)^\perp$, see [Lou89]. If K is not injective, infinitely many minimizers exist. By choosing the minimizer with minimum norm we get a uniquely determined element $f^\dagger \in X$ which is called the *generalized solution* of $Kf = g$. The mapping

$$K^\dagger : \mathcal{D}(K^\dagger) = \mathrm{rg}\,(K) \oplus \mathrm{rg}\,(K)^\perp \subset Y \to X$$

with $K^\dagger g = f^\dagger$ is called *generalized inverse* or *pseudoinverse* or *Moore-Penrose inverse* of K. The generalized solution is also given as solution of the normal equation

$$K^*Kf = K^*g \tag{2.4}$$

in $\overline{\operatorname{rg}(K^*)}$, see [Lou89].
The concept of the generalized inverse can be considered as partial solution to the problem of existence and uniqueness, properties i) and ii) as given above. It remains to assure the continuous dependence of the solution on the data, property iii) as given above. This turns out to be the most challenging part. The generalized inverse is continuous only if the range of the operator is closed. This gives raise to the following definition of ill-posedness according to Nashed, see [Nas87].

Definition 2.1 (Nashed)**.** *For given normed spaces X and Y and an operator $K : X \to Y$ the problem of solving $Kf = g$ for given $g \in Y$ is called* ill-posed according to Nashed *if the image of K is not closed in Y. Otherwise the problem is called* well-posed according to Nashed.

In the following we consider ill-posed problems according to Definition 2.1. In Section 2.3 we see that *compact* operators can be considered as prototypes of ill-posed operators. In the next section we quantify ill-posedness by introducing the *degree of ill-posedness*. As an illustration of the difficulties described by properties i) – iii) we consider the *integration operator*. We return to this operator in Example 2.10 and in Section 5.1, where we compute some Sobolev smoothness properties of its singular system.

Example 2.2. *Given a function f we define the* integration operator K *as*

$$g(x) = Kf(x) := \int_0^x f(t)\mathrm{d}t.$$

If the right hand side g of $Kf = g$ is continuously differentiable with $g(0) = 0$ the (continuous) solution is given by $f = g'$. If we modify the data, e.g., $g(0) = \delta \neq 0$, the problem is no longer solvable. Another difficulty arises for approximate data. Let g^δ be an approximate or noisy version of g given as

$$g^\delta(x) = g(x) + \delta \sin nx, \quad \delta > 0,\ n \in \mathbb{N}_+.$$

The noisy function g^δ is continuously differentiable with $g^\delta(0) = 0$. The data error is bounded by

$$\|g - g^\delta\|_\infty \leq \delta.$$

The solution f^δ is given by

$$f^\delta(x) = f(x) + n\delta \cos nx.$$

The error of the solution depends on n according to

$$\|f - f^\delta\|_\infty = n\delta,$$

which can get arbitrarily large. But even this case is still an ideal one. Noisy data which are continuously differentiable with initial value zero are not very likely to occur as result of a measurement process. In order to take this into account we enlarge the space of possible results Y. In Example 2.10 we consider the integration operator defined between L_2-spaces.

We now give some general ideas on solving ill-posed problems. We consider different types of noise and present results on optimal rates.

2.2 Solving Ill-posed Inverse Problems

We assume a linear operator K acting between Hilbert spaces X and Y. We study the ill-posed problem of recovering some element $f \in X$ from noisy data g^δ near $g = Kf$. The noisy data is modeled as described in (2.2)

$$g^\delta = Kf + \delta\xi,$$

where ξ denotes the normalized noise (deterministic or stochastic) and δ is a small positive number used for measuring the noise level. We further assume that the unknown solution f belongs to $\mathcal{A} \subset X$. The subset $\mathcal{A} \subset X$ describes a-priori knowledge on the exact solution f, e.g., a smoothness condition like $f \in H^s(\mathbb{R}^d)$. We follow the lines of [MP01] and characterize the ill-posed problem by the triple $(K : X \to Y, \mathcal{A}, \delta)$. Here X indicates the spaces in which we measure the accuracy and Y indicates the space in which noisy measurements are given. Examples 2.2 and 2.10 of the integration operator show that the problem of solving an operator equation changes with the considered spaces X and Y.
Solving a given inverse problem $(K : X \to Y, \mathcal{A}, \delta)$ is done with some *method* m. As admissible method we consider at the moment any mapping from Y to X. We denote the class of methods by $\mathcal{M}$. For a method $m \in \mathcal{M}$ the approximate solution is given as $m(g^\delta)$. As introduced in Chapter 1 we study *two-step methods* $m = T_{\alpha,\lambda} = R_\alpha S_\lambda$ with

$$S_\lambda : Y \to Y \quad \text{and} \quad R_\alpha : Y \to X.$$

We repeat that in (2.2) the noise ξ is not restricted to be of deterministic or stochastic nature. There is a considerable amount of literature concerned with deterministic noise. We mention [TA77, EHN96, Lou89].

The study of inverse problems with stochastic noise, in which case we deal with a *statistical inverse problem*, was initiated in the 1960-ies [SH64, Bak69]. Several methods for the solution of inverse problems with random noise were proposed in recent years. We give a – by far not complete – list. Linear ill-posed problems with integral operators are studied in [Wah77, NC89, Luk98]. Nonlinear methods for linear problems using wavelet theory are studied in [Don95, AS98, CHR04]. In [MR96, NP99, MP01] approaches to inverse problems using Hilbert scales are studied. Other papers concerned with statistical inverse problems are [JS91] and [CGLT03]. The last one considers linear operator equations with special assumptions on the singular values. We return to this in Section 5.1 where we have a close look on the singular system of the integration operator.

2.2.1 Measuring Reconstruction Errors

The quality of a method is judged by its reconstruction error, i.e., the difference between the exact solution and the reconstructed solution. The necessary notations are given here. We develop the error terms for deterministic and stochastic noise similarily. In Section 2.3.2 we point out the correspondence to the error for regularization methods as defined in [Lou89].

We start with the case of deterministic noise. For any given method $m \in \mathcal{M}$, its error at the exact solution f is measured as

$$e^{\det}(K, m, f, \delta) := \sup\{\|f - m(Kf + \delta\xi)\|_X,\ \|\xi\|_Y \leq 1\}.$$

When using the short notation g^δ for the noisy data, the error reads as $e^{\det}(K, m, f, \delta) = \sup\{\|f - m(g^\delta)\|_X,\ \|\xi\|_Y \leq 1\}$.

For the stochastic case it is widely accepted to take white noise as model for the error [Don95, CIK99, CT02, CGLT03, CHR04]. In Appendix A the necessary notations and definitions from stochastics are given. The statistical model is written symbolically as given in (2.2),

$$g^\delta = Kf + \delta\xi.$$

Therein ξ denotes a stochastic process, see Definition A.2, such that for any test function $v \in Y$ observable quantities are given as the random variable

$$\langle g^\delta, v\rangle = \langle Kf, v\rangle + \delta\langle \xi, v\rangle.$$

In the case of ξ being white noise this means that the inner product $\langle \xi, v\rangle$ is a Gaussian random variable on a probability space $(\Omega, \mathcal{S}, P)$ with mean zero and variance $\|v\|^2$. That is, for white noise we obtain

$$\langle g^\delta - g, v\rangle = \delta \|v\| \mathcal{X}, \tag{2.5}$$

with $\mathcal{X}$ a normal random variable with mean zero and variance 1, i.e., $\mathcal{X} \sim \mathcal{N}(0,1)$. For the exact definition of a *white noise process* see Definition A.4. We refer to the white noise process as formal derivative of a standard Brownian motion, $\xi = \mathrm{d}W$, see Definition A.3 and Definition A.5. Hence we write our noisy data as

$$g^\delta = Kf + \delta \mathrm{d}W. \tag{2.6}$$

Within the framework of white noise the error of any method $m \in \mathcal{M}$ is measured by the risk

$$e^{\mathrm{ran}}(K, m, f, \delta) := (E\|f - m(Kf + \delta \mathrm{d}W)\|_X^2)^{1/2}$$

where E denotes the expectation operator, see (A.1). As short notation we use $e^{\mathrm{ran}}(K, m, f, \delta) = (E\|f - m(g^\delta)\|_X^2)^{1/2}$. For a-priori information $f \in \mathcal{A} \subset X$ the uniform error is defined as the supremum over pointwise errors with respect to $\mathcal{A}$. For stochastic noise we define $e^{\mathrm{ran}}(K, m, \mathcal{A}, \delta) := \sup_{f\in\mathcal{A}} e^{\mathrm{ran}}(K, m, f, \delta)$. For deterministic noise we obtain analogously for $f \in \mathcal{A}$

$$e^{\mathrm{det}}(K, m, \mathcal{A}, \delta) := \sup_{f\in\mathcal{A}} \{\|f - m(g^\delta)\|_X,\ \|\xi\|_Y \le 1\}. \tag{2.7}$$

The quantities $e^{\mathrm{det}}(K, m, \mathcal{A}, \delta)$ and $e^{\mathrm{ran}}(K, m, \mathcal{A}, \delta)$ measure the quality of some specific method m. Varying m within a class $\mathcal{M}$ of methods allows to consider the minimal error within this class, $e^{\mathrm{ran}}(K, \mathcal{M}, \mathcal{A}, \delta) := \inf_{m\in\mathcal{M}} e^{\mathrm{ran}}(K, m, \mathcal{A}, \delta)$, and the corresponding version $e^{\mathrm{det}}(K, \mathcal{M}, \mathcal{A}, \delta)$ for deterministic noise. Considering all admissible mappings between Y and X yields a lower bound for $e^{\mathrm{ran}}(K, \mathcal{M}, \mathcal{A}, \delta)$ as

$$\mathcal{E}^{\mathrm{ran}}(K, \mathcal{A}, \delta) := \inf_{m: Y\to X} \sup_{f\in\mathcal{A}} (E\|f - m(g^\delta)\|_X^2)^{1/2} \tag{2.8}$$

with the corresponding version for deterministic noise.

2.2.2 Degree of Ill-posedness and Optimality

Following [MP01] we introduce the *degree of ill-posedness* of an inverse problem $(K, \mathcal{A}, \delta)$ for deterministic and stochastic noise, the latter being restricted to white noise. For doing this the problem is formulated in Hilbert

scales. Assuming a-priori information on the exact solution we define *optimality* of a method.
Given an inverse problem $(K, \mathcal{A}, \delta)$ where δ describes the noise level of the data, we assume that we want to ensure any possible solution f^δ to vary in a ball with radius proportional to δ. This is the case if the norm of the inverse operator K^{-1} is bounded, $\|K^{-1} : Y \to X\| < \infty$. Otherwise the problem is ill-posed and we lose accuracy when recovering f from noisy data.
We now introduce Hilbert scales of spaces for X and Y and define the *degree of ill-posedness* of an inverse problem as largest scale parameter ν for which the norm of the inverse operator considered on the parameterized scale stays bounded. The degree of ill-posedness was first introduced by [Wah77]. For an introduction to Hilbert scales see [KP66].

Definition 2.3 (Hilbert scale)**.** *Let L be a densely defined selfadjoint strictly positive operator in a Hilbert space X which fulfills $\|Lx\| \geq \|x\|$ on its domain. Then a Hilbert space norm $\|\cdot\|_\nu$ is induced by the inner product*

$$\langle x, y\rangle_\nu := \langle L^\nu x, L^\nu y\rangle.$$

For $\nu \geq 0$ let X_ν be the completion of $\bigcap_{k=0}^{\infty} D(L^k)$ with respect to the $\|\cdot\|_\nu$-norm. For $\nu < 0$ let X_ν be the dual space of $X_{-\nu}$. Then the family of Hilbert spaces $\{X_\nu\}_{\nu\in\mathbb{R}}$ is called a Hilbert scale *(induced by the operator L).*

Remark 2.4. *If we speak of Hilbert scales we often think of "the scale of" Sobolev spaces. It has been shown in [KP66] that the Sobolev spaces $H^s(\mathbb{R}^d)$ build a Hilbert scale. In [Neu88b] it is shown that this is no longer true for $H^s(\Omega)$, if Ω is an open bounded subset of $\mathbb{R}^d$. It is also shown that Sobolev spaces with certain boundary conditions again build a Hilbert scale.*

Hilbert scales are invariant with respect to rescaling $\nu \to a\nu + b$, for $a > 0$, $b \in \mathbb{R}$. Often the Hilbert scale $\{X_\nu\}_{\nu\in\mathbb{R}}$ is a specific scale like the scale of Sobolev spaces. According to [MP01] the scaling of $\{X_\nu\}_{\nu\in\mathbb{R}}$ fits the usual Sobolev scale if the $n-th$ *approximational number* of the canonical embedding $J_\nu : X_\nu \to X,\ \nu > 0$ fulfills

$$a_n(J_\nu) \simeq n^{-\nu}, \tag{2.9}$$

where $\simeq$ means equivalent in order. The *n-th approximational number* of any linear operator T between Banach spaces X and Y is defined as, see [Pie80],

$$a_n(T) := \inf\{\, \|T - U\|_{X\to Y},\ U \in \mathcal{L}(X, Y),\ \operatorname{rank} U < n\}.$$

In [MP01] condition (2.9) is used to calculate the degree of ill-posedness for statistical inverse problems. We repeat their result in Proposition 2.7. In Section 3.1 we use equation (2.9) to connect Hilbert scales and some results of [RT04a] on the Sobolev embedding operator.
For the inverse problem $(K : X \to Y, \mathcal{A}, \delta)$ we assume that both the domain X and the target Y belong to appropriate Hilbert scales, i.e., $X \in \{X_\nu\}$ and $Y \in \{Y_\nu\}$. Often only a finite segment of parameters $\nu \in \mathbb{R}$ will be used. For example, in Section 3.3 we consider $\nu \in [0, 1]$. Further we assume that the scales are "linked" somehow, e.g., by $X = X_0 = Y_0$ and that the scaling for $\{X_\nu\}$ and $\{Y_\nu\}$ is the same. Norms in the spaces X_ν and Y_ν are only indicated with the parameter while the space is surpressed. We discuss the "link" of the Hilbert scales for an example operator defined on Sobolev spaces. Let $K : L_2 \to H^r$ be continuously invertible for $r = t$ but compact for $r < t$. In the notation $(K : X \to Y, \mathcal{A}, \delta)$ of an inverse problem, Y indicates the space in which noisy measurements are given. Hence, for L_2-errors $\|g - g^\delta\|_{L_2} \leq \delta$ the link $X = X_0 = Y_0$ implies that $Y_0 = Y$ and also $X_\nu = Y_\nu$. But for H^r-errors $\|g - g^\delta\|_{H^r} \leq \delta$ the link $X = X_0 = Y_0$ implies that $Y_0 = X_0 = L_2$. The data space $H^r = Y$ is of course contained in the Hilbert scale but not as Y_0. The link $X = X_0 = Y_0$ of the Hilbert scales $\{X_\nu\}$ and $\{Y_\nu\}$ therefore allows a more sophisticated view on properties of the error.
The basic assumption concerning the operator K is that the operator acts along the Hilbert scales with a step t as isomorphism between pairs $X_{\nu-t}$ and Y_ν. More precisely we assume that for some parameter $t > 0$ and for any $\nu \in \mathbb{R}$ the range of $K : X_{\nu-t} \to Y_\nu$ coincides with Y_ν and we have for all $x \in X_{\nu-t}$ the norm equivalence

$$\|x\|_{\nu-t} \simeq \|Kx\|_\nu. \tag{2.10}$$

For the case of Sobolev spaces this means that the operator K smoothes in Sobolev scales.
If an injective operator K does not meet condition (2.10) for some standard Hilbert scale, one can construct a scale adapted to the given operator K. For example in [NP99] this is done for an injective operator K with the generator $L = (K^*K)^{-1}$ which results in condition (2.10) with $t = \frac{1}{2}$. In Section 3.3 we discuss classical regularization methods as data smoothing operation. For this we introduce in (2.26) the Y_ν-spaces generated by the operator $L_2 = (KK^*)^{-1/2}$. The analogous X_ν-spaces are generated by $L_1 = (K^*K)^{-1/2}$ and meet the source condition used for classical regularization methods.

At this point we remind the reader of the topic of this thesis: two-step methods consisting of a data smoothing step and a reconstruction step. To combine methods like wavelet shrinkage and classical regularization we need a common theoretical framework. Hilbert scales are used in Lemma 4.6 to translate source condition given in ν-spaces into Sobolev smoothness and vice versa. In Section 2.4 we motivate this translation by comparing rates for problems defined in different frameworks.
Studying inverse problems with operators acting along Hilbert scales was initiated in [Nat84]. Since then inverse problems in Hilbert scales with deterministic noise have been studied intensively [Neu88a, Neu92, Mai94, Heg95, Tau96, DM96, Neu00, EN05]. The Hilbert scale approach is also used within the field of statistical inverse problems [MR96, MP01].

Degree of ill-posedness

In the following we define the *degree of ill-posedness* of an inverse problem. Later on we often speak of the degree of ill-posedness of an operator, which is to be understood as a pars pro totum. When computing the degree of ill-posedness it turns out that it depends on the noise model. White noise increases the degree of ill-posedness by $d/2$, where d denotes the dimension of the argument space.
We consider the inverse problem $(K : X \to Y, \mathcal{A}, \delta)$ where X and Y belong to Hilbert scales $\{X_\nu\}$ and $\{Y_\nu\}$. Accuracy is measured in $X = X_0 = Y_0$. We assume that there is a scale parameter $\tilde{\mu}$ such that $K : X \to Y_{\tilde{\mu}}$ has a bounded inverse. Then we can formally rewrite equation (2.2) and obtain

$$f = K^{-1} g^\delta - \delta K^{-1} \xi. \tag{2.11}$$

If the noise ξ is deterministic with $\|\xi\|_{\tilde{\mu}} \leq 1$ we can recover the unknown solution f up to order δ. If for $\mu < \tilde{\mu}$ this is not the case, i.e., if $\|K^{-1} : Y_\mu \to X\|$ is not bounded, the problem is ill-posed. We remark that there is a smallest parameter ϑ for which K has a bounded inverse as mapping from $X \to Y_\vartheta$. In order to tie up to the degree of ill-posedness as defined in [MP01] we set $\tilde{\vartheta} = -\vartheta$ and consider $K : X_{\tilde{\vartheta}} \to Y_0$ instead of $K : X \to Y_\vartheta$. This is possible since the scaling for $\{X_\nu\}$ and $\{Y_\nu\}$ is the same and they are linked by $X_0 = Y_0 = X$. Now we use the largest parameter $\tilde{\vartheta}$ for which the operator as mapping from $X_{\tilde{\vartheta}}$ to Y_0 has a bounded inverse to define the *degree of ill-posedness.*

Definition 2.5. *Let $(K : X \to Y, \mathcal{A}, \delta)$ be an inverse problem with deterministic noise ξ with $\|\xi\|_Y \leq 1$. Let X and Y belong to Hilbert scales $\{X_\mu\}$*

and $\{Y_\mu\}$ *with* $X_0 = Y_0 = X$*. Accuracy is measured in* X_0*. Let* $\tilde{\mu}$ *be a parameter such that* $K : X \to Y_{\tilde{\mu}}$ *has a bounded inverse. Then the quantity*

$$\vartheta := \sup\{\, \mu,\ \|K^{-1} : Y \to X_\mu\| < \infty\}$$

is less than or equal to 0*. The value* $(-\vartheta)_+$ *is called the* degree of ill-posedness *of the inverse problem.*

Remark. *In [MP01] the* degree of ill-posedness *is defined analogously for the case that accuracy is measured in* X_ν*. Then it is* $\vartheta \le \nu$ *and the* degree of ill-posedness *is defined as* $(\nu - \vartheta)_+$*.*

We consider the special case that K fulfills (2.10) for some t, i.e., $\|x\|_{\mu-t} \simeq \|Kx\|_\mu$. For $\mu = 0$ this means that $K : X_{-t} \to Y_0$ is continuously invertible. In this case the supremum in Definition 2.5 is attained and $(\nu-\vartheta)_+ = (\nu+t)$ is the degree of ill-posedness of the operator K. In this work we consider only the case that $\nu = 0$. For operators defined on Sobolev spaces this means that the reconstruction error is measured in L_2. For $\nu = 0$ the notion of the degree of ill-posedness coincides with the one introduced by [Wah77].
The situation is more complicated in the presence of stochastic noise. Even if we assume that the operator K has a bounded inverse and we arrive at the formally inverted equation (2.11) we cannot carry on in the same way as in the case of deterministic noise. For stochastic noise the expectation is not guaranteed to be bounded, i.e., $E\|K^{-1}\xi\|_\mu^2$ may be unbounded, and then accuracy is lost again. In analogy to the deterministic setting a statistical inverse problem is ill-posed if the above expectation is unbounded. We define the degree of ill-posedness of a statistical inverse problem according to [MP01].

Definition 2.6. *Let* $(K : X \to Y, \mathcal{A}, \delta)$ *be an inverse problem with stochastic noise* ξ *with* $E(\|\xi\|_Y) \le 1$*. Let* X *and* Y *belong to Hilbert scales* $\{X_\mu\}$ *and* $\{Y_\mu\}$ *with* $X_0 = Y_0 = X$*. Accuracy is measured in* X_ν*. Then the quantity* $(\nu - \vartheta)_+$ *is called the* degree of ill-posedness *of the operator* K *where* ϑ *is defined as*

$$\vartheta := \sup\{\, \mu,\ E\|K^{-1}\xi\|_\mu^2 < \infty\}.$$

For white noise the degree of ill-posedness is computed in [MP01] under the additional assumption that the operator K and the generator L of the Hilbert scale $\{X_\mu\}$ are properly related. This is to say, it is assumed that the eigenvectors of the operator L coincides with the eigenvectors of K^*K. As a consequence the operator L^{-1} as well as the operator K can be represented in the form

$$L^{-1}f = \sum_{k=1}^{\infty} \ell_k \langle f, u_n \rangle u_n, \qquad Kf = \sum_{k=1}^{\infty} \sigma_k \langle f, u_n \rangle v_n \tag{2.12}$$

where $\{u_n\}$, $\{v_n\}$ are orthonormal bases of X.

Proposition 2.7. *Let $(K, \mathcal{A}, \delta)$ be a statistical inverse problem with white noise. Let the operator K fulfill* (2.9)*,* (2.10) *and* (2.12)*. Then the operator has degree of ill-posedness $(\nu + t + d/2)_+$.*

Proof. The proof can be found in [MP01]. According to a remark of the authors the same result can be proved using only conditions (2.9) and (2.10). □

We summarize the results on the degree of ill-posedness of an inverse problem. For an operator which smoothes with stepsize $t > 0$ in a Hilbert scale the degree of ill-posedness is t for deterministic noise and $t + d/2$ for white noise. Thus the stochastic nature of white noise introduces an additional degree $d/2$ of ill-posedness. This influence of white noise on the degree of ill-posedness for the special case when the operator K denotes t-fold integration considered in Sobolev scales was observed by Nussbaum and explained in [NP99].

Optimality

We now present results on lower bounds for the error caused by solving the inverse problem $(K : X \to Y, \mathcal{A}, \delta)$. The best possible accuracies for $\mathcal{E}^{\text{det}}(K, \mathcal{A}, \delta)$ and $\mathcal{E}^{\text{ran}}(K, \mathcal{A}, \delta)$ are given. The set $\mathcal{A}$ denotes the additional information on the exact solution f of the inverse problem. We fix $\mathcal{A}$ to be a ball of radius $\rho > 0$ in some Hilbert space X_s of the Hilbert scale $\{X_\mu\}$, i.e., $\mathcal{A} := X_s(\rho) \subset X_s$ with $f \in X_s(\rho) \Leftrightarrow \|f\|_s \le \rho$.
From the discussion above we have well-posedness in the deterministic setting as long as we measure accuracy in the space X_ν with $\nu \le -t$. This yields $\mathcal{E}^{\text{det}}(K, \mathcal{A}, \delta) \simeq \delta$. Since we fixed ν to equal 0 and t is assumed to fulfill $t > 0$, we deal with an ill-posed problem. For this ill-posed problem the asymptotic is well-known [Nat84, DM96, EHN96],

$$\mathcal{E}^{\text{det}}(K, X_s(\rho), \delta) \simeq \delta^{\frac{s}{s+t}}. \tag{2.13}$$

The stochastic analog of estimate (2.13) is taken from [MP01].

Proposition 2.8. *Let $(K, X_s(\rho), \delta)$ be a statistical inverse problem with white noise. Additional information is given as $f \in X_s(\rho)$. Let K be an operator with smoothing property $t > 0$ in a Hilbert scale, see* (2.10)*. If the operator also fulfills* (2.9) *and* (2.12) *we have*

$$\mathcal{E}^{\mathrm{det}}(K, X_s(\rho), \delta) \simeq \delta^{\frac{s}{s+t+d/2}}. \tag{2.14}$$

Proof. We refer to [MP01]. □

In Section 2.4 we compare these rates with the one for classical regularization methods using ν-spaces, see Definition 2.17. The next section presents classical regularization methods induced by filter functions.

2.3 Classical Theory

We present here the very basic theory of classical regularization as given in [Lou89]. We introduce the Picard condition as well as the conditions for optimal filter functions. These conditions are used later in this thesis when we discuss the *reduced regularization methods*, see Section 3.3 and Section 4.1.

An operator K is called *compact* if the image (Kf_n) of a bounded sequence $(f_n) \subset X$ contains a convergent subsequence. Every compact operator is continuous. The generalized inverse of a *compact* operator is unbounded if the range of the operator is infinite-dimensional [Heu92]. We therefore consider operator equations with a compact operator with infinite-dimensional range as model for an ill-posed problem as introduced in Definition 2.1.

For symmetric compact operators on Hilbert spaces we have a decomposition in orthonormal eigenvectors, see standard books on functional analysis, e.g., [Heu92]. For a given operator K we denote by K^* its adjoint. If K is compact so is K^* and K^*K. By λ_n, u_n, $n \geq 0$ we denote the eigenvalues and the normalized eigenvectors of the symmetric operator K^*K. The eigenvalues are sorted such that $\lambda_n \geq \lambda_{n+1}$. We define $\sigma_n := +\sqrt{\lambda_n}$ and $v_n := \sigma_n^{-1} K u_n$. The σ_n are called *singular values*, the u_n, v_n are called *singular vectors* and the tripel $(\sigma_n, u_n, v_n)_{n \geq 0}$ is called the *singular system* of the operator K. We remark that the singular values fulfill $\sigma_n \to 0$ for $n \to \infty$. The action of the operator can be described by the *singular value decomposition*

$$Kf = \sum_{n \geq 0} \sigma_n \langle f, u_n \rangle v_n. \tag{2.15}$$

The range of a compact operator is characterized by the Picard condition.

Proposition 2.9 (Picard condition). *Let $K : X \to Y$ be a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$. For given $g \in \overline{\mathrm{rg}\,(K)}$ the equation $Kf = g$ has a solution iff the series*

$$\sum_{\sigma_n \geq 0} \sigma_n^{-2} |\langle g, v_n \rangle|^2 \tag{2.16}$$

converges.

Proof. The proof is by straightforward calculation. □

The generalized inverse of a compact operator can be decomposed in terms of its singular system, see [Rie03]. For a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$ and $g \in \mathcal{D}(K^\dagger)$ it is

$$K^\dagger g = \sum_{\sigma_n > 0} \sigma_n^{-1} \langle g, v_n \rangle u_n. \tag{2.17}$$

Equation (2.17) involves the reciprocals of the singular values. Since $\sigma_n \to 0$ for $n \to \infty$, errors in the data belonging to small singular values are magnified in the inversion process. Section 2.3.1 presents classical regularization theory which controls the influence of data errors by damping small singular values.
We now continue with Example 2.2 of the integration operator.

Example 2.10. *The integration operator*

$$K : L_2(0,1) \to L_2(0,1)$$

with

$$Kf(x) = \int_0^x f(t)\mathrm{d}t$$

is a compact operator with the singular system

$$\sigma_n = \frac{1}{(n+1/2)\pi},$$

$$u_n(t) = \sqrt{2}\cos(n+1/2)\pi t,$$

$$v_n(t) = \sqrt{2}\sin(n+1/2)\pi t,$$

see [Lou89].

We return to the singular system of the integration operator in Section 5.1 when we dicuss a special class of operators characterized by properties of their singular system.

2.3.1 Regularization Methods

We recollect that methods for solving the inverse problem $(K : X \to Y, \mathcal{A}, \delta)$ are defined as mappings from Y to X. In this section we present the special class of regularization methods induced by filter functions.
We concentrate on *non-degenerated* operators, i.e., operators with infinite-dimensional and non-closed range. Since the generalized inverse $K^\dagger$ of non-degenerated operators is not continuous, the generalized inverse is replaced by a *regularization*.
We consider noisy data of deterministic type with error level $\delta > 0$, i.e., we have g^δ with $\|g - g^\delta\| \le \delta$.

Definition 2.11. *A family of operators*

$$\{R_\alpha\}_{\alpha>0}, \ R_\alpha : Y \to X$$

is called a regularization *of* $K^\dagger$ *if the following holds. A real-valued mapping* $\alpha : \mathbb{R}_+ \times Y \to \mathbb{R}_+$ *exists such that* $g \in \mathcal{D}(K^\dagger), \ g^\delta \in Y$ *with* $\|g - g^\delta\| \le \delta$ *implies*

$$\lim_{\delta\to 0} R_{\alpha(\delta,g^\delta)} g^\delta = K^\dagger g.$$

If all R_α *are linear, the family* $\{R_\alpha\}$ *is called a* linear regularization*. The* regularization parameter α *is chosen to fulfill*

$$\lim_{\delta\to 0} \alpha(\delta, g^\delta) = 0.$$

If α *is independent of* g^δ *the parameter choice is called* a-priori*, otherwise it is called* a-posteriori*.*

For compact operators we get special regularization methods. We remember the spectral decomposition

$$Kf = \sum_n \sigma_n \langle f, u_n\rangle v_n$$

and the representation of the generalized inverse (2.17),

$$K^\dagger g = \sum_{\sigma_n>0} \sigma_n^{-1} \langle g, v_n\rangle u_n.$$

Let F_α be a real-valued function defined on the spectrum of the operator K. For $\alpha > 0$ we define the family of operators

$$R_\alpha g := \sum_{\sigma_n>0} F_\alpha(\sigma_n, g)\sigma_n^{-1} \langle g, v_n\rangle u_n. \tag{2.18}$$

The function F_α is called a *filter function*. For F_α independent of g the operator R_α is linear, otherwise R_α is nonlinear. If the filter function is chosen properly it damps the influence of error terms belonging to small singular values, see the following definition. The amount of damping is controlled by the regularization parameter α.

Definition 2.12. *Let K be a linear compact operator with singular value system $(\sigma_n, u_n, v_n)_{n\geq 0}$. Let the filter function F_α be independent of g. The filter is called* regularizing *(for the operator K) if the following holds*

$$\sup_n |F_\alpha(\sigma_n)\sigma_n^{-1}| = c(\alpha) < \infty, \tag{2.19a}$$

$$\lim_{\alpha\to 0} F_\alpha(\sigma_n) = 1 \quad \text{pointwise in } \sigma_n, \tag{2.19b}$$

$$|F_\alpha(\sigma_n)| \leq c \quad \forall \alpha, \sigma_n. \tag{2.19c}$$

We have the following proposition, see [Lou89].

Proposition 2.13. *The family of operators R_α induced by a regularizing filter is a regularization of $K^\dagger$ with $\|R_\alpha\| = c(\alpha)$.*

Well-known examples of filter-induced regularizations are the *truncated singular value decomposition*, the *Tikhonov regularization* and the *Landweber iteration*.

Example 2.14 (TSVD)**.** *The function F_α with*

$$F_\alpha(\sigma) := \begin{cases} 1, & \sigma \geq \alpha, \\ 0, & \sigma < \alpha. \end{cases} \tag{2.20}$$

is a regularizing filter. The induced regularization is called truncated singular value decomposition.

The effect of the filter function F_α given in (2.20) is as follows: All components belonging to singular values smaller than the parameter α are cut off, the others are kept unchanged.
Another regularizing filter is the following *Tikhonov* filter.

Example 2.15 (Tikhonov)**.** *The function F_α with*

$$F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha} \tag{2.21}$$

is a regularizing filter, see [Lou89]. The induced regularization is called Tikhonov(-Phillips) regularization.

The effect of the Tikhonov filter can be described as follows: For $\sigma_n^2 \gg \alpha$ we have $F_\alpha(\sigma_n) \approx 1$. The corresponding parts of the solution are modified only slightly. For σ_n small we have $F_\alpha(\sigma_n)$ small. The parts of the solution which amplify data error are damped but not cut off as in the TSVD case. The Tikhonov regularization can be formulated as special case of the minimization problem

$$\min_f J_\alpha(f) \quad \text{with} \quad J_\alpha(f) = \|Kf - g\|^2 + \alpha P(f).$$

The first term is the defect as given in (2.3). The second term is a so-called *penalty term* and represents desired properties of the solution. By chosing a smoothness norm for $P(f)$ one punishes a non-smooth solution. The parameter α weights the two terms: For α small, the defect $\|Kf - g\|^2$ has more influence on $J_\alpha(f)$ and vice versa. The choice of $P(f) = \|f\|^2$, as was done by Tikhonov, corresponds to the filter (2.21). The minimizing element f_α is then computed as solution of the *regularized normal equation*

$$(K^*K + \alpha I)f_\alpha = K^*g. \tag{2.22}$$

Here I denotes the identity operator. For $\alpha = 0$ the functional $J_\alpha(f)$ coincides with the defect and the regularized normal equation with the normal equation (2.4). An example for an *iterative* regularization method is given by the following filter.

