An Outline of Adaptive Wavelet Galerkin Methods for Tikhonov Regularization of Inverse Parabolic Problems

Stephan Dahlke*
Philipps–Universität Marburg
Fachbereich Mathematik und Informatik
Hans Meerwein Strasse, Lahnberge
35032 Marburg
Germany

Peter Maaß †
Universität Bremen
Zentrum für Technomathematik
Postfach 330440
28334 Bremen
Germany

Abstract

In this paper, we discuss some ideas how adaptive wavelet schemes can be applied to the treatment of certain inverse problems. The classical Tikhonov–Phillips regularization produces a numerical scheme which consists of an inner and an outer iteration. In its normal form, the inner iteration can be interpreted as a boundedly invertible operator equation which can be handled very efficiently by using a stable wavelet basis. This general framework is illustrated by an application to the inverse heat equation.

Key Words: Inverse parabolic problems, Tikhonov regularization, wavelet bases, adaptive refinements.

AMS Subject classification: 42C40, 65J20, 65N12, 65N30, 65M32

*The work of this author has been partially supported by Deutsche Forschungsgemeinschaft, Grant Da 360/4–1.

†Partially supported by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie under grant number BMBF-03-MSM1HB.

1 Introduction

Due to its theoretical challenges and its practical importance for many industrial applications the theory of regularization methods for inverse problems has gained increasing interest in the mathematical community over the last two decades. Excellent introductions to this field can be found e.g. in [12, 14, 16]. In this article we aim at presenting a framework for adaptive Tikhonov regularization and its realization by adaptive wavelet methods for parabolic differential equations. Moreover, in order to highlight the main ideas we will only consider inverse problems with a linear or an affine linear operator, e.g., parameter estimation problems for heat transfer equations. Hence we consider a compact operator A between Hilbert spaces X and Y and a corresponding operator equation

$$Ax = y \ , \tag{1.1}$$

where x is the searched for function and y denotes perfect data, however we assume that only some observed data y^δ with a known error bound $||y - y^\delta|| \leq \delta$ is given.

Tikhonov–Phillips regularization of such an ill–posed problem is achieved by replacing the linear equation (1.1) by the minimization problem

$$\begin{aligned} &\textit{find } x_\alpha^\delta \in X \textit{ which minimizes} \\ &T_\alpha(x) \ = \ \| \, Ax - y^\delta \, \|_Y^2 \ + \ \alpha \, \|x\|_X^2 \ . \end{aligned} \tag{1.2}$$

The idea of Tikhonov–Phillips regularization (1.2) is to control the influence of the data error in the regularized solution x_α^δ by adding a penalty term. The unique minimizer of (1.2) is given as the unique solution of the regularized normal equation

$$(A^*A \, + \, \alpha \, I) \, x_\alpha^\delta \ = \ A^* y^\delta \ . \tag{1.3}$$

Early results on the convergence of Tikhonov regularization methods were usually entirely based in function spaces, the additional influence of an appropriate discretization of the operator was hardly mentioned. For some exceptions see, e.g., [19, 20, 21]. However, any numerical scheme for solving inverse problems by Tikhonov regularization depends on at least two parameters (regularization parameter α, a parameter determining the discretization of the operator) and a stopping rule. Characterizing a numerical scheme for operator equations as adaptive usually refers to a nonlinear dependence of these ingredients on the given data y^δ. In this sense, any a posteriori stopping rule leads to an adaptive scheme. In this paper, we address adaptive schemes in a stronger sense: we analyze methods where the regularization parameter and the discretization spaces depend on the unknown solution and are chosen adaptively during the solution procedure without using a priori information. More precisely, we will consider the following framework for Tikhonov regularization:

- given data: $A, y^\delta, \delta, 0 < q < 1, \alpha_0$;
- outer iteration for determining the regularization parameter: choose iteratively $\alpha_n = q^n \alpha_0$, for each α_n determine a critical level of approximation $\epsilon = \epsilon(\alpha_n, \delta, y^\delta)$

for the solution. This parameter has to be chosen, such that the over all scheme realizes optimal convergence rates;

