
Nonlinear Inverse Problems: Theoretical Aspects and Some Industrial Applications

Heinz W. Engl^{1,2} and Philipp Kügler²

¹ Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, A-4040 Linz, Austria heinz.engl@oeaw.ac.at

² Institut für Industriemathematik, Johannes Kepler Universität, A-4040 Linz, Austria kuegler@indmath.uni-linz.ac.at

1 Introduction

Driven by the needs from applications both in industry and other sciences, the field of inverse problems has undergone a tremendous growth within the last two decades, where recent emphasis has been laid more than before on nonlinear problems. This is documented by the wide current literature on regularization methods for the solution of nonlinear ill-posed problems. Advances in this theory and the development of sophisticated numerical techniques for treating the direct problems allow to address and solve industrial inverse problems on a level of high complexity.

Inverse problems arise whenever one searches for causes of observed or desired effects. Two problems are called inverse to each other if the formulation of one problem involves the solution of the other one. These two problems then are separated into a direct and an inverse problem. At first sight, it might seem arbitrary which of these problems is called the direct and which one the inverse problem. Usually, the direct problem is the more classical one. E.g., when dealing with partial differential equations, the direct problem could be to predict the evolution of the described system from knowledge of its present state and the governing physical laws including information on all physically relevant parameters while a possible inverse problem is to estimate (some of) these parameters from observations of the evolution of the system; this is called "parameter identification". Sometimes, the distinction is not so obvious: e.g., differentiation and integration are inverse to each other, it would seem arbitrary which of these problems is considered the direct and the inverse problem, respectively. But since integration is stable and differentiation is unstable, a property common to most inverse problems, one usually considers integration the direct and differentiation the inverse problem. Note also that integration is a smoothing process, which is inherently connected with the instability of differentiation.

Other important classes of inverse problems are

- *(Computerized) tomography* (cf. [Nat86]), which involves the reconstruction of a function, usually a density distribution, from values of its line integrals and is important both in medical applications and in nondestructive testing [ELR96b]. Mathematically, this is connected with the inversion of the Radon transform.
- *Inverse scattering* (cf. [CK92], [Ram86]), where one wants to reconstruct an obstacle or an inhomogeneity from waves scattered by those. This is a special case of *shape reconstruction* and closely connected to shape optimization [HN88]: while in the latter, one wants to construct a shape such that some outcome is optimized, i.e., one wants to reach a *desired* effect, in the former, one wants to determine a shape from measurements, i.e., one is looking for the cause for an *observed* effect. Here, uniqueness is a basic question, since one wants to know if the shape (or anything else in some other kind of inverse problem) can be determined uniquely from the data ("identifiability"), while in a (shape) optimization problem, it might even be advantageous if one has several possibilities to reach the desired aim, so that one does not care about uniqueness there.
- *Inverse heat conduction problems* like solving a heat equation backwards in time or "sideways" (i.e., with Cauchy data on a part of the boundary) (cf. [ER95], [BBC85]).
- *Geophysical inverse problems* like determining a spatially varying density distribution in the earth from gravity measurements (cf. [ELR96a]).
- Inverse problems in *imaging* like deblurring and denoising (cf. [BB98]).
- *Identification of parameters* in (partial) differential equations from interior or boundary measurements of the solution (cf. [BK89], [Isa98]), the latter case appearing e.g. in *impedance tomography*. If the parameter is piecewise constant and one is mainly interested in the location where it jumps, this can also be interpreted as a shape reconstruction problem (cf. [IN99]).

Detailed references for these and many more classes of inverse problems can be found e.g. in [EHN96], [Eng93], [EG87], [Gro93], [Lou89], [Kir96], [Hof99], [CER90].

The mathematical formulation of inverse problems leads to models that typically are *ill-posed*: According to Hadamard, a mathematical problem is called well-posed if

- for all admissible data, a solution exists,
- for all admissible data, the solution is unique and
- the solution depends continuously on the data.