Example 2.16 (Landweber)**.** *Let $m \in \mathbb{N}_+$. Let K be a compact operator and $0 < \beta < \frac{2}{\|K\|^2}$. Then the function F_m with*

$$F_m(\sigma) = 1 - (1 - \beta\sigma^2)^m \tag{2.23}$$

is a regularizing filter with parameter $\alpha = \frac{1}{m}$, see [Lou89]. The corresponding regularization is linear and is called Landweber regularization.

Some of our results in this work are about the Tikhonov and the Landweber filter F_α. In Section 3.3 and in Section 4.1 we consider the functions

$$F_\alpha^\gamma(\sigma) := (F_\alpha(\sigma))^\gamma$$

with $\gamma \in [0, 1]$. Applying the exponent γ to the filter function results in what we call *filter reduction*. In Section 4.1 we prove that as long as $\gamma > 1/2$ the reduced filter functions (Tikhonov and Landweber) are regularizing filters. We remark that filter reduction does not work with the TSVD filter. At the end of Section 4.1 we present a conjecture on the question of reducibility of filter functions.

2.3.2 Order Optimality for Regularization Methods

In this section we present the theory of order optimality for filter-induced regularization methods. For two reasons we follow again closely the basic lines of theory as given in [Lou89]. In the first place the filter reduction for the Tikhonov and the Landweber method as studied in Section 3.3 and Section 4.1 belongs to this setting, enlarged by the theory of Hilbert scales. And in the second place we connect the classical results to the theory of optimality as introduced in Section 2.2.2.

Convergence rates for classical regularization methods are based on a-priori assumptions on the exact data or equivalently, on the exact solution f of $Kf = g$. Usually it is assumed that the exact solution belongs to some subset of X. In [Lou89] for $\nu > 0$ the *ν-spaces* X_ν are defined as

$$X_\nu := \mathrm{rg}\,((K^*K)^{\nu/2}) = \mathcal{N}(K)^\perp \cap \mathcal{D}((K^*K)^{-\nu/2}), \tag{2.24}$$

where $\mathcal{N}(K)$ denotes the null space of the operator K. These sets are often called *source sets*, for $f \in X_\nu$ it is said that f fulfills a *source condition.* For compact operators K with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$ the spaces X_ν can be characterized via the singular values. It is $f \in \mathrm{rg}\,((K^*K)^{\nu/2})$ if and only if the series

$$\sum_{\sigma_n>0} \sigma_n^{-2\nu} |\langle f, u_n\rangle|^2$$

converges. If so, we define

$$\|f\|_\nu := \Big(\sum_{\sigma_n>0} \sigma_n^{-2\nu} |\langle f, u_n\rangle|^2\Big)^{1/2} \tag{2.25}$$

and call $\|f\|_\nu$ the *ν-norm* of f. We give another interpretation of (2.25): For $f \in X_\nu$ the Fourier coefficients of f with respect to u_n decay fast enough to compensate the growth of $\sigma_n^{-\nu}$.

Equations (2.24) and (2.25) can be seen as scale of conditions. Corresponding to this we redefine the scale of spaces X_ν as a Hilbert scale, see Definition 2.3 and the remarks following condition (2.10). We assume the operator K to be injective (if not we restrict it to the orthogonal complement of its null space). The Hilbert scale X_ν is generated by the operator $(K^*K)^{-1/2}$ and fulfills $X_\nu := \mathrm{rg}\,((K^*K)^{\nu/2})$ as in (2.24). We remark that also the definition of X_ν as $\mathrm{rg}\,((K^*K)^{\nu})$ is customary [EHN96]. This results in a different scaling of the spaces and in different exponents whenever source conditions are used. In terms of the generator of the Hilbert scale this means $L = (K^*K)^{-1}$ and was also mentioned following the condition (2.10).

We further introduce the scale of spaces Y_ν generated by $(KK^*)^{-1/2}$ as

$$Y_\nu := \operatorname{rg}((KK^*)^{\nu/2}). \tag{2.26}$$

The scales Y_ν are adapted to the image space of the operator K. With $X_0 = Y_0 = X$ in mind the smoothing property of the operator K in these scales of spaces can be expressed as

$$\begin{aligned} \|x\|_{-1} = \|Kx\|_0 \quad &\text{for} \quad X_\nu - \text{spaces}, \\ \|x\|_0 = \|Kx\|_1 \quad &\text{for} \quad Y_\nu - \text{spaces}. \end{aligned}$$

We remark that the scale of spaces Y_ν can be seen as condition on the decay rate of $\{\langle g, v_n\rangle\}$. Especially for $\nu = 1$ the condition $y \in Y_1$ is known as Picard condition, see (2.16), which characterizes the range of the operator.
At this point we make a short remark on the topic of this thesis: two-step methods $T_{\alpha,\lambda} = R_\alpha S_\lambda$ for the solution of ill-posed problems. The operator S_λ is defined on the data side and describes a smoothing step,

$$S_\lambda : Y \to Y.$$

In Section 3.3 we regard part of the filter function as data smoothing operation S_λ. We particularily show that smoothing in this way is done with respect to the scale Y_ν as introduced in (2.26),

$$S_\lambda : Y \to Y_\nu.$$

Results are given in Proposition 3.9 for the Tikhonov regularization and in Proposition 3.12 for the Landweber regularization.
After this short detour we turn to the question of optimality for classical regularization methods. In particular there are conditions on the filter function which guarantee the best possible rate of the induced regularization.
We consider noisy data $g^\delta = g + \delta\xi$ with deterministic normalized noise $\|\xi\|_Y \le 1$. We then have that $\|g - g^\delta\|_Y = \delta\|\xi\|_Y \le \delta$. Let the exact solution f be in $X_\nu(\rho)$, i.e., $f \in X_\nu$ with $\|f\|_\nu \le \rho$. Then for the inverse problem $(K : X \to Y, X_\nu(\rho), \delta)$ and an arbitrary mapping $R : Y \to X$ the error of R is defined as in (2.7)

$$e^{\det}(K, R, X_\nu(\rho), \delta) := \sup_{f \in X_\nu(\rho)} \{\|Rg^\delta - f^\dagger\|_X \ : \ \|g - g^\delta\|_Y \le \delta\}.$$

We rewrite $f \in X_\nu(\rho)$ as $\|f^\dagger\|_\nu \le \rho$ and arrive at

$$e^{\det}(K, R, X_\nu(\rho), \delta) = \sup\{\|Rg^\delta - f^\dagger\|_X \ : \ \|g - g^\delta\|_Y \le \delta, \ \|f^\dagger\|_\nu \le \rho\}.$$

In [Lou89] this error is denoted by $E_\nu(\delta, \rho, R)$. Since the following results are from the book [Lou89] we present them with the notation used therein. The classical theory of optimality reads as follows. Solving $Kf = g$ with noisy data yields the unavoidable error

$$E_\nu(\delta, \rho) := \inf_R E_\nu(\delta, \rho, R).$$

For a motivation of the following definition of optimality we refer the reader to [Lou89].

Definition 2.17. *Let R_α be a regularization method with parameter α. The method is called* optimal for $\nu > 0$ *if for all $\delta > 0$ and $\rho > 0$ a parameter $\alpha = \alpha(\delta, \rho)$ exists such that*

$$E_\nu(\delta, \rho, R_\alpha) \leq \delta^{\frac{\nu}{\nu+1}} \cdot \rho^{\frac{1}{\nu+1}}.$$

The method R_α is called order optimal for $\nu > 0$ *if a constant c exists such that for all $\delta > 0$ and $\rho > 0$ a parameter $\alpha = \alpha(\delta, \rho)$ exists with*

$$E_\nu(\delta, \rho, R_\alpha) \leq c\delta^{\frac{\nu}{\nu+1}} \cdot \rho^{\frac{1}{\nu+1}}.$$

In Section 2.4 we compare this rate with the ones given in (2.13) and (2.14). The following conditions on a regularizing filter F_α assure that the induced regularization is order optimal.

Proposition 2.18. *Let F_α be a regularizing filter, $\beta > 0$ and c, c_ν constants such that*

$$\sup_{0<\sigma\leq\sigma_1} |F_\alpha(\sigma)\sigma^{-1}| \leq c\alpha^{-\beta}, \tag{2.27a}$$

$$\sup_{0<\sigma\leq\sigma_1} |(1 - F_\alpha(\sigma))\sigma^{\nu^*}| \leq c_{\nu^*}\alpha^{\beta\nu^*}. \tag{2.27b}$$

With the parameter choice

$$\alpha = \kappa \left(\frac{\delta}{\rho}\right)^{1/\beta(\nu+1)}, \ 0 < \kappa$$

the induced regularization R_α is order optimal for all ν with $0 \leq \nu \leq \nu^*$. The error bound gets minimal for

$$\alpha = \left(\frac{c}{\nu c_\nu} \frac{\delta}{\rho}\right)^{1/\beta(\nu+1)}.$$

The minimal error bound is

$$\|R_\alpha g^\delta - K^\dagger g\| \leq (c\delta)^{\nu/(\nu+1)} (c_\nu \rho)^{1/(\nu+1)} (\nu + 1)\nu^{-\nu/(\nu+1)}.$$

Proof. The proof can be found in [Lou89]. □

The examples of regularization methods given in Section 2.3.1 are order optimal. Proofs of the following propositions can be found in [Lou89].

Proposition 2.19. *The truncated singular value decomposition with the parameter choice*

$$\alpha = \kappa \left(\frac{\delta}{\rho}\right)^{1/(\nu+1)}, \kappa \in \mathbb{R}_+,$$

is order optimal for all $\nu > 0$.

Proposition 2.20. *The Tikhonov regularization with the parameter choice*

$$\alpha = \kappa \left(\frac{\delta}{\rho}\right)^{1/2(\nu+1)}, \kappa \in \mathbb{R}_+,$$

is order optimal for all $0 < \nu < \nu^* = 2$.

Proposition 2.21. *The Landweber regularization is order optimal for all* $\nu > 0$ *if the iteration is stopped for*

$$m = \left\lfloor \left(\frac{\nu^2}{\beta}\right)^{(\nu+1)} \left(2\frac{\beta}{\nu}e\right)^{-\nu/(\nu+1)} \left(\frac{\rho}{\delta}\right)^{2/(\nu+1)} \right\rfloor,$$

where $\lfloor x \rfloor$ *denotes the biggest number smaller or equal to* x.

The choice of the regularizing parameter is crucial for the error bound. The last propositions on TSVD, Tikhonov and Landweber regularization show that the regularization parameter depend on source conditions and norm estimates of the exact solution which are usually not known in advance. A possibility to get optimal rates without such a-priori information is to determine the regularization parameter by *Morozov's discrepancy principle*. The basic idea for this is to choose the parameter in such a way that the defect of the regularized solution is approximately equal to the error level of the data, $\|Kf_\alpha^\delta - g\| \simeq \delta$.

So far three different optimal rates are given, see equation (2.13), equation (2.14) and Definition 2.17. The next section discusses how these rates are connected.

2.4 Rates for ν-spaces and Sobolev Spaces

This section summarizes and discusses the so far presented rates for ill-posed problems, (2.13), (2.14) and Definition 2.17. All rates depend on a-priori information on the exact solution f given in specific theoretical frameworks. We use Hilbert scales as a common framework for ν-spaces and Sobolev spaces and compare the corresponding rates.
In classical regularization theory it is usually assumed that for some $\nu > 0$ the exact solution f of $Kf = g$ fulfills a source condition. That is, $f \in \operatorname{rg}((K^*K)^{\nu/2})$, or there is an $\omega \in X$ such that

$$f = (K^*K)^{\nu/2} w \quad \text{with} \quad \|w\|_X \leq \rho.$$

We assume noisy data g^δ with error level δ and deterministic noise, hence

$$\|g - g^\delta\| \leq \delta.$$

Then, according to Definition 2.17, for an optimal regularization method the following convergence result holds,

$$\|f - f_{\text{reg}}\|_X = \mathcal{O}(\delta^{\frac{\nu}{\nu+1}}). \tag{2.28}$$

In many cases the spaces X and Y are Sobolev spaces. We assume that for some $t > 0$ the operator $K : L_2 \to H^t$ is continuously invertible but compact as $K : L_2 \to L_2$. The exact solution f fulfills $f \in H^s,\ s > 0$ and the noise is deterministic with $\|g - g^\delta\|_{L_2} \leq \delta$. In this context an optimal regularization method achieves according to equation (2.13) the following rate,

$$\|f - f_{\text{reg}}\|_{L_2} = \mathcal{O}(\delta^{\frac{s}{s+t}}). \tag{2.29}$$

This result corresponds to (2.28) if we translate range conditions into Sobolev smoothness according to

$$f \in H^s \Leftrightarrow f \in \operatorname{rg}((K^*K)^{\nu/2}) \quad \text{with} \qquad \nu = \frac{s}{t}. \tag{2.30}$$

For operators with smoothing property t in Sobolev scales we prove this translation in Lemma 4.6. In order to incorporate other than L_2-data errors we consider K as the mapping $K : L_2 \to H^r$ with K compact for $r < t$ and $\|g - g^\delta\|_{H^r} \leq \delta$. In this case an optimal method achieves the following rate

$$\|f - f_{\text{reg}}\|_{L_2} = \mathcal{O}(\delta^{\frac{s}{s+t-r}}), \tag{2.31}$$

see [DM96]. Considering data in H^r with $r < 0$ is interpreted by Mathé and Pereverzev in [MP01] as additional degree of ill-posedness of the operator. A translation of the source conditions equivalent to (2.30) is given by

$$f \in H^s \Leftrightarrow f \in \mathrm{rg}((K^*K)^{\nu/2}) \quad \text{with} \quad \nu = \frac{s}{t-r}.$$

Ramlau and Teschke present in [RT04a] a similar result on the translation of range condition and Sobolev smoothness for the Sobolev embedding operator $i : H^s(\mathbb{R}) \to L_2(\mathbb{R})$.
We now assume that the noise is stochastic. That is, we assume that the noisy data are generated by adding a white noise process of variance δ^2 to the exact data. We model white noise as the derivative of a standard Brownian motion or standard Wiener process W and refer to it by $\mathrm{d}W$. Hence the noisy data g^δ are given by, see (2.6),

$$g^\delta = Kf + \delta \mathrm{d}W.$$

In this context an optimal regularization method achieves the following convergence result (equation (2.14))

$$E(\|f - f_{\mathrm{reg}}\|_{L_2}) = \mathcal{O}(\delta^{\frac{s}{s+t+d/2}}). \tag{2.32}$$

The optimal rates therefore depend on the noise model. For L_2-errors, which means $r = 0$, (2.29) follows from (2.31). One can also say that (2.31) follows from (2.29) by substracting the smoothness of the noise in the denominator of the exponent. In Proposition 3.6 we compute that the expectation of white noise can be measured in the space H^r with $r < -d/2$. Inserting $r = -d/2$ in (2.31) yields the rate (2.32).

We now turn to the first step of the proposed two-step methods. Section 3 considers three different possibilities for the data smoothing step: regularization of the Sobolev embedding operator, wavelet shrinkage and classical regularization methods.

Chapter 3

Data Estimation

This chapter is concerned with data estimation as the first step of two-step methods $T_{\alpha,\lambda} = R_\alpha S_\lambda$. In the following we present some realizations of the smoothing operator

$$S_\lambda : Y \to Y.$$

Hence we concentrate on the right hand side, or data side, g of the equation $Kf = g$. We assume that we only have a noisy version g^δ of g. Noise can be due to a measurement process. We further assume that we deal with an operator which smoothes with a certain step size in some Hilbert scale, see (2.10). Later on we particularly focus on operators smoothing in Sobolev and Besov scales. In this case the right hand side g gains smoothness by the application of the operator to a function f. The smoothness of the exact data is thoroughly altered in the presence of noise. In Section 3.2.1 we show that for white noise the smoothness of the noisy data is reflected by the Sobolev space $H^{-d/2}$ with d being the dimension of the problem.
Data estimation aims at partially recovering the smoothness properties of the exact data. We present three methods for data estimation. Section 3.1 deals with Sobolev embedding operators. Section 3.2 introduces the necessary theoretical background and results on wavelet shrinkage. In Section 3.3 we reinterpret classical regularization of ill-posed problems as combination of a data smoothing and a reconstruction step.

3.1 Regularization of the Sobolev Embedding Operator

In this section we describe a smoothing technique for noisy data based on the inversion of the Sobolev embedding operator as studied in [RT04a]. We consider the special case that the operator

$$K : X \to Y$$

is defined for Sobolev spaces X and Y. We use the definition of Sobolev spaces based on the decay properties of the Fourier transform.

Definition 3.1. *Let $s \geq 0$. The Hilbert spaces*

$$H^s(\mathbb{R}^d) = \{f \in L_2(\mathbb{R}^d) : \|f\|_{H^s(\mathbb{R}^d)} < \infty\}$$

with

$$\|f\|_{H^s(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} (1 + |\omega|^2)^s |\hat{f}(\omega)|^2 \mathrm{d}\omega \right)^{1/2} \tag{3.1}$$

are called Sobolev spaces*.*

We remark that Sobolev spaces can also be defined for a negative smoothness index. This is done by means of duality, see [Ada75]. We assume as in (2.10) that the operator has a smoothing property of order $t > 0$ in the scale of Sobolev spaces. At the moment we assume that for all $s \geq 0$

$$K : H^s \to H^{s+t}.$$

The exact data $g = Kf$ therefore satisfies a smoothness condition, we assume that $g \in H^\theta$ with $\theta > 0$. The noisy data g^δ is modeled as $g^\delta = Kf + \delta\xi$, see (2.2). Usually the noise belongs to H^r with $r < \theta$. We assume that ξ is in H^r and is normalized according to $\|\xi\|_{H^r} \leq 1$.
Let i denote the *Sobolev embedding operator*

$$i : H^\theta \to H^r$$
$$ix = x.$$

A typical measurement process can be described as follows

$$H^\theta \ni x \longmapsto ix + \delta\xi = x^\delta \in H^r.$$

Hence one possibility to construct an approximation $\tilde{x}$ to x in H^θ is to solve the equation $ix = x$ from noisy data x^δ. Regularization methods for the ill-posed embedding operator are studied extensively in [RT04a]. The authors use the Tikhonov method to compute the approximation $\tilde{x}$. For solving the related minimization problem, wavelet and Fourier based techniques are used. We note the results of [RT04a] on

1. the translation of source condition and Sobolev smoothness.
2. the smoothing property of the adjoint embedding operator.
3. the singular system of the embedding operator.

Properties of the Sobolev embedding operator

In order to get convergence results for the Tikhonov method the authors of [RT04a] require a *source* or *smoothness condition* $x = (i^*i)^{\nu/2}z$. They prove that for the embedding operator

$$i : H^s(\mathbb{R}) \to L_2(\mathbb{R}),\ \ s > 0 \tag{3.2}$$

a smoothness condition $x = (i^*i)^{\nu/2}\omega$ holds if and only if $x \in H^{s+s\nu}$. The next lemma is taken from [RT04a] and gives a representation of the adjoint operator

$$i^* : L_2(\mathbb{R}) \to H^s(\mathbb{R}).$$

We use this representation to show that the adjoint operator i^* actually smoothes with stepsize $2s$.

Lemma 3.2. *For i defined as in* (3.2) *the adjoint operator*

$$i^* : L_2(\mathbb{R}) \to H^s(\mathbb{R})$$

is given in the Fourier space by

$$\widehat{(i^*z)}(\omega) = \frac{\hat{z}(\omega)}{(1+|\omega|^2)^s}. \tag{3.3}$$

Proof. The proof can be found in [RT04a]. □

Corollary 3.3. *Let i be the embedding operator defined as in* (3.2)*. Then the adjoint embedding operator smoothes with stepsize* $2s$*, i.e.,*

$$i^* : L_2(\mathbb{R}) \to H^{2s}(\mathbb{R}).$$

Proof. The statement follows by computing the Sobolev norm of i^*z for $z \in L_2(\mathbb{R})$. For $\kappa \geq 0$ it is

$$\|i^*z\|^2_{H^\kappa} = \int_{\mathbb{R}} |\widehat{(i^*z)}(\omega)|^2(1+|\omega|^2)^\kappa \mathrm{d}\omega.$$

Inserting (3.3) yields

$$\|i^*z\|^2_{H^\kappa} = \int_{\mathbb{R}} |\hat{z}(\omega)|^2(1+|\omega|^2)^{\kappa-2s} \mathrm{d}\omega.$$

Hence the Sobolev norm stays bounded for $\kappa \leq 2s$. For $\kappa = 2s$ we have $\|i^*z\|^2_{H^\kappa} \simeq \|z\|^2_{L_2}$. □

We use this result on the smoothing property of the adjoint embedding operator at the end of Section 4.1. In Section 4.1 we show that filter reduction is possible for the Tikhonov and the Landweber regularization method. Our general conjecture is that filter reduction is possible whenever the regularization method uses the adjoint operator. This conjecture is motivated by the two examples considered in Section 4.1 and the result on the smoothing property of the adjoint embedding operator.
At this point we remind the reader of Hilbert scales as introduced in Definition 2.3. We connect a result from [RT04a] on the singular system of the embedding operator to the theory of Hilbert scales. For this we use equation (2.9) which involves the n-th approximational number. For compact operators on a Hilbert space the approximational numbers equals the singular values, i.e., for K compact we have

$$a_n(K) = \sigma_n(K).$$

In [RT04a] the singular value system of the Sobolev embedding operator considered for periodic functions over the intervall $\Omega = [0, 2\pi)$ is computed. In this case the embedding operator

$$i : H^s(\Omega) \to L_2(\Omega)$$

is compact with singular values

$$\sigma_n = (1 + n^2)^{-s/2}.$$

Thus the canonical embeddings

$$i_s : H^s(\Omega) \to L_2(\Omega)$$

of the Sobolev spaces $H^s((0, 2\pi])$, considered as Hilbert scale with the smoothness index as scale parameter, fulfill $\sigma_n(i_s) \simeq n^{-s}$ and hence (2.9).

3.2 Data Estimation by Wavelet Shrinkage

Another method for data estimation is *wavelet shrinkage*. An approximation $\tilde{g}$ of g from noisy data g^δ is constructed in three steps.

1. The noisy signal g^δ is transformed into the wavelet domain, i.e., g^δ is expanded in a series with respect to a wavelet basis.

2. The coefficients computed by this means are filtered.

3. The estimator $\tilde{g}$ for the exact data g is obtained by transforming the filtered series back.

The filtering can be done in different ways: cutting off the wavelet expansion at a certain level (linear filtering) or keeping only those coefficients which are greater than a certain value, the so called *threshold*, (nonlinear filtering). In this thesis wavelet shrinkage with hard thresholding is used. The next section presents the necessary theoretical background on wavelet analysis. Lemma 3.8 provides the reader with a result on wavelet shrinkage fitted to the situation considered later in Section 4.2. For a comprehensive treatment of wavelet analysis we refer the reader to [Dau92, SN96, LMR97, Mal98]. For a short introduction to wavelet shrinkage see [Tas00], for details see [DJ94, DJKP95, DJ98] and the references given later on.

3.2.1 Wavelet Decomposition

We consider wavelet bases of $L_2(\Omega)$ which are defined by two types of functions: a scaling function φ and a wavelet function ψ. The basis functions are scaled and shifted versions $\varphi_{j,k}$ and $\psi_{j,k}$ of φ and ψ. For standard wavelet bases on $\mathbb{R}$ the functions $\psi_{j,k}$ and $\varphi_{j,k}$ are defined as $\psi_{j,k} = 2^{j/2}\psi(2^j \cdot -k)$ and $\varphi_{j,k} = 2^{j/2}\varphi(2^j \cdot -k)$. The index j denotes the scale and the index k denotes the space parameter. With little effort wavelet bases can be adapted to fairly general domains $\Omega \subset \mathbb{R}^d$, see [Coh03] for a survey of these adaptations. The functions φ and ψ may then change form near the boundary of the domain and the indices j and k need a more sophisticated view. For simplicity we neglect these effects in our short overview.
At level j, which in the standard case corresponds to resolution 2^{-j}, the scaling functions $(\varphi_{j,k})_{k\in\mathbb{Z}}$ span a space V_j within a hierarchy $V_0 \subset V_1 \subset \cdots \subset L_2(\Omega)$ of nested approximation spaces. The wavelets $(\psi_{j,k})_{k\in\mathbb{Z}}$ span a complement W_j of V_j in V_{j+1}. The wavelet decomposition of a function f is given by

$$f = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\varphi_{j_0,k} + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} f_{j,k}\psi_{j,k}.$$

We suppose that the functions $(\varphi_{j_0,k}, \psi_{j,k})_{j\geq j_0, k\in\mathbb{Z}}$ are mutually orthogonal and normed. This is the case for the class of Daubechies wavelets, see [Dau92]. For orthonormal functions the approximation and the detail coefficients $\alpha_{j_0,k}$ and $f_{j,k}$ of f are given by

$$\alpha_{j_0,k} = \langle f, \varphi_{j_0,k}\rangle \quad \text{and} \quad f_{j,k} = \langle f, \psi_{j,k}\rangle.$$

To simplify the notation we will always take $j_0 = 0$ and incorporate the first layer of scaling functions $(\varphi_{0,k})_{k\in\mathbb{Z}}$ into the wavelet layer $(\psi_{-1,k})_{k\in\mathbb{Z}}$. The wavelet decomposition of a function f is then given by

$$f = \sum_{j\geq -1}\sum_{k\in\mathbb{Z}} f_{j,k}\psi_{j,k}. \tag{3.4}$$

The wavelet functions $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ as well as the combination of the basic level scaling function with all the finer level wavelet functions $(\varphi_{0,k}, \psi_{j,k})_{j\geq 0,k\in\mathbb{Z}}$ build a basis of $L_2(\mathbb{R})$. For the class of Daubechies wavelets this is an orthonormal basis.

If we truncate the wavelet expansion (3.4) at level j we obtain the projector onto the space V_j,

$$P_j f = \sum_{-1\leq\tilde{j}\leq j}\sum_{k\in\mathbb{Z}} f_{\tilde{j},k}\psi_{\tilde{j},k}. \tag{3.5}$$

To estimate the approximation power of V_j we introduce the *degree of polynomial reproduction* or *order of polynomial exactness*. It is said that the space V_0 spanned by the functions $\varphi(\cdot - k)$, $k \in \mathbb{Z}$, has *degree L of polynomial reproduction* if all polynomials of degree less than or equal to L are "contained" in the space V_0. From the invariance of Π_L to the change of scale, these polynomials are also contained in all spaces V_j, $j \in \mathbb{Z}$. For the exact definition and further results, like the connection to the so-called *Strang-Fix conditions*, we refer to [Coh03]. The approximation power of V_j is expressed by the following *direct estimate*.

Lemma 3.4. *Let $0 \leq \eta < \theta$, $f \in H^\theta$ and $\varphi \in H^\eta$. For $\theta < m \in \mathbb{N}$ where $m-1$ is the degree of polynomial reproduction in V_j one has the* direct estimate

$$\|f - P_j f\|_{H^\eta} \lesssim 2^{-j(\theta-\eta)}\|f\|_{H^\theta}. \tag{3.6}$$

Proof. The proof can be found in [Coh03] (separately for $\eta > 0$ and $\eta = 0$). □

Methods for estimating the Sobolev smoothness of wavelets can be found in [Vil93] or [Thi04].

Smoothness of white noise

As an application of wavelet analysis we compute the smoothness of white noise. For this we use wavelet norm equivalences for Sobolev and Besov norms. We show that white noise is reflected by the Sobolev space $H^{-d/2}$ or the Besov spaces $B^{-d/2}_{p,p}$ of negative order, where d is the dimension of the problem. With dW we denote a white noise process and the noisy data are given as in (2.6),

$$g^\delta = g + \delta \mathrm{d}W.$$

According to (2.5), for $v \in L_2$ we have

$$\langle g^\delta - g, v \rangle = \delta \|v\|_{L_2} \mathcal{X}$$

where $\mathcal{X}$ is a random variable with a standard normal or Gaussian distribution, $\mathcal{X} \sim \mathcal{N}(0,1)$. The next lemma shows that white noise cannot be considered as an element of the space L_2.

Lemma 3.5. *Let* dW *be a white noise process,* $g \in L_2$*,* $\delta > 0$ *and* $g^\delta = g + \delta \mathrm{d}W$*. Then*

$$E(\|g^\delta - g\|_{L_2}) = \infty.$$

Proof. Let $\{v_n\}$ be an orthonormal basis of L_2. For $g \in L_2$ we have

$$\|g\|^2_{L_2} = \sum_n |\langle g, v_n \rangle|^2.$$

For $\delta \mathrm{d}W = g^\delta - g$ we compute analogously

$$\begin{aligned} E(\|g^\delta - g\|^2_{L_2}) &= E\big(\sum_n |\langle g^\delta - g, v_n \rangle|^2\big) \\ &= \delta^2 \|v\|^2_{L_2} E\big(\sum_n |\mathcal{X}|^2\big) \\ &= \delta^2 \sum_n E(|\mathcal{X}|^2). \end{aligned}$$

Since $E(|\mathcal{X}|^2)$ equals some constant c the assertion is proved. □

Motivated by the proof of Lemma 3.5 we look for a weighted norm $\|\cdot\|_W$ characterized by a sequence of weights $\{w_n\}_n$, i.e.,

$$\|\cdot\|^2_W = \sum w_n^2 |\langle \cdot, v_n \rangle|^2.$$

We ask whether the weights $\{w_n\}_n$ can be chosen such that

$$E(\|g^\delta - g\|^2_W) = \delta^2 E(|\mathcal{X}|^2) \sum_n w_n^2$$

converges. Especially wavelet bases $(\psi_{j,k})_{j,k}$ offer representations of Sobolev and Besov norms as weighted series of coefficients. We define Besov spaces

through finite differences according to [Coh03]. For equivalent definitions and more details on Besov spaces and their connection to wavelet bases the reader is referred to [Ada75, Tri83, Coh03] and references given therein.
The Besov space $B^{\kappa}_{p,q}(\mathbb{R}^d)$ is essentially a space of functions on $\mathbb{R}^d$ which have "κ derivatives in $L_p(\mathbb{R}^d)$"; the index q provides some fine-tuning. The precise definition involves the moduli of smoothness of the function, defined by finite differencing instead of derivatives. Moreover the behaviour of these moduli is combined at different scales. The result is that functions that are mostly smooth but have a few local irregularities can still belong to a Besov space of high smoothness index.
We denote the *n-th order finite difference operator* by $\triangle^n_h$. It is defined recursively as $\triangle^1_h g(x) = g(x+h) - g(x)$ and $\triangle^n_h g(x) = \triangle^1_h(\triangle^{n-1}_h)g(x)$. For example, the second order difference of a function g is $\triangle^2_h g(x) = g(x+2h) - 2g(x+h) + g(x)$. For $\Omega \subset \mathbb{R}^d$ and h a vector of the Euclidean space $\mathbb{R}^d$ with norm less than some bound τ we consider the set

$$\Omega_{h,n} := \{x \in \Omega : x + kh \in \Omega, k = 0, \ldots, n\}.$$

The *n-the order L_p-modulus of smoothness* of a function g is defined by

$$\omega_n(g, \tau, \Omega)_p = \sup_{|h| \leq \tau} \|\triangle^n_h g\|_{L_p(\Omega_{h,n})}.$$

For $p, q \geq 1$ and $0 < \kappa < n \in \mathbb{N}$ the Besov space $B^{\kappa}_{p,q}(\Omega)$ consists of those functions $g \in L_p(\Omega)$ for which

$$(2^{\kappa j}\omega_n(g, 2^{-j}, \Omega)_p)_{j\geq 0} \in \ell_q.$$

A natural norm for this space is

$$\begin{aligned} \|g\|_{B^{\kappa}_{p,q}} &:= \|g\|_{L_p} + |g|_{B^{\kappa}_{p,q}} \quad \text{with} \\ |g|_{B^{\kappa}_{p,q}} &:= \|(2^{\kappa j}\omega_n(g, 2^{-j}, \Omega)_p)_{j\geq 0}\|_{\ell_q}. \end{aligned}$$

The space $B^{\kappa}_{p,q}$ represents "κ order of smoothness measured in L_p". The parameter q allows a finer tuning on the degree of smoothness: it is

$$B^{\kappa}_{p,q_1} \subset B^{\kappa}_{p,q_2} \ \text{ if } \ q_1 \leq q_2.$$

However, the parameter q plays a minor role in comparison with κ. Regardless of the values of q_1 and q_2 it is

$$B^{\kappa_1}_{p,q_1} \subset B^{\kappa_2}_{p,q_2} \ \text{ if } \ \kappa_1 \geq \kappa_2.$$

For $p = q = 2$ the Besov spaces coincides with the Sobolev spaces H^κ. The relations between Besov, Sobolev and numerous other smoothness spaces as well as their relation to approximation spaces are a topic of extensive studies, e.g., [Lin05]. Besov spaces can also be defined for p and q less than 1, see [Coh03]. For $\kappa > 0$ Besov spaces with negative smoothness are identified by means of duality as

$$B^{-\kappa}_{p',q'} = (B^\kappa_{p,q})'$$

with $1 \leq p,\ q \leq \infty$ such that $\frac{1}{p} + \frac{1}{p'} = 1 = \frac{1}{q} + \frac{1}{q'}$.
The norm equivalence used in the following is from [Coh03, Corollary 3.6.2]. We remark that the orthonormal compactly supported wavelet bases used throughout this work fulfill all necessary assumptions and so we do not note the assumptions in detail. The approximation basis should fulfill a direct estimate as given in Lemma 3.4 and a so-called *Whitney estimate*, for details see [Coh03]. For the Besov norm in $B^\kappa_{p,q}$ of $g = \sum_{j \geq -1, k \in \mathbb{Z}^d} \langle g, \psi_{j,k} \rangle \psi_{j,k}$ the following equivalence holds

$$\|g\|_{B^\kappa_{p,q}} \simeq \|(2^{\kappa j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|(\langle g, \psi_{j,k} \rangle)_{k \in \mathbb{Z}^d}\|_{\ell_p})_{j \geq -1}\|_{\ell_q}. \tag{3.7}$$

As a special case we consider $q = p \neq 2$. The norm equivalence then becomes

$$\|g\|^p_{B^\kappa_{p,p}} \simeq \sum_{j \geq -1} 2^{pj\left(\kappa + d(\frac{1}{2}-\frac{1}{p})\right)} \sum_{k \in \mathbb{Z}^d} |\langle g, \psi_{j,k} \rangle|^p.$$

The norm equivalence for Sobolev spaces H^κ follows from (3.7) by setting $p = q = 2$. It is

$$\|g\|^2_{H^\kappa} \simeq \sum_{j \geq -1} 2^{2j\kappa} \sum_{k \in \mathbb{Z}^d} |\langle g, \psi_{j,k} \rangle|^2. \tag{3.8}$$

Using this norm equivalence we show that white noise is reflected by Sobolev spaces of order less than $-d/2$.