- inner iteration for determining the minimizer $x^{\delta}_{\alpha,\Lambda_\epsilon}$ of (1.3): $x^{\delta}_{\alpha,\Lambda_\epsilon}$ will be determined by suitable wavelet Galerkin approximations of the forward operator A^*A, these wavelet approximations will be chosen adaptively by using local a posteriori error estimates and an appropriate refinement strategy.

The paper is organized as follows. Section 2 contains the description of a model problem, which describes a parameter estimation problem for a heat equation. Section 3 deals with the approximation requirements of the outer iteration and the resulting adaptive approximation levels $\epsilon = \epsilon(\alpha, \delta, y^\delta)$. Finally Section 4 analyzes how to construct an adaptive wavelet Galerkin method which realizes the required levels of approximation.

2 A Model Problem

In this paper, we just aim at outlining a general approach for adaptive Tikhohonov regularization via wavelet discretizations. Hence we will not present any numerical results. However, in order to focus our ideas we will introduce a simple model problem, which serves as motivation for the subsequent sections. We do not present any new results in this section, to the contrary the content is rather classical and elementary, see, e.g., [23, 25]. Since we want to merge results from inverse problems and wavelet analysis, which have developed some conflicting notations and which sometimes even give different meanings to the same expressions, we would like to introduce some basic concepts in detail.

We consider inverse heat problems, the underlying differential equation is hence given by

$$u_t = \operatorname{div}\{\sigma \nabla u\}$$

on $x \in \Omega$, $t \in [0, T]$, where $\Omega \subset I\!R^2$ denotes a region with piecewise smooth boundary $\Gamma = \partial\Omega$.

The construction of wavelet Galerkin methods and their convergence properties have only recently been analyzed successfully, these results will be described in Section 4.

The inverse problems we consider will differ in terms of the given and/or the measured data: initial data $\mu = u(\cdot, 0)$; boundary data $a(x,t) = u(x,t)$ for $x \in \Gamma$, $t \in [0,T]$; observation at a fixed time instant $g(x) = u(x,T)$, observation on an interior region $b(x,t) = u(x,t)$ for $x \in \tilde{\Omega} \subset \Omega$, $t \in [0,T]$.

Let us first consider the standard inverse heat problem:

$$\text{given data: } a, g \text{ , searched for quantity: } \mu \text{ .}$$

For this model problem the forward operator $A = A(\mu)$ is defined as follows: For a fixed a let L denote the solution operator of the parabolic problem

$$u_t = \operatorname{div}\{\sigma \nabla u\} \text{ for } x \in \Omega$$

with initial data μ and boundary values a, i.e.,

$$L(\mu)(x,t) = u(x,t) \text{ for } x \in \Omega,\ t \in [0,T] \ .$$

Then

$$A(\mu)(x) = L(\mu)(x,T) \quad , \tag{2.1}$$

which leads to the formal description of the operator equation for the inverse problem

$$A(\mu) = g \ .$$

In order to allow the modelling of measurement error, A is considered as a mapping from $L_2(\Omega) \longrightarrow L_2(\Omega)$.

For non-zero boundary data a, the operator A is nonlinear. However, introducing $u^{\#}$ and $g^{\#} = u^{\#}(\cdot, T)$, where $u^{\#}$ denotes the solution with zero initial and non-zero boundary data, i.e.,

$$u_t = \mathrm{div}\{\sigma \nabla u\} \text{ for } x \in \Omega \quad , u(\cdot, 0) = 0 \ , \ a(x,t) = u(x,t) \text{ for } x \in \Gamma,\ t \in [0,T] \ ,$$