If one of these properties is violated, the problem is called ill-posed. Neither existence nor uniqueness of a solution to an inverse problem are guaranteed. As mentioned, non-uniqueness is sometimes of advantage, then allowing to choose among several strategies for obtaining a desired effect. In practical

applications, one never has exact data, but only data perturbed by noise are available due to errors in the measurements or also due to inaccuracies the model itself. Even if their deviation from the exact data is small, algorithms developed for well-posed problems then fail in case of a violation of the third Hadamard condition if they do not address the instability, since data as well as round-off errors may then be amplified by an arbitrarily large factor. In order to overcome these instabilities one has to use *regularization* methods, which in general terms replace an ill-posed problem by a family of neighboring well-posed problems.

In this survey paper, we concentrate on regularization techniques for solving inverse and ill-posed problems that are nonlinear. We formulate these problems in functional analytic terms as nonlinear operator equations. Nevertheless, we start with the theory for linear problems in order to familiarize the reader with basic properties and definitions, then also relevant for the discussion of nonlinear problems. The latter will address both theoretical and computational aspects of two popular classes of regularization methods, namely Tikhonov regularization and iterative techniques. Finally, we present in some detail examples for nonlinear inverse problems appearing in iron and steel production as well as in quantitative finance and show how regularization methods were used for solving them in a numerically stable way.

2 Regularization Methods

A prototype for linear inverse problems are linear integral equations of the first kind such as

$$\int_G k(s, t)x(t) dt = y(s) \quad (s \in G) \quad (1)$$

with $k \in L^2(G \times G)$, $y \in L^2(G)$. A case of special importance is that k actually depends on $s - t$, i.e., y is a convolution of x and k ; solving (1) is then called *deconvolution*. For this and a collection of other linear inverse problems from various application fields we refer to [Eng93]. A simple parameter identification problem serves as our prototype example of a nonlinear inverse problem: In physical or technical applications, the physical laws governing the process may be known in principle, while actual values of some of the physical parameters in these laws are often unknown. For instance, in describing the heat conduction in a material occupying a three dimensional domain Ω whose temperature is kept zero at the boundary, the temperature distribution u after a sufficiently long time is modeled by

$$\begin{aligned} -\nabla \cdot (q(x)\nabla u) &= f(x) \quad x \text{ in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \quad (2)$$

where f denotes internal heat sources and q is the spatially varying heat conductivity. If one cannot measure q directly, one can try to determine q from internal measurements of the temperature u or from boundary measurements of the heat flux $q \frac{\partial u}{\partial n}$. One refers to this inverse problem also as an *indirect measurement* problem. Parameter identification problems like (2) appear, e.g., in geophysical applications and in non-destructive material testing. Note that (2) with unknown q is nonlinear since the relation between this parameter and the solution u , which serves as the data in the inverse problem, is nonlinear even if the direct problem of computing u with given q is linear.

Both these inverse problems turn out to be ill-posed, for obtaining a solution in a (numerically) stable way, one has to develop regularization methods. We will use (1) and (2) for illustrating regularization for linear and nonlinear inverse problems, respectively. Regularization methods replace an ill-posed problem by a family of well-posed problems, their solution, called *regularized solutions*, are used as approximations to the desired solution of the inverse problem. These methods always involve some parameter measuring the closeness of the regularized and the original (unregularized) inverse problem, rules (and algorithms) for the choice of these *regularization parameters* as well as convergence properties of the regularized solutions are central points in the theory of these methods, since only they allow to finally find the right balance between stability and accuracy (see below).

While the theory of regularization methods for linear ill-posed problems is by now rather comprehensive, it is still evolving and far from complete in the nonlinear case. Though we mainly focus on nonlinear inverse problems, we begin our survey with the theory for linear problems in order to give a first introduction into basic perceptions and terminologies.