Proposition 3.6. *Let $g \in L_2(\mathbb{R}^d)$ with finite support and $g^\delta = g + \delta \mathrm{d}W$ with $\mathrm{d}W$ white noise. Then the expectation of the error can be measured in Sobolev spaces H^κ with order $\kappa < -d/2$. Thus,*

$$E(\|g^\delta - g\|^2_{H^\kappa}) < \infty$$

if and only if $\kappa < -d/2$.

Proof. Without loss of generality we assume $\operatorname{supp} g \subset [0,1]^d$. Let ψ be a compactly supported wavelet and $(\psi_{j,k})_{j\in\mathbb{Z},k\in\mathbb{Z}^d}$ an orthonormal wavelet basis of $L_2([0,1]^d)$. We assume that we have $j_0 = 0$ and that the first layer of scaling functions is incorporated into the first wavelet layer as described in Section 3.2.1. Thus the basis is described by $(\psi_{j,k})_{j\geq -1,k\in\mathbb{Z}^d}$.
We use the norm equivalence (3.8) to compute the expectation of $\|g^\delta - g\|^2_{H^\kappa}$.

$$E(\|g^\delta - g\|^2_{H^\kappa}) \simeq E\Big(\sum_{j\geq -1} 2^{2j\kappa} \sum_{k\in\mathbb{Z}^d} |\langle g^\delta - g, \psi_{j,k}\rangle|^2\Big)$$
$$= \sum_{j\geq -1} 2^{2j\kappa}\delta^2 \sum_{k\in\mathbb{Z}^d} \|\psi_{j,k}\|^2 E(|\mathcal{X}|^2).$$

For a fixed level $j \geq -1$ the sum over $k \in \mathbb{Z}^d$ involves 2^{jd} translations of wavelets on $[0,1]^d$. Hence it is

$$E(\|g^\delta - g\|^2_{H^\kappa}) \simeq \delta^2 E(|\mathcal{X}|^2) \sum_{j\geq 0} 2^{j(2\kappa+d)}.$$

The geometric series $\sum_j 2^{j(2\kappa+d)}$ exists for $\kappa < -d/2$. This completes the proof. □

A similar result holds for Besov spaces.

Proposition 3.7. *Let $g \in L_2(\mathbb{R}^d)$ with finite support and $g^\delta = g + \delta \mathrm{d}W$ with $\mathrm{d}W$ white noise. Then the expectation of the error can be measured in Besov spaces $B^\kappa_{p,p}$ with order $\kappa < -d/2$. It is*

$$E(\|g^\delta - g\|^p_{B^\kappa_{p,p}}) < \infty$$

if and only if $\kappa < -d/2$.

Proof. We assume without loss of generality that $\operatorname{supp} g \subset [0,1]^d$. With ψ we denote a wavelet with compact support such that $(\psi_{j,k})_{j\in\mathbb{Z},k\in\mathbb{Z}^d}$ is an orthonormal wavelet basis of $L_2([0,1]^d)$. We assume further that $j_0 = 0$ and that the first layer of scaling functions is incorporated into the first wavelet layer as described in Section 3.2.1. Hence the wavelet basis is described by $(\psi_{j,k})_{j\geq -1,k\in\mathbb{Z}^d}$. We use the norm equivalence (3.7) for $p = q$ to compute $\|g\|^p_{B^\kappa_{p,p}}$,

$$\|g\|^p_{B^\kappa_{p,p}} \simeq \sum_{j\geq -1} 2^{\kappa jp} 2^{djp(\frac{1}{2}-\frac{1}{p})} \sum_{k\in\mathbb{Z}^d} |\langle g, \psi_{j,k}\rangle|^p.$$

For $E(\|g^\delta - g\|^p_{B^\kappa_{p,p}})$ we get

$$E(\|g^\delta - g\|^p_{B^\kappa_{p,p}}) \simeq E\Big(\sum_{j\geq -1} 2^{jp(\kappa+d(\frac{1}{2}-\frac{1}{p}))} \sum_{k\in\mathbb{Z}^d} |\langle g^\delta - g, \psi_{j,k}\rangle|^p\Big)$$
$$= \sum_{j\geq -1} 2^{jp(\kappa+d(\frac{1}{2}-\frac{1}{p}))} \delta^p E\big(\sum_{k\in\mathbb{Z}^d} \|\psi_{j,k}\|^p |\mathcal{X}|^p\big).$$

Since $\|\psi_{j,k}\| = 1$ and for fixed level $j \geq -1$ the sum over $k \in \mathbb{Z}^d$ involves 2^{jd} translations of wavelets on $[0,1]^d$ it follows

$$E(\|g^\delta - g\|^p_{B^\kappa_{p,p}}) \simeq \delta^p E(|\mathcal{X}|^p) \sum_{j\geq -1} 2^{jp(\kappa+d(\frac{1}{2}-\frac{1}{p}))} 2^{jd}$$
$$= c\delta^p \sum_{j\geq -1} 2^{jp(\kappa+d/2)}.$$

The geometric series $\sum_j 2^{jp(\kappa+d/2)}$ exists for $\kappa < -d/2$. This completes the proof. $\square$

We end the discussion of the appropriate smoothness space for white noise with another motivation for the Sobolev space of order smaller than $-d/2$. White noise is named according to white light. White light is light of all frequencies and is therefore characterized by a constant spectrum. Analogously a white noise process is defined as (stationary) process with constant spectral density, see Definition A.4. In this case the existence of the integral in (3.1) is assured for a Sobolev smoothness index smaller than $-d/2$.

3.2.2 Wavelet Shrinkage

We turn back to the task of constructing an approximation to the exact function g when only noisy data g^δ are known. For this we estimate the wavelet coefficients of g from the wavelet coefficients of g^δ. The noisy wavelet coefficients undergo a thresholding procedure defined by a threshold λ and a thresholding operator S_λ. A possible thresholding operator is

$$S_\lambda(x) := \begin{cases} x & \text{if } |x| \geq \lambda, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

The noisy coefficients are shrunk to zero whenever their moduli are smaller than the threshold λ, otherwise they are kept unchanged. This is called *hard* *thresholding* and is used throughout this thesis. Another realization of the thresholding operator is

$$S_\lambda(x) := \begin{cases} x - \lambda & \text{if } x \geq \lambda, \\ x + \lambda & \text{if } -x \geq \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

The noisy coefficients are again shrunk to zero whenever their moduli are smaller than the threshold λ. Otherwise they are shrunk towards zero by the threshold. This is called *soft thresholding*. In [Lor05] it is shown that hard and soft shrinkage can be treated in a common framework and it is possible to construct a natural interpolation between both of them. For more details on wavelet shrinkage including different ways of choosing the threshold λ we refer the reader to [DJ94, DJKP95, DJ98].

We suppose that the exact function g is contained in some Besov space $B^\theta_{p,p}$. The noisy version is given by $g^\delta = g + \delta \, \mathrm{d}W$.We choose the smoothness η of the approximation space H^η and an appropriate wavelet basis. Since the exact function g is assumed to have Besov smoothness θ we bound the smoothness of the approximation space by θ, i.e., $\eta \leq \theta$. The noisy data g^δ are projected on the subspace $V_j \subset H^\eta(\Omega)$ and then the (hard) shrinkage operator (3.9) is applied to the coefficients:

$$\tilde{g} = g^\delta_{\lambda,j} = S_\lambda P_j g^\delta := \sum_{\tilde{j} \leq j} \sum_{k \in \mathbb{Z}^d} S_\lambda(\langle g^\delta, \psi_{\tilde{j},k}\rangle)\psi_{\tilde{j},k}.$$

An error estimate for this approximation method is given by the next lemma. This estimate becomes important in Section 4.2 where we combine wavelet shrinkage and general regularization methods.

Lemma 3.8. *Let $\theta > 0$, d the dimension of the argument space and $\eta_{\min} := \frac{d}{2}\frac{1}{2\theta+d-1}$. Let $\eta_{\min} < \eta \leq \theta$ and $p > 0$ with $\frac{1}{p} = \frac{1}{2} \cdot \frac{2\theta+d}{2\eta+d}$. Let g be in the Besov ball $B := \{g \in B^\theta_{p,p}(\Omega) : \|g\|_{B^\theta_{p,p}} \leq M\}$ and g^δ given by* (2.6). *Let $\varphi \in H^\eta(\Omega)$ induce a wavelet basis with the projection operator P_j as in* (3.5) *and the shrinkage operator S_λ as in* (3.9).
Choosing the threshold $\lambda = C\delta\sqrt{|\log \delta|}$ and the projection level $j \leq j_1$ with $j_1 = j_1(\delta)$ defined by

$$2^{j_1} \simeq \frac{1}{\delta^2 \, |\log \delta|} \tag{3.10}$$

yields the error estimate

$$\mathrm{E}(\|S_\lambda P_j g^\delta - P_j g\|^2_{H^\eta}) = \mathcal{O}((\delta\sqrt{|\log \delta|})^{\frac{4(\theta-\eta)}{2\theta+d}}). \tag{3.11}$$

Proof. This result follows from [CDKP01, Theorem 4]. □

Remark. *The condition $\eta > \eta_{\min}$ is imposed to simplify notations. For $0 < \eta \leq \eta_{\min} = \frac{d}{2}\frac{1}{2\theta+d-1}$ the condition $g \in B^{\theta}_{p,p}$ has to be supplemented, but the rate* (3.11) *is also achieved. For details we refer the reader to [CDKP01, Theorem 4]. In the following we assume Lemma 3.8 to be valid for $\eta > 0$. Lemma 3.8.*

In Lemma 3.8 an upper bound on the number of detail levels is given. The index j of the detail level corresponds to the size of the support of the wavelet: The greater j gets the smaller is the support of the indicated wavelet. The direct estimate of Lemma 3.4 tells us that the approximation error gets smaller when more detail levels are used. In the case of noisy data, wavelets corresponding to small details are "suspected" to carry more noise than signal information. Due to this, details which are smaller than the ones given by the relation (3.10) are cut off.
Wavelet shrinkage is introduced for Sobolev and Besov spaces and is adapted to smoothness properties of the data. We continue with data estimation by regularization methods based on the singular value decomposition as described in Section 2.3.1.

3.3 Data Estimation by Classical Regularization

In this section we consider a data estimation $S_\lambda : Y \to Y$ based on classical regularization methods for linear compact operators as introduced in (2.18),

$$R_\alpha g := \sum_{\sigma_n > 0} F_\alpha(\sigma_n)\sigma_n^{-1}\langle g, v_n\rangle u_n.$$

We have a close look on the filter function F_α. For $\gamma_1, \gamma_2 \in [0,1]$ with $\gamma_1 + \gamma_2 = 1$ we split the filter $F_\alpha(\sigma)$ according to

$$F_\alpha(\sigma) = F_\alpha(\sigma)^{\gamma_2} F_\alpha(\sigma)^{\gamma_1}. \tag{3.12}$$

The mapping $R_\alpha : X \to Y$ is split analogously in two parts which fits in the proposed scheme of a two-step method. We define

$$R_\alpha = T_{\alpha,\gamma_2} S_{\alpha,\gamma_1}$$

with

$$\begin{aligned} S_{\alpha,\gamma_1} &: Y \to Y \\ T_{\alpha,\gamma_2} &: Y \to X. \end{aligned}$$

The operators S_{α,γ_1} and T_{α,γ_2} are given by

$$g^\delta_{\alpha,\gamma_1} = S_{\alpha,\gamma_1} g^\delta = \sum_{\sigma_n \geq 0} F_\alpha(\sigma_n)^{\gamma_1} \langle g^\delta, v_n \rangle v_n, \tag{3.13}$$

$$f^\delta_{\alpha,\gamma_2} = T_{\alpha,\gamma_2} g^\delta_{\alpha,\gamma_1} = \sum_{\sigma_n > 0} F_\alpha(\sigma_n)^{\gamma_2} \sigma_n^{-1} \langle g^\delta_{\alpha,\gamma_1}, v_n \rangle u_n. \tag{3.14}$$

As long as we use the same parameter α in (3.13) and (3.14) and maintain the condition $\gamma_1 + \gamma_2 = 1$ this is a classical regularization method split in two steps. The data smoothing step is given by S_{α,γ_1}, acting on the data side Y. The reconstruction step is given by T_{α,γ_2}, the "rest" of the regularization R_α acting from Y to X, see Figure 3.1(a).

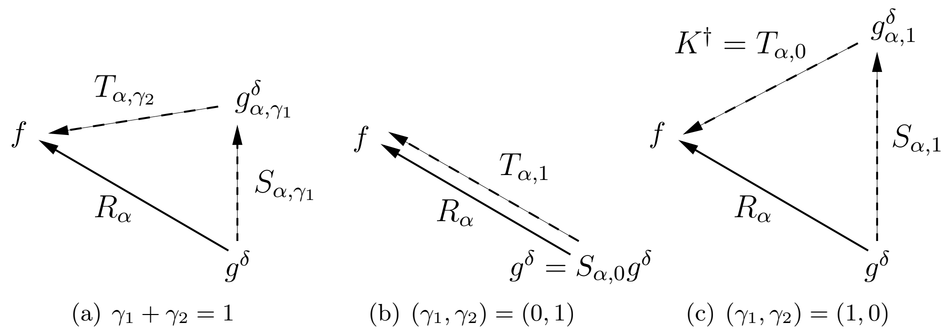


Figure 3.1: Classical regularization as a two-step method.

We remark that for the couple of parameters $(\gamma_1, \gamma_2) = (0, 1)$ the operator S_{α,γ_1} is the identity and the operator T_{α,γ_2} is the initial regularization operator R_α, see Figure 3.1(b). For the parameter couple $(\gamma_1, \gamma_2) = (1, 0)$ the operator S_{α,γ_1} smoothes with the complete filter on the data side and the operator T_{α,γ_2} is the generalized inverse of the operator K, see Figure 3.1(c). These ideas connect to the work of Jonas and Louis [JL01]. The authors show that linear regularization methods can be interpreted as a combination of the pseudoinverse and a smoothing operator in either order. In the context of [JL01] the case described by the parameter couple $(\gamma_1, \gamma_2) = (1, 0)$ belongs to the *prewhitening* or *range mollification* methods. We remark that the approach introduced in this work is more general. The range of γ_1 and γ_2 in the parameter couple (γ_1, γ_2) allows every possible combination of data smoothing and reconstruction in terms of filter functions used for classical regularization methods.
We remark that the two steps (3.13) and (3.14) still share the parameter α. Another interesting idea in this context is to consider a couple (α_1, α_2) of

regularization parameters. This is another possibility to weight the two operators S_{α_1,γ_1} and T_{α_2,γ_2} and thereby weight the influence of data estimation and reconstruction. But we do not follow up this idea any further.
We concentrate on the data-side operator S_{α,γ_1} for $0 < \gamma_1 < 1$. The spaces $Y_\nu = \mathrm{rg}\,((KK^*)^{\nu/2})$ as introduced in (2.26) measure the decay of the Fourier coefficients of $g \in Y$ with respect to the singular value sequence $\sigma_n^{-\nu}$. We prove that for both the Tikhonov and the Landweber method the data estimate $S_{\alpha,\gamma_1} g^\delta$ is contained in the space Y_ν as long as $\gamma_1 > \nu/2$, see Propositions 3.9 and 3.12. We restrict ourselves to the scale of spaces Y_ν for $\nu \in [0,1]$. For $\nu = 0$ the space Y_ν equals the space Y in which data accuracy is measured. For the upper bound, $\nu = 1$, the Fourier coefficients $\langle g, v_n\rangle$ of $g \in Y$ are weighted against σ_n^{-1}. This is also the case for the Picard condition (2.16) which in addition to $g \in \overline{\mathrm{rg}\,(K)}$ characterizes the range of a compact operator. That means, if the data estimate is obtained by a filter function with exponent $\gamma_1 > 1/2$ no further regularization is necessary; the generalized inverse could be applied directly and stable. In other words for $\gamma_1 > 1/2$ the operator $R_\alpha := K^\dagger S_{\alpha,\gamma_1}$ already defines an optimal regularization method. This is discussed in detail in Section 4.1. It explains the well-known effect of oversmoothing when using the classical Tikhonov or Landweber method, see [Lam01].
We continue with the classical Tikhonov regularization interpreted as data estimation.

3.3.1 Tikhonov Method

A filter frequently used for solving ill-posed problems is the Tikhonov filter $F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2+\alpha}$ with parameter α. For $\gamma_1, \gamma_2 \in [0,1]$ with $\gamma_1 + \gamma_2 = 1$ we split the Tikhonov filter as introduced in (3.12),

$$F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2+\alpha} = \left(\frac{\sigma^2}{\sigma^2+\alpha}\right)^{\gamma_2} \left(\frac{\sigma^2}{\sigma^2+\alpha}\right)^{\gamma_1}.$$

The Tikhonov operator $R_\alpha : X \to Y$ is split in two parts according to $R_\alpha = T_{\alpha,\gamma_2} S_{\alpha,\gamma_1}$. The operators $S_{\alpha,\gamma_1} : Y \to Y$ and $T_{\alpha,\gamma_2} : Y \to X$, see (3.13) and (3.14), are given by

$$\begin{aligned} g^\delta_{\alpha,\gamma_1} &= S_{\alpha,\gamma_1} g^\delta = \sum_{n\geq 0} \left(\frac{\sigma_n^2}{\sigma_n^2+\alpha}\right)^{\gamma_1} \langle g^\delta, v_n\rangle v_n, \qquad (3.15) \\ f^\delta_{\alpha,\gamma_2} &= T_{\alpha,\gamma_2} g^\delta_{\alpha,\gamma_1} = \sum_{\sigma_n>0} \left(\frac{\sigma_n^2}{\sigma_n^2+\alpha}\right)^{\gamma_2} \sigma_n^{-1} \langle g^\delta_{\alpha,\gamma_1}, v_n\rangle u_n. \end{aligned}$$

In the following we are concerned with the smoothing capability of the data side operator $S_{\alpha,\gamma} : Y \to Y$, where we omit the index of the parameter γ. The smoothing capability is measured in terms of the Hilbert scale Y_ν introduced in (2.26). We particularly show that $g_{\alpha,\gamma} = S_{\alpha,\gamma} g$ fulfills the Picard-condition as long as $\gamma > 1/2$.
For $\gamma < 1$ we call the application of the exponent γ to the filter function a *filter reduction*. This notation is misleading: The filter $F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2+\alpha}$ is smaller than 1. Since $\gamma \in (0, 1)$ we take a root and enlarge the value, i.e., $F_\alpha(\sigma) \leq F_\alpha^\gamma(\sigma)$. This can be seen in Figure 3.2 where the filter $F_\alpha^\gamma(\sigma)$ is shown for different values of γ. Figure 3.3 shows the influence of the Tikhonov filter on the reciprocal singular values – again for different values of γ. The *filter reduction* is represented by an increase: The smaller γ gets the greater the values of $\sigma^{-1}(\sigma^2/(\sigma^2 + \alpha))^\gamma$ become, and the less damping is done by the filter function. Hence, the parameter γ causes a *reduction of filtering*. Data errors belonging to small singular values gain more influence. In Section 4.1 we show that as long as the exponent γ is greater than 1/2 the reduced filter still defines an optimal regularization method.

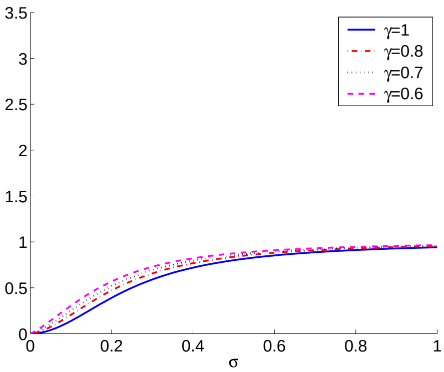


Figure 3.2: Classical and reduced Tikhonov filter $(\sigma^2/(\sigma^2+\alpha))^\gamma$ for $\alpha = 1/16$ and $\gamma = 1, 0.8, 0.7, 0.6$.

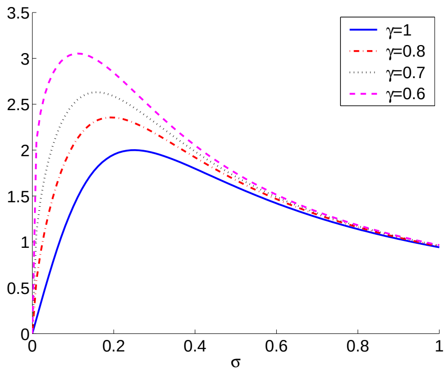


Figure 3.3: Data error $\sigma^{-1}(\sigma^2/(\sigma^2 + \alpha))^\gamma$ of classical and reduced Tikhonov filter for $\alpha = 1/16$ and $\gamma = 1, 0.8, 0.7, 0.6$.

The next proposition presents the result on the smoothing capability of the operator $S_{\alpha,\gamma}$.

Proposition 3.9. *Let $K : X \to Y$ be a compact operator with singular system (σ_n, u_n, v_n). Let $\{Y_\nu\}$ with $Y_\nu := \operatorname{rg}((KK^*)^{\nu/2})$ denote the Hilbert scale generated by the operator $L = (KK^*)^{-1/2}$. Let $\nu \in (0, 1]$ and $g \in Y =$*

Y_0 be given. Let $\alpha > 0$ and $S_{\alpha,\gamma}$ be the Tikhonov smoothing operator as in (3.15)*. For $\alpha > 0$ and $\gamma > \frac{\nu}{2}$ we have*

$$g_{\alpha,\gamma} := S_{\alpha,\gamma} g \in Y_\nu \tag{3.16}$$

with

$$\|g_{\alpha,\gamma}\|_{Y_\nu} \lesssim \alpha^{-\nu/2}. \tag{3.17}$$

For the reduced Tikhonov filter *F_α^γ the following holds*

$$\sup_{0<\sigma\le\sigma_1} |\sigma^{-\nu} F_\alpha^\gamma(\sigma)| \le c_{\gamma,\nu}\alpha^{-\nu/2}. \tag{3.18}$$

Proof (of Proposition 3.9). For

$$g_{\alpha,\gamma} = S_{\alpha,\gamma} g = \sum_{n\ge 0} \left(\frac{\sigma_n^2}{\sigma_n^2+\alpha}\right)^\gamma \langle g, v_n\rangle v_n$$

we compute the Y_ν-norm.

$$\begin{aligned} \|g_{\alpha,\gamma}\|_{Y_\nu}^2 &= \sum_{n\ge 0} \sigma_n^{-2\nu} \left(\frac{\sigma_n^2}{\sigma_n^2+\alpha}\right)^{2\gamma} |\langle g, v_n\rangle|^2 \\ &= \sum_{n\ge 0} \underbrace{\frac{\sigma_n^{2(2\gamma-\nu)}}{(\sigma_n^2+\alpha)^{2\gamma}}}_{=:\varphi(\sigma_n,\gamma,\nu)} |\langle g, v_n\rangle|^2 \\ &\le \sup_{\sigma>0} \varphi(\sigma,\gamma,\nu)\, \|g\|_{Y_0}^2. \end{aligned}$$

The statement follows by maximizing φ with respect to σ. It is

$$\varphi'(\sigma) = [(2\gamma-\nu)(\sigma^2+\alpha) - 2\sigma^2\gamma] \cdot h(\sigma)$$

with a function $h \ne 0$. Hence we get as condition for critical points

$$\sigma^2 = \alpha\frac{2\gamma-\nu}{\nu}. \tag{3.19}$$

Existence is assured by $2\gamma > \nu$. For $2\gamma > \nu$ the function φ is continuous. Since $\varphi \ge 0$, $\varphi(0) = 0$ and $\lim_{\sigma\to\infty}\varphi(\sigma) = 0$ we get the maximum point $\sigma_*(\gamma,\nu) = \sqrt{\alpha}\sqrt{(2\gamma-\nu)/\nu}$. This proves assertion (3.16). Inserting σ_* in φ yields the maximum

$$\varphi(\sigma) \le \varphi(\sigma_*) = \alpha^{-\nu} c_{\gamma,\nu}$$

with the constant

$$c_{\gamma,\nu} = \left(\frac{2\gamma-\nu}{\nu}\right)^{2\gamma-\nu} \cdot \left(\frac{\nu}{2\gamma}\right)^{2\gamma}$$

depending on γ and ν. This proves assertions (3.17) and (3.18). $\square$

Corollary 3.10. *Let all conditions of Proposition 3.9 be fulfilled. For $g \in Y$ and $\gamma > \frac{1}{2}$ we have that $g_{\alpha,\gamma} := S_{\alpha,\gamma} g$ fulfills the Picard condition* (2.16).

Proof. Inserting $\nu = 1$ in $\|\cdot\|_{Y_\nu}^2 = \sum_{\sigma_n>0} \sigma_n^{-2\nu} |\langle \cdot, v_n\rangle|^2$ yields (2.16) and the assertion follows immediately from Proposition 3.9. $\square$

In the next section we show that for the Landweber filter a property according to (3.18) is valid. This property is discussed at the end of Section 3.3. In Section 4.2 the property turns out to be important for the combination of regularization methods with wavelet shrinkage.

3.3.2 Landweber Method

In this section we consider the Landweber method as data smoothing operation. The Landweber filter is introduced in (2.23) as

$$F_m(\sigma) = 1 - (1 - \beta\sigma^2)^m.$$

The Landweber regularization is an iterative method with index $m \in \mathbb{N}_+$. The reciprocal $\alpha := 1/m$ is considered as regularization parameter. We split the filter as introduced in (3.12). The Landweber smoothing operator is defined as in (3.13)

$$g_{m,\gamma} = S_{m,\gamma} g = \sum_{n\geq 0} (1 - (1 - \beta\sigma_n^2)^m)^\gamma \langle g, v_n\rangle v_n. \tag{3.20}$$

We call $(1 - (1 - \beta\sigma^2)^m)^\gamma$ the *reduced Landweber filter* with parameter γ. Figure 3.4 shows the classical Landweber filter and the reduced filter for different exponents γ. Figure 3.5 shows the influence of the classical and the reduced Landweber filter on the reciprocals of the singular values. The amount of damping decreases with γ. Hence the parameter γ causes a *reduction of filtering*.

In Proposition 3.12 we prove that for $\gamma > \nu/2$ and $g \in Y$ we have $S_{m,\gamma} g \in Y_\nu$. For the special case $\nu = 1$ this implies that the Picard condition is fulfilled for $\gamma > 1/2$. In Section 4.1 we show that as long as the exponent γ is greater than $1/2$ the reduced Landweber filter defines an order optimal

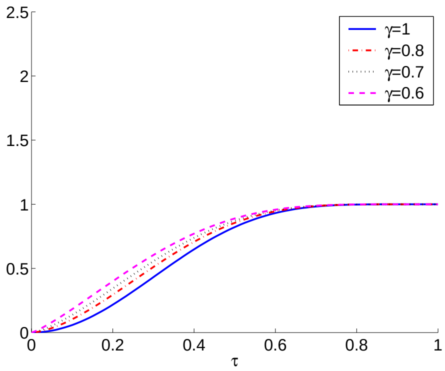


Figure 3.4: Classical and reduced Landweber filter $(1-(1-\tau^2)^m)^\gamma$ for $m=6$ and $\gamma = 1, 0.8, 0.7, 0.6$.

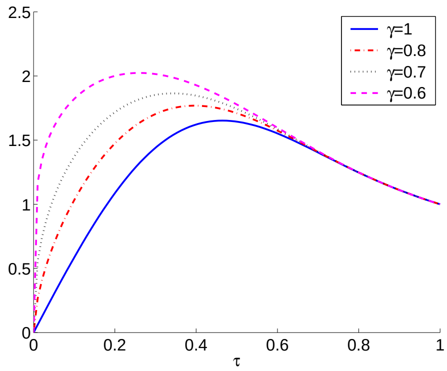


Figure 3.5: Data error $\tau^{-1}(1-(1-\tau^2)^m)^\gamma$ of classical and reduced Landweber filter for $m=6$ and $\gamma = 1, 0.8, 0.7, 0.6$.

regularization method. The following lemma is an auxiliary result used in Proposition 3.12. It proves that the reduced Landweber filter controls the function $\tau^{-\nu}$ for an exponent γ with $\gamma > \nu/2$.

Lemma 3.11. *Let $\nu > 0$ and $m \in \mathbb{N}_+$. For $\gamma > \nu/2$ the function*

$$\phi(\tau) = \tau^{-2\nu}[1-(1-\tau^2)^m]^{2\gamma}$$

is continuous in $[0,\infty)$. For $\gamma > \nu/2$ and $m \geq 2$ the function ϕ restricted to $[0, \sqrt{2}]$ has a maximum and is bounded by

$$\phi(\tau) \leq m^\nu. \tag{3.21}$$

For $m = 1$ it is $\phi(\tau) \leq \phi(\sqrt{2}) = 2^{2\gamma-\nu}$.

Proof. We start with proving the continuity of ϕ. The expansion $(1-\tau^2)^m = \sum_{k=0}^m (-1)^k \tau^{2k}\binom{m}{k}$ yields

$$\phi(\tau) = \tau^{-2\nu}\left(\sum_{k=1}^m (-1)^{k-1}\tau^{2k}\binom{m}{k}\right)^{2\gamma}.$$

Thus for $\gamma > \nu/2$ there is no singularity and the function ϕ is continuous. Hence the function restricted to the interval $[0, \sqrt{2}]$ has a maximum. For $m = 1$ we have $\phi(\tau) = \tau^{2(2\gamma-\nu)}$ and ϕ is bounded by the value at the upper bound for the argument, $\phi(\sqrt{2}) = 2^{2\gamma-\nu}$.

For $m > 1$ the proof of assertion (3.21) is outlined as follows. We maximize not ϕ itself but ϕ restricted to all its critical points. The description of the critical points is gained by seeking zeros of the derivative of ϕ. Differentiating ϕ with respect to τ yields

$$\begin{aligned}
\phi'(\tau) &= (-2\nu)\tau^{-1}\tau^{-2\nu}[1-(1-\tau^2)^m]^{2\gamma} \\
&\quad + \tau^{-2\nu}\left(4\gamma m\tau[1-(1-\tau^2)^m]^{2\gamma}[1-(1-\tau^2)^m]^{-1}(1-\tau^2)^{m-1}\right) \\
&= (-2\nu)\tau^{-1}\phi(\tau) + 4\gamma m\tau[1-(1-\tau^2)^m]^{-1}(1-\tau^2)^{m-1}\phi(\tau) \\
&= \phi(\tau)\left(-2\nu\tau^{-1} + 4\gamma m\tau[1-(1-\tau^2)^m]^{-1}(1-\tau^2)^{m-1}\right).
\end{aligned}$$

Hence the derivative of ϕ can be written as

$$\phi'(\tau) = \phi(\tau)\cdot h(\tau)$$

with

$$h(\tau) = -2\nu\tau^{-1} + 4\gamma m\tau(1-\tau^2)^{m-1}[1-(1-\tau^2)^m]^{-1}.$$

Seeking zeros of ϕ' amounts to seeking zeros of ϕ and h,

$$\phi'(\tau) = 0 \;\Leftrightarrow\; \phi(\tau) = 0 \;\vee\; h(\tau) = 0.$$

Since ϕ is nonnegative with $\phi(1) = 1 > 0$ the zeros of ϕ are no candidates for maxima of ϕ. Hence we concentrate on the zeros of h. Without loss of generality we assume $\tau \neq 0$ and $1-(1-\tau^2)^m \neq 0$ (all τ with $1-(1-\tau^2)^m = 0$ and $\tau = 0$ are zeros of ϕ). Then the condition $h(\tau) = 0$ is equivalent to

$$(1-\tau^2)^m = \frac{\nu(1-\tau^2)}{\nu(1-\tau^2) + 2\gamma m\tau^2}. \tag{3.22}$$

This is a description of the critical points of ϕ. We use (3.22) to define the function ϕ_{critical} according to

$$\phi_{\text{critical}} := \phi_{|\text{critical points}}.$$

We insert (3.22) in $\phi(\tau) = \tau^{-2\nu}(1-(1-\tau^2)^m)^{2\gamma}$ and get

$$\phi_{\text{critical}}(\tau) = \tau^{-2\nu}\left(\frac{2\gamma m\tau^2}{\nu(1-\tau^2) + 2\gamma m\tau^2}\right)^{2\gamma}.$$

If ϕ_{critical} has a maximum in $[0, \sqrt{2}]$ we have

$$\phi \leq \max_{[0,\sqrt{2}]} \phi_{\text{critical}}(\tau).$$

The rest of the proof is concerned with maximizing ϕ_{critical}. Figure 3.6 shows the functions ϕ and ϕ_{critical} as well as the maximum of ϕ_{critical}. Differentiating ϕ_{critical} with respect to τ yields

$$\begin{aligned}\phi'_{\text{critical}}(\tau) &= 2\nu\tau^{-(2\nu+1)}\left(\frac{2\gamma m\tau^2}{\nu(1-\tau^2)+2\gamma m\tau^2}\right)^{2\gamma}\left[\frac{2\gamma}{\nu(1-\tau^2)+2\gamma m\tau^2}-1\right] \\ &= 2\nu\tau^{-1}\phi_{\text{critical}}(\tau)\left[\frac{2\gamma}{\nu(1-\tau^2)+2\gamma m\tau^2}-1\right].\end{aligned}$$

Because of $\phi_{\text{critical}} > 0$ seeking zeros of ϕ_{critical} results in solving

$$\frac{2\gamma}{\nu(1-\tau^2)+2\gamma m\tau^2} = 1. \tag{3.23}$$

This is equivalent to

$$2\gamma - \nu = \tau^2(2\gamma m - \nu).$$

Thus for $2\gamma > \nu$ critical points exist and are given by

$$\tau_*^2 = \frac{2\gamma - \nu}{2\gamma m - \nu}.$$

For $m > 1$ we have $\tau_*^2 < 1$ and the positive root is in the interval $(0,1) \subset [0,\sqrt{2}]$. In order to check whether $\tau_* = +\sqrt{\frac{2\gamma-\nu}{2\gamma m-\nu}}$ is a maximum of ϕ_{critical} we compute the second derivative. We remark that

$$\begin{aligned}\phi'_{\text{critical}}(\tau) &= 2\nu\tau^{-1}\phi_{\text{critical}}(\tau)\left[\frac{2\gamma}{\nu(1-\tau^2)+2\gamma m\tau^2}-1\right] \\ &= 2\nu\tau^{-1}\phi_{\text{critical}}(\tau)f(\tau).\end{aligned}$$

The second derivative is thus given by

$$\begin{aligned}\phi''_{\text{critical}}(\tau) &= -2\nu\tau^{-2}\phi_{\text{critical}}(\tau)f(\tau) \\ &\quad +2\nu\tau^{-1}[\phi'_{\text{critical}}(\tau)f(\tau) + \phi_{\text{critical}}(\tau)f'(\tau)].\end{aligned}$$

With $\phi'_{\text{critical}}(\tau_*) = 0$ and $f(\tau_*) = 0$ we get

$$\begin{aligned}\phi''_{\text{critical}}(\tau_*) &= 2\nu\tau_*^{-1}\phi_{\text{critical}}(\tau_*)f'(\tau_*) \\ &= -8\gamma\nu^2 \cdot \frac{2\gamma m - \nu}{(\nu(1-\tau_*^2)+2\gamma m\tau_*^2)^2} \\ &< 0.\end{aligned}$$

The last line follows from $\nu > 0$ and $\gamma > \nu/2$.