leads to an affine decomposition

$$A(\mu) = \tilde{A}\mu + g^{\#} \ ,$$

where $\tilde{A}$ is the linear operator, which solves

$$u_t = \mathrm{div}\{\sigma \nabla u\} \text{ for } x \in \Omega \quad , \ u(\cdot, 0) = \mu \ , \ 0 = u(x,t) \text{ for } x \in \Gamma,\ t \in [0,T] \ ,$$

and restricts the solution to its values at time T. Hence by combining the originally measured data g with the particular solution $g^{\#}$ via

$$\tilde{g} = g - g^{\#}$$

leads to a linear inverse problem $\tilde{A}\mu = \tilde{g}$.

A similar affine decomposition also holds for the inverse problem posed by

$$\text{given data: } b, \ \text{ searched for quantities: } (\mu, a) \ .$$

In all these cases including many variations, we are finally lead to consider an exponentially ill-posed linear operator equation.

3 A Framework for Adaptive Tikhonov Regularization

We consider Tikhonov regularization for solving a linear operator equation (1.1), i.e., we consider

$$x_\alpha^\delta \ = \ (A^*A + \alpha I)^{-1} A^* y^\delta \quad , \tag{3.1}$$

where $\|y - y^\delta\| \le \delta$ and A is a compact operator between Hilbert spaces X, Y

$$A : X \to Y \quad .$$

Now let us incorporate an adaptive Galerkin discretization of $(A^*A + \alpha I)$ in (3.1). I.e., we fix an approximation tolerance ϵ and construct an index set Λ_ϵ such that the corresponding approximate solution $x^\delta_{\alpha,\Lambda_\epsilon}$ satisfies a guaranteed error estimate

$$||x^\delta_\alpha - x^\delta_{\alpha,\Lambda_\epsilon}|| \le \text{const.}\ \frac{\epsilon}{\sqrt{\alpha}} \ . \tag{3.2}$$

An adaptive scheme, which realizes this condition will be described in Section 4.

The choice of α and ϵ determines the approximation properties of $x^\delta_{\alpha,\Lambda_\epsilon}$. So far we have discussed the solution of (1.2) for a fixed value of α. Let us now discuss how to determine a suitable value of α. We will choose α according to a discrepancy principle of the form (or some modification thereof)

$$\|Ax^\delta_{\alpha,\Lambda_\epsilon} - y^\delta\| = \tau\delta + \sigma\epsilon, \tag{3.3}$$

where $\tau > 1$ and σ sufficiently large, for a precise statement see Theorem 3.1. This still describes an idealized situation: in practice one never aims at solving (3.3) precisely, one rather chooses α from a sequence of test parameters and determines $\alpha_N \in \{\alpha_n = q^n\alpha_0 |\ n = 0, 1, 2, ...\}$, for a fixed $0 < q < 1$ by requiring

$$\|Ax^\delta_{\alpha_N,\Lambda_\epsilon} - y^\delta\| \ \le \ \tau\delta + \sigma\epsilon, \tag{3.4}$$

$$\|Ax^\delta_{\alpha_n,\Lambda_\epsilon} - y^\delta\| \ > \ \tau\delta + \sigma\epsilon \ \text{ for } \ n < N \ . \tag{3.5}$$

Hence the overall algorithm for computing $x^\delta_{\alpha,\Lambda_\epsilon}$ requires to solve $(N + 1)$ operator equations of type (3.1).

Of course the number of iterations N is a priori unknown. Thus an efficient procedure for obtaining sparse approximations of $(A^*A+\alpha I)$ in connection with a reliable strategy for selecting the approximation level ϵ will greatly reduce the numerical cost of the algorithm. Our main objective in this section is to determine an approximation level $\epsilon(\delta, \alpha)$ such that $x^\delta_{\alpha,\Lambda_\epsilon}$ exhibits optimal convergence rates. Note that the approximation level $\epsilon(\delta, \alpha)$ may change with α during the search process for the optimal regularization parameter α_N. This will later be used to choose coarser approximations for larger values of α.