2.1 Linear Inverse Problems

Starting point for our discussion is the operator equation

$$Tx = y, \quad (3)$$

where T denotes a bounded linear operator acting between Hilbert spaces X and Y . As concept of solution we use that of a best-approximate solution, which is the minimizer of the residual $\|Tx - y\|$, i.e., a least squares solution, that minimizes $\|x\|$ among all minimizers of the residual. Our goal is to approximate the best-approximate solution of (3) in the situation that the exact data y are possibly not known precisely and only perturbed data y^δ with

$$\|y - y^\delta\| \leq \delta \quad (4)$$

are available. Here, δ is called the noise level. Note that we use a deterministic error concept by assuming a bound in the Hilbert space norm and also

considering convergence there. A different approach which is followed recently in connection with "uncertainty" is to use a stochastic error concept and to consider convergence with respect e.g. to the Prokhorov metric on a space of probability measures, see e.g. [EW85], [BB01].

The operator T^\dagger which maps the exact data $y \in D(T^\dagger)$ to the best-approximate solution of (3) is known as the Moore-Penrose (generalized) inverse of T (see, e.g., [Gro77], [Nas76]) and its domain is given by

$$D(T^\dagger) = R(T) \dot{+} R(T)^\perp.$$

The Moore-Penrose inverse is bounded, i.e., the problem of determining the best-approximate solution of (3) is stable if and only if $R(T)$ is closed. Otherwise, solving (3) is ill-posed, the first and the third of Hadamard's conditions being violated even if we consider the best-approximate solution since $D(T^\dagger)$ is only dense in Y . The range $R(T)$ is non-closed especially if T is compact with $\dim R(T) = \infty$, for which the injectivity of a compact operator T is a sufficient condition if X is infinite dimensional. Since an integral operator like the one in (1) is compact under the conditions on k mentioned there, integral equations of the first kind with non-degenerate kernels are a prototype of ill-posed problems.

In the ill-posed case, $T^\dagger y^\delta$ cannot serve as a reliable approximation of $T^\dagger y$ due to the unboundedness. Instead, we are looking for a regularized solution which depends continuously on the noisy data (such that it can be computed in a stable way) and converges to $T^\dagger y$ as the noise level tends to zero with the regularization parameter properly chosen. We explain the construction of a regularization method for the important special case of a compact operator T and refer to [EHN96] for the non-compact situation.

If T is a compact operator, there exists a singular system $(\sigma_i; u_i, v_i)_{i \in N}$, which is defined as follows: With $T^* : Y \rightarrow X$ denoting the adjoint operator of T (introduced via the requirement that for all $x \in X$ and $y \in Y$, $\langle Tx, y \rangle = \langle x, T^*y \rangle$ holds), the $(\sigma_i^2)_{i \in N}$ are the non-zero eigenvalues of the self-adjoint operator T^*T (and also of TT^*), written down in decreasing order with multiplicity, $\sigma_i > 0$. Furthermore, the $(u_i)_{i \in N}$ are a corresponding complete orthonormal system of eigenvectors of T^*T (which spans $\overline{R(T^*)} = \overline{R(T^*T)}$), and the $(v_i)_{i \in N}$ are defined via

$$v_i := \frac{Tu_i}{\|Tu_i\|}.$$

As in the finite-dimensional situation (recall the singular value decomposition of a matrix), the $(v_i)_{i \in N}$ are a complete orthonormal system of eigenvalues of TT^* and span $\overline{R(T)} = \overline{R(TT^*)}$. This translates into the formulas

$$Tu_i = \sigma_i v_i \tag{5}$$

$$T^*v_i = \sigma_i u_i \quad (6)$$

$$Tx = \sum_{i=1}^{\infty} \sigma_i \langle x, u_i \rangle v_i \quad (x \in X) \quad (7)$$

$$T^*y = \sum_{i=1}^{\infty} \sigma_i \langle y, v_i \rangle u_i \quad (y \in Y), \quad (8)$$

where these infinite series converge in the Hilbert space norms of X and Y , respectively; (7) and (8) are called “singular value expansion” and are the infinite-dimensional analogues of the singular value decomposition.