Hence we evaluate ϕ_{critical} at τ_* and get as maximum

$$\phi_{\text{critical}}(\tau_*) = \left(\frac{2\gamma - \nu}{2\gamma m - \nu}\right)^{2\gamma-\nu} m^{2\gamma}. \tag{3.24}$$

It remains to show that

$$\left(\frac{2\gamma - \nu}{2\gamma m - \nu}\right)^{2\gamma-\nu} m^{2\gamma} \leq m^{\nu}.$$

Since $\gamma > \nu/2 > 0$ and $m > 1$ this is equivalent to $(2\gamma - \nu)^{2\gamma-\nu} m^{2\gamma-\nu} \leq (2\gamma m - \nu)^{2\gamma-\nu}$. By the monotony of the power function this is equivalent to $2\gamma m - m\nu \leq 2\gamma m - \nu$ and assertion (3.21) is proved. $\square$

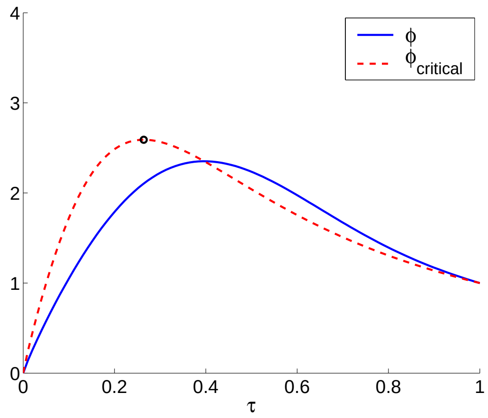


Figure 3.6: The functions ϕ (blue, solid) and ϕ_{critical} (red, dashed) for the parameters $\gamma = 0.7$, $\nu = 0.75$ and $m = 6$.

The smoothing capability of the reduced Landweber filter with respect to the Hilbert scale $\{Y_\nu\}$ with $Y_\nu := \operatorname{rg}((KK^*)^{\nu/2})$ is described in the following proposition.

Proposition 3.12. *Let $K : X \to Y$ be a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$. Then $\{Y_\nu\}$ with $Y_\nu := \operatorname{rg}((KK^*)^{\nu/2})$ is a Hilbert scale generated by the operator $L = (KK^*)^{-1/2}$. Let $g \in Y = Y_0$ be given. Let $m \in \mathbb{N}_+$ and $\alpha = 1/m$. Let $S_{\alpha,\gamma}$ be given as in* (3.13) *with the Landweber filter*

$$F_m(\sigma) = 1 - (1 - \beta\sigma^2)^m.$$

Let $0 < \beta < \frac{2}{\|K\|^2}$ and $\nu \in (0, 1]$ fixed. Then for $\gamma > \frac{\nu}{2}$ it is

$$g_{\alpha,\gamma} := S_{\alpha,\gamma} g \in Y_\nu \tag{3.25}$$

and

$$\|g_{\alpha,\gamma}\|_{Y_\nu} \lesssim \beta^{\nu/2} m^{\nu/2}. \tag{3.26}$$

For the reduced Landweber filter F_α^γ it is

$$\sup_{0<\sigma\le\sigma_1} |\sigma^{-\nu} F_m^\gamma(\sigma)| \le \beta^{\nu/2} m^{\nu/2}. \tag{3.27}$$

Proof. With $1/\alpha = m$ we compute $g_{m,\gamma} = S_{m,\gamma} g$ as

$$g_{m,\gamma} = \sum_{n\ge 0} [1-(1-\beta\sigma_n^2)^m]^\gamma \langle g, v_n\rangle v_n.$$

For the (squared) Y_ν-norm of $g_{m,\gamma}$ it is

$$\begin{aligned}\|g_{m,\gamma}\|_{Y_\nu}^2 &= \sum_{n\ge 0} \underbrace{\sigma_n^{-2\nu}[1-(1-\beta\sigma_n^2)^m]^{2\gamma}}_{=:\varphi(\sigma_n,\gamma,\nu)} |\langle g, v_n\rangle|^2 \\ &\le \sup_{0<\sigma\le\sigma_1} \varphi(\sigma,\gamma,\nu)\, \|g\|_{Y_0}^2. \end{aligned} \tag{3.28}$$

In $\varphi(\sigma) = \sigma^{-2\nu}[1-(1-\beta\sigma^2)^m]^{2\gamma}$ we substitute $\tau^2 := \beta\sigma^2$. This yields $\sigma^{-2\nu}[1-(1-\beta\sigma^2)^m]^{2\gamma} = \beta^\nu \tau^{-2\nu}[1-(1-\tau^2)^m]^{2\gamma} = \beta\tilde\varphi(\tau)$ with

$$\tilde\varphi(\tau) = \tau^{-2\nu}[1-(1-\tau^2)^m]^{2\gamma}.$$

Since $\beta < \frac{2}{\|K\|^2} = \frac{2}{\sigma_1^2}$ we maximize $\tilde\varphi$ on the interval $[0,\sqrt{2}]$. For $\gamma > \nu/2$ we get from Lemma 3.11

$$\tilde\varphi(\tau) \le m^\nu.$$

With the factor β^ν this yields

$$\sup_{0<\sigma\le\sigma_1} \varphi(\sigma,\gamma,\nu) \le \beta^\nu m^\nu.$$

We insert the estimate for φ in (3.28) and get

$$\begin{aligned}\|g_{m,\gamma}\|_{Y_\nu}^2 &\le \sup_{0<\sigma\le\sigma_1} \varphi(\sigma,\gamma,\nu)\, \|g\|_{Y_0}^2 \\ &\lesssim \beta^\nu m^\nu\end{aligned}$$

which proves the assertions (3.25), (3.26) and (3.27). □

Corollary 3.13. *Let all conditions of Proposition 3.12 be satisfied. For $g \in Y$ and $\gamma > \frac{1}{2}$ we have that $g_{\alpha,\gamma} := S_{\alpha,\gamma} g$ fulfills the Picard condition* (2.16)*.*

Proof. The assertion follows immediately from Proposition 3.12 by inserting $\nu = 1$ in $\|\cdot\|_{Y_\nu}^2 = \sum_{\sigma_n>0} \sigma_n^{-2\nu} |\langle\cdot, v_n\rangle|^2$. □

We end the discussion of using classical regularization methods as data estimation with some remarks on the truncated singular value decomposition and a discussion on filter conditions.

Truncated singular value decomposition

Another order optimal regularization method is the truncated singular value decomposition, see (2.20) and Proposition 2.21. Regularization is achieved by truncating those components which belong to singular values smaller than a certain value. Performing the truncation of the singular value expansion in two steps simply amounts to cutting off components to singular values in two steps. There are no reduction results similar to the ones gained for the Tikhonov and the Landweber filter by application of an exponent γ. We postpone the further discussion to the next chapter where we – among others things – present a conjecture on the question which properties of a filter function allow to use a reduced filter function without losing the order optimality of the induced methods.
For the TSVD we show that it has a property similiar to (3.18) for the Tikhonov and (3.27) for the Landweber filter. This property is used later in Section 4.2.

Lemma 3.14. *Let $K : X \to Y$ be a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$ and let $F_\alpha(\sigma)$ be the filter given by* (2.20)*. We then have for $\alpha > 0$ and $\nu > 0$ that*

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu} F_\alpha(\sigma)| \leq \alpha^{-\nu}.$$

Proof. The filter function $F_\alpha(\sigma)$ is given by $F_\alpha(\sigma) := \begin{cases} 1, & \sigma \geq \alpha, \\ 0, & \sigma < \alpha. \end{cases}$ Hence, for $\nu > 0$ we immediately have

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu} F_\alpha(\sigma)| = \sup_{\alpha\leq\sigma\leq\sigma_1} |\sigma^{-\nu}| = \alpha^{-\nu}.$$

□

Discussion of filter condition

So far we have shown that the (reduced) Tikhonov filter, the (reduced) Landweber filter and the TSVD filter fulfills the following condition

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu} F_\alpha^\gamma(\sigma)| \leq c\alpha^{-\beta\nu} \quad \text{for } \nu \in (0, 1]. \tag{3.29}$$

(This was shown in Proposition 3.9, Proposition 3.12 and in Lemma 3.14.)

Condition (3.29) is similar to condition (2.27a) which assures optimality for the induced regularization method and reads as follows

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-1}F_\alpha^\gamma(\sigma)| \leq c\alpha^{-\beta}.$$

In Section 4.1 we show that the reduced Tikhonov and the reduced Landweber filter $F_\alpha^\gamma(\sigma)$ fulfills the optimality condition (2.27a) as long as $\gamma > 1/2$. Here we ask how (2.27a) and (3.29) are linked.
Condition (2.27a) follows from (3.29) by setting $\nu = 1$. The other way round we consider equation (3.29) as variation of condition (2.27a) in the following sense. In (3.29) the damping property of the filter function is given not only with respect to σ^{-1} but also with respect to $\sigma^{-\nu}$ for all $\nu \in (0, 1]$.
The following lemma shows that for $|F_\alpha^\gamma(\sigma)| \leq c$ for all α and all $\sigma \in (0, \sigma_1]$ the classical condition (2.27a) implies (3.29). We remark that in this lemma we omit the index γ for convenience of notation and F_α is any optimal filter (optimality of the reduced filter functions F_α^γ is treated in the very next Section 4.1).

Lemma 3.15. *Let F_α be an optimal filter with $|F_\alpha(\sigma)| \leq c$ for all α and all $\sigma \in (0, \sigma_1]$. Let $\nu \in (0, 1]$. Then*

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-1}F_\alpha(\sigma)| \leq c\alpha^{-\beta}$$

implies

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu}F_\alpha(\sigma)| \leq c\alpha^{-\beta\nu}.$$

Proof. i) Let $\alpha^\beta\sigma^{-1} \geq 1$. For $\nu \in (0, 1]$ it is $(\alpha^\beta\sigma^{-1})^\nu \leq \alpha^\beta\sigma^{-1}$. Hence $\alpha^{\beta\nu}\sigma^{-\nu}|F_\alpha(\sigma)| \leq \alpha^\beta\sigma^{-1}|F_\alpha(\sigma)| \leq c$. Multiplying by $\alpha^{-\beta\nu}$ yields

$$\sigma^{-\nu}|F_\alpha(\sigma)| \leq c\alpha^{-\beta\nu}.$$

ii) Let $\alpha^\beta\sigma^{-1} < 1$. For $\nu \in (0, 1]$ it follows that $\alpha^{\beta\nu}\sigma^{-\nu} < 1$. For $\sigma \in (0, \sigma_1]$ we have $\alpha^{\beta\nu}\sigma^{-\nu}|F_\alpha(\sigma)| < |F_\alpha(\sigma)| \leq c$. Multiplying by $\alpha^{-\beta\nu}$ yields

$$\sigma^{-\nu}|F_\alpha(\sigma)| \leq c\alpha^{-\beta\nu}.$$

For $\sigma \in (0, \sigma_1]$ we take the supremum and get

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu}F_\alpha(\sigma)| \leq c\alpha^{-\beta\nu}.$$

□

Chapter 4

Two-step Regularization Methods

In this chapter we present our results on solving linear ill-posed problems in two steps according to $T_{\alpha,\lambda} = R_\alpha S_\lambda$. The chapter is divided in two sections. In Section 4.1 we perform classical regularization methods in two steps, whereas in Section 4.2 classical regularization methods and wavelet shrinkage are combined. This is related to the work of Cohen et al. [CHR04] in which the authors study the combination of wavelet-Galerkin projection and wavelet shrinkage.
The main result of Section 4.1 is that for the Tikhonov and the Landweber filter a *filter reduction* as introduced in Section 3.3 is possible without loosing the order optimality of the induced regularization methods, see Proposition 4.2 and Proposition 4.4. We also present an assumption on the conditions which enable filter reduction. The main result of Section 4.2 is given in Theorem 4.12. We prove optimality for the combined method of regularization and wavelet shrinkage. Moreover in Section 4.2.3 we show that using wavelet shrinkage as data preprocessing allows to enlarge the class of reconstruction operators which achieves the optimal convergence results of Theorem 4.12.

4.1 Reduced Classical Methods

We recall the interpretation of a regularization operator R_α as a two-step operation as given in Section 3.3,

$$R_\alpha = T_{\alpha,\gamma_2} S_{\alpha,\gamma_1}$$

with

$$\begin{aligned} S_{\alpha,\gamma_1} &: Y \to Y, \\ T_{\alpha,\gamma_2} &: Y \to X. \end{aligned}$$

The operators S_{α,γ_1} and T_{α,γ_2} are defined in (3.13) and (3.14). In Section 3.3 we have considered the data side operator S_{α,γ_1} with different filter functions. From Corollary 3.10 and Corollary 3.13 we know that the data estimate obtained from the reduced Tikhonov as well as from the reduced Landweber filter fulfills the Picard condition as long as $\gamma_1 > 1/2$. Since the Picard condition characterizes the range of a compact operator we suggest to use as reconstruction operator $T_{\alpha,\gamma_2} : Y \to X$ the generalized inverse $K^\dagger$. In doing so we get the following two-step method:

$$R_{\alpha,\gamma_1} = K^\dagger S_{\alpha,\gamma_1}.$$

We recall the representations (3.13) for S_{α,γ_1} and (2.17) for $K^\dagger$,

$$S_{\alpha,\gamma_1} g^\delta = \sum_{\sigma_n \geq 0} F_\alpha(\sigma_n)^{\gamma_1} \langle g^\delta, v_n \rangle v_n,$$

$$K^\dagger g = \sum_{\sigma_n > 0} \sigma_n^{-1} \langle g, v_n \rangle u_n.$$

For the compound operator $K^\dagger S_{\alpha,\gamma_1}$ we get

$$K^\dagger S_{\alpha,\gamma_1} g^\delta = \sum_{\sigma_n > 0} F_\alpha(\sigma_n)^{\gamma_1} \sigma_n^{-1} \langle g^\delta, v_n \rangle u_n.$$

The same mapping was already given in (3.14) as the second step T_{α,γ_2} of classical regularization methods performed in two steps. Hence the two-step method $K^\dagger S_{\alpha,\gamma_1}$ can be performed in one step according to

$$T_{\alpha,\gamma_2} = K^\dagger S_{\alpha,\gamma_1} \quad \text{with } \gamma_2 = \gamma_1.$$

From now on we consider the mapping $T_{\alpha,\gamma} : Y \to X$ where we omit the index of γ. We show that for the Tikhonov and for the Landweber filter the operator $T_{\alpha,\gamma}$ defines an order optimal regularization method in the sense of Definition 2.17 as long as $\gamma > 1/2$. We remark that $T_{\alpha,\gamma}$ is not a two-step method any more, but that the result on the possibility of filter reduction is gained by considering classical regularization as a two-step method.

Definition 4.1. *Let $K : X \to Y$ be a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$. Let $F_\alpha(\sigma)$ denote the Tikhonov filter or the Landweber filter as given in* (2.21) *or* (2.23)*. For $\gamma \in (0,1)$ the mapping*

$$T_{\alpha,\gamma} : Y \to X$$

with

$$T_{\alpha,\gamma} g = \sum_{\sigma_n > 0} F_\alpha(\sigma_n)^\gamma \sigma_n^{-1} \langle g, v_n \rangle u_n.$$

is called reduced Tikhonov method *or* reduced Landweber method*.*

We remark that for $\gamma = 1$ the operator $T_{\alpha,\gamma}$ is the classical Tikhonov or Landweber method. For $\gamma = 0$ the operator $T_{\alpha,\gamma}$ is the generalized inverse. We prove as one of our main results that for $\gamma \in (1/2, 1)$ the reduced methods define order optimal regularization methods. Propositions and proofs are given separately for the Tikhonov and the Landweber filter. Source conditions on the exact solution f are given in terms of the X_ν-spaces as defined in (2.24).

4.1.1 Reduced Tikhonov Method

Proposition 4.2. *Let $K : X \to Y$ be a compact operator with singular system $(\sigma_n, u_n, v_n)_{n\geq 0}$. Let the exact solution $f^\dagger$ of $Kf = g$ fulfill $\|f^\dagger\|_\nu \leq \rho$ in the space $X_\nu = \operatorname{rg}((K^*K)^{\nu/2})$. Then for $\gamma \in (1/2, 1)$ the reduced Tikhonov method with the parameter choice*

$$\alpha = \kappa \left(\frac{\delta}{\rho}\right)^{1/2(\nu+1)}, \quad \kappa > 0 \text{ constant}$$

is order optimal for all $0 < \nu < \nu^* = 2$.

Proof. The filter of the reduced Tikhonov method is

$$F_\alpha^\gamma(\sigma) = \left(\frac{\sigma^2}{\sigma^2 + \alpha}\right)^\gamma.$$

A regularizing filter is described by the conditions (2.19a)-(2.19c). Order optimality is assured by the conditions (2.27a)-(2.27b). We note that the conditions (2.19b) and (2.19c) are fulfilled (with constant $c = 1$). Condition (2.19a) and (2.27a) demand

$$\sup_n |F_\alpha^\gamma(\sigma_n)\sigma_n^{-1}| = c(\alpha) < \infty$$

and

$$\sup_{0<\sigma\leq\sigma_1} |F_\alpha^\gamma(\sigma)\sigma^{-1}| \leq c\alpha^{-\beta}$$

respectively. Both of them follow from the proof of Proposition 3.9. It is shown there that the function

$$\varphi(\sigma,\gamma,\mu) = \sigma^{-2\mu}\left(\frac{\sigma^2}{\sigma^2+\alpha}\right)^{2\gamma}$$

is bounded if $2\gamma > \mu$. Here $\mu \in (0,1]$ indicates a condition in the space $Y_\mu = \operatorname{rg}((KK^*)^{\mu/2})$ and is not to be mixed up with ν indicating a source condition in X_ν. The bound for φ is given by

$$\varphi(\sigma,\gamma,\mu) \leq \alpha^{-\mu} c_{\gamma,\mu}$$

with $c_{\gamma,\mu} = \left(\frac{2\gamma-\mu}{\mu}\right)^{2\gamma-\mu} \cdot \left(\frac{\mu}{2\gamma}\right)^{2\gamma}$. We remark that

$$\varphi(\sigma,\gamma,\mu) = \sigma^{-2\mu} F_\alpha^{2\gamma}(\sigma).$$

For $\mu = 1$ it is $\sigma^{-1}F_\alpha^\gamma(\sigma) = \sqrt{\varphi(\sigma)} \leq \alpha^{-1/2}\sqrt{c_{\gamma,1}}$. Hence for $\gamma > \mu/2 = 1/2$ it follows from the proof of Proposition 3.9 that the reduced filter fulfills (2.19a) and (2.27a) with $\beta = 1/2$ and constant $c_\gamma = (2\gamma-1)^{(2\gamma-1)}(2\gamma)^{-2\gamma}$. It remains to show that the reduced Tikhonov filter fulfills (2.27b), which postulates

$$\sup_{0<\sigma\leq\sigma_1} |(1-F_\alpha^\gamma(\sigma))\sigma^{\nu^*}| \leq c_{\nu^*}\alpha^{\frac{\nu^*}{2}}.$$

Since for $\sigma > 0$ we have $0 < F_\alpha(\sigma) = \frac{\sigma^2}{\sigma^2+\alpha} \leq 1$ it follows $F_\alpha^\gamma(\sigma) \geq F_\alpha(\sigma)$ for $\gamma > 0$. Hence

$$(1-F_\alpha^\gamma(\sigma))\sigma^{\nu^*} \leq (1-F_\alpha(\sigma))\sigma^{\nu^*}$$

and condition (2.27b) follows from the optimality of the classical Tikhonov filter with $\nu^* = 2$. □

In the proof we use conditions (2.27a) and (2.27b) on a filter function which assure order optimality of the induced regularization method. By means of these conditions we describe some effects of filter reduction. Condition (2.27a) is

$$\sup_{0<\sigma\leq\sigma_1} |F_\alpha^\gamma(\sigma)\sigma^{-1}| \leq c\alpha^{-\beta}.$$

It assures that the considered filter function is able to compensate the growth of σ^{-1}. In Figure 3.3 the influence of the reduction parameter γ on $\sigma^{-1}F_\alpha^\gamma(\sigma)$

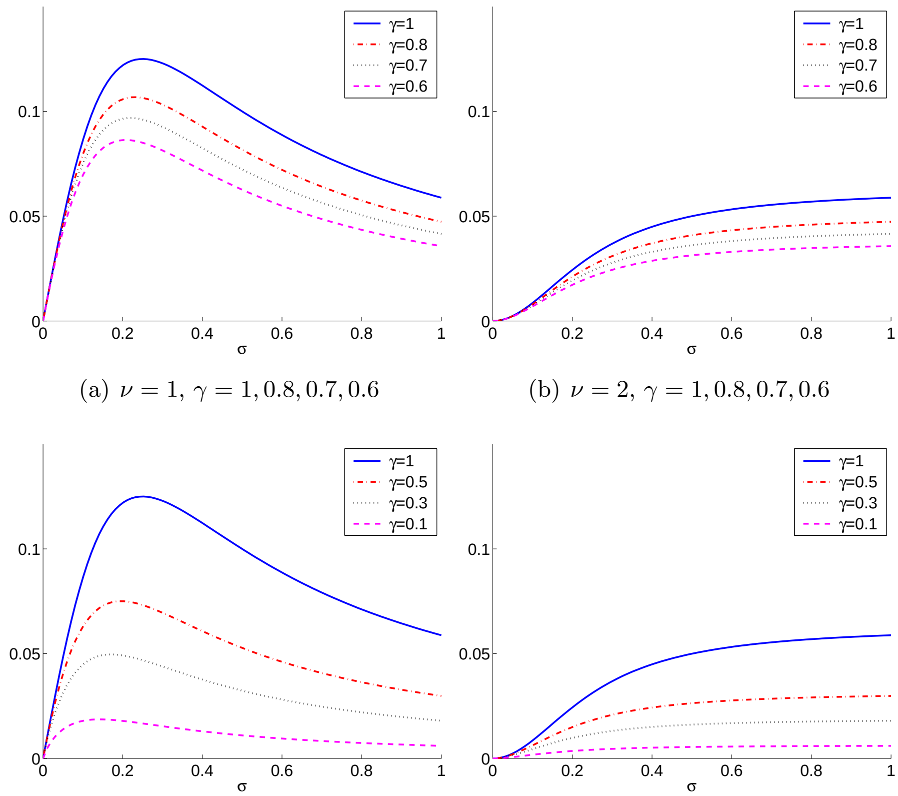
is shown. The supremum also determines the norm of the induced regularization operator. Hence as long as the condition $\gamma > 1/2$ is fulfilled the norm of the induced regularization operator stays bounded. This is not the case for $\gamma \leq 1/2$.
From the proof of Proposition 4.2 we know that $\sigma^{-1} F_\alpha^\gamma(\sigma) \leq \alpha^{-1/2}\sqrt{c_\gamma}$ with $c_\gamma = (2\gamma-1)^{(2\gamma-1)}(2\gamma)^{-2\gamma}$. Proposition 2.18 on the order optimality of filter induced regularization methods describes the influence of c_γ. The minimal bound on the error as well as the size of the corresponding regularization parameter depend on c_γ. Hence we consider the "constant" c_γ as function of the parameter γ. This gives some insight on the effect of filter reduction. The function $c(\gamma)$ of γ decays monotonously. The optimal regularization parameter α_{opt} given in Proposition 2.18 depends on c_γ according to $\alpha_{\text{opt}} \sim c_\gamma^{1/\beta(\nu+1)}$. We suppose that the decrease of the filter exponent γ is compensated by a greater regularization parameter.
We proceed with condition (2.27b) which measures the approximation error caused by using the regularization operator instead of the generalized inverse, i.e., it measures the difference $(1 - F_\alpha^\gamma(\sigma))$ and weights it against σ^{ν^*}. We discuss the influence of filter reduction on this condition. For $\gamma = 0$ no filtering is done and the difference equals zero. For $\gamma = 1$ we have the original Tikhonov filter. Hence reducing the filter by $\gamma \in (0, 1)$ decreases the distance. Figure 4.1 (a) and (b) show the filter error of the classical and reduced method for some parameters $\gamma > 1/2$ and $\nu = 1,\ 2$.
The condition $\gamma > 1/2$ is not necessary to control the approximation error as can be seen in Figure 4.1 (c) and (d). This condition is only imposed to control the growth of σ^{-1}, see Proposition 3.9.
We remark that so far we do not exploit the fact that the approximation error decreases with γ. The estimate used in the proof of Proposition 4.2 is quite coarse since it is the one for the classical Tikhonov filter. The restriction of source conditions to $\nu < \nu^* = 2$ originates in this estimate and might be relaxed for the reduced Tikhonov method.

4.1.2 Reduced Landweber Method

This section is concerned with the reduced Landweber method.
The following technical lemma is used in the proof of Proposition 4.4 to show that the approximation error of the reduced Landweber method in bounded by the approximation error of the classical Landweber method.

Lemma 4.3. *Let $m \in \mathbb{N}_+$ and $\gamma \in [0, 1]$. Then for $-1 \leq x \leq 1$ it is*

$$|1 - (1 - x^m)^\gamma| \leq |x|^m.$$



(a) $\nu = 1,\ \gamma = 1, 0.8, 0.7, 0.6$

(b) $\nu = 2,\ \gamma = 1, 0.8, 0.7, 0.6$

(c) $\nu = 1,\ \gamma = 1, 0.5, 0.3, 0.1$

(d) $\nu = 2,\ \gamma = 1, 0.5, 0.3, 0.1$

Figure 4.1: Approximation error $\sigma^\nu(1 - (\sigma^2/(\sigma^2 + \alpha))^\gamma)$ for classical and reduced Tikhonov filter with $\alpha = 1/16$. For (a) and (b) the condition $\gamma > 1/2$ from Proposition 4.2 is fulfilled whereas this is not the case for (c) and (d). It is shown that the approximation error decays with respect to γ. This is not the case for the data error as it is shown in Figure 3.3.

Proof. Since $|y| \le S$ is equivalent to $y \le S$ and $-y \le S$ we consider two cases.

1. We prove that $1 - (1 - x^m)^\gamma \le |x|^m$.

   This is equivalent to $1 - |x|^m \le (1 - x^m)^\gamma$. Since $-1 \le x \le 1$ this is true for $|x|^m \ne x^m$. For $|x|^m = x^m$ it follows from $0 \le 1 - |x|^m \le 1$ and $\gamma \in [0, 1]$.

2. We prove that $-1 + (1 - x^m)^\gamma \le |x|^m$.

   This is equivalent to $(1 - x^m)^\gamma \le 1 + |x|^m$. For $|x|^m \ne x^m$ it is

$(1 - x^m)^\gamma = (1 + |x|^m)^\gamma \leq 1 + |x|^m$ since $\gamma \leq 1$. For $|x|^m = x^m$ it follows from $|x| \leq 1$ and $\gamma \leq 1$.

□

The following proposition states that for $\gamma > 1/2$ the reduced Landweber method is an order optimal regularization method.

Proposition 4.4. *Let all conditions of Proposition 4.2 be satisfied. Then for* $0 < \beta_{\mathrm{LW}} < \frac{2}{\|K\|^2}$ *and for* $\gamma \in (1/2, 1)$ *the reduced Landweber method is a regularization method. It is order optimal for all* $\nu > 0$ *if the iteration is stopped for*

$$m = \left\lfloor \left(\frac{\nu^2}{\beta_{\mathrm{LW}}}\right)^{(\nu+1)} \left(2\frac{\beta_{\mathrm{LW}}}{\nu}e\right)^{-\nu/(\nu+1)} \left(\frac{\rho}{\delta}\right)^{2/(\nu+1)} \right\rfloor,$$

where $\lfloor x \rfloor$ *denotes the biggest number smaller or equal to* x.