As usual we assume that the generalized solution x^+ lies in the range of $(A^*A)^\nu$, that is,

$$x^+ = (A^*A)^\nu v, \quad \|v\| \le \varrho \ . \tag{3.6}$$

Moreover we restrict ourselves to smoothness assumptions of the order

$$0 < \nu \le \frac{1}{2} \ ,$$

since higher order regularity of x^+ does not further improve the convergence rate of $\|x^{\delta}_{\alpha,\Lambda_\epsilon} - x^+\|$. This is consistent with the theory of a posteriori parameter selection for classical Tikhonov regularization since – even when using the exact operator A – applying a discrepancy functional of type (3.3) limits optimal convergence rates to the range $0 < \nu \leq 1/2$. To avoid unnecessary notation we furthermore assume that

$$\overline{\text{range}\,(A)} = Y, \quad \|y^\delta\| > \delta, \quad \|A\| \leq 1 \ . \tag{3.7}$$

The starting for this investigation is a basic estimate which reveals the three error contributions in estimating $\|x^{\delta}_{\alpha,\Lambda_\epsilon} - x^+\|$. This result is a small adaptation of previously published standard estimates, see, e.g.,[19, 21].

Lemma 3.1 *Let x^+ be the generalized solution of $Ax = y$ and let $x^{\delta}_{\alpha,\Lambda_\epsilon}$ be defined by the discretized version of* (3.1). *Assume that $\|y - y^\delta\| \leq \delta$ and that x^+ obeys* (3.6). *Then,*

$$\|x^{\delta}_{\alpha,\Lambda_\epsilon} - x^+\| \leq \frac{\delta}{2\sqrt{\alpha}} + \frac{\epsilon\|x^+\|}{\sqrt{\alpha}} + \alpha^\nu c_{\nu,\alpha}(v)$$

where

$$c^2_{\nu,\alpha}(v) = \sum_{n\geq 0} \left\{ \frac{\alpha^{1-\nu}\sigma_n^{2\nu}}{(\sigma_n^2 + \alpha)} \ \ \langle v, u_n \rangle \right\}^2 \leq \{(1-\nu)^{1-\nu}\nu^\nu \varrho\}^2 \ .$$

In connection with the modified discrepancy principle (3.4) this result gives an optimal convergence rate.

Theorem 3.1 *If $\epsilon = O(\delta^p \alpha^q)$, with $0 < p, q,\ p + q = 1$, and if α is chosen by the modified discrepancy principle* (3.4) *with $\tau > 2/q$, $\sigma > 9||x^+||/4q$, then*

$$\|x^{\delta}_{\alpha,\Lambda_\epsilon} - x^+\| = O(\delta^{2\nu/(2\nu+1)}) \ .$$

The above theorem shows that we can e.g. choose $p = q = 1/2$ and still obtain optimal convergence rates. Such a choice is preferable for large values of α which is the case in the beginning of our iterative search for the optimal regularization parameter.

Optimal convergence rates cannot be achieved in general if $p + q < 1$.

4 Wavelet Galerkin Methods for Operator Equations

In recent years, much effort has been spent to design efficient numerical schemes based on wavelets. The most far-reaching results were obtained for operator equations of the form

$$\mathcal{A}u = f, \tag{4.1}$$

where $\mathcal{A} : H \to H'$ is a linear operator from a Hilbert space H into its normed dual H'. In our applications, H will typically be a Sobolev space H^t on some domain $\Omega \subset I\!\!R^d$ or on a closed manifold. We assume that $\mathcal{A}$ is *boundedly invertible* so that

$$\|\mathcal{A}v\|_{H'} \sim \|v\|_H, \ v \in H \tag{4.2}$$

holds. This setting fits perfectly to the normal equation (3.1) arising in the inner iteration, i.e., to the problem

$$x_\alpha^\delta \;=\; (A^*A+\alpha I)^{-1}A^*y^\delta \quad , \tag{4.3}$$

since, as already stated above, $\mathcal{A}=(A^*A+\alpha I)$ is boundedly invertible on $L_2(\Omega)$.