If (and only if) T has a finite-dimensional range, T has only finitely many singular values, so that all infinite series involving singular values degenerate to finite sums. However, if there are infinitely many singular values (the generic case), they accumulate (only) at 0, i.e.,

$$\lim_{i \rightarrow \infty} \sigma_i = 0. \quad (9)$$

Since for $y \in D(T^\dagger)$, which holds if and only if the Picard Criterion

$$\sum_{n=1}^{\infty} \frac{|\langle y, v_n \rangle|^2}{\sigma_n^2} < \infty$$

is satisfied, the best-approximate solution of (3) has the series representation

$$T^\dagger y = \sum_{n=1}^{\infty} \frac{\langle y, v_n \rangle}{\sigma_n} u_n, \quad (10)$$

we see why (9) turns (3) into an ill-posed problem: errors in the Fourier components of y with respect to v_n , i.e., in $\langle y, v_n \rangle$, are multiplied by $\frac{1}{\sigma_n}$, a factor growing to infinity for $n \rightarrow \infty$ due to (9) (if $\dim R(T) = \infty$). Thus, especially errors in Fourier components of the data for large n , usually termed as “high frequency errors”, are strongly amplified. Also, the faster the decay of the σ_n , the stronger the error amplification, which also allows to quantify ill-posedness: one usually distinguishes between mildly, i.e., $\sigma_n = \mathcal{O}(n^{-\alpha})$ (with $\alpha > 1$), and severely, i.e., $\sigma_n = \mathcal{O}(e^{-n})$, ill-posed problems. Although, as we will see, the connection between the ill-posedness of a nonlinear problem and of its linearization is not as close as one might expect, this is usually also used as a (rough) quantification of ill-posedness in the nonlinear situation via the decay rate of the singular values of the linearization.

Regularization methods now are techniques that can handle these problems. In the linear compact case, they can be constructed and analyzed based on the singular value expansion: From (5)–(8) and (10) we see that

$$T^\dagger y = \sum_{n=1}^{\infty} \frac{\langle T^* y, u_n \rangle}{\sigma_n^2} u_n$$

holds. The basic idea for deriving a regularization method is to replace the amplification factors $\frac{1}{\sigma_n^2}$ by a filtered version $U(\alpha, \sigma_n^2)$, where the filter function $U(\alpha, \cdot)$ is piecewise continuous on $[0, +\infty[$ for a regularization parameter $\alpha > 0$ and converges to $\frac{1}{\lambda}$ as $\alpha \rightarrow 0$. This allows to introduce the regularized solution

$$x_\alpha := \sum_{n=1}^{\infty} U(\alpha, \sigma_n^2) \cdot \langle T^* y, u_n \rangle u_n \quad (11)$$

or

$$x_\alpha^\delta := \sum_{n=1}^{\infty} U(\alpha, \sigma_n^2) \cdot \langle T^* y^\delta, u_n \rangle u_n \quad (12)$$

in case of perturbed data y^δ fulfilling (4).

The conditions on the family $\{(U(\alpha, \lambda), \alpha > 0)\}$ under which x_α in fact converges to $T^\dagger y$ are stated in

Theorem 1. Let, for an $\varepsilon > 0$, $U : R^+ \times [0, \sigma_1^2 + \varepsilon] \rightarrow R$ fulfill the following assumptions

- for all $\alpha > 0$, $U(\alpha, \cdot)$ is piecewise continuous;
- there is a $C > 0$ such that for all
- (α, λ) , $|\lambda \cdot U(\alpha, \lambda)| \leq C$ holds
- for all $\lambda \neq 0$, $\lim_{\alpha \rightarrow 0} U(\alpha, \lambda) = \frac{1}{\lambda}$.

Then, for all $y \in D(K^\dagger)$,

$$\lim_{\alpha \rightarrow 0} U(\alpha, T^* T) T^* y = T^\dagger y.$$

holds.

Furthermore, we have the following stability estimate for the regularized solutions:

Theorem 2. Let U be as in Theorem 1, x_α and x_α^δ be defined by (11)–(12). For $\alpha > 0$, let

$$g_U(\alpha) := \sup\{|U(\alpha, \lambda)| / \lambda \in [0, \sigma_1^2]\}.$$

Then

$$\|x_\alpha - x_\alpha^\delta\| \leq \delta \cdot \sqrt{C g_U(\alpha)}$$

holds.