Proof. A regularizing filter is described by conditions (2.19a)-(2.19c). Order optimality is assured by conditions (2.27a)-(2.27b). The filter of the reduced Landweber method is

$$F_m^\gamma(\sigma) = (1 - (1 - \beta_{\mathrm{LW}}\sigma^2)^m)^\gamma.$$

Here the index "LW" indicates that the parameter β_{LW} belongs to the Landweber filter. It is introduced to avoid confusion with the parameter β of the filter condition (2.27a).
For $\sigma \in (0, \sigma_1]$ the restriction $0 < \beta_{\mathrm{LW}} < \frac{2}{\|K\|}$ yields $-1 < 1 - \beta_{\mathrm{LW}}\sigma^2 < 1$ and conditions (2.19b) and (2.19c) follow immediately.
Condition (2.19a) and (2.27a) demand

$$\sup_n |F_\alpha^\gamma(\sigma_n)\sigma_n^{-1}| = c(\alpha) < \infty$$

and

$$\sup_{0<\sigma\leq\sigma_1} |F_\alpha^\gamma(\sigma)\sigma^{-1}| \leq c\alpha^{-\beta}$$

respectively. Both conditions follow from the proof of Proposition 3.12. There it is shown that the function

$$\varphi(\sigma, \gamma, \mu) = \sigma^{-2\mu}[1 - (1 - \beta_{\mathrm{LW}}\sigma^2)^m]^{2\gamma}$$

is bounded if $2\gamma > \mu$. Herein $\mu \in (0,1]$ denotes the scale of spaces $Y_\mu = \mathrm{rg}\,((KK^*)^{\mu/2})$. The bound is given by

$$\varphi(\sigma, \gamma, \mu) \leq \beta_{\mathrm{LW}}^{\mu} m^{\mu}.$$

For $\mu = 1$ we have $\sqrt{\varphi(\sigma)} = \sigma^{-1} F_m^\gamma(\sigma)$. Hence for $\gamma > \mu/2 = 1/2$ we get

$$\sigma^{-1} F_m^\gamma(\sigma) \leq \beta_{\mathrm{LW}}^{1/2} m^{1/2}.$$

In case of the iterative Landweber method the regularization parameter is $\alpha = 1/m$. This yields condition (2.19a). We also have the validity of condition (2.27a) with exponent $\beta = 1/2$ (not to be mixed up with the Landweber parameter β_{LW}). The constant of (2.27a) is $c_\gamma = \beta_{\mathrm{LW}}^{1/2}$.
It remains to show that the reduced Landweber filter fulfills condition (2.27b) which postulates

$$\sup_{0<\sigma\leq\sigma_1} |(1 - F_m^\gamma(\sigma))\sigma^{\nu^*}| \leq c_{\nu^*} m^{\frac{-\nu^*}{2}}.$$

Inserting the formula for the reduced Landweber filter we have to estimate

$$\sup_{0<\sigma\leq\sigma_1} |(1 - [1 - (1 - \beta_{\mathrm{LW}}\sigma^2)^m]^\gamma)\sigma^{\nu^*}|.$$

With $x = 1 - \beta_{\mathrm{LW}}\sigma^2$ it is $|1 - [1 - (1 - \beta_{\mathrm{LW}}\sigma^2)^m]^\gamma| = |1 - (1 - x^m)^\gamma|$. Since $0 < \beta_{\mathrm{LW}} < \frac{2}{\sigma_1^2}$ it follows that $|x| \leq 1$. With Lemma 4.3 we get $|1 - (1 - x^m)^\gamma| \leq |x|^m$ which means that the filter error of the reduced Landweber method is bounded by the filter error of the classical Landweber method:

$$\sup_{0<\sigma\leq\sigma_1} |(1 - [1 - (1 - \beta_{\mathrm{LW}}\sigma^2)^m]^\gamma)\sigma^{\nu^*}| \leq \sup_{0<\sigma\leq\sigma_1} |(1 - \beta_{\mathrm{LW}}\sigma^2)^m||\sigma^{\nu^*}|.$$

Hence condition (2.27b) follows from the optimality of the classical Landweber filter for all $\nu > 0$. The constant c_ν for the classical Landweber filter is given in [Lou89] as

$$c_\nu = e^{-\nu/2}\left(\frac{\nu}{2\beta_{\mathrm{LW}}}\right)^{\nu/2}.$$

□

Analogously to the discussion after the proof of order optimality of the reduced Tikhonov method we use the conditions (2.27a) and (2.27b) to discuss some effects of filter reduction. Condition (2.27a) postulates

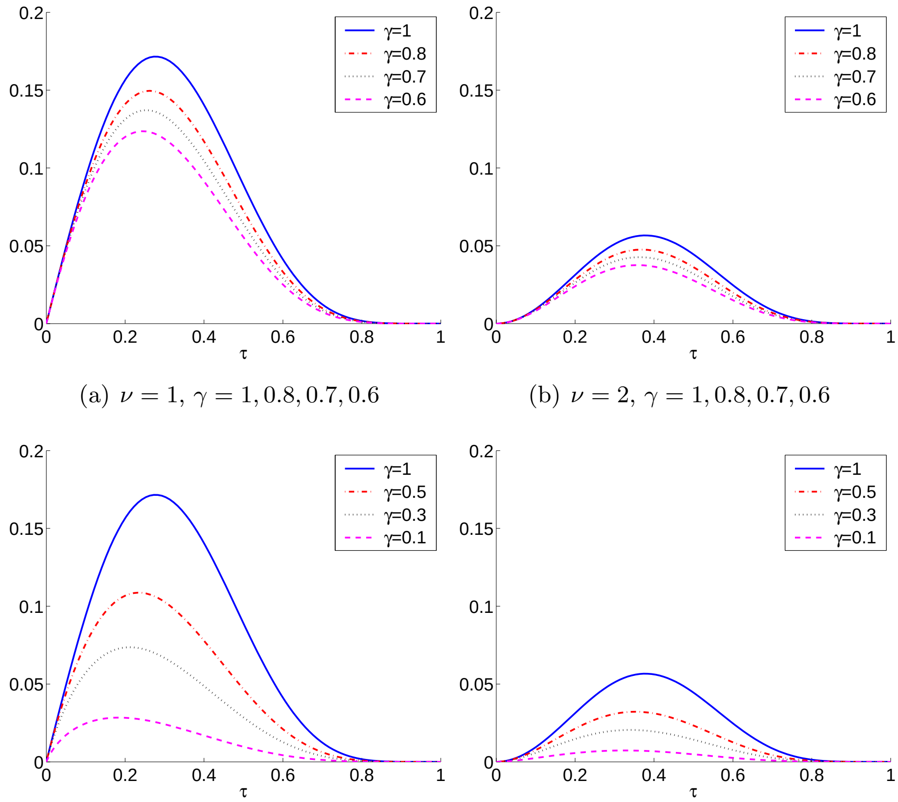
$$\sup_{0<\sigma\leq\sigma_1} |F_\alpha^\gamma(\sigma)\sigma^{-1}| \leq c\alpha^{-\beta}.$$

It assures that the considered filter function is able to compensate the growth of σ^{-1}, moreover the supremum determines the norm of the induced regularization operator. Hence, the condition $\gamma > 1/2$ is necessary to assure the boundedness of the induced regularization operator.

A different situation occurs for condition (2.27b) which measures the approximation error caused by using the regularization operator instead of the generalized inverse: the difference $(1 - F_\alpha^\gamma(\sigma))$ is weighted against σ^{ν^*}. For $\gamma = 0$ no filtering is done and the difference equals zero. For $\gamma = 1$ we have the original Landweber filter. Applying the exponent $\gamma \in (0, 1)$ to the filter function decreases the distance. In other words, less filtering is done or the amount of filtering is *reduced*. Figure 4.2 (a) and (b) show the approximation error of the classical and reduced method for some parameters $\gamma > 1/2$ and $\nu = 1,\ 2$. Figure 4.2 (c) and (d) show the approximation error of the classical and reduced method for some parameters $\gamma \leq 1/2$. The curves stay bounded which illustrates that the condition $\gamma > 1/2$ is not necessary to control the approximation error.

We end the discussion of the reduced Landweber method with some remarks on the stopping index. The stopping index for the reduced Landweber iteration given in Proposition 4.4 is identical to the one for the classical Landweber iteration as taken from [Lou89], see Proposition 2.21. As discussed after the proof of Proposition 4.2 the situation is different for the regularization parameter of the reduced Tikhonov method: One of the constants depends on the parameter γ and influences the size of the regularizing parameter. Actually on a closer look at the calculations in [Lou89] we remark that this is also true for the reduced Landweber method. The bound on the data error $\sigma^{-1}F_m(\sigma)$ as given in [Lou89] is greater than the one which follows from the proof of Proposition 3.12. We remind the reader of estimate (3.24): For $\mu = 1$ it yields $\phi_{\text{critical}}(\tau_*) = \frac{2\gamma-1}{2\gamma m-1}^{2\gamma-1} m^{2\gamma}$. Inserting $\gamma = 1$ we obtain for the classical Landweber method $\phi_{\text{critical}}(\tau_*) = \frac{1}{2m-1}m^2$. For $m \geq 2$ we have $\frac{1}{2m-1}m^2 \leq m/2$. Following the lines of the proof of Proposition 4.4 this results in the estimate $\sigma^{-1}F_m(\sigma) \leq \sqrt{\beta_{\text{LW}}/2}\sqrt{m}$. This bound is by the factor $1/\sqrt{2}$ smaller as the constant given in [Lou89].

The same procedure but for $\gamma = 1/2$ yields $\phi_{\text{critical}}(\tau_*) = m$. Hence the "constant" considered as function of γ seems again to decay monotonously. For the reduced Landweber iteration this results in dependency of the stopping index on the parameter γ, see the parameter choice rule of Proposition 2.18. We suppose that the influence of γ is as follows: When less filtering is performed the iteration has to be stopped earlier to avoid a blow up due to the influence of noise.



(a) $\nu = 1$, $\gamma = 1, 0.8, 0.7, 0.6$

(b) $\nu = 2$, $\gamma = 1, 0.8, 0.7, 0.6$

(c) $\nu = 1$, $\gamma = 1, 0.5, 0.3, 0.1$

(d) $\nu = 2$, $\gamma = 1, 0.5, 0.3, 0.1$

Figure 4.2: Approximation error $\tau^\nu(1 - (1 - (1 - \tau^2)^m)^\gamma)$ for the classical and the reduced Landweber filter with $m = 6$. For (a) and (b) the condition $\gamma > 1/2$ from Proposition 4.2 is fulfilled whereas this is not the case for (c) and (d). In both cases, $\gamma \in [0, 1)$, the approximation error decays with respect to γ, opposite to the data error as shown in Figure 3.5.

4.1.3 Numerical Realization

The use of the reduced methods is of particular interest for problems where classical methods generally oversmoothes the solutions. In this case sharp or fine features of the solution are lost. This is particularly troublesome in imaging or tomographic applications, where it is of high priority to recover sharp or fine features of the solutions [Lam01].

Unfortunately the numerical implementation of the reduced methods is not as straightforward as for the classical ones. This section presents some ideas on the computation of $g_{\alpha,\gamma} = \sum_{\sigma_n > 0} F_\alpha^\gamma(\sigma_n)\sigma_n^{-1}\langle g, v_n\rangle u_n$ according

to Definition 4.1. We start with the formulation of the reduced Tikhonov method using the operator and its adjoint. This formulation involves roots of operators which have to be computed. For this we have in mind an approximation obtained from a cut off of a Taylor expansion.
For the classical Tikhonov method the solution f_α can be computed as solution of the regularized normal equation, see (2.22),

$$(K^*K + \alpha I) f_\alpha = K^* g.$$

Also the classical Landweber method can be computed using the operator and its adjoint. The mth iterate f^m can be represented as

$$f^m = \sum_{j=0}^{m-1} (I - \beta K^*K)^j K^* g, \tag{4.1}$$

see [Lou89]. In the following we concentrate on the reduced Tikhonov method.

Proposition 4.5. *The solution $f_{\alpha,\gamma}$ of the reduced Tikhonov method as given in Definiton 4.1 can be computed according to*

$$(K^*K + \alpha I)^\gamma (K^*K)^{1-\gamma} f_{\alpha,\gamma} = K^* g. \tag{4.2}$$

Proof. The reduced Tikhonov method with parameter γ defines a regularized solution $f_{\alpha,\gamma}$ as

$$f_{\alpha,\gamma} = \sum_{n>0} \sigma_n^{-1} \left(\frac{\sigma_n^2}{\sigma_n^2 + \alpha} \right)^\gamma \langle g, v_n \rangle u_n.$$

We rewrite (4.2) according to

$$f_{\alpha,\gamma} = (K^*K)^{\gamma-1} (K^*K + \alpha I)^{-\gamma} K^* g.$$

For the Fourier coefficient $\langle f_{\alpha,\gamma}, u_n \rangle$ we get

$$\begin{aligned}
\langle f_{\alpha,\gamma}, u_n \rangle &= \langle (K^*K)^{\gamma-1} (K^*K + \alpha I)^{-\gamma} K^* g, u_n \rangle \\
&= \langle g, K (K^*K + \alpha I)^{-\gamma} (K^*K)^{\gamma-1} u_n \rangle \\
&= \langle g, \sigma_n^{2(\gamma-1)} (\sigma_n^2 + \alpha)^{-\gamma} K u_n \rangle \\
&= \sigma_n^{2\gamma} \sigma_n^{-2} (\sigma_n^2 + \alpha)^{-\gamma} \sigma_n \langle g, v_n \rangle \\
&= \sigma_n^{-1} \left(\frac{\sigma_n^2}{\sigma_n^2 + \alpha} \right)^\gamma \langle g, v_n \rangle.
\end{aligned}$$

The assertion follows with $f_{\alpha,\gamma} = \sum_n \langle f_{\alpha,\gamma}, u_n \rangle u_n$. $\square$

The representation (4.2) involves roots of the operators $K^*K+\alpha I$ and K^*K. In order to compute these roots we use the Taylor expansion of a suitable function h

$$h(x+\alpha) = h(x) + h'(x)\alpha + \frac{1}{2}h''(x)\alpha^2 + \dots . \tag{4.3}$$

We set $h(x) = x^\gamma$ and $x = I + \alpha K^*K$ and obtain the following Taylor expansion for $(I + \tilde{\alpha}K^*K)^\gamma$,

$$\begin{aligned}
(I+\alpha K^*K)^\gamma &= h(I) + h'(I)\alpha K^*K + \frac{1}{2}h''(I)(\alpha K^*K)^2 + \dots \\
&= I + \gamma\alpha K^*K + \frac{\gamma(\gamma-1)}{2}\alpha^2(K^*K)^2 \\
&\qquad + \frac{\gamma(\gamma-1)(\gamma-2)}{3!}\alpha^3(K^*K)^3 + \dots .
\end{aligned}$$

With the coefficients $c_0 = 1$ and

$$c_j = \alpha^j \frac{1}{j\,!}\prod_{m=0}^{j-1}(\gamma - m)$$

for $j > 0$ the expansion reads as

$$(I+\alpha K^*K)^\gamma = \sum_{j\geq 0} c_j (K^*K)^j. \tag{4.4}$$

We obtain a computable approximation for $(I + \tilde{\alpha}K^*K)^\gamma$ by taking a finite sum of the series expansion. Estimates for the remainder can be used to assess the quality of the approximation (for that it might be necessary to use some rescaling, e.g., to assure $\|K^*K\| < 1$).
For the second factor $(K^*K)^{1-\gamma}$ in (4.2) we consider $(K^*K)^{1-\gamma} = (I + (K^*K - I))^{1-\gamma}$ and use again (4.3). This yields

$$\begin{aligned}
(K^*K)^{1-\gamma} &= (I + (K^*K-I))^{1-\gamma} \\
&= h(I) + h'(I)(K^*K - I) + \dots \\
&= I + (1-\gamma)(K^*K - I) + \frac{(1-\gamma)(-\gamma)}{2}(K^*K-I)^2 \\
&\quad + \frac{(1-\gamma)(-\gamma)(-\gamma-1)}{3!}(K^*K-I)^3 + \dots .
\end{aligned}$$

With $d_0 = 1$ and

$$d_j = \frac{1}{j\,!}\prod_{m=2-j}^{1}(-\gamma + m)$$

for $j > 0$ this yields the series representation

$$(K^*K)^{1-\gamma} = \sum_{j\geq 0} d_j (K^*K - I)^j. \tag{4.5}$$

A computable approximation for $(K^*K)^{1-\gamma}$ is given as finite sum of the series expansion. Estimates for the remainder can be used to assess the quality of the approximation (for that it might be necessary to use some rescaling, e.g., to assure $\|K^*K - I\| < 1$).
An approximation for $(K^*K + \alpha I)^\gamma (K^*K)^{1-\gamma}$ can be obtained as follows. First the series expansion (4.5) is transformed into a series in $(K^*K)^j$. Then the series expansions (4.4) and (4.5) are multiplied. The last step requires more sophisticated studies of convergence conditions and computations of the according coefficients. This is not persecuted any further in this work. But we note the approximations A_N of $(K^*K + \alpha I)^\gamma (K^*K)^{1-\gamma}$ obtained by using the first $N = 1, 2, 3$ terms of each series in (4.4) and (4.5).
For $N = 1$ we get

$$A_1 = I.$$

For $N = 2$ we get

$$\begin{aligned} A_2 &= (I + \alpha\gamma K^*K) \cdot (\gamma I + (1-\gamma) K^*K) \\ &= \gamma I + (\alpha\gamma^2 - \gamma + 1) K^*K + \alpha\gamma(1-\gamma)(K^*K)^2. \end{aligned}$$

For $N = 3$ we get

$$\begin{aligned} 4 \cdot A_3 &= \Big(2I + 2\alpha\gamma K^*K + \gamma(\gamma-1)\alpha^2 (K^*K)^2\Big) \\ &\quad \cdot \Big(\gamma(1+\gamma) I + 2(1-\gamma)(\gamma+1) K^*K - \gamma(1-\gamma)(K^*K)^2\Big) \\ &= 2\gamma(1+\gamma) I + [2\alpha\gamma^2 + (\gamma-1)(\gamma^2-4)] K^*K \\ &\qquad + 2\gamma(1-\gamma)[4\alpha(1+\gamma) - 1](K^*K)^2 \\ &\qquad + 2\alpha\gamma(1-\gamma)[\alpha(\gamma^2-1) - \gamma](K^*K)^3 \\ &\qquad + 2\alpha^2\gamma^2(\gamma-1)^2 (K^*K)^4. \end{aligned}$$

We end the analysis of the reduced methods with some remarks on the question which properties allow a filter reduction.

Discussion of filter reduction

We claim that filter reduction is possible whenever the adjoint operator is involved in the computation of the regularization. This is the case for the

Tikhonov regularization, see (2.22), and for the Landweber regularization, see (4.1).
This conjecture is also motivated by the work of [RT04a] on the smoothing properties of the Sobolev embedding operator. It is shown in Corollary 3.3 that for the Sobolev embedding operator $i : H^s \to L_2$ the adjoint operator smoothes with step size $2s$, i.e., $i^* : L_2 \to H^{2s}$. Hence whenever the adjoint embedding operator is used, there is an additional smoothing amount which might be responsible for oversmoothing.
We return to an operator K which acts continuously invertible between Sobolev scales of stepsize t, i.e., $K : L_2 \leftrightarrow H^t$ but is compact as operator $K : L_2 \to H^r$ with $r < t$. The compact mapping $\tilde{K} : L_2 \to H^r$ can be written as $\tilde{K} = iK$ where only the Sobolev embedding operator $i : H^t \to H^r$ introduces the ill-posedness. But as for $\tilde{K} = iK$ the adjoint of the embedding operator is a part of $\tilde{K}^* = K^* i^*$, this leads to an oversmoothing. We emphasize that this is a conjecture and its motivation. So far we do not have any strict results.

4.2 Regularization and Wavelet Shrinkage

In this section we consider the combination of wavelet shrinkage and classical regularization methods for the solution of linear ill-posed problems. We combine wavelet shrinkage as a method adjusted to smoothness properties of the data and regularization operators which are based on the singular value decomposition of the operator and hence are adjusted to properties of the operator.
We construct a regularized solution f_{reg} for $Kf = g$ from noisy data $g^\delta = Kf + \delta \mathrm{d}W$, see (2.6), in two steps. Firstly we estimate the data by nonlinear wavelet shrinkage S_λ, secondly we apply a regularization operator R_α,

$$f_{\mathrm{reg}} = f^\delta_{\alpha,\lambda} = R_\alpha S_\lambda g^\delta.$$

Since wavelet shrinkage as introduced in Section 3.2 also involves a linear projection we specify the construction of a solution according to

$$f_{\mathrm{reg}} = f^\delta_{\alpha,\lambda,j} := R_\alpha S_\lambda P_j g^\delta.$$

Our approach is related to the one in [CHR04]. The authors examine the combination of wavelet shrinkage and wavelet-Galerkin projection for the solution of linear ill-posed problems. This combination fits in our two-step setting with wavelet shrinkage as data estimation and wavelet-Galerkin projection as reconstruction. The authors have proved that their method achieves optimal convergence results. In Theorem 4.12 we prove that the combination of wavelet shrinkage and classical regularization methods also achieves optimal convergence results. Both methods are related in the data estimation part and hence in the function classes for which the optimal rate is achieved. They differ in the reconstruction part: In [CHR04] the reconstruction is done with the same wavelet basis as used for the shrinkage step whereas in our method the reconstruction is based on the singular system of the operator. The idea behind this is to keep using the singular system since it describes the action of the operator perfectly (remember the spectral decomposition (2.15)). To make this combination possible we need a translation of Sobolev smoothness and source conditions. This translation is accomplished with the unifying framework of Hilbert scales, see Lemma 4.6. Though we do use wavelet shrinkage as it is done in [CHR04] we use it with an additional degree of freedom. In [CHR04] an operator is considered which smoothes with stepsize t in Sobolev and Besov scales, i.e., for all $s \geq 0$

$$\begin{aligned} K &: H^s \to H^{s+t} \\ K &: B^s_{p,q} \to B^{s+t}_{p,q}. \end{aligned} \tag{4.6}$$

In addition K as given in (4.6) is assumed to be continuously invertible. Accuracy of the reconstructed solution is measured in L_2. Wavelet shrinkage is done with a basis of Sobolev smoothness t. Therefore the data estimate in [CHR04] is an element of H^t which is also the expected smoothness of the operator K applied to some function $f \in L_2$, and hence the inverse operator of K could be applied directly.

We perform wavelet shrinkage with a wavelet basis of Sobolev smoothness H^η with $\eta \in [0, t]$. The regularization operator R_α is considered as operator $R_\alpha : H^\eta \to L_2$. We interpret the parameter η as a weighting parameter for the two steps of data estimation and reconstruction.

The main result is given in Theorem 4.12. We show that for a proper parameter choice the optimal convergence rate (2.32) is achieved.

In classical regularization theory the error of the regularized solution $f_{\text{reg}} = R_\alpha g^\delta$ is split into the data error and the regularization error

$$\|f_{\text{reg}} - f\| \leq \|R_\alpha(g^\delta - g)\| + \|R_\alpha g - f\|.$$

In our method the additional step of data estimation is reflected in a third error term. The regularization error is kept unchanged while the data error is split again into the skrinkage error and the approximation error due to the projection of the data onto a wavelet subspace,

$$\|f_{\text{reg}} - f\| \leq \|R_\alpha(S_\lambda P_j g^\delta - P_j g)\| + \|R_\alpha(P_j g - g)\| + \|R_\alpha g - f\|. \quad (4.7)$$

The shrinkage error and the approximation error are estimated according to Lemma 3.4 and Lemma 3.8. Both error terms depend on the smoothness η of the underlying wavelet basis. The resulting smoothness of the data estimator $S_\lambda P_j g^\delta$ is reflected in the norm of the regularization operator, see Lemma 4.8.

4.2.1 Auxiliary Results

We start with the translation of source condition in ν-scales and smoothness conditions in Sobolev spaces as introduced in (2.30). We assume that we deal with a linear compact operator $K : L_2 \to L_2$ with smoothing property $t > 0$ expressed by

$$\|Kf\|_{L_2} \simeq \|f\|_{H^{-t}}. \quad (4.8)$$

In (2.10) we introduced this property as basic assumption in Hilbert scales.

Lemma 4.6. *Let $K : L_2 \to L_2$ be a compact linear operator with smoothing property $t > 0$ in Sobolev spaces, see* (4.8)*. Let $s \leq t$. Then for a function $f \in L_2$ the following equivalence is valid,*

$$f \in H^s \Leftrightarrow f \in \operatorname{rg}((K^*K)^{\nu/2}) \quad \textit{for} \quad \nu = s/t.$$

Proof. We consider the Hilbert scale $(H_\tau)_{\tau\in\mathbb{R}}$ of Sobolev spaces as introduced in Definition 2.3 and Remark 2.4. For $\omega \in H_0 = L_2$ we use the Hilbert scale norm equivalence, see [Nat84],

$$\|(K^*K)^{\nu/2}\omega\|_0 \sim \|\omega\|_{-\nu t} \quad \text{for} \quad |\nu| \leq 1. \quad (4.9)$$

Let f be in $\operatorname{rg}((K^*K)^{\nu/2})$. Then an $\omega \in L_2$ exists with $f = (K^*K)^{\nu/2}\omega$ and $\|\omega\| \leq \rho$ for some constant ρ. With $\nu = s/t$, $s \leq t$ it is

$$\|f\|_{H^s} = \|f\|_s = \|f\|_{\nu t} \sim \|(K^*K)^{-\nu/2}f\|_0 = \|\omega\|_0 \leq \rho$$

and therefore $f \in H^s$.
For $f \in H^s$ it follows analogously that $\omega := (K^*K)^{-\nu/2}f$ exists for $\nu = s/t$, $s \leq t$. Hence f is an element of $\operatorname{rg}((K^*K)^{\nu/2})$. □

Lemma 4.6 connects source conditions and Sobolev smoothness. For an operator of smoothing property t only Sobolev smoothness $s \leq t$ of the exact solution can be considered. Of course, the solution could be of arbitrary smoothness, but only for smoothness $s \leq t$ a corresponding ν can be found. This results in $\nu \leq 1$ which means that only a small range of source condition can be considered. A relaxation of this condition might be possible when considering special classes of compact operators. We illustrate some consequences of Lemma 4.6 for the class of linear compact operators for which the smoothing property $t > 0$ in Sobolev scales is expressed as property of the singular values, namely $\sigma_n \sim n^{-t}$. In terms of classification of ill-posed problems as given in [Lou89] this corresponds to ill-posedness of order t. For the singular functions u_n it follows from Lemma 4.6 that $\|u_n\|_{H^s} = \|u_n\|_{\nu t} \sim \|(K^*K)^{-\nu/2}u_n\|_0 = \sigma_n^{-\nu}\|u_n\|_0$. With $\sigma_n \sim n^{-t}$ and $\nu = s/t$ we get $\|u_n\|_{H^s} \sim n^s$. In Section 5.1 we calculate these properties of the singular system explicitly for some example operators. We suppose that from these properties of the singular functions we can obtain a norm equivalence similar to (4.9). This is however not tackled further in this work.

In the following we consider a regularization method induced by an optimal filter $F_\alpha(\sigma)$, i.e., a filter which fulfills conditions (2.19) and (2.27). The notation $a \lesssim b$ is used for $a \leq c \cdot b$ with a constant $c \geq 0$ or equivalently for $a = \mathcal{O}(b)$. We estimate the three error terms of (4.7). For the regularization error we get the following result.

Lemma 4.7. *Let $K : L_2(\Omega) \to L_2(\Omega)$ be a compact operator with smoothing property $t > 0$. Let R_α be a regularization method induced by a filter $F_\alpha(\sigma)$ with* (2.27b),

$$\sup_{\sigma<0\leq\sigma_1} |(1 - F_\alpha(\sigma))\sigma^{\nu^*}| \leq c_{\nu^*}\alpha^{\beta\nu^*}$$

Let $s \leq t$ and $f \in H^s$ with $\|f\|_{H^s} \leq \rho$ and $g = Kf$. The regularization error is then given by

$$\|R_\alpha g - f\|_{L_2} \lesssim c_{s/t}\alpha^{\beta\frac{s}{t}}\rho.$$

Proof. Estimate (2.27b) is valid for all $\nu \in [0, \nu^*]$ if we replace ν^* by ν, see [Lou89]. We translate the Sobolev smoothness information on f into a source condition as described in Lemma 4.6. For $s \leq t$ we get $f \in \mathrm{rg}((K^*K)^{\nu/2})$

with $\nu = s/t$. It is

$$\begin{aligned}
\|R_\alpha g - f\|_{L_2}^2 &= \sum_{\sigma_n>0} (1 - F_\alpha(\sigma))^2 \left|\langle f, u_n\rangle\right|^2 \sigma_n^{-2s/t}\sigma_n^{2s/t} \\
&\leq \sup_{\sigma\geq 0}(1 - F_\alpha(\sigma))^2\sigma^{2s/t} \underbrace{\sum_{\sigma_n>0} \sigma_n^{-2s/t} \left|\langle f, u_n\rangle\right|^2}_{\simeq\|f\|_{H^s}^2} \\
&\lesssim c_{s/t}^2\alpha^{2\beta s/t}\rho^2.
\end{aligned}$$

□

Inserting the values for β and ν^* immediately yields the results for specific methods like the (reduced) Tikhonov method, the (reduced) Landweber method or the TSVD.
As a result of data preprocessing with a wavelet basis of smoothness η, the regularization operator R_α is defined as

$$R_\alpha : H^\eta \to L_2$$

instead of

$$R_\alpha : L_2 \to L_2.$$

The effect on the norm of the regularization operator is given by the following lemma. The smoothness of the data estimate and hence the fact that R_α is defined as discussed above, is taken into account by property (3.29)

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu} F_\alpha^\gamma(\sigma)| \leq c\alpha^{-\beta\nu} \quad \text{for } \nu \in (0, 1].$$

Lemma 4.8. *Let $K : L_2(\Omega) \to L_2(\Omega)$ be a compact operator with smoothing property $t > 0$ and $0 \leq \eta < t$. Let R_α be a regularization method induced by a filter $F_\alpha(\sigma)$ with* (3.29)*. The norm of the operator $R_\alpha : H^\eta \to L_2$ fulfills*

$$\|R_\alpha\|_{H^\eta\to L_2} \lesssim \alpha^{\beta\frac{\eta-t}{t}}. \tag{4.10}$$

Proof. According to Lemma 4.6 we translate the Sobolev smoothness $h \in H^\eta$ into a source condition in Y_ν-scales and get $h \in \mathrm{rg}((KK^*)^{\eta/2t})$. For the norm of $R_\alpha : H^\eta \to L_2$ we compute

$$\begin{aligned}
\|R_\alpha\|^2_{H^\eta \to L_2} &= \sup_{\|h\|_{H^\eta}=1} \sum_{\sigma_n>0} \sigma_n^{-2} F_\alpha^2(\sigma_n) \left|\langle h, v_n\rangle_{L_2}\right|^2 \sigma_n^{2\eta/t} \sigma_n^{-2\eta/t} \\
&\leq \sup_{\|h\|_{H^\eta}=1} \sup_{0<\sigma\leq\sigma_1} \sigma^{2(\frac{\eta}{t}-1)} F_\alpha^2(\sigma) \underbrace{\sum_{n>0} \sigma_n^{-2\eta/t} \left|\langle h, v_n\rangle_{L_2}\right|^2}_{\simeq \|h\|^2_{H^\eta}=1} \\
&\lesssim \sup_{0<\sigma\leq\sigma_1} \sigma^{2(\frac{\eta}{t}-1)} F_\alpha^2(\sigma).
\end{aligned}$$

Since $0 \leq \eta < t$ we have $-1 \leq \frac{\eta}{t} - 1 < 0$. With (3.29) we get

$$\sup_{0<\sigma\leq\sigma_1} \sigma^{2(\frac{\eta}{t}-1)} F_\alpha^2(\sigma) \lesssim \alpha^{2\beta(\frac{\eta}{t}-1)}$$

and the assertion is proved. □

Corollary 4.9. *Lemma 4.8 is valid with $\beta = 1/2$ for the (reduced) Tikhonov method and the (reduced) Landweber method. For the truncated singular value decomposition it is valid with $\beta = 1$.*

Proof. The (reduced) Tikhonov method fulfills condition (3.29) with $\beta = 1/2$, see (3.18) of Proposition 3.9. The (reduced) Landweber method fulfills condition (3.29) with $\beta = 1/2$, see (3.27) of Proposition 3.12. The truncated singular value decomposition fulfills condition (3.29) with $\beta = 1$, see Lemma 3.14. □

We want to dicuss Lemma 4.8 for the special cases $\eta = 0$ and $\eta = t$. For $\eta = 0$ we have $R_\alpha : L_2 \to L_2$ and equation (4.10) becomes the classical result on regularization operators as given in [EHN96] or [Lou89]. For $\eta = t$ we have $R_\alpha : H^t \to L_2$ and equation (4.10) becomes $\|R_\alpha\|_{H^t \to L_2} \lesssim$ const. In this case the data estimate given by $S_\lambda P_j$ already has the Sobolev smoothness which a L_2-function would get from the application of the operator K. Hence the generalized inverse could be applied directly to the data estimate $S_\lambda P_j g^\delta$. For the latter case we refer again to the article [JL01] and its discussion in Section 3.3.
We now turn to the second error term of (4.7). The effect of a regularization operator on the approximation error due to projection on a wavelet subspace is described in the following lemma.

Lemma 4.10. *Let $K : L_2(\Omega) \to L_2(\Omega)$ be a compact operator with smoothing property $t > 0$. Let $s \geq 0$, $s + t \in \mathbb{N}$, $s + t < m$ and g be in the Sobolev* *ball* $B := \{g \in H^{s+t}(\Omega) : \|g\|_{H^{s+t}} \leq \rho\}$. *Let* $\eta \in \mathbb{N}$ *with* $0 \leq \eta < s+t$ *and* $\varphi \in H^{\eta}(\Omega)$ *induce a wavelet basis with degree* $m-1$ *of polynomial reproduction.* P_j *denotes the projection operator* (3.5). *Let* R_α *be a regularization operator induced by an order optimal filter with* (3.29). *Then for* $\alpha > 0$ *the following estimate holds*

$$\|R_\alpha(P_j g - g)\|_{L_2}^2 \lesssim \alpha^{2\beta\frac{\eta-t}{t}} 2^{-2j(s+t-\eta)}.$$

Proof. The proof follows directly from Lemma 4.8 and Lemma 3.4. □

The last step, before we present the main result on regularization combined with wavelet shrinkage, is to give an estimate of the remaining first error term in (4.7). The effect of regularization on the error of the data estimator $S_\lambda P_j g^\delta$ is given in the next lemma.

Lemma 4.11. *Let all conditions of Lemma 3.8 hold with* $\theta = s+t$. *For the regularization operator* R_α *induced by an order optimal filter with* (3.29) *and parameter* $\alpha > 0$ *we get*

$$\mathrm{E}(\|R_\alpha(S_\lambda P_j g^\delta - P_j g)\|_{L_2}^2) \lesssim \alpha^{2\beta\frac{\eta-t}{t}} (\delta\sqrt{|\log\delta|})^{\frac{4(s+t-\eta)}{2s+2t+d}}$$

for the projection level $j \leq j_1$ *with* $j_1 = j_1(\delta)$ *defined by* $2^{j_1} \simeq \frac{1}{\delta^2|\log\delta|}$.

Proof. The proof follows directly from Lemma 4.8 and Lemma 3.8. □

We are now ready to state and prove the optimal convergence rate for the combination of wavelet shrinkage and regularization methods.

4.2.2 Optimal Convergence Rate

After the auxiliary results given in the last section we present Theorem 4.12 as one of our main results. The two-step method consisting of wavelet shrinkage and a regularization operator,

$$f_{\mathrm{reg}} = f_{\alpha,\lambda,j}^{\delta} = R_\alpha S_{\lambda,j} g^\delta$$

achieves – up to a logarithmic factor – the optimal rate given in (2.14) for stochastic noise. Parameter choice rules are given for the wavelet detail level in (4.11) and the regularizing parameter in (4.12). The existence of a detail level is discussed subsequent to the proof of Theorem 4.12.