Before we discuss later on the specific problems arising in the numerical treatment of (4.3), let us briefly recall the basic numerical concepts. We are especially interested in *adaptive* schemes, and we shall focus on numerical algorithms based on *wavelets*, i.e., the basis functions are taken from a family $\Psi=\{\psi_\lambda,\lambda\in\mathcal{J}\}$ satisfying the following fundamental assumptions:

- Ψ induces *norm equivalences* for a whole scale of Sobolev spaces, $\|\sum_{\lambda\in\mathcal{J}}d_\lambda\psi_\lambda\|_{H^s}\sim(\sum_{\lambda\in\mathcal{J}}2^{2|\lambda|s}|d_\lambda|^2)^{1/2},\ s_0\le s\le s_1$;
- ψ_λ possesses the *cancellation property* $|\langle v,\psi_\lambda\rangle|\ \lesssim\ 2^{-|\lambda|m}|v|_{H^m(\mathrm{supp}\psi_\lambda)}$;
- the wavelets are *local* in the sense that $\mathrm{diam}(\mathrm{supp}\psi_\lambda)\sim 2^{-|\lambda|},\lambda\in\mathcal{J}$.

Nowadays, several constructions of bases satisfying these assumptions are available [4, 7, 8, 9]. Our goal is to develop a suitable Galerkin scheme to approximate the solution of (4.3). Therefore we consider subspaces of the form

$$S_\Lambda:=\{\psi_\lambda:\ \lambda\in\Lambda\},\qquad \Lambda\subset J, \tag{4.4}$$

and project our problem onto these spaces, i.e., the Galerkin approximation u_Λ is defined by

$$\langle\mathcal{A}u_\Lambda,v\rangle=\langle f,v\rangle\ ,\qquad v\in S_\Lambda. \tag{4.5}$$

In an adaptive scheme, the goal is always to find a *possibly small* set $\Lambda\subset J$ such that the actual error is below some given tolerance. In principle, such a scheme consists of the following three steps:

- compute the current Galerkin approximation u_Λ;
- estimate the error $\|u-u_\Lambda\|$ in some suitable norm;
- add wavelets if necessary which yields a new index set $\hat{\Lambda}$.

For the second step, one clearly needs an *a posteriori error estimator* since the exact solution u is unknown, and for the third step one has to develop a suitable *refinement strategy* so that the whole algorithm converges. In the wavelet setting, an error estimator can be easily constructed by employing assumption (4.2), norm equivalences, and Galerkin orthogonality, i.e.,

$$\begin{aligned}\|u-u_\Lambda\|_{H^t} &\sim \|\mathcal{A}(u-u_\Lambda)\|_{H^{-t}}\sim\|f-\mathcal{A}u_\Lambda\|_{H^{-t}}\\ &= \|r_\Lambda\|_{H^{-t}}\sim\left(\sum_{\mathcal{J}\backslash\Lambda}2^{-2t|\lambda|}|\langle r_\Lambda,\psi_\lambda\rangle|^2\right)^{1/2}.\end{aligned} \tag{4.6}$$

In our example for the inverse heat problem we have $\mathcal{A} : L_2(\Omega) \to L_2(\Omega)$, i.e. $t = 0$. From (4.6), we observe that the current error can be estimated by computing the wavelet coefficients of the *residual* $r_\Lambda = f - \mathcal{A}u_\Lambda$. Intuitively, the *residual weights* $\rho_\lambda := 2^{-t|\lambda|}|\langle r_\Lambda, \psi_\lambda\rangle|$ serve as local error indicators. Therefore a suitable refinement strategy can be derived by adding those wavelets which produce large entries in the expansion of the residual, i.e., we define the new index set $\hat{\Lambda}$ in such a way that