Different choices of the filter function $U(\alpha, \lambda)$ now lead to different regularization methods. The probably simplest choice satisfying the assumptions of Theorem 1 is

$$U_\alpha(\lambda) := \frac{1}{\alpha + \lambda},$$

then leading to *Tikhonov regularization*

$$x_\alpha^\delta = \sum_{n=1}^{\infty} \frac{\sigma_n}{\alpha + \sigma_n^2} \langle y^\delta, v_n \rangle u_n = (\alpha I + T^*T)^{-1} T^* y^\delta.$$

Here, the regularized solution x_α^δ can also be characterized in variational form as minimizer of the functional

$$x \rightarrow \|Tx - y^\delta\|^2 + \alpha \|x\|^2, \quad (13)$$

which allows to carry this method over to nonlinear problems (although the proofs of its properties are then different since the spectral theoretic foundation is lacking there). Without the additional penalty term $\alpha \|x\|^2$, this would be called "output least squares minimization", but would be unstable. Instead of $\alpha \|x\|^2$, one uses also penalty terms of the more general form $\alpha \|Lx\|^2$ with a suitable (usually differential) operator L ; the method will then approximate a least squares solution minimizing $\|Lx\|$.

Another widely used method is the truncated singular value expansion with

$$U(\alpha, \lambda) = \begin{cases} \frac{1}{\lambda} & \lambda \geq \alpha \\ 0 & \lambda < \alpha. \end{cases}$$

This yields

$$x_\alpha^\delta = \sum_{\substack{n=1 \\ \sigma_n^2 \geq \alpha}} \frac{\langle y^\delta, v_n \rangle}{\sigma_n} u_n, \quad (14)$$

where the small singular values ("high frequencies") are filtered out by a "low-pass filter". This method is in some sense optimal (cf. [EHN96]), but can only be used if the singular value expansion is readily available. Finally, we consider the method of asymptotic regularization which is based on the initial value problem

$$\begin{aligned} u'_\delta(t) + T^*T u_\delta(t) &= T^* y^\delta, \quad t \in R_0^+, \\ u_\delta(0) &= 0, \end{aligned} \quad (15)$$

where $u_\delta : R_0^+ \rightarrow X$. A motivation for looking at this initial value problem is that when the solution becomes stationary, it solves the (*Gaussian*) *normal equation*

$$T^*T x = T^* y. \quad (16)$$

Thus, in the noise free case, one expects the solution u of this initial value problem to tend, as $t \rightarrow \infty$, to the best-approximate solution $T^\dagger y$, which is characterized by (16). The regularized solution is now defined as

$$x_\alpha^\delta = u_\delta\left(\frac{1}{\alpha}\right)$$

and can be put into the general framework via

$$x_\alpha^\delta = U(\alpha, T^*T)T^*y^\delta$$

with

$$U(\alpha, \lambda) = \int_0^{\frac{1}{\alpha}} e^{-\lambda s} ds.$$

The regularization effect in this method is obtained by integrating the initial value problem not up to infinity but only up to an abscissa $1/\alpha$. The method can also be understood as a continuous version of the iterative Landweber method to be discussed in Section 2.2 since the latter can be derived from solving (15) by the forward Euler method with step size α , i.e.,

$$u_\delta(t + \alpha) \sim u_\delta(t) + \alpha T^*(y^\delta - T u_\delta(t)). \quad (17)$$

We will see that there, the regularization is achieved by stopping the iteration at a specific iteration index, which is a discrete analogue to stopping the integration of (15) early.