Theorem 4.12. *Let $K : L_2(\Omega) \to L_2(\Omega)$ be a linear compact operator with smoothing property of order $t > 0$ with respect to any Besov space $B^\tau_{p,p}(\mathbb{R}^d)$ and any Sobolev space $H^\tau(\mathbb{R}^d)$, see* (4.6)*. Let $s \geq 0$, $s + t > 1/2$, $s \leq t$ and η with $0 \leq \eta \leq t$ be given. We assume that f belongs to $H^s(\mathbb{R}^d) \cap B^s_{p,p}(\mathbb{R}^d)$ with $\frac{1}{p} = \frac{1}{2} \cdot \frac{2t+d}{2\eta+d} + \frac{s}{2\eta+d}$.*
Let $\varphi \in H^\eta$ define an orthonormal wavelet basis of $L_2(\mathbb{R}^d)$ such that the degree m of polynomial reproduction in V_j satisfies $m+1 > t$. Let P_j be the L_2-projection on V_j and let S_λ denote the shrinkage operator with threshold $\lambda = C\delta\sqrt{|\log\delta|}$, see (3.9)*. Let R_α denote an order optimal regularization method induced by a filter with* (3.29)*. We choose*

i) the projection level j with

$$2^{-j} \leq (\delta\sqrt{|\log\delta|})^{\frac{1}{\eta+d/2}}, \tag{4.11}$$

ii) the regularization parameter α for $\eta < t$ as

$$\alpha \simeq (\delta\sqrt{|\log\delta|})^{\frac{1}{\beta}\frac{2t}{2s+2t+d}} \tag{4.12}$$

with the constant β from condition (3.29)*.*
For $\eta = t$ we choose $\alpha \simeq (\delta\sqrt{|\log\delta|})^{2/\beta}$.

Then the two-step method $R_\alpha S_{\lambda j}$ consisting of wavelet shrinkage and regularization achieves, up to a logarithmic factor, the optimal rate (2.14)*, i.e.,*

$$\mathrm{E}(\|f^\delta_{\alpha\lambda j} - f\|^2_{L_2}) = \mathcal{O}((\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}). \tag{4.13}$$

Proof. We first split the error term into three parts: one each for the effects of smoothing, projection and regularization:

$$\begin{aligned} \mathrm{E}(\|f^\delta_{\alpha\lambda j} - f\|^2_{L_2}) &= \mathrm{E}(\|R_\alpha S_\lambda P_j g^\delta - f\|^2_{L_2}) \\ &\lesssim \mathrm{E}(\|R_\alpha(S_\lambda P_j g^\delta - P_j g)\|^2_{L_2}) \\ &\qquad + \|R_\alpha(P_j g - g)\|^2_{L_2} \; + \; \|R_\alpha g - f\|^2_{L_2}. \end{aligned} \tag{4.14}$$

The first term is estimated by Lemma 4.11,

$$\mathrm{E}(\|R_\alpha(S_\lambda P_j g^\delta - P_j g)\|^2_{L_2}) \lesssim \alpha^{2\beta\frac{\eta-t}{t}}(\delta\sqrt{|\log\delta|})^{\frac{4(s+t-\eta)}{2s+2t+d}}, \tag{4.15}$$

with $j \leq j_1$, j_1 given by (3.10).

The second term of (4.14) is estimated by Lemma 4.10,

$$\|R_\alpha(P_j g - g)\|_{L_2}^2 \lesssim \alpha^{2\beta\frac{\eta-t}{t}} 2^{-2j(s+t-\eta)}. \tag{4.16}$$

The parameter j determines the detail level of the wavelet decomposition and is chosen independent of the smoothness s as given in equation (4.11). We remark that the choice of the detail level does not conflict with the upper bound given in (3.10) and that for sufficiently small δ an integer level j can be chosen. A detailed discussion of the level choice is given subsequent to the proof. For $\delta\sqrt{|\log\delta|} \leq 1$ we have

$$2^{-j} \lesssim (\delta\sqrt{|\log\delta|})^{\frac{1}{\eta+d/2}} \leq (\delta\sqrt{|\log\delta|})^{\frac{2}{2s+2t+d}}. \tag{4.17}$$

Thus for $\eta < t$ the smoothing error (4.15) and the approximation error (4.16) are estimated as

$$\mathrm{E}(\|R_\alpha(S_\lambda P_j g^\delta - P_j g)\|_{L_2}^2) + \|R_\alpha(P_j g - g)\|_{L_2}^2 \lesssim \alpha^{2\beta\frac{\eta-t}{t}} (\delta\sqrt{|\log\delta|})^{\frac{4(s+t-\eta)}{2s+2t+d}}.$$

For $\eta = t$ the norm of $R_\alpha : H^t \to L_2$ is constant and it is

$$\mathrm{E}(\|R_\alpha(S_\lambda P_j g^\delta - P_j g)\|_{L_2}^2) + \|R_\alpha(P_j g - g)\|_{L_2}^2 \lesssim (\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}.$$

It remains to consider the effect of regularization which is given by the third term of (4.14). With Lemma 4.7 we get

$$\|R_\alpha g - f\|_{L_2}^2 \lesssim \alpha^{2\beta\frac{s}{t}}. \tag{4.18}$$

For $\eta < t$ we use the parameter choice rule (4.12). Assembling all three error bounds we get

$$\mathrm{E}(\|R_\alpha S_\lambda P_j g^\delta - f\|_{L_2}^2) \lesssim (\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}.$$

For $\eta = t$ we insert $\alpha \simeq (\delta\sqrt{|\log\delta|})^{2/\beta}$ and get for the regularization error

$$\|R_\alpha g - f\|_{L_2}^2 \lesssim \alpha^{2\beta\frac{s}{t}} \simeq (\delta\sqrt{|\log\delta|})^{\frac{4s}{t}} \leq (\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}.$$

Thus assertion (4.13) is proved. □

We make a few remarks on Theorem 4.12. We start with the *quasi-optimal* rate (4.13), where *quasi-optimal* is due to the logarithmic factor. For any $\kappa > 0$ and $\delta \to 0$ we have with an $\varepsilon > 0$ the inequality

$$(\delta\sqrt{|\log\delta|})^{\kappa} \lesssim \delta^{\kappa-\varepsilon}.$$

Hence the logarithmic factor can be neglected by means of a (very) small decrease in the exponent.
We continue with the parameter choice rule (4.12). For $\eta = t$ we can choose the regularizing parameter independent of s and t. This corresponds to the remarks made subsequent to Lemma 4.8. Smoothing in H^t results in $\|R_\alpha\|_{H^t \to L_2} \leq c$. We remind the reader of [CHR04] where a combination of wavelet shrinkage and wavelet-Galerkin projection with wavelets in H^t is examined. For $\eta = t$ the rule for choosing the detail level, (4.11), nearly coincides with the one from [CHR04]. The only difference is the term $d/2$ in the denominator. This condition is imposed to assure the existence of a detail level, see the following discussion.
The result of Theorem 4.12 can be generalized to operators which do not define optimal regularization methods as defined by (2.19) and (2.27). Examples for such methods are the reduced Tikhonov or Landweber method with $\gamma \leq 1/2$. For $\gamma \leq 1/2$ the norm of the operator as a mapping from L_2 into L_2 is no longer bounded. But depending on the size of γ – or on the size of η – the induced operator is bounded from H^η into L_2. Hence, if the underlying wavelet basis complies with a smoothness condition, the combination of wavelet shrinkage and the reduced methods with $\gamma \leq 1/2$ achieves the convergence rate (4.13), see Corollary 4.14 in Section 4.2.3.

Existence of detail level

In Theorem 4.12, (4.12), a rule for the detail level is given as

$$2^{-j} \leq (\delta \sqrt{|\log \delta|})^{\frac{1}{\eta + d/2}}.$$

From the estimate of the shrinkage error there is an upper bound for the detail level: Equation (3.10) of Lemma 3.8 limits the detail level j from above by j_1 with

$$2^{-j_1} \simeq \delta^2 |\log \delta|$$

or

$$j_1 \simeq -\log_2(-\delta^2 \log \delta).$$

In the two-step regularization method consisting of wavelet shrinkage and classical regularization the additional step of data estimation results in two error terms: one each for the shrinkage error and the approximation error. The approximation error gets smaller the more levels are used, see Lemma 3.4. This is opposed to the upper bound from the shrinkage estimate. For $\delta\sqrt{|\log \delta|} \leq 1$ we know from equation (4.17) that

$$2^{-j} \lesssim (\delta \sqrt{|\log \delta|})^{\frac{1}{\eta + d/2}} \leq (\delta \sqrt{|\log \delta|})^{\frac{2}{2s + 2t + d}}.$$

The right hand side of this inequality gives a lower bound j_0 for the number of used detail levels,

$$2^{-j_0} \leq (\delta^2 |\log \delta|)^{\frac{1}{2s+2t+d}}$$

or

$$j_0 = \frac{1}{2s+2t+d} \cdot (-\log_2(-\delta^2 \log \delta)) = \frac{1}{2s+2t+d} j_1.$$

Since $\frac{1}{2s+2t+d} \leq 1$ we have at least that $j_0 \lesssim j_1$. Therefore all detail levels in $[j_0, j_1]$ are admissible. We postpone the question whether there is an admissible integer detail level, i.e., $[j_0, j_1] \cap \mathbb{N} \neq \emptyset$, to Lemma 4.13. We discuss three possible choices for the detail level.

1. Choose the lower bound j_0.

   Since the lower bound depends heavily on specific parameters of the problem, in particular on the smoothness s of the exact solution, we try to avoid this choice.

2. Choose the upper bound j_1.

   As the upper bound only depends on the error level this is the easiest choice. Since it also involves the computation of the most wavelet coefficient we try for a third alternative.

3. Choose the s-independent detail level j_η given by (4.11).

   This is a compromise between the upper bound j_1 and the lower bound j_0. It takes into account the smoothness η of the underlying wavelet basis which is chosen in advance and thus well-known. We have $j_\eta \simeq \frac{1}{2\eta+d}(-\log_2(-\delta^2 \log \delta))$ and it follows $j_0 \lesssim j_\eta \lesssim j_1$, see Figure 4.3.

Up to now we have not taken into account that the detail level should be an integer. The next lemma shows that as long as the parameter values for s, t and d fulfill $2s + 2t + d \geq \ln 2 + 1$, the minimal difference between the lower and the upper bound exceeds the unit steplength. Then the two bounds include an integer number. We remark that $2s + 2t + d \geq \ln 2 + 1$ is not an additional condition: Theorem 4.12 postulates $s + t > 1/2$ and we have $d \geq 1$.

Lemma 4.13. *For $\delta \in (0, 1)$, $j_0 := -\log_2(-\delta^2 \log \delta)$, $j_1 = \frac{1}{\kappa} j_0$ we have*

$$|j_0 - j_1| > 1$$

as long as $\kappa > \ln 2 + 1 \approx 1.694$.

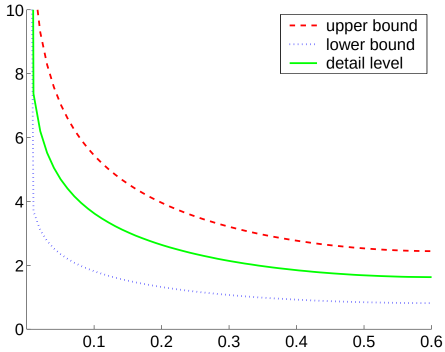


Figure 4.3: Detail level: upper bound (red, dashed), lower bound (blue, dotted) and choice rule (green, solid) from Theorem 4.12 for the parameter values $s = t = 0.5$, $d = 1$, $\eta = 0.25$ and the error levels $\delta \in (0, 0.6]$.

Proof. The assertion is proved straightforward by minimizing the distance function $d(\delta) := j_0 - j_1$ and imposing the condition $\min d(\delta) \geq 1$. $\square$

The integer detail level could thus be chosen as $\lceil j_1 \rceil$ or $\lfloor j_0 \rfloor$, see Figure 4.4.

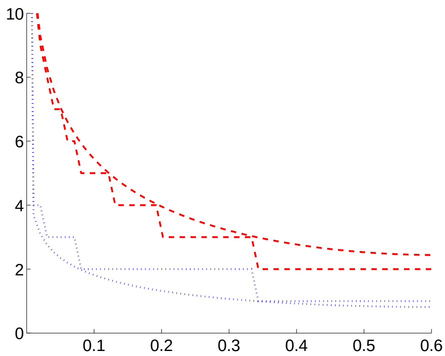
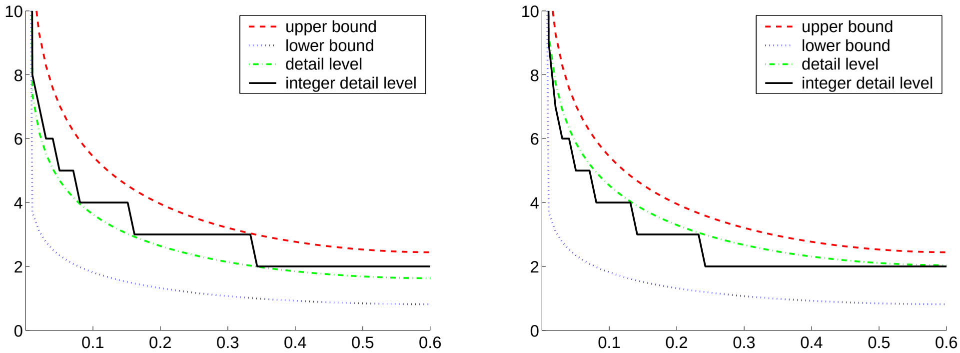


Figure 4.4: Detail level: upper bound and integer upper bound (red, dashed), lower bound and integer lower bound (blue, dotted) for the parameter values $s = t = 0.5$, $d = 1$, $\eta = 0.25$ and the error levels $\delta \in (0, 0.6]$.

One can also choose an integer level with the detail level choice rule (4.11) as a starting point. Figure 4.5 shows the situation for two different sets of parameter values. For $d = 1$ and $\eta \to 0$ the choice rule (4.11) approaches the condition for the upper bound j_1 and one has to choose $\lfloor j_\eta \rfloor$ as detail level. Whereas for $\eta \to s + t$ the choice rule (4.11) approaches the condition for the lower bound j_0 and one must use $\lceil j_\eta \rceil$. To be more precise we use again Lemma 4.13 and conclude that there is an integer level between the upper bound j_1 and the level j_η whenever $2\eta + d \geq \ln 2 + 1$. This is not the case for small η which is depicted in Figure 4.5 (b). Figure 4.5 (a) shows that the estimate of Lemma 4.13 is quite coarse. The condition $2\eta + d \geq \ln 2 + 1 \approx 1.694$ is missed slightly by the given values of the parameter but there is still an integer level in between.



(a) $s = t = 0.5,\ d = 1,\ \eta = 0.25$

(b) $s = t = 0.5,\ d = 1,\ \eta = 0.1$

Figure 4.5: Integer detail level: upper bound (red, dashed), lower bound (blue, dotted), choice rule (green, dash-dotted) and integer choice rule (black, solid) for two parameter sets and for error levels $\delta \in (0, 0.6]$.

4.2.3 Generalization of Reconstruction Methods

In Section 4.2.2 we proved the order optimality of the combination of wavelet shrinkage and filter induced regularization methods. In this section we discuss some conditions of Theorem 4.12 and arrive at a generalization for the admissible reconstruction operators.

For this the smoothness η of the wavelet basis is important. In Theorem 4.12 the only lower bound on η is zero. Hence the wavelets should at least be L_2-functions. The filter function F_α is assumed to be order optimal and to fulfill condition (3.29). In this thesis we have shown that the reduced Tikhonov and the reduced Landweber filter F_α^γ are such filter functions. They induce order optimal methods

$$T_{\alpha,\gamma} g = \sum_{\sigma_n > 0} F_\alpha(\sigma_n)^\gamma \sigma_n^{-1} \langle g, v_n \rangle u_n.$$

as long as $\gamma > 1/2$, see Proposition 4.2 and Proposition 4.4. The condition $\gamma > 1/2$ guarantees that the norm of

$$T_{\alpha,\gamma} : L_2 \to L_2$$

is bounded. However, if wavelet shrinkage with wavelets of Sobolev smoothness η is used as data preprocessing, we consider

$$T_{\alpha,\gamma} : H^\eta \to L_2.$$

In this case the mapping $T_{\alpha,\gamma}$ is defined on a space of Sobolev smoothness η and we ask whether the condition $\gamma > 1/2$ is still necessary. In the following we show that also reconstruction operators with $\gamma \leq 1/2$ achieve the optimal convergence rate of Theorem 4.12, if they are combined with wavelet shrinkage where the wavelet basis complies with a smoothness condition.
We denote the data estimate in H^η by g_η. The Sobolev smoothness η is translated into a source condition according to Lemma 4.6. This yields $g_\eta \in \operatorname{rg}((KK^*)^{\mu/2})$ with $\mu = \eta/t$. Therefore we have $g_\eta = (KK^*)^{\eta/2t}\omega$ with $\omega \in L_2$. We compute $T_{\alpha,\gamma} g_\eta$ as

$$\begin{aligned} T_{\alpha,\gamma} g_\eta &= \sum_{\sigma_n>0} F_\alpha^\gamma(\sigma_n)\sigma_n^{-1}\langle g_\eta, v_n\rangle u_n \\ &= \sum_{\sigma_n>0} F_\alpha^\gamma(\sigma_n)\sigma_n^{-1}\sigma_n^{\eta/t}\langle \omega, v_n\rangle u_n \\ &= \sum_{\sigma_n>0} F_\alpha^\gamma(\sigma_n)\sigma_n^{\eta/t-1}\langle \omega, v_n\rangle u_n. \end{aligned}$$

Hence, to compute $\|T_{\alpha,\lambda}\|_{H^\eta \to L_2}$ we have to estimate

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{\frac{\eta}{t}-1} F_\alpha^\gamma(\sigma)|. \tag{4.19}$$

In order to get convergence rates we aim at an estimate which depends in some way on α and η. In this context we remind the reader of condition (3.29) which reads as

$$\sup_{0<\sigma\leq\sigma_1} |\sigma^{-\nu} F_\alpha^\gamma(\sigma)| \leq c\alpha^{-\beta\nu} \quad \text{for } \nu \in (0,1].$$

This condition was originally obtained from analyzing the smoothing capability of regularization methods with respect to Y_ν-spaces. The condition $2\gamma > \nu$ assures that the data estimate defined by that means belongs to the space Y_ν.

We now "turn the tables" as follows. We consider a given γ and ask for which scale parameter ν – and for which corresponding η – equation (4.19) fulfills an estimate according to (3.29). For $\nu = -(\frac{\eta}{t} - 1) = \frac{t-\eta}{t}$ the term in (4.19) becomes the left hand side of equation (3.29). Equation (4.19) is valid for $2\gamma > \nu = \frac{t-\eta}{t}$. We solve the last inequality for η and get the following condition on the smoothness η of the wavelet basis,

$$\eta > t(1 - 2\gamma).$$

Hence for $\gamma \in (0, 1]$ fixed and a wavelet-based data estimate of smoothness $\eta > t(1 - 2\gamma)$ the norm of $T_{\alpha,\gamma} : H^\eta \to L_2$ stays bounded with

$$\|T_{\alpha,\gamma}\|_{H^\eta \to L_2} \lesssim \alpha^{\beta(\frac{\eta}{t} - 1)}.$$

We remark that this is not a new result. It was already shown in Lemma 4.8. But here we consider a fixed $\gamma \in [0, 1]$ and searched for the smoothness η which assures optimality of the combined method. We summarize the result in the following corollary.

Corollary 4.14. *Let all conditions of Theorem 4.12 hold. The reconstruction operator R_α is given by the reduced Tikhonov or the reduced Landweber method with parameter $\gamma \leq 1/2$. With a wavelet basis of smoothness $\eta > t(1 - 2\gamma)$ the optimal convergence rate (4.13) is valid for the combined method of wavelet shrinkage and reconstruction R_α.*

We discuss the condition $\eta > t(1 - 2\gamma)$ for some special values of γ.

1. $\gamma = 0$

   In this case we get $\eta > t$. For $\gamma = 0$ we have $F_\alpha^\sigma \equiv 1$ and no regularization is performed. This is the extreme case of reduction. In this case the shrinkage step must map at least into $H^{t+\varepsilon}$.

   An interesting question is whether the ε could be neglected. If this is the case, we arrive at the result that no regularization has to be done for data in H^t. Since the operator was supposed to have smoothing property t and to be continously invertible as operator from L_2 in H^t this would fit the setting.

2. $\gamma \in (0, 1/2]$

   In this case we get $\eta > t(1 - 2\gamma) \geq 0$. The boundedness of the reduced "regularization" method $T_{\alpha,\gamma} : H^\eta \to L_2$ is assured by the shrinkage with wavelets of Sobolev smoothness η.

We remind the reader that $T_{\alpha,\gamma}$ with $\gamma \leq 1/2$ is not bounded as operator from L_2 into L_2 and therefore does not define a regularization in the classical sense. But as a *reconstruction* method together with wavelet shrinkage it defines an order optimal method.

3. $\gamma \in (1/2, 1]$

   In this case the reduced methods induced by F_α^γ are order optimal. We have $t(1-2\gamma) < 0$. Since Theorem 4.12 postulates $\eta > 0$ the condition $\eta > t(1-2\gamma)$ is always fulfilled and no additional condition is applied.

   Nevertheless, if we allow η to be negative, there are some interesting questions. The classical regularization methods are obtained for $\gamma = 1$. For this case the condition $\eta > t(1-2\gamma)$ becomes $\eta > -t$. We remind the reader of the results for filter reduction. The filter function F_α^γ assures optimality of the induced method as long as $\gamma > 1/2$. The "rest" of the filter function amplify the filtering effect, sometimes causing oversmoothing. We suppose that this amplification is expressed in the condition $\eta > t(1-2\gamma)$, but this needs further research.

Chapter 5

Examples and Application

In this chapter we study examples of compact operators in Sobolev scales. Smoothness properties of the singular systems are computed both analytically and numerically. Furthermore for an example from medical imaging based on the attenuated Radon transform example reconstructions are done with the combined method of wavelet shrinkage and Tikhonov regularization.

5.1 Examples

In Lemma 4.6 a translation of Sobolev smoothness into source conditions was given. For compact operators whose smoothing property t in Sobolev spaces is expressed by the singular values as $\sigma_n \sim n^{-t}$, the translation implies the asymptotics $\|u_n\|_{H^s} \sim n^s$ for the Sobolev norm of the singular functions. In this section we check this behaviour for two example operators.

5.1.1 Integration Operator

The integration operator is introduced in Example 2.2 and Example 2.10. This operator smoothes with stepsize one in the scale of Sobolev spaces, i.e., $K : H^s \to H^{s+t}$ with $t = 1$. We consider the operator between $L_2(0,1)$ and $L_2(0,1)$, i.e., we deal with L_2-errors.

Proposition 5.1. *Let $0 < s < 1/2$. The singular functions u_n of the integration operator $K : L_2(0,1) \to L_2(0,1)$ with $Kf(x) := \int_0^x f(t)\mathrm{d}t$ fulfill $\|u_n\|_{H^s} \lesssim n^s$.*

Proof. The singular system of the integration operator $K : L_2(0,1) \to L_2(0,1)$ is given in Example 2.10. The singular values are $\sigma_n = \frac{1}{(n+1/2)\pi}$, hence they behave like $\sigma_n \simeq 1/n$. The singular functions are given as $u_n(t) = \sqrt{2}\cos((n+1/2)\pi t)$. We realize the restriction of K on $L_2(0,1)$ by computing the Sobolev norm of $u_n(t)\cdot\chi_{[0,1]}(t)$. The characteristic function $\chi_{[0,1]}$ is in H^s only for $s < 1/2$. This restriction is artificial for the singular functions, see also the remarks after the proof. The Fourier transform of u_n on $[0,1]$ is

$$\begin{aligned}\hat{u}_n(\eta) &= \sqrt{2}\int_0^1 \cos\Big(\underbrace{(n+\frac{1}{2})\pi\, t}_{\tilde{n}}\Big)e^{-i\eta t}\mathrm{d}t \\ &= \sqrt{2}\begin{cases} \frac{\tilde{n}}{\tilde{n}^2-\eta^2}e^{-i\eta}\sin\tilde{n} + \frac{i\eta}{\tilde{n}^2-\eta^2}, & |\eta| \neq \tilde{n} \\ \frac{1}{2} - \frac{i}{2\tilde{n}}, & |\eta| = \tilde{n}. \end{cases}\end{aligned}$$

The squared absolute value of the Fourier transform is

$$|\hat{u}_n(\eta)|^2 = 2\begin{cases} \frac{1}{(\tilde{n}^2-\eta^2)^2}\left(\tilde{n}^2+\eta^2-2\eta\tilde{n}\sin\tilde{n}\sin\eta\right), & |\eta| \neq \tilde{n} \\ \frac{1}{4} + \frac{1}{4\tilde{n}^2}, & |\eta| = \tilde{n}. \end{cases}$$

From the definition of the H^s-norm,

$$\|u_n\|_{H^s}^2 = \int_{-\infty}^{\infty}(1+|\eta|^2)^s|\hat{u}_n(\eta)|^2\mathrm{d}\eta,$$

it follows immediately that the integral only exists for $s < 1/2$.
Due to symmetry we concentrate on the integration over $[0,\infty)$. We split the integral at $|\eta| = \tilde{n}$.

$$\|u_n\|_{H^s}^2 = \underbrace{2\int_0^{\tilde{n}}(1+|\eta|^2)^s|\hat{u}_n(\eta)|^2\mathrm{d}\eta}_{I_1} + \underbrace{2\int_{\tilde{n}}^{\infty}(1+|\eta|^2)^s|\hat{u}_n(\eta)|^2\mathrm{d}\eta}_{I_2}.$$

We consider each integral separately. For the first integral we get

$$\begin{aligned} I_1 &= 2\int_0^{\tilde{n}} \frac{(1+|\eta|^2)^s}{(\eta+\tilde{n})^2}\cdot\frac{\tilde{n}^2+\eta^2-2\eta\tilde{n}\sin\tilde{n}\sin\eta}{(\eta-\tilde{n})^2}\mathrm{d}\eta \\ &\le 2\frac{(1+\tilde{n}^2)^s}{\tilde{n}^2}\cdot\int_0^{\tilde{n}}\frac{\tilde{n}^2+\eta^2-2\eta\tilde{n}\sin\tilde{n}\sin\eta}{(\eta-\tilde{n})^2}\mathrm{d}\eta. \end{aligned}$$

We substitute $\theta = \eta - \tilde{n}$. With $\cos\tilde{n} = 0$ and $\sin^2\tilde{n} = 1$ we get

$$I_1 \le 2\frac{(1+\tilde{n}^2)^s}{\tilde{n}^2}\cdot\int_0^{\tilde{n}}\Big(1 + 2\tilde{n}\underbrace{\frac{(\tilde{n}-\theta)(1-\cos\theta)}{\theta^2}}_{h(\theta)}\Big)\mathrm{d}\theta.$$

The cosine expansion $\cos\theta = \sum_{k=0}^{\infty}(-1)^k \frac{\theta^{2k}}{(2k)!}$ yields $h(0) = \tilde{n}/2$. With $0 \le 1 - \cos\ \theta \le \frac{1}{2}\theta^2$ we have for $\theta \ge 0$

$$h(\theta) = \frac{(\tilde{n}-\theta)(1-\cos\theta)}{\theta^2} \le \frac{\tilde{n}(1-\cos\theta)}{\theta^2} \le \frac{\tilde{n}}{2}.$$

This yields for the first integral

$$I_1 \le 2\frac{(1+\tilde{n}^2)^s}{\tilde{n}^2}\left(\tilde{n}+\tilde{n}^2\right) \le 4(1+\tilde{n}^2)^s.$$

For the second integral we compute

$$\begin{aligned}
I_2 &= \int_{\tilde{n}}^{\infty} \frac{(1+|\eta|^2)^s}{(\eta+\tilde{n})^2(\eta-\tilde{n})^2}(\tilde{n}^2+\eta^2-2\eta\tilde{n}\underbrace{\sin\tilde{n}\sin\eta}_{\in[-1,1]})\mathrm{d}\eta \\
&\le \int_{\tilde{n}}^{\infty} \frac{(1+|\eta|^2)^s}{(\eta-\tilde{n})^2}\mathrm{d}\eta \\
&\lesssim \int_0^{\infty} \frac{(1+\theta+\tilde{n})^{2s}}{\theta^2}\mathrm{d}\theta \\
&\le (1+\tilde{n})^{2s}\int_0^{\infty} \frac{(1+\theta)^{2s}}{\theta^2}\mathrm{d}\theta \\
&\lesssim (1+\tilde{n})^{2s}.
\end{aligned}$$

With $\|u_n\|_{H^s}^2 = 2(I_1+I_2)$ we obtain asymptotically for the H^s-norm of the singular functions that

$$\|u_n\|_{H^s} \lesssim n^s.$$

□

In summary, the smoothing property of the integration operator is expressed by the decay of the singular values, $\sigma_n \simeq n^{-1}$. The H^s-norm of the singular functions is – for $s < 1/2$ – bounded by $\|u_n\|_{H^s} \lesssim n^s$. The restriction $s < 1/2$ is caused by the function $\chi_{[0,1]}(t)$ which is used to realize the domain $[0,1]$ of u_n for the computation of the Fourier transform. If this is done with a function of more smoothness we expect the restriction $s < 1/2$ to be relaxed. In [HvW05] the composition of the integration operator as given above and an injective multiplier function, e.g., a weight function, is studied. A special class of multiplier functions are power functions $m(t) = t^\alpha$ with $\alpha > -1$ which fulfill for some constants $-1 \le \alpha_1 \le \alpha_2$ and $c, C > 0$,

$$ct^{\alpha_2} \le |m(t)| \le Ct^{\alpha_1} \quad \text{a.e. on } [0,1].$$

The modified integration operator $\tilde{K}$ is then defined by means of a multiplier function according to

$$\tilde{K}f(x) = m(x) \int_0^x f(t)\mathrm{d}t, \quad (0 \le x \le 1).$$

In [HvW05] the authors show that $\sigma_n(\tilde{K}) \simeq n^{-1}$. Hence this is another class of integration operators for which the Sobolev smoothness of the singular functions is of interest.

5.1.2 Radon Transform

Let f be a real-valued function in $\mathbb{R}^2$ with compact support. We assume without loss of generality that the support lies in the unit disc $\Omega = \{x \in \mathbb{R}^2 : \|x\| \le 1\}$. With S^1 we denote $\delta\Omega$, the 1-dimensional unit sphere, and elements $\omega \in S^1$ are called directions. The unit cylinder in $\mathbb{R}^2$ is given by $Z = \mathbb{R} \times S^1$. The function

$$Rf : Z \to \mathbb{R},$$

defined by

$$Rf(s, \omega) := \int f(s\omega + t\omega^{\perp})\mathrm{d}t$$

is called the *Radon transform* of f. The Sobolev space $H^s(Z)$ is defined with the norm

$$\|g\|^2_{H^s(Z)} = \int_{S^1} \int_{\mathbb{R}} (1 + |\rho|^2)^s |\hat{g}(\rho, \omega)|^2 \mathrm{d}\rho \mathrm{d}\omega,$$

where the Fourier transform is taken with respect to the first argument. We have the following result [Nat01b].

Lemma 5.2. *The Radon transform is a bounded operator from $H^s(\Omega)$ into $H^{s+1/2}(Z)$ for real s.*

Hence we consider the Radon transform as an operator with smoothing property $t = 1/2$ in Sobolev scales. The Radon transform as an operator between L_2-spaces is compact. The singular system as well as the Fourier transform of the singular functions u_n are given by the following theorem, see [Lou84].

Theorem 5.3. *The Radon transform*

$$R : L_2(\Omega) \to L_2(Z, \omega^{-1})$$

with weighting function $\omega(s) = \sqrt{1-s^2}$ *has the singular system*

$$\{(\sigma_{ml}, u_{ml}, v_{ml}) : m \geq 0, l \in \mathbb{Z} : |l| \leq m, m+l \in 2\mathbb{Z}\}.$$

The singular values fulfill

$$\sigma_{ml} = \sigma_m \simeq m^{-1/2}.$$

The singular functions are given according to

$$u_{ml}(s \cdot \omega) = \begin{cases} \sqrt{\frac{m+1}{\pi}} s^{|l|} P^{0,|l|}_{(m-|l|)/2}(2s^2-1) Y_l(\omega), & s \in [-1,1],\ \omega \in S^1 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$v_{ml}(s, \omega) = \begin{cases} \frac{1}{\pi}\omega(s) U_m(s) Y_l(\omega), & s \in [-1,1],\ \omega \in S^1 \\ 0 & \text{elsewhere.} \end{cases}$$

Here $P_n^{\alpha,\beta}$ *are the Jacobi-polynomials of degree* n *with parameters* α, β, *the functions* Y_l *are the spherical harmonics of degree* l *and* U_m *are the Tschebyscheff-polynomials.*
The Fourier transform of the singular function u_{ml} *is given by*

$$\hat{u}_{ml}(\sigma \cdot \omega) = \sqrt{\frac{m+1}{\pi}} (-1)^m \sigma^{-1} J_{m+1}(\sigma) Y_l(\omega),$$

where J_m *denotes the Bessel function of the first kind of order* m.

In order to estimate the Sobolev smoothness of the singular functions of the Radon transform we have to evaluate the integral

$$\|u_{ml}\|^2_{H^s} = \int_{\mathbb{R}^2} (1+|\xi|^2)^s |\hat{u}_{ml}(\xi)|^2 \mathrm{d}\xi.$$

We start with some facts on Bessel functions, see [AS84] and [Wat66].

Lemma 5.4. *Let* J_m *denote the Bessel function of the first kind of order* m. *Then the following holds*

i)
$$|J_0(x)| \leq 1, \quad |J_m(x)| \leq 1/\sqrt{2} \quad \text{for } m > 0.$$

ii)
$$\mathrm{Re}(m) > 0,\ \alpha\ \textit{constant}: \quad \int_0^\infty J_m^2(\alpha x) \frac{\mathrm{d}x}{x} = \frac{1}{2m}.$$

iii) For m fixed and $|x| \to \infty$

$$J_m(x) = \sqrt{\frac{2}{\pi x}}\left(\cos(x - \frac{2m+1}{4}\pi) + \mathcal{O}(|x|^{-1})\right).$$

In order to show that the Radon transform fits in our proposed scheme we estimate the Sobolev norm of its singular functions.