$$\left(\sum_{\lambda\in\hat{\Lambda}\setminus\Lambda} 2^{-2t|\lambda|}|\langle r_\Lambda, \psi_\lambda\rangle|^2\right)^{1/2} \geq \beta \left(\sum_{\lambda\in\mathcal{J}\setminus\Lambda} 2^{-2t|\lambda|}|\langle r_\Lambda, \psi_\lambda\rangle|^2\right)^{1/2} \tag{4.7}$$

for some suitable parameter β. However, this strategy is not directly numerically realizable since catching the bulk of the residual requires knowing *all* its wavelet coefficients. Nevertheless, in [6], it was shown that a judicious variant of this idea exploiting the cancellation property of wavelets indeed leads to an implementable and convergent algorithm, i.e., given a tolerance ϵ, the adaptive scheme produces a final index set $\tilde{\Lambda}_\epsilon$ such that

$$\|u - u_{\tilde{\Lambda}_\epsilon}\| \leq \epsilon \tag{4.8}$$

by using only information on the given data. Moreover, in [5], subtle generalizations have been derived which yield asymptotically optimal schemes in the sense that (within a certain range) the convergence rate of best N–term approximation is achieved at a computational expense which stays proportional to the number $N = |\Lambda_\epsilon|$ of degrees of freedom. Furthermore, in [1], a first efficient numerical realization is documented.

As already stated above, we suggest to use this strategy for the numerical treatment of the basic problem (4.3),

$$x_\alpha^\delta \;=\; (A^*A + \alpha I)^{-1}A^*y^\delta \quad . \tag{4.9}$$

Clearly this problem fits perfectly into the framework described above. However, as explained in detail in [5, 6], the design of an implementable refinement strategy requires some *compressibility* properties of the underlying operator. For the special operators considered here, this issue will be further analyzed in the near future. Moreover, for an efficient implementation, the problem remains how to compute the entries of the associated stiffness matrix

$$(\mathcal{A}_\Lambda)_{\lambda,\lambda'} := \langle \mathcal{A}\psi_{\lambda'}, \psi_\lambda\rangle = \langle A\psi_{\lambda'}, A\psi_\lambda\rangle + \alpha\langle\psi_{\lambda'}, \psi_\lambda\rangle \tag{4.10}$$

and of the right–hand side

$$(A^*y^\delta)_\lambda = \langle y^\delta, A\psi_\lambda\rangle. \tag{4.11}$$

Fortunately, the adjoint operator A^* is not needed, but nevertheless the task is nontrivial since the operator A is induced by the forward problem (2.1), i.e., it is given as a parabolic equation. We intend to solve this problem with another fully adaptive scheme as we shall now explain. Following the basic investigations in [2, 3], we treat our parabolic equation as an abstract Cauchy problem

$$\begin{aligned} u'(t) + \mathcal{B}u(t) &= 0, \quad t \in (0, T], \\ u(0) &= u_0. \end{aligned} \tag{4.12}$$

Usually, this problem is treated by the *method of lines*. Discretization in space first leads to a block system of ordinary differential equations. However, as already outlined in [2, 3], for an adaptive approach the other discretization sequence, *first time then space*, which is classically known as the *method of Rothe* [24] seems to be preferable. Then (4.12) is viewed as an ordinary differential equation in some suitable Hilbert space which, due to stability reasons, is solved by an implicit scheme with time-step control. Then, in each step, a certain elliptic subproblem has to be solved. However, since these subproblems are boundedly invertible in the sense of (4.2), they can again be efficiently discretized by employing the well–known adaptive wavelet algorithm. Clearly, the convergence and efficiency of this strategy has to be analyzed in detail. This will be performed in the near future.

References

[1] A. Barinka, T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, and K. Urban, Adaptive wavelet schemes for elliptic problems – Implementation and numerical experiments, *SIAM J. Scientific Comp.* **23(3)**, 910–939 (2001).