In any regularization method, the *regularization parameter* α plays a crucial role. As can be seen for instance from (14), its choice always represents a compromise between accuracy and stability: if α is too large, the series (10) is truncated too early, leading to a poor approximation of $T^\dagger y$. On the other hand, if α is chosen too small, possible data errors may already be amplified too strongly. For choosing the parameter, there are two general classes of options: *A-priori rules* define the regularization parameter as a function of the noise level only, i.e., $\alpha = \alpha(\delta)$, while in *a-posteriori rules*, α depends both on the noise level and the actual data, i.e., $\alpha = \alpha(\delta, y^\delta)$. An example for the latter is the so-called *discrepancy principle*, where α is chosen such that

$$\|Tx_\alpha^\delta - y^\delta\| = C\delta \quad (18)$$

holds (with some $C > 1$). Note that in (18) the determination of the parameter amounts to solving a nonlinear equation.

One can show (see [EHN96]) that *error-free* strategies, where $\alpha = \alpha(y^\delta)$ does not depend on δ , cannot lead to convergence as $\delta \rightarrow 0$ in the sense that $\lim_{\delta \rightarrow 0} x_\alpha^\delta = T^\dagger y$ for all y^δ satisfying (4) and all $y \in D(T^\dagger)$. Since this is only an asymptotic statement, these techniques may still occasionally work well for a

fixed noise level $\delta > 0$, see [HH93]. However, the knowledge and use of a bound for the data error as in (4) is necessary for the construction of regularization methods based on a sound theoretical foundation. The error-free strategies include the popular methods of generalized cross-validation ([Wah90]) and the *L-curve method* ([HO93]); for its non-convergence, see [EG94] and [Vog96].

Crucial questions in applying regularization methods are convergence *rates* and how to choose regularization parameters to obtain optimal convergence rates. By convergence rates we mean rates for the worst-case error

$$\sup\{\|x_\alpha^\delta - T^\dagger y\| \mid \|y - y^\delta\| \leq \delta\}$$

for $\delta \rightarrow 0$ and $\alpha = \alpha(\delta)$ or $\alpha = \alpha(\delta, y^\delta)$ chosen appropriately. For an ill-posed problem, no uniform rate valid for all $y \in Y$ can be given, convergence of any method for an ill-posed problem can be arbitrarily slow ([Sch85]), rates can only be obtained on compact subsets of X (cf. [LY98]), i.e., under additional assumptions on the solution $T^\dagger y$. For instance, under a *source condition* (with $\nu > 0$)

$$T^\dagger y \in R((T^*T)^\nu), \quad (19)$$

which can be (due to the fact that usually T is smoothing) thought of as an (abstract) a-priori smoothness condition, Tikhonov regularization converges with the rate

$$\|x_\alpha^\delta - T^\dagger y\| = O(\delta^{\frac{2\nu}{1+2\nu}}) \quad (20)$$

for the a-priori choice

$$\alpha \sim \delta^{\frac{2}{1+2\nu}}$$

and $\nu \leq 1$. This (as it turns out, optimal under (19)) rate is also achieved with the a-posteriori parameter choice (18), but only for $\nu \leq \frac{1}{2}$. For a-posteriori parameter choice rules that always lead to optimal convergence rates see [EG88] and [Rau84].

The typical total error behavior of a regularization method is shown in Figure 1: the regularization error $\|x_\alpha - T^\dagger y\|$ goes to 0 as $\alpha \rightarrow 0$, while the propagated data error $\|x_\alpha - x_\alpha^\delta\|$ grows without bound as $\alpha \rightarrow 0$. The difficulty in optimally choosing the regularization parameter is that the curves in Figure 1 are not computable.

For numerically solving an inverse problem, any regularization method has to be realized in finite-dimensional spaces. In fact, a regularization effect can already be obtained by a finite-dimensional approximation of the problem, where the approximation level plays the role of the regularization parameter. Projection methods based on this regularizing property contain the least squares projection, where the minimum norm solution of (3) is sought in a finite-dimensional subspace X_n of X , and the dual least-squares method, where (3) is projected onto a finite-dimensional subspace Y_n of Y before the