Proposition 5.5. *Let $0 < s < 1/2$. The singular functions u_{ml} of the Radon transform fulfill $\|u_{ml}\|_{H^s} \lesssim m^s$.*

Proof. From Lemma 5.3 we know

$$|\hat{u}_{ml}(\sigma \cdot \omega)|^2 = \frac{m+1}{\pi}\sigma^{-2}|J_{m+1}(\sigma)|^2|Y_l(\omega)|^2.$$

We rewrite the integration on $\mathbb{R}^2$ to use the special structure of $\hat{u}_{ml}$. Let $\delta K_\sigma(0)$ denote the circle with radius σ around zero. We integrate over circles around zero with radius in $\mathbb{R}_{\geq 0}$. It follows

$$\begin{aligned}\|u_{ml}\|_{H^s}^2 &= \int_{\mathbb{R}^2}(1+|\xi|^2)^s|\hat{u}_{ml}(\xi)|^2\mathrm{d}\xi \\ &= \int_{\mathbb{R}_{\geq 0}}\int_{\delta K_\sigma(0)}(1+|\omega|^2)^s|\hat{u}_{ml}(\omega)|^2\mathrm{d}\omega\mathrm{d}\sigma.\end{aligned}$$

For $\omega \in \delta K_\sigma(0)$ it is $|\omega| = \sigma$ and we write $\omega = \sigma\tilde{\omega}$ with $|\tilde{\omega}| = 1$. For the spherical harmonic Y_l it is $\int_{\delta K_\sigma(0)}|Y_l(\omega)|^2\mathrm{d}\omega = \sigma\int_{\delta K_1(0)}|Y_l(\tilde{\omega})|^2\mathrm{d}\tilde{\omega} = 2\pi\sigma$. We get

$$\|u_{ml}\|_{H^s}^2 = 2(m+1)\int_0^\infty(1+\sigma^2)^s\frac{J_{m+1}^2(\sigma)}{\sigma}\mathrm{d}\sigma.$$

We split the integral at $a := m+1 > 0$ and consider the part for large arguments first. We remark that for large orders m the asymptotical behaviour $J_m(x) \sim 1/\sqrt{x}$ given in Lemma 5.4 iii) already holds for $x > m+1 = a$, see [Wat66]. For smaller orders we use the boundedness of the Bessel functions and take $J_m(x) \leq c\sqrt{\frac{1}{x}}$ with the constant c independent of the order. For $s < 1/2$ we get

$$\begin{aligned}I_2 &= 2(m+1)\int_a^\infty \underbrace{(1+\sigma^2)^s}_{\sim\sigma^{2s}}\sigma^{-1}\underbrace{J_{m+1}^2(\sigma)}_{\lesssim\frac{1}{\sigma}}\mathrm{d}\sigma \\ &\lesssim (m+1)\int_a^\infty\sigma^{2s-2}\mathrm{d}\sigma \\ &= (m+1)\frac{1}{1-2s}a^{2s-1} \\ &\simeq m^{2s}.\end{aligned}$$

For the first part of the integral we use Lemma 5.4 and get

$$
\begin{aligned}
I_1 &= 2(m+1)\int_0^a (1+\sigma^2)^s \sigma^{-1} J_{m+1}^2(\sigma)\mathrm{d}\sigma \\
&\leq 2(m+1)(1+a^2)^s \int_0^\infty \frac{J_{m+1}^2(\sigma)}{\sigma}\mathrm{d}\sigma \\
&= (1+a^2)^s \\
&\simeq m^{2s}.
\end{aligned}
$$

The assertion follows by assembling the estimates for the integrals I_1 and I_2. □

In the next section we return to the Radon transform as mathematical model of an application from medical imaging.

5.2 Single Photon Emission Computerized Tomography

In this section we apply the combined method of wavelet shrinkage and Tikhonov regularization, short *TikShrink*, to simulated data from medical imaging. We deal with the inversion of the SPECT (Single Photon Emission Computerized Tomography) operator. Single photon emission computerized tomography provides three-dimensional images of the concentration of a radiopharmaceutical within the body. The mathematical model for this is the *attenuated Radon transform*

$$
R(f,\mu)(s,\omega) = \int_{\mathbb{R}} f(s\omega^\perp + t\omega)\underbrace{e^{-\int_t^\infty \mu(s\omega^\perp+\tau\omega)\mathrm{d}\tau}}_{\text{attenuation}}\mathrm{d}t.
$$

In Section 5.2.2 we present a numerical smoothness analysis for the singular system. Test computations for TikShrink are performed in Matlab and presented in Section 5.2.3. They confirm the theoretical results of Theorem 4.12.

SPECT in mathematical literature

Single photon emission computerized tomography is widely studied throughout literature and often used to assess the quality of reconstruction methods. Exact inversion formulae as well as approximate solution methods can be classified by their assumptions on the attenuation function μ. We start with some remarks on exact inversion formulae. For $\mu = 0$, the problem is solved by the Radon inversion formula. If μ is constant and known inside a convex set, the problem can be reduced to the exponential Radon transform for which various inversion formulae are available [TM80, Pal96]. For known attenuation function μ, an exact inversion formula for the attenuated Radon transform is given in [Nov02, Nat01a]. In [Nat01a] the class of functions f for which the inversion formula is valid is not determined exactly. It is, however, remarked that the inversion formula is valid for continuously differentiable functions f of compact support.

We remark that for all inversion formulae it is assumed that μ is known. In this case the attenuated Radon transform is a linear operator. If μ is not known, reconstructions have to be done simultaneously for the activity function f and the attenuation function μ. In this case the attenuated Radon transform is a nonlinear operator.

In the following we cite some of the recent works on SPECT. We start with the case that the attenuation function μ is assumed to be known and hence the operator is linear. Reconstruction algorithms based on different inversion formulae of the Radon transform are studied in [CN04, WD02, Kun01]. We remind the reader that throughout this thesis we use additive white noise to model the noisy data. In [GN04] the SPECT data are modelled as the attenuated ray transform with Poisson noise. Among other things the authors present formulae for describing statistical properties of this model and propose new possibilities for improving stability. The situation that reconstruction for the attenuated Radon transform is done not only for full measurements but also for partial measurements is studied in [Bal04]. The authors of [FM04] study SPECT together with PET (positron emission tomography) as modern techniques in brain imaging. A reconstruction algorithm based on a Tschebyscheff approximation of the data functions is presented. In [RT04b, RT04a] a data-driven scheme for an optimal choice of the regularization parameter is studied. SPECT is used as one example to show the applicability of the proposed method. Another approach for regularization, considered as a minimization problem, is to vary the penalty term. In [ACC+02] Total Variation based reconstructions of 3D SPECT images are studied.

In the case that the attenuation function μ is not known, a simultaneous reconstruction of activity and attenuation functions has to be done. We refer the reader to [Ram03, Ram02a, RCNB00, Dic99]. In [Dic99] the author demonstrates that an approximation of both f and μ from SPECT data alone is feasible, leading to quantitatively more accurate SPECT images. The result is based on nonlinear Tikhonov regularization techniques

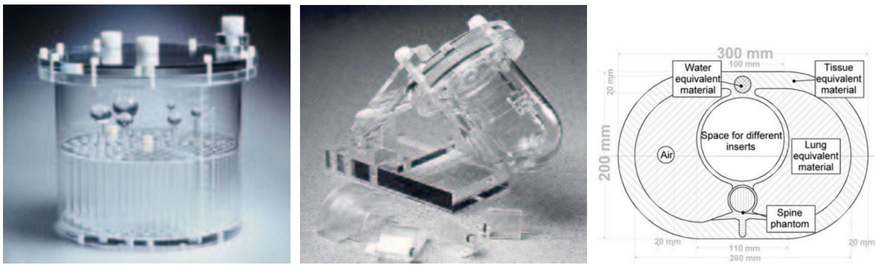
for parameter estimation problems in differential equations combined with Gauss-Newton conjugate gradient minimization.
In [Ram03] the author presents a sophisticated numerical analysis of a combination of Tikhonov regularization and the gradient method for solving nonlinear ill-posed problems. Numerical results are presented for SPECT. In [Ram02b] Morozov's discrepancy principle for Tikhonov regularization of nonlinear operator equations is studied. SPECT is investigated as one example of a practically relevant case. In [Ram02a] a new approach for finding a global minimizer of the Tikhonov functional is given and applied to SPECT.
In this thesis we do not consider the simultaneous reconstruction but assume the attenuation map μ to be known. We reconstruct the activity function f with the two-step method of wavelet shrinkage and Tikhonov regularization.

We continue with a short, non-mathematical, description of single photon emission computerized tomography. Mathematical details follow in Section 5.2.2.

5.2.1 Description of the Technology

Single photon emission computerized tomography is a special form of computerized tomography (CT). In general, CT is a non-invasive diagnosis technique in medicine used to get information of the morphology of a patient. The patient is scanned from the outside and the desired characteristics are computed from the scan. SPECT is an imaging method designed to provide information about the *function level* of a part of the body. SPECT involves the injection of a low-level radioactive chemical, called a *radiotracer* or *radiopharmaceutical*, into the bloodstream. The radioactive material ejects photons which travel through the body and are measured outside. Scans are made with a device that can detect radioactivity in the body by measuring its decay products, e.g., a SPECT camera, gamma camera or tomograph. The obtained images reflect the manner in which the tracer is processed by the body. Thus this technology provides functional information in contrast to the structural information provided by CT, MRI (magnetic resonance imaging) and ultrasound.
Each radiotracer used with SPECT is a radiative substance which is used alone or which is attached to an element appropriate for obtaining specific information. For example, certain types of proteins called antibodies attach to specific types of tumors. The radiotracer can be attached to antibodies which bind to the tumors, and thus can be identified and located. SPECT

can provide information about the level of chemical or cellular activity within an organ or system, as well as provide structural information. This process may show areas of increased activity such as the inflammation in an abscess, or decreased activity such as the diminished blood flow to the heart in the presence of coronary artery disease.
Test computations are often performed using phantoms. In Figure 5.1 photos of a torso phantom (a) and a heart phantom (b) are shown. Figure 5.1 (c) shows a draft of a torso phantom.



(a) Jasczcak torso phantom

(b) Cardiac insert

(c) Sketch of torso phantom

Figure 5.1: Torso and heart phantoms, photos and sketch.

In summary, the radiotracer used for a scan is specific to the disease process being investigated. It circulates in the bloodstream and binds to specific target cells. The emitted radiation from the radiotracer travels through the body, interferes with the surrounding material and is measured outside the body by a gamma camera as a series of projections.
The measurement process involves the gamma camera rotating around the patient, acquiring images at various positions. The number of images and the rotation angle covered varies depending on the type of investigation required. For the test computations in Section 5.2.3 we use 79 angles over the whole circle and 80 samples.

5.2.2 Mathematical Background

The radiopharmaceutical injected for a SPECT scan enriches due to the blood flow in a certain area of the body. The distribution of the radiopharmaceutical is interpreted as activity function f. From the measurements g of the gamma camera the activity function f is to be reconstructed. The measured data g is modelled as follows.
While travelling through the body the photons interact with the body tissue and loose energy. The interaction is described by a so-called *damping* or *at-*

tenuation function μ. In emission tomography the resulting data is modelled by the *attenuated Radon transform*:

$$\begin{aligned} g &= R(f,\mu)(s,\omega) \\ &= \int_{\mathbb{R}} f(s\omega^{\perp} + t\omega) e^{-\int_t^{\infty} \mu(s\omega^{\perp}+\tau\omega)\mathrm{d}\tau} \mathrm{d}t, \qquad s \in \mathbb{R},\ \omega \in S^1, \end{aligned} \tag{5.1}$$

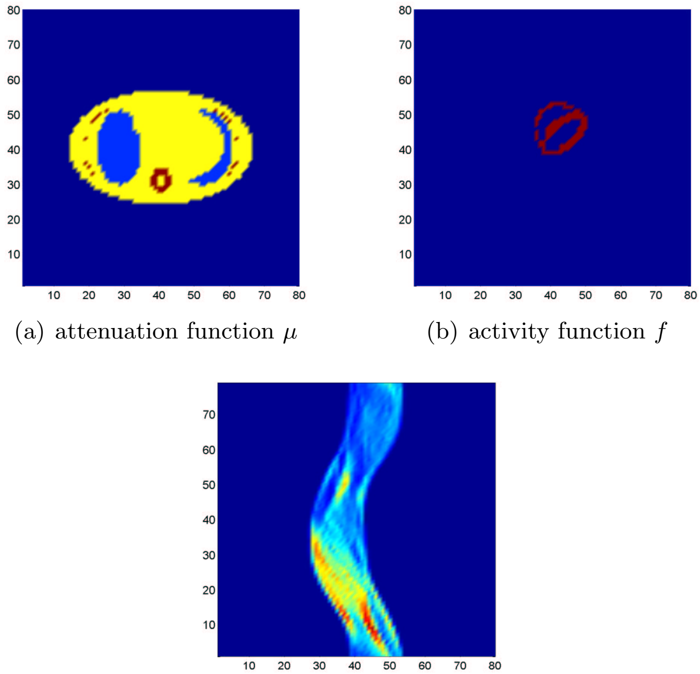
where s and ω describe the geometry. We refer the reader to [Nat01b] for a sophisticated treatment of the mathematics of computerized tomography. The attenuation due to the body density μ is described by the term

$$e^{-\int_t^{\infty} \mu(s\omega^{\perp}+\tau\omega)\mathrm{d}\tau}.$$

We assume the *attenuation map* μ to be known and consider the linear operator $R_\mu(f)$. We assume further that for known μ the smoothing property of R_μ is the same as for the Radon transform without attenuation. Hence from Lemma 5.2 we assume $t = 1/2$ as smoothing property of the attenuated Radon transform.

5.2.3 Numerical Results

The test computations presented here are not performed with real data but with simulated data. The activity function f and the attenuation function μ are given as cross sections of a heart phantom, Figure 5.2 (b), and of a torso phantom, Figure 5.2 (a). To generate the data, the attenuated Radon transform (5.1) with given attenuation μ is discretized. For this 79 angles over the whole circle and 80 samples are used. The resulting values constitute a matrix which we refer to as *SPECT matrix*. This matrix depends on the attenuation function μ. To estimate the influence of the attenuation μ on the matrix, we set $\mu \equiv 0$ and generate a discretized version of the classical Radon transform which we refer to as *CT matrix*.
The SPECT data is then generated by a matrix-vector multiplication. For this the activity function f is transformed into a vector. The resulting sinogram data is shown in Figure 5.2 (c).
The singular systems of the SPECT matrix and of the CT matrix are computed with MATLAB. According to Theorem 5.3 we expect the singular values of the CT matrix to build clusters of increasing length. To be more precise, for each index m in σ_{ml} there are $m+1$ indices l and hence $m+1$ singular values σ_{ml} which should fulfill $\sigma_{ml} \simeq m^{-1/2}$. For the CT matrix the first 15 to 20 (of 6320) singular values show the described behaviour, whereas for the SPECT matrix this seems an arbitrary classification. For



(a) attenuation function μ (b) activity function f

(c) sinogram data g

Figure 5.2: (a) Attenuation function μ. (b) Activity function f. The attenuation and the activity functions are given as cross sections of a body and a heart phantom. (c) Resulting sinogram data g.

the singular functions we expect, according to Theorem 5.3, a radial symmetry caused by the sperical harmonics Y_l. This is true for the singular functions of the CT matrix but not for the ones of the SPECT matrix, see Figure 5.9 for plots of some singular functions of the SPECT matrix.

We do assume that for known attenuation function μ the smoothing property $t = 1/2$ of the Radon transform, see Lemma 5.2, carries over to R_μ. However, as described above, the additional inhomogenity caused by the exponential attenuation term damages the clustering of the singular values as well as the radial symmetrie of the singular functions. Hence, we do not take into account the multiplicity of the singular functions u_{ml} when computing the decay behaviour of the singular values of the SPECT matrix.

In Figure 5.3 the singular values are shown together with a least-square fit of the first 4250 values. For this restriction the singular values decay asymptotically as $\sigma_n \simeq n^{-0.492}$. If we include all 6320 singular values in the computation of the least-square fit we get $\sigma_n \simeq n^{-1.33}$.

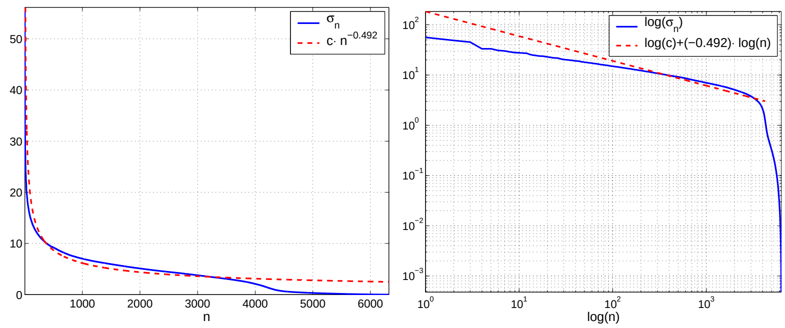


Figure 5.3: Left: singular values of SPECT matrix (blue, solid) and least-square polynomial fit (red, dashed). Right: log-log plot of the singular values (blue, solid) and polynomial fit with slope -0.492 (red, dashed).

For the smoothness analysis of the singular system we assume that we have an exact solution f of Sobolev smoothness s. Accordingly, we estimate the Sobolev smoothness of the singular functions u_n. We compute the H^s-norm of u_n as discretized integral of the properly weighted Fourier transform. Figure 5.10 shows the absolute value of the Fourier transform of u_n for the same singular functions as in Figure 5.9.
According to the remarks following Lemma 4.6 the H^s-norm of the singular functions u_n should fulfill $\|u_n\|_{H^s} \simeq n^s$. For medical applications the activity function f belongs to H_0^s with s close to $1/2$. We assume the unknown activity f to be in $H_0^{1/2}$. Hence we compute $\|u_n\|_{H^s}$ with $s = 1/2$. The result is shown in Figure 5.4.
The decay characteristic of $\|u_n\|_{H^s}$ is determined by a least-square polynomial fit to the computed data. We remark that the behaviour $\|u_n\|_{H^s} \simeq n^s$ is to be understood as an upper bound for the norm of the singular functions. For $s = 1/2$ the H^s-norm of the singular functions follows $\|u_n\|_{H^s} \simeq n^{0.117}$. Hence the sequence of norms of the singular functions could grow even faster.

Reconstruction of the activity function

Test computations with simulated data are performed in Matlab. To simulate the measurement process, white noise with different error levels is added to the data. In Figure 5.5 the data model as well as the two-step reconstruction is visualized. In the upper row the simulation of data according to

$$g^\delta = Kf + \text{noise}$$

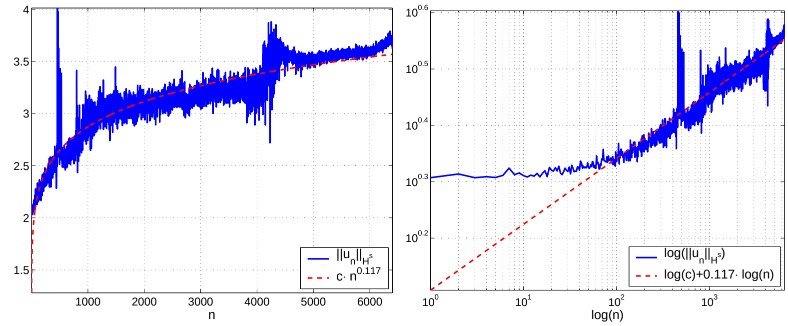


Figure 5.4: Left: H^s-norm for s=1/2 of the singular functions u_n of the SPECT matrix (blue, solid) and least-square polynomial fit (red, dashed). Right: log-log plot of the H^s-norm for s=1/2 of the singular functions u_n (blue, solid line) and polynomial fit with slope 0.117 (red, dashed).

is shown. In the lower row one sees the construction of a solution by a two-step method $T_{\alpha,\lambda}$ consisting of the data estimation S_λ and the reconstruction R_α,

$$f_{\text{rec}} = f_{\alpha,\lambda} = T_{\alpha,\lambda} g^\delta = R_\alpha S_\lambda g^\delta.$$

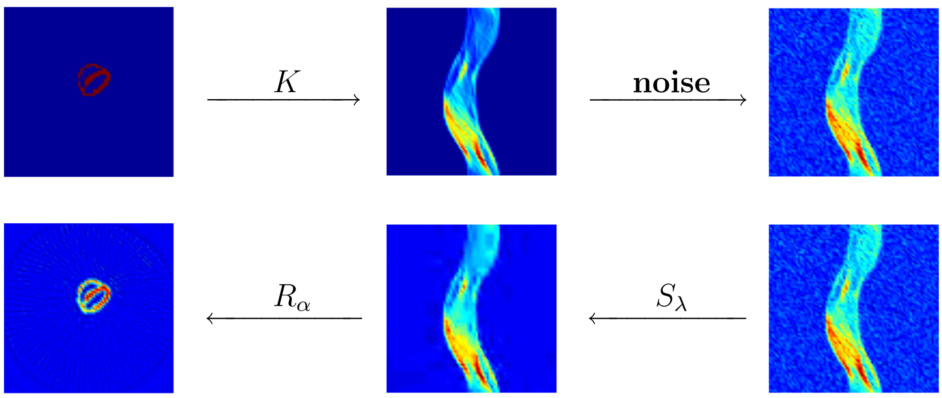


Figure 5.5: Upper row: modeling of SPECT data as application of the operator K followed by adding noise; lower row: two-step solution for SPECT consisting of a data estimation S_λ and a reconstruction R_α.

For the data estimation step S_λ we use wavelet shrinkage. The reconstruction R_α is done with the classical Tikhonov method. Test computations are performed for fifty relative error levels of the data from 1% up to 50%. Figure 5.6 shows an example reconstruction by TikShrink from noisy data with 5% relative error amount. For the shrinkage step the wavelet db2 was used.

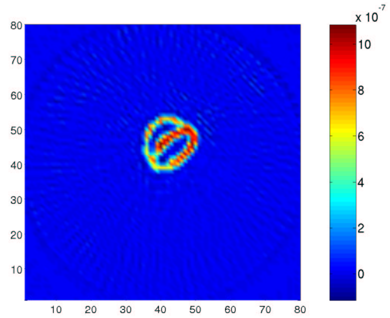


Figure 5.6: Reconstruction by TikShrink of the activity function f from noisy sinogram data g^δ with 5% relative error.

In the following we compare the results of classical Tikhonov regularization and the TikShrink method. For this, reconstructions are done with and without shrinkage. To assess the results of Theorem 4.12 the regularization parameter is determined by Morozov's discrepancy principle on the one hand and by the parameter choice rule (4.12) on the other hand. The discrepancy principle is also used to determine the constant in the parameter choice rule (4.12). Figures 5.7 and 5.8 show some results of the test computations. Figure 5.7 shows the reconstruction error and a log-log plot of the reconstruction error together with a least-square polynomial fit.

We compare the computed results and the theoretical results given in Theorem 4.12. We expect the error to behave as given in (4.13) which reads as

$$\|f_{\text{reg}}^\delta - f\|_{L_2}^2) \leq c(\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}.$$

We insert the parameter values $t = 1/2$ for the smoothing property, $s = 1/2$ for the smoothness of the exact solution and $d = 2$ for the dimension of the problem. This yields $\frac{4s}{2s+2t+d} = \frac{1}{4}$ and we expect the error to behave like $(\delta\sqrt{|\log\delta|})^{\frac{1}{4}}$. The least-square fit to the computed data of the Tikhonov method yields the exponent 0.2372 which fits the expected rate of 0.25 quite well. The difference in the expected and computed rate might be due to numerical effects since the variance of the added white noise is very small for the low error levels.

In Figure 5.8 we compare the results for the regularization parameter chosen by Morozov's discrepancy principle and chosen by the parameter choice

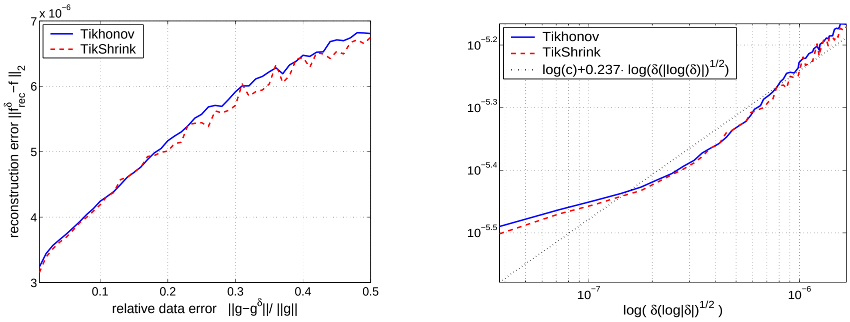


Figure 5.7: Left: absolute error of the Tikhonov method (blue, solid) and the TikShrink method (red, dashed) for relative data error from 1% to 50%. Right: log-log plot of the absolute error of the Tikhonov method (blue, solid) and the TikShrink method (red, dashed) and polynomial fit (black, dotted) to the TikShrink data with slope 0.2372.

rule (4.12) of Theorem 4.12. The choice rule is

$$\alpha \simeq (\delta\sqrt{|\log\delta|})^{\frac{1}{\beta}\frac{2t}{2s+2t+d}}$$

with the parameter β depending on the regularization method. We insert the parameter values $\beta = 1/2$ for the Tikhonov method, $t = 1/2$ for the smoothing property, $s = 1/2$ for the smoothness of the exact solution and $d = 2$ for the dimension of the problem. This yields $\frac{1}{\beta}\frac{2t}{2s+2t+d} = \frac{1}{2}$ and we choose α as $c(\delta\sqrt{|\log\delta|})^{\frac{1}{2}}$. The constant c is determined using a fixed error level and comparing for this level the parameter from the direct choice rule and from Morozov's discrepancy principle.

Computational results for different combinations of wavelet shrinkage and Tikhonov regularization are listed in Table 5.1 for 10% relative data error and in Table 5.2 for 20% relative data error. There are several options for the shrinkage step and the reconstruction step by Tikhonov regularization. For the shrinkage step, the entry "none" indicates that the shrinkage operator equals the identity and the data are not preprocessed at all. For the reconstruction step, the entry "none" indicates that the regularizing parameter equals zero and that the Tikhonov operator equals the generalized inverse. In this case, the error of the reconstructed solution demonstrates the ill-posedness of the problem.

If shrinkage is performed, the corresponding tabel entry consists of the used wavelet and the number of used levels compared to the possible number of

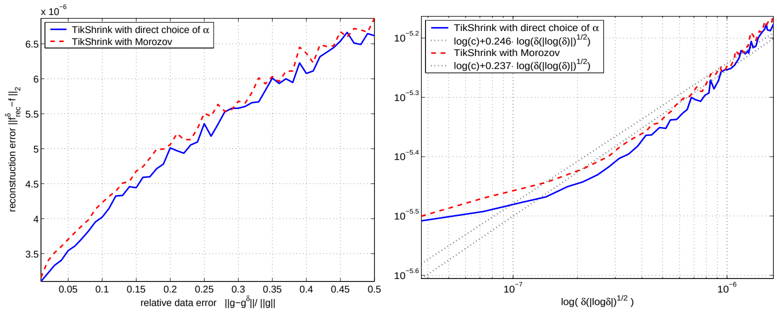


Figure 5.8: Left: absolute error of the TikShrink method with direct choice of α (blue, solid) and α computed by Morozov's discrepancy principle (red, dashed) for relative data error from 1% to 50%. Right: log-log of the same data with polynomial fit (black, dotted) with slope 0.246 for direct choice of α and slope 0.237 for Morozov's discrepancy principle.

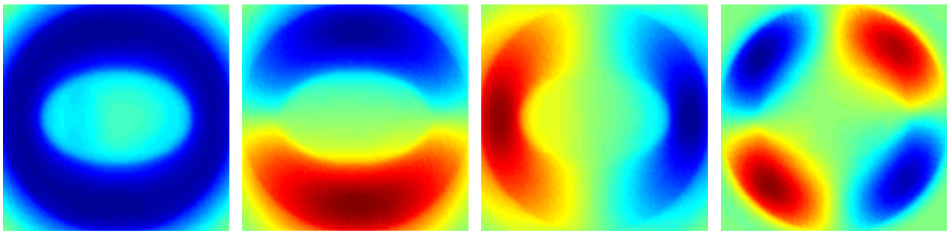
levels. From the listed results we conclude that the use of shrinkage alone already improves the reconstruction.
If regularization is performed, the corresponding tabel entry consists of the used parameter choice strategy and the resulting value of the regularizing parameter. The options for the parameter choice strategy are "direct" for the rule (4.12) and "Morozov" for the discrepancy principle. From the listed results we conclude that the use of shrinkage reduces the size of the regularizing parameter as determined by the discrepancy principle. That means, when wavelet shrinkage is used, less regularization has to be performed.

data error in %	**shrinkage**		**regularization**		**reconstruction error in %**
	ψ	j/j_{up}	method	value	
10	none		none		104.78
10	db1	42 / 42	none		75.13
10	db2	22 / 42	none		78.95
10	db3	16 / 42	none		77.98
10	none		direct	0.589	42.36
10	db1	42 / 42	direct	0.5875	40.4
10	db2	22 / 42	direct	0.5907	41.45
10	db3	16 / 42	direct	0.5886	41.26
10	none		Morozov	3.3554	44.48
10	db1	42 / 42	Morozov	2.5600	43.32
10	db2	22 / 42	Morozov	2.8823	44.81
10	db3	13 / 42	Morozov	2.7488	44.81

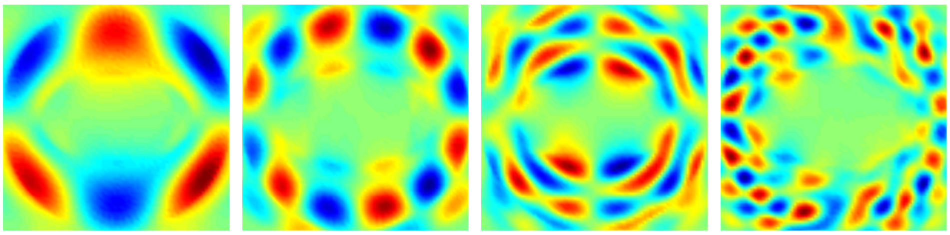
Table 5.1: Results for TikShrink reconstructions from SPECT data with 10% relative error in the data. Reconstructions are done using different combinations of wavelet shrinkage and Tikhonov regularization. The last column lists the relative error in the reconstructed activity function.

data error in %	**shrinkage**		**regularization**		**reconstruction error in %**
	ψ	j/j_{up}	method	value	
20	none		none		205.95
20	db1	41 / 41	none		139.89
20	db2	21 / 41	none		124.52
20	db3	15 / 41	none		120.41
20	none		direct	0.8256	53.66
20	db1	40 / 40	direct	0.8222	48.32
20	db2	21 / 40	direct	0.8274	48.48
20	db3	15 / 40	direct	0.8269	47.94
20	none		Morozov	8.7961	54.46
20	db1	40 / 40	Morozov	6.0446	52.04
20	db2	21 / 40	Morozov	6.1897	52.79
20	db3	15 / 41	Morozov	6.0446	52.63
20	db4	12 / 40	Morozov	5.9030	52.25

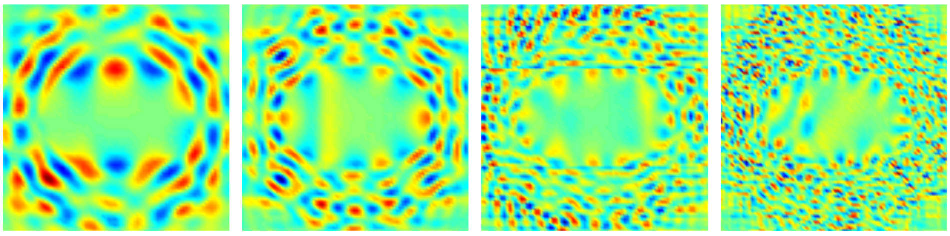
Table 5.2: Results for TikShrink reconstructions from SPECT data with 20% relative error in the data. Reconstructions are done using different combinations of wavelet shrinkage and Tikhonov regularization. The last column lists the relative error in the reconstructed activity function.