[2] F. Bornemann, An adaptive multilevel approach to parabolic equations I. General theory and 1D implementations, *Impact Comput. Sci. Engrg.* **2**, 279–317 (1990).

[3] F. Bornemann, An adaptive multilevel approach to parabolic equations II. Variable–order time discretization based on a multiplicative error correction, *Impact Comput. Sci. Engrg.* **3**, 93–122 (1991).

[4] C. Canuto, A. Tabacco, and K. Urban, The wavelet element method, part II: Realization and additional features in 2d and 3d, *Appl. Comp. Harm. Anal.* **8**, 123-165 (2000).

[5] A. Cohen, W. Dahmen, and R. DeVore, Adaptive wavelet methods for elliptic operator equations – Convergence rates, *Math. Comp.* **70**, 22–75 (2001).

[6] S. Dahlke, W. Dahmen, R. Hochmuth, and R. Schneider, Stable multiscale bases and local error estimation for elliptic problems, *Appl. Numer. Math.* **23**, 21–47 (1997).

[7] W. Dahmen and R. Schneider, Composite wavelet bases for operator equations, *Math. Comput.* **68**, 1533–1567 (1999).

[8] W. Dahmen and R. Schneider, Wavelets on manifolds I: Construction and domain decomposition, *SIAM J. Math. Anal.* **31**, 184-230 (1999).

[9] W. Dahmen and R. Schneider, Wavelets with complementary boundary conditions — function spaces on the cube, *Res. in Math.* **34**, 255–293 (1998).

[10] V. Dicken and P. Maaß, Wavelet–Galerkin methods for ill–posed problems, *J. Inv. and Ill-posed Probl.* **4(3)**, 203–222 (1996).

[11] H.W. Engl, Discrepancy principles for Tikhonov regularization of ill–posed problems leading to optimal convergence rates, *J. Opti. Theory Appl.* **52**, 209–215 (1987).

[12] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Boston, (1996).

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

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

[15] J.T. King and A. Neubauer, A variant of finite-dimensional Tikhonov regularization with a–posteriori parameter choice, *Computing* **40**, 91–109 (1988).

[16] A.K. Louis, *Inverse und schlechtgestellte Probleme*, Teubner, Stuttgart (1989).

[17] A.K. Louis, P. Maaß, and A. Rieder, *Wavelets – Theorie und Anwendungen*, Teubner, Stuttgart (1994). English version: Wiley, Chichester.

[18] P. Maaß and R. Ramlau, Wavelet accelerated regularization methods for hyperthermia treatment planning, *Int. J. Imag. Sys. and Tech.*, **7**, 191–199 (1996).

[19] P. Maaß and A. Rieder, Wavelet–accelerated Tikhonov–regularisation with applications, in “Inverse Problems in Medical Imaging and Nondestructive Testing”, eds. H.W. Engl, A.K. Louis, and W. Rundell, Springer, Wien, New York, pp. 134–159 (1997).

[20] A. Neubauer, An a posteriori parameter selection choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates, *SIAM J. Numer. Anal.* **25**, 1313–1326 (1988).

[21] A. Neubauer, An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error, *Appl. Num. Math.* **4**, 507–519 (1988).

[22] S.V. Pereverzev, Optimization of projection methods for solving ill–posed problems, *Computing* **55** (1995).

[23] J. Reinhardt, On a sideways parabolic equation, *Inverse Problems* **13**, 297–309 (1997).

[24] E. Rothe, Zweidimensionale parabolische Randwertaufgabe als Grenzfall eindimensionaler Randwertaufgaben, *Math. Ann.* **102**, 650–670 (1930).

[25] M. Yamamoto and J. Zou, Simultaneous reconstruction of the initial temperature and heat radiation coefficient, *Inverse Problems* **17**, 1181–1202 (2001).