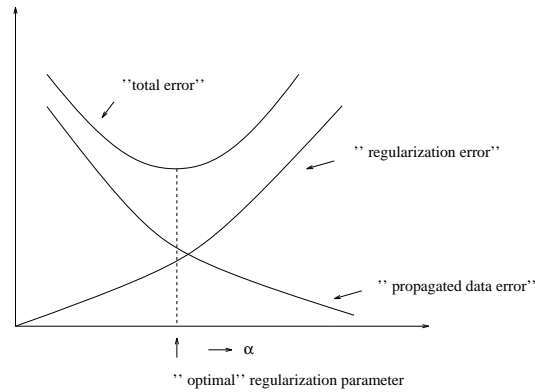


Fig. 1. Typical error behaviour

minimum norm solution of the resulting equation is computed, see [Nat77], [GN88], [Eng82]. However, error estimates for the case of noisy data and numerical experience show that at least for severely ill-posed problems the dimension of the chosen subspace has to be low in order to keep the total error small. Hence, for obtaining a reasonable accuracy, projection methods should be combined with an additional regularization method, e.g., with one of those discussed before, see [EN88], [PV90].

Methods which are closely related to regularization methods are *mollifier methods*, where one looks for a smoothed version of the solution, cf. [LM90]. Based on these ideas, one can develop methods where solutions of linear inverse problems can be computed fast by applying an "approximate inverse", which is usually an integral operator whose kernel, the "reconstruction kernel", can be precomputed (see EHN184). This is in turn closely related to the "linear functional strategy" from e.g. [And86]. While these are still linear methods, there are also *nonlinear methods* for solving linear ill-posed problems, e.g., the *Backus–Gilbert method* ([BG67], [KSB88]) and the *conjugate gradient method* ([Bra87], [Lou87], [Han95]). For details about these methods and other aspects we do not touch here (like the non-compact case and numerical aspects in the framework of combining regularization in Hilbert space with projection into finite-dimensional spaces) see [EHN96]. We mention in closing that there is a close connection between regularization and approximation by neural networks (cf. [BE00] and the references quoted there).

2.2 Nonlinear Inverse Problems

After this first introduction to the field of inverse and ill-posed problems we now turn to the nonlinear situation. There we will again meet many of the topics mentioned in the previous section, i.e., we will address rules for the choice

of the regularization parameter, observe again the typical error behavior from Figure 1 and discuss convergence rates. However, since the tools of spectral theory are no longer available, the construction and especially the analysis of regularization methods for nonlinear ill-posed problems becomes much harder.

Nonlinear inverse problems can be cast into the abstract framework of nonlinear operator equations

$$F(x) = y, \quad (21)$$

where F acts between two Hilbert spaces X and Y . The basic assumptions for a reasonable theory are that F is continuous and is weakly sequentially closed, i.e., for any sequence $x_n \in \mathcal{D}(F)$, $x_n \rightharpoonup x$ in X and $F(x_n) \rightharpoonup y$ in Y imply that $x \in \mathcal{D}$ and $F(x) = y$. (cf. [EKN89]). As opposed to the linear case, F is usually not explicitly given, but represents the operator describing the direct (also sometimes called "forward") problem. Considering for instance the parameter identification problem (2), the *parameter-to-output map* F maps the parameter q onto the solution u_q of the state equation or the heat flux $q \frac{\partial u_q}{\partial n}$. In inverse scattering, the operator F maps the shape of a scattering body onto the scattered far field pattern.

Neither existence nor uniqueness of a solution to (21) are guaranteed. Assuming for simplicity that the exact data y are attainable, i.e., that (21) in fact admits a solution and that the underlying model is thus correct, we again introduce a generalized solution concept (see [BEGNS94] for the non-attainable case): For $x^* \in X$, we call a solution x^\dagger of (21) which minimizes $\|x - x^*\|$ among all solutions an *x^* -minimum-norm solution* ($x^* - MNS$) of (21). The element x^* should include available a-priori information like positions of singularities in x if they happen to be available and will also be part of solution algorithms, see below.

In the following, we slur over the issue of uniqueness and consider problem (21) to be ill-posed if its solution does not depend continuously on the data y . Although