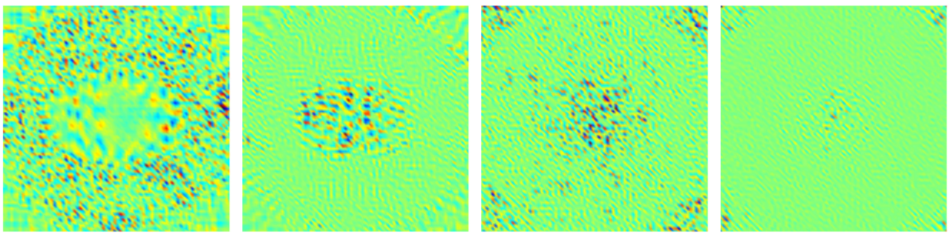
(a) $u_1,\ u_2,\ u_3,\ u_4$



(b) $u_{10},\ u_{25},\ u_{50},\ u_{75}$

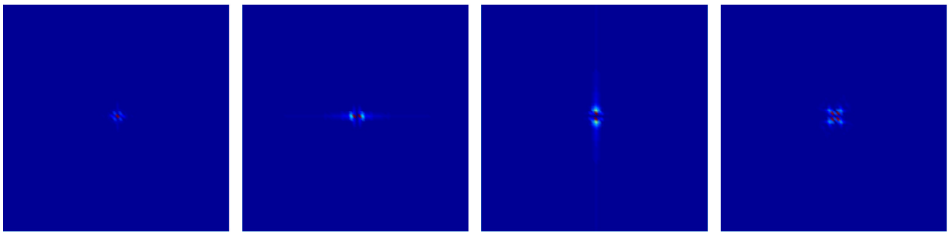


(c) $u_{100},\ u_{250},\ u_{500},\ u_{1000}$

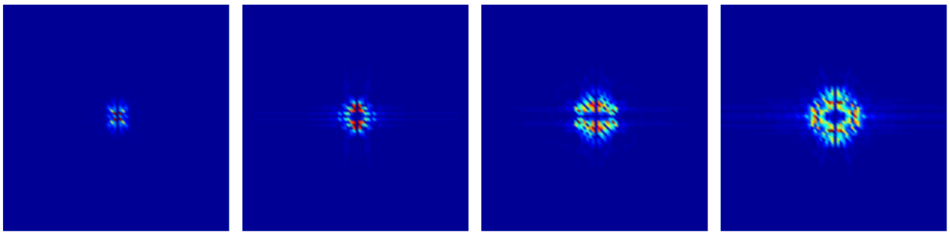


(d) $u_{2000},\ u_{4000},\ u_{6000},\ u_{6400}$

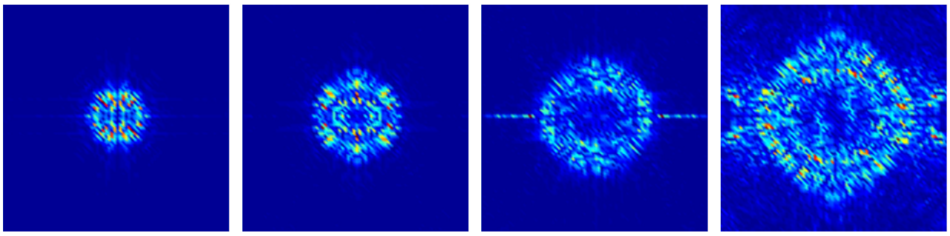
Figure 5.9: Some singular functions u_n of the SPECT matrix; green indicates vanishing values, red positive values and blue negative values.



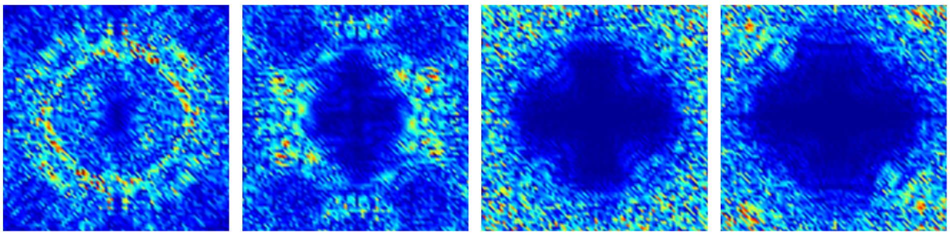
(a) $|\hat{u}_1|$, $|\hat{u}_2|$, $|\hat{u}_3|$, $|\hat{u}_4|$



(b) $|\hat{u}_{10}|$, $|\hat{u}_{25}|$, $|\hat{u}_{50}|$, $|\hat{u}_{75}|$



(c) $|\hat{u}_{100}|$, $|\hat{u}_{250}|$, $|\hat{u}_{500}|$, $|\hat{u}_{1000}|$



(d) $|\hat{u}_{2000}|$, $|\hat{u}_{4000}|$, $|\hat{u}_{6000}|$, $|\hat{u}_{6400}|$

Figure 5.10: Absolute value of Fourier transform, $|\hat{u}_n|$, of some singular functions u_n of the SPECT matrix; values are for all images increasing from blue (zero) over green to red; the saling is different for each image.

Chapter 6

Summary and Outlook

In this thesis we have studied two-step methods $T_{\alpha,\lambda} = R_\alpha S_\lambda$ for the solution of the inverse problem $Kf = g$ with linear operator K and noisy data g^δ. Two-step methods were introduced as the combination of a data estimation operator $S_\lambda : Y \to Y$ and a reconstruction operator $R_\alpha : Y \to X$. Hilbert scales were used as unifying theoretical framework. The amount of data estimation by S_λ on the data side space Y was measured in a Hilbert-scale $\{Y_\mu\}$, i.e., for a parameter $\nu \in [0, 1]$ we have considered the operator

$$S_\lambda : Y \to Y_\nu.$$

The reconstruction operator is then defined on the space which contains the data estimate $S_\lambda g$, i.e.,

$$R_\alpha : Y_\nu \to X.$$

As main instances for two-step methods we examined filter-induced regularization methods (Tikhonov, Landweber, TSVD), and the combination of nonlinear wavelet shrinkage and filter-induced regularization methods. The first instance resulted in the so-called *reduced methods*.

We briefly summarize the results of this thesis.

1. The reduced Tikhonov as well as the reduced Landweber method with parameter $\gamma \in [0, 1]$ are order optimal if $\gamma > 1/2$.

   (Propositions 4.2 and 4.4)

2. Wavelet shrinkage and filter-induced regularization methods were combined succesfully with the unifying framework of Hilbert scales. The combination is an asymptotically order optimal method.

   (Theorem 4.12)

3. The combination of wavelet shrinkage and reduced methods with parameter $\gamma \leq 1/2$ is asymptotically order optimal if the wavelet basis complies with the smoothness condition $\eta > t(1-2\gamma)$.

   (Corollary 4.14)

The *reduced methods* were defined for filter-induced regularization methods. An additional parameter $\gamma \in [0,1]$ was applied to the filter function F_α according to

$$F_\alpha^\gamma := (F_\alpha)^\gamma.$$

This results in a *reduction of filtering* since it decreases the amount of damping. For the classical Tikhonov as well as for the classical Landweber filter F_α it was shown that the reduced filter induces an order optimal regularization method $R_{\alpha,\gamma} : Y \to X$ as long as $\gamma > 1/2$. The fact that not the whole filter is needed to achieve order optimality explains the well-known effect of oversmoothing when classical Tikhonov or Landweber regularization is used.

Furthermore the capabilities of the filter function for data smoothing were explored. A data smoothing operator $S_{\alpha,\gamma}$ was defined by application of F_α^γ on the data side. It was shown that $S_{\alpha,\gamma}$ maps into the Hilbert-scale space $Y_\nu = \operatorname{rg}((KK^*)^{\nu/2})$ as long as $\gamma > \nu/2$.

For the combination of wavelet shrinkage and filter based regularization methods we translated the additional information on the exact solution f given as Sobolev smoothness, i.e., $f \in H^s$, into source conditions given by operator-adapted spaces, i.e., $f \in \operatorname{rg}((K^*K)^{\nu/2})$. Under the additional assumption that the operator K smoothes with stepsize t in the Sobolev scale we proved that

$$f \in H^s \Leftrightarrow f \in \operatorname{rg}((K^*K)^{\nu/2}) \quad \text{for} \quad \nu = s/t.$$

With this translation the combination of wavelet shrinkage and classical regularization methods was accomplished. The following quasi-optimal convergence rate has been shown:

$$\mathrm{E}(\|f_{\alpha\lambda j}^\delta - f\|_{L_2}^2) = \mathcal{O}((\delta\sqrt{|\log\delta|})^{\frac{4s}{2s+2t+d}}).$$

As next step we considered the combination of wavelet shrinkage and reduced methods with parameter $\gamma \leq 1/2$. If the wavelet basis is of Sobolev smoothness H^η, the reconstruction operator R_α has to be bounded as operator

$$R_\alpha : H^\eta \to L_2$$

but not as operator

$$R_\alpha : L_2 \to L_2.$$

It was shown that for $\gamma \le 1/2$ the condition

$$\eta > t(1 - 2\gamma)$$

on the Sobolev smoothness η of the wavelet basis assures the order optimality of the combined method. We empasize that for ill-posed problems a certain amount of “special treatment” is necessary. The results on the combination of wavelet shrinkage and reduced methods demonstrate that this “special treatment” can be divided into data estimation and reconstruction.

The contribution of this thesis to the field of regularization methods for linear ill-posed problems might be summarized as follows: On the one hand, the two-step approach allows a deeper understanding of filter based regularization methods, on the other hand, the combination of different methods like wavelet shrinkage and regularization methods was accomplished using the theory of Hilbert scales.

Last but not least several questions have arised during the work on this thesis which we note as starting points for further research.

Outlook

We present some questions which could be starting points to continue with the research on two-step methods.

1. How does the degree of ill-posedness in Hilbert scales change with the scale parameter?

   We assume that an inverse problem is solved by a two-step method where the data estimation operator S_λ smoothes with respect to a scale of spaces $\{Y_\nu\}$,
   $$S_\lambda : Y \to Y_\nu.$$
   The degree of ill-posedness is defined for an inverse problem
   $$(K : X \to Y, \mathcal{A}, \delta).$$
   If the operator S_λ is applied we consider the “new” inverse problem
   $$(K : X \to Y_\nu, \mathcal{A}, \tilde{\delta}).$$

The degree of ill-posedness depends on the space Y_ν which contains the data. The space Y_ν is in turn changed by data estimation. We are interested in a quantification of this dependency. Related to this question is the next one.

2. How does the singular system and the adjoint operator of $K : X \to Y$ change with Y?

   We assume the following setting: the operator $K : L_2 \to H^t$ is continuously invertible but compact as operator $K : L_2 \to L_2$. The singular values of the compact operator behave like $\sigma_n \simeq n^{-t}$.

   We solve $K : L_2 \to L_2$ by a two-step method where the first step results in solving $K : L_2 \to H^\eta$ with $\eta \leq t$. Does this result in $\sigma_n \simeq n^{-t+\eta}$? And how does the adjoint operator $K^* : H^\eta \to L_2$ change with η?

3. What do we gain from a better estimate for the approximation error of the reduced methods?

   For the approximation error we have used the estimates of the classical methods. The main work in this thesis was done in proving that, as long as $\gamma > 1/2$, the reduced operator $R_{\alpha,\gamma} : L_2 \to L_2$ stays bounded. The fact that filter reduction reduces the amount of damping and hence reduces the approximation error was not exploited.

4. Reduced filter functions: Conditions and Numerical Realization?

   During the course of this thesis we have claimed that filter reduction is possible whenever the adjoint operator is involved in the computation of the regularization. For the numerical realization we have suggested an approximation based on a Taylor expansion.

5. Application of two-step methods to nonlinear operator equations:

   We regard two-step methods with their ability of weighting the influence of data estimation and regularization as a promising approach also for nonlinear operator equations. However, this demands additional skills not yet acquired.

Appendix A

Review of Stochastic

We present the basic notations necessary to define stochastic processes. Special emphasize is given to different approaches to white noise as a stochastic process.

A.1 Basic Definitions

Probability theory is concerned with the study of experiments whose outcomes are random. That means the outcomes cannot be predicted with certainty. The collection Ω of all possible outcomes of a random experiment is called a *sample space*. An element ω of Ω is called an *elementary event* (or a *sample point*). The following definitions are given for completeness.

Definition A.1 (Probability space)**.** *Let Ω be a given set and $\mathcal{A}$ a σ-*algebra *on Ω, i.e., $\mathcal{A}$ is a family of subsets of Ω with the following properties:*

1. $\emptyset \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$, *where* $A^C = \Omega \backslash A$ *is the complement of* A *in* Ω,
3. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair $(\Omega, \mathcal{A})$ is called a measurable space*. A* probability measure *P on a measurable space $(\Omega, \mathcal{A})$ is a function $P : \mathcal{A} \to [0,1]$ such that*

1. $P(\emptyset) = 0, \; P(\Omega) = 1$,
2. *if $A_1, A_2, \ldots \in \mathcal{A}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$) then*
$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{A}, P)$ *consisting of a measurable space and a probability measure is called a* probability space.

From now on $(\Omega, \mathcal{A}, P)$, or short Ω, denotes a *complete* probability space. We do not define completeness here, but we remark that every probability space can be made complete [Oks03].
The subsets A of Ω which belong to $\mathcal{A}$ are called $\mathcal{A}$*-measurable* sets. In a probability context these sets are called *events*. A function $Y : \Omega \to \mathbb{R}^d$ is called $\mathcal{A}$*-measurable* if

$$Y^{-1}(C) := \{\omega \in \Omega; Y(\omega) \in C\} \in \mathcal{A}$$

for all open sets $C \subset \mathbb{R}^d$.

A.1.1 Random Variables

A *random variable* X is an $\mathcal{A}$-measurable function $X : \Omega \to \mathbb{R}^d$. Every random variable induces a probability measure μ_X on $\mathbb{R}^d$ defined by

$$\mu_X(B) = P(X^{-1}(B)).$$

μ_X is called the *distribution* of X. If $\int_\Omega \left|X(\omega)\right| \mathrm{d}P(\omega) < \infty$ then the number

$$E[X] := \int_\Omega \left|X(\omega)\right| \mathrm{d}P(\omega) = \int_{\mathbb{R}^d} x \mathrm{d}\mu_X(x) \tag{A.1}$$

is called the *expectation* of X (with respect to P). The expectation is a linear operator.
Let $X : \Omega \to \mathbb{R}$ be a random variable. In particular the outcome of X is a real number. The *distribution function* F of X is defined by

$$F_X(x) = P[X \leq x] = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

If $F_X(x)$ is differentiable, then X is called a *continuous random variable* and $f_X = \frac{\mathrm{d}F_X}{\mathrm{d}x}$ is a non-negative function, called the *density function* of X. For the expectation of a continuous random variable X we have

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x.$$

We assume that a random variabale Y is defined by a function g according to $Y := g(X)$. The expectation of Y can be computed by $E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$.

We define the *nth moment* of a random variable to be the expectation $\mathrm{E}[X^n]$. Related to the second moment are the central second moment, or *variance*, and the *standard deviation.* The variance is defined by

$$\operatorname{var} X = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

The standard deviation σ_X is defined by

$$\sigma_X = \sqrt{\operatorname{var} X}.$$

Perhaps the most important continuous random variables are those whose density is of the form

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big].$$

For fixed parameters σ and μ the according random variable is called a *normal* random variable of *mean* μ and variance σ^2 and is denoted by $\mathcal{N}(\mu, \sigma^2)$. To verify that this formula defines a density we refer the reader to [IC88].

A.1.2 Families of Random Variables, Independence

If more than one random variable is defined on a probability space, the random variables are said to be *jointly distributed.* In addition to the distribution function of each random variable, there is a distribution function for the collection of the random variables. This is called the *joint distribution function* of the family of random variables. Let $X_1, \ldots, X_n$ be a family of jointly distributed random variables. The joint distribution function of $\mathbf{X} = (X_1, \ldots, X_n)$ is denoted by $F_{\mathbf{X}}$ and it is defined according to

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = P[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

Two jointly distributed random variables X and Y are said to be *independent* if

$$P[X \leq x, Y \leq y] = P[X \leq x]P[Y \leq y]$$

or $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

Definition A.2 (Stochastic process)**.** *Let T be a parameter space. For $t \in T$ let $X_t : \Omega \to \mathbb{R}^d$ be a random variable on the probability space $(\Omega, \mathcal{A}, P)$. The parameterized collection*

$$\{X_t\}_{t \in T}$$

of random variables is called a stochastic process *X.*

The *parameter space* T often is the halfline $[0, \infty)$, but it may also be an interval $[a, b]$, the non-negative integers or subsets of $\mathbb{R}^d$ for $d \geq 1$.
Sometimes stochastic processes are written as $X(t, \omega)$ instead of $X_t(\omega)$. In doing so we regard the process as a function of two variables

$$(t, \omega) \to X(t, \omega)$$

from $T \times \Omega$ into $\mathbb{R}^d$. For $t \in T$ fixed, the process is a random variable

$$\omega \to X(t, \omega) = X_t(\omega), \qquad \omega \in \Omega.$$

For $\omega \in \Omega$ fixed, we consider the process as function on T,

$$t \to X(t, \omega) = X_t(\omega), \qquad t \in T$$

which is called the *sample path* or *trajectory* of X_t. The collection of all sample paths is called the *sample path space* of the process X. In general the sample path space is a subspace of all real-valued functions defined on T.
A stochastic process is called *stationary* if its joint distribution is unchanged under time translations, which means that $X(t)$ and $Y(t) = X(t + \tau)$ have the same joint distribution. One example of a stochastic process is the so-called *Brownian motion*.

Definition A.3 (Brownian motion). *A stochastic process $\{X(t)\}_{t \geq 0}$ is called a* standard Brownian motion *if it fulfills the following properties:*

1. $X(0) = 0$ *with probability* 1.
2. *Every increment $X(t) - X(s)$ is normally distributed with mean zero and variance $\sigma^2 |t - s|$ for a fixed variance parameter σ.*
3. *For $0 = t_0 < t_1 < \ldots < t_n < \infty$, the increments $(X(t_i) - X(t_{i-1}))$, $1 \leq i \leq n$, are independent and distributed as in 2.*

The Brownian motion process is also associated with the names Bachelier, Einstein, Levy and Wiener.
Within the calculus of stochastic processes there are several notions of continuity, differentiability and integration. We mention *mean square differentiability* since this is used for a formal representation of white noise. Differentiability is defined with respect to the space parameter $t \in T$. Because of this we use the notation $X(t, \omega)$ which takes into account both parameters equally.

A stochastic process $X(t,\omega)$, $t \in T$ is called a *second order process* if for every $t \in T$ fixed it is $E[(X(t,\cdot))^2] < \infty$. A second-order process $X(t)$ is said to be *mean-square differentiable* at $t \in T$ if there is a second-order process $Y(s)$, $s \in T$, such that

$$\lim_{h \to 0} E[(h^{-1}(X(t+h) - X(t)) - Y(t))^2] = 0.$$

In the case of existence we write $Y(t) = X'(t)$.
We remark that the Brownian motion is not mean-square differentiable.

A.2 White Noise Processes

White noise processes are defined in several ways. We give a sketch of some of them. We remind the reader of the δ-functional, see standard books on functional analysis. Let f be a continuous function on $\mathbb{R}$. A generalized function $\delta(x)$ defined by the evaluation function

$$\int_{\mathbb{R}} f(x)\delta(x - x_0)\mathrm{d}x = f(x_0)$$

is called the *(Dirac)δ-functional*.
White noise is named according to white light. White light is light of all frequencies and is therefore characterized by a constant spectrum. With this analogy in mind we present the following definition of a *white noise process* as a stationary process with constant *spectral density*. For the definition of spectral density and covariance we refer the reader to [Kan79].

Definition A.4 (White noise process)**.** *A white noise process $W(t)$, $t \in \mathbb{R}$, is a stationary process either with*

1. *constant spectral density, or*

2. *the Dirac δ-functional as covariance, or*

3. *covariance and spectral density given by*

$$K(t) = Ne^{-iN|t|} \qquad \textit{and} \qquad f(x) = \frac{1}{\pi}\frac{1}{1 + (x/N)^2}$$

respectively. Here N is infinitely large.

A.2.1 A Formal Representation of White Noise

Let $X(t)$, $-\infty < t < \infty$, be a standard Brownian motion. This process is not mean-square differentiable. But if one formally considers the mean-square derivative $\dot{X}(t)$ of $X(t)$ the covariance function of $\dot{X}(t)$ turns out to be the Dirac δ-functional (see [Kan79] for details). Thus $\dot{X}(t)$ can be treated as white noise process and $W(t)$ can be formally represented as

$$W(t) = \frac{\mathrm{d}X(t)}{\mathrm{d}t}.$$

This representation suggests another definition of the white-noise processes.

Definition A.5. *A* white noise process $W(t)$ *is the formal derivative of a standard Brownian motion* $X(t)$. *In this case* $W(t) = \dot{X}(t)$ *is to be treated as functional that acts on continuously differentiable functions as follows:*

$$\int_a^b f(t)W(t)\mathrm{d}t = \int_a^b f(t)\mathrm{d}X(t) = f(t)X(t)|_a^b - \int_a^b X(t)\mathrm{d}f(t).$$

Since almost no sample path of $X(t)$ is differentiable, the definition of the integral $\int_a^b f(t)\mathrm{d}X(t)$ gives also a definition of integration with respect to a Brownian motion. We simply call this a *stochastic integral* and remark that there is extensive work available on such integrals.

We end our review on stochastic with the remark that stochastic processes can be considered as functions in Hilbert spaces. We refer to [Roz87] for details.

Bibliography

[ACC$^+$02] L. Antonelli, L. Carracciuolo, M. Ceccarelli, L. D'Amore, and A. Murli. Total variation regularization for edge preserving 3D SPECT imaging in high performance computing environments. In *Sloot, Peter M. A. (ed.) et al., Computational science - ICCS 2002. 2nd international conference, Amsterdam, the Netherlands, April. 21-24, 2002. Proceedings. Part 2. Berlin: Springer. Lect. Notes Comput. Sci. 2330, 171-180.* 2002.

[Ada75] Robert A. Adams. *Sobolev spaces.* Pure and Applied Mathematics, a Series of Monographs and Textbooks. Vol. 65. New York - San Francisco - London: Academic Press, Inc., a subsidiary of Harcourt Brace Jovanovich, Publishers. XVIII, 268 p., 1975.

[AS84] M.(ed.) Abramowitz and I.A.(ed.) Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables. Reprint of the 1972 ed.* A Wiley-Interscience Publication. Selected Government Publications. New York: John Wiley & Sons, Inc; Washington, D.C.: National Bureau of Standards. XIV, 1046 p., 1984.

[AS98] Felix Abramovich and Bernard W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129, 1998.

[Bak69] A.B. Bakusinskii. On the construction of regularizing algorithms in the presence of random noise. *Sov. Math., Dokl.*, 10:1376–1378, 1969.

[Bal04] Guillaume Bal. On the attenuated Radon transform with full and partial measurements. *Inverse Probl.*, 20(2):399–418, 2004.

[CDKP01] Albert Cohen, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191, 2001.

[CGLT03] L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory Probab. Appl.*, 48(3):426–446, 2003.

[CH05] Laurent Cavalier and Nicolas W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.

[CHR04] Albert Cohen, Marc Hoffmann, and Markus Reiss. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numerical Analysis*, 42(4), 2004.

[CIK99] Pao-Liu Chow, Ildar A. Ibragimov, and Rafail Z. Khasminskii. Statistical approach to some ill-posed problems for linear partial differential equations. *Probab. Theory Relat. Fields*, 113(3):421–441, 1999.

[CN04] Rolf Clackdoyle and Frédéric Noo. A large class of inversion formulae for the 2D Radon transform of functions of compact support. *Inverse Probl.*, 20(4):1281–1291, 2004.

[Coh03] Albert Cohen. *Numerical analysis of wavelet methods.* Studies in Mathematics and its Applications. 32. Amsterdam: North-Holland. xviii, 336 p., 2003.

[CT02] Laurent Cavalier and Alexandre Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123(3):323–354, 2002.

[Dau92] Ingrid Daubechies. *Ten lectures on wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. 61. Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics. xix, 357 p. , 1992.

[DDD04] Ingrid Daubechies, Michel Defrise, and Christine DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[Dic99] Volker Dicken. A new approach towards simultaneous activity and attenuation reconstruction in emission tomography. *Inverse Probl.*, 15(4):931–960, 1999.

[DJ94] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[DJ98] David L. Donoho and Iain M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Stat.*, 26(3):879–921, 1998.

[DJKP95] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: Asymptopia? *J. R. Stat. Soc., Ser. B*, 57(2):301–369, 1995.

[DM96] Volker Dicken and Peter Maass. Wavelet–galerkin methods for ill–posed problems. *J. Inv. Ill–Posed Prob.*, 1996.

[Don95] David L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126, 1995.

[EHN96] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems.* Mathematics and its Applications (Dordrecht). 375. Dordrecht: Kluwer Academic Publishers. viii, 321 p., 1996.

[EN05] Herbert Egger and Andreas Neubauer. Preconditioning Landweber Iteration in Hilbert Scales. *Numerische Mathematik*, Sept. 2005.

[FM04] Athanassios S. Fokas and Vangelis Marinakis. Reconstruction algorithm for the brain imaging techniques of PET and SPECT. In *Lipitakis, Elias A. (ed.), HERCMA 2003. Proceedings of the 6th Hellenic-European conference on computer mathematics and its applications, Athens, Greece, September 25-27, 2003. 2 volumes. Athens: LEA. 257-273.* 2004.

[GN04] J.-P. Guillement and R.G. Novikov. A noise property analysis of single-photon emission computed tomography data. *Inverse Probl.*, 20(1):175–198, 2004.

[HA98] M. Hegland and R.S. Anderssen. A mollification framework for improperly posed problems. *Numer. Math.*, 78(4):549–575, 1998.

[Heg95] Markus Hegland. Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization. *Appl. Anal.*, 59(1-4):207–223, 1995.

[Heu92] Harro Heuser. *Funktionalanalysis. Theorie und Anwendung. (Functional Analysis. Theory and Applications). 3., durchges. Aufl.* Mathematische Leitfäden. Stuttgart: Teubner. 696 S., 1992.

[HR04] Marc Hoffmann and Markus Reiss. Nonlinear estimation for linear inverse problems with error in the operator. Technical report, WIAS Preprint No. 990, 2004.

[HvW05] Bernd Hofmann and Lothar von Wolfersdorf. Some results and a conjecture on the degree of ill-posedness for integration operators with weights. *Inverse Probl.*, 21(2):427–433, 2005.

[IC88] Reza Iranpour and Paul Chacon. *Basic stochastic processes. The Mark Kac lectures.* New York: Macmillan Publishing Company; London: Collier Macmillan Publishers. xiv, 258 p., 1988.

[JL01] P. Jonas and A.K. Louis. A Sobolev space analysis of linear regularization methods for ill-posed problems. *J. Inverse Ill-posed Probl.*, 9, 2001.

[JS91] Iain M. Johnstone and Bernard W. Silverman. Discretization effects in statistical inverse problems. *J. Complex.*, 7(1):1–34, 1991.

[Kan79] D. Kannan. *An introduction to stochastic processes.* North Holland Series in Probability and Applied Mathematics. New York, Oxford: North Holland. XIII, 296 p., 1979.

[KP66] S.G. Krein and J.I. Petunin. Scales of Banach spaces. 1966.

[Kun01] L.A. Kunyansky. A new SPECT reconstruction algorithm based on the Novikov explicit inversion formula. *Inverse Probl.*, 17(2):293–306, 2001.

[Lam01] Lamm, Patricia K. Some recent developments and open problems in solution methods for mathematical inverse problems. Technical report, Preprint, August 2001, to appear in Proceedings, XXIV National Congress of Apllied and Computational Mathematics, Belo Horizonte, Brazil, Sept. 10-13, 2001.

[Lin05] Mathias Lindemann. *Approximation Properties of Non-Separable Wavelet-Bases with Isotropic Scaling Properties and their Relation to Besov Spaces.* PhD thesis, University of Bremen, 2005.

[LMR97] Alfred K. Louis, Peter Maass, and Andreas Rieder. *Wavelets. Theory and applications.* Pure and Applied Mathematics. A Wiley-Interscience Series of Texts, Monographs & Tracts. Chichester: John Wiley & Sons. xvii, 324 p., 1997.

[Lor05] Dirk A. Lorenz. *Wavelet shrinkage in signal & image processing: an investigation of relations and equivalences.* PhD thesis, University of Bremen, 2005.

[Lou84] Alfred K. Louis. Orthogonal function series expansions and the null space of the Radon transform. *SIAM J. Math. Anal.*, 15:621–633, 1984.

[Lou89] Alfred K. Louis. *Inverse und schlecht gestellte Probleme. (Inverse and ill-posed problems).* Teubner Studienbücher: Mathematik. Stuttgart: B. G. Teubner. 205 S., 1989.

[Luk98] Mark A. Lukas. Comparisons of parameter choice methods for regularization with discrete noisy data. *Inverse Probl.*, 14(1):161–184, 1998.

[Mai94] Bernard A. Mair. Tikhonov regularization for finitely and infinitely smoothing operators. *SIAM J. Math. Anal.*, 25(1):135–147, 1994.

[Mal98] Stéphane Mallat. *A wavelet tour of signal processing.* San Diego, CA: Academic Press. xxiv, 577 p., 1998.

[MP01] Peter Mathé and Sergei V. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38(6):1999–2021, 2001.

[MR96] Bernard A. Mair and Frits H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[Nas87] M.Z. Nashed. A new approach to classification and regularization of ill-posed operator equations. In *Inverse and ill-posed prob-*

lems, Alpine-U.S. Semin. St. Wolfgang/Austria 1986, Notes Rep. Math. Sci. Eng. 4, 53-75. 1987.

[Nat84] Frank Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.*, 18:29–37, 1984.

[Nat01a] Frank Natterer. Inversion of the attenuated Radon transform. *Inverse Probl.*, 17(1):113–119, 2001.

[Nat01b] Frank Natterer. *The mathematics of computerized tomography.* Classics in Applied Mathematics. 32. Philadelphia, PA: SIAM. xvii, 222 p., 2001.

[NC89] Douglas W. Nychka and Dennis D. Cox. Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Stat.*, 17(2):556–572, 1989.

[Neu88a] Andreas Neubauer. An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM J. Numer. Anal.*, 25(6):1313–1326, 1988.

[Neu88b] Andreas Neubauer. When do Sobolev spaces form a Hilbert scale ?. *Proc. Am. Math. Soc.*, 103(2):557–562, 1988.

[Neu92] Andreas Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Appl. Anal.*, 46(1-2):59–72, 1992.

[Neu00] Andreas Neubauer. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. *Numer. Math.*, 85(2):309–328, 2000.

[Nov02] Roman G. Novikov. An inversion formula for the attenuated X-ray transformation. *Ark. Mat.*, 40(1):145–167, 2002.

[NP99] Michael Nussbaum and Sergei Pereverzev. The Degree of Ill-Posedness in Stochastic and Deterministic Models. Technical report, Preprint 509, Weierstraß-Institut, Berlin, 1999.

[Oks03] Bernt Oksendal. *Stochastic differential equations. An introduction with applications. 6th ed.* Universitext. Berlin: Springer. xxiii, 360 p., 2003.

[Pal96] Victor P. Palamodov. An inversion method for an attenuated x-ray transform. *Inverse Probl.*, 12(5):717–729, 1996.

[Pie80] Albrecht Pietsch. *Operator ideals. Licenced ed.* North-Holland Mathematical Library, Vol. 20. Amsterdam - New York - Oxford: North-Holland Publishing Company. XVIII, 432 p., 1980.

[Ram02a] Ronny Ramlau. A steepest descent algorithm for the global minimization of the Tikhonov functional. *Inverse Probl.*, 18(2):381–403, 2002.

[Ram02b] Ronny Ramlau. Morozov's discrepancy principle for Tikhonov-regularization of nonlinear operators. *Numer. Funct. Anal. Optimization*, 23(1-2):147–172, 2002.

[Ram03] Ronny Ramlau. TIGRA – an iterative algorithm for regularizing nonlinear ill-posed problems. *Inverse Probl.*, 19(2):433–465, 2003.

[RCNB00] Ronny Ramlau, Rolf Clackdoyle, R. Noo, and Guillaume Bal. Accurate attenuation correction in SPECT imaging using optimization of bilinear functions and assuming an unknown spatially-varying attenuation distribution. *ZAMM, Z. Angew. Math. Mech.*, 80(9):613–621, 2000.

[Res05] Elena Resmerita. Regularization of ill-posed problems in Banach spaces. Convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[Rie03] Andreas Rieder. *Keine Probleme mit inversen Problemen.* Vieweg, 2003.

[Roz87] Yurij A. Rozanov. *Introduction to random processes. Transl. from the Russian by Birgit Röthinger.* Springer Series in Soviet Mathematics. Berlin etc.: Springer-Verlag. VIII, 117 p., 1987.

[RT04a] Ronny Ramlau and Gerd Teschke. Regularization of Sobolev Embedding Operators and Applications to Medical Imaging and Meteorological Data.Part I: Regularization of Sobolev Embedding Operators. *Sampling Theory in Signal and Image Processing*, 3(2), 2004.

[RT04b] Ronny Ramlau and Gerd Teschke. Regularization of Sobolev Embedding Operators and Applications to Medical Imaging and Meteorological Data.Part II: Data Driven Regularization and Applications. *Sampling Theory in Signal and Image Processing*, 3(3), 2004.

[SH64] V.N. Sudakov and L.A. Halfin. Statistical approach to the correctness of problems in mathematical physics. *Sov. Math., Dokl.*, 5:1094–1096, 1964.

[SN96] Gilbert Strang and Truong Nguyen. *Wavelets and filter banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.

[TA77] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems. Translation editor Frity John.* Scripta Series in Mathematics. New York etc.: John Wiley & Sons; Washington, D.C.: V. H. Winston & Sons. XIII, 258 p., 1977.

[Tas00] Carl Taswell. The what, how, and why of wavelet shrinkage denoising. *Computing in Science and Engineering*, 2(3):12–19, May/June 2000.

[Tau96] Ulrich Tautenhahn. Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.*, 33(6):2120–2130, 1996.

[Thi04] Henning Thielemann. Bounds for smoothness of refinable functions. *Linear Algebra and its Applications*, 391:245–260, November 2004.

[TM80] Oleh Tretiak and Charles Metz. The exponential Radon transform. *SIAM J. Appl. Math.*, 39:341–354, 1980.

[Tri83] Hans Triebel. *Theory of function spaces.* Monographs in Mathematics, Vol. 78. Basel-Boston-Stuttgart: Birkhäuser Verlag, 1983.

[Vil93] Lars F. Villemoes. Sobolev regularity of wavelets and stability of iterated filter banks. In *Progress in wavelet analysis and applications (Toulouse, 1992)*, pages 243–251. Frontières, Gif, 1993.

[Wah77] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.

[Wat66] G.N. Watson. *A treatise on the theory of Bessel functions. 2nd ed.* London: Cambridge University Press. VII, 804 p. , 1966.

[WD02] J. Wang and J. Du. The analytic study on the Radon transform and the attenuated Radon transform. *J. Math., Wuhan Univ.*, 22(4):369–373, 2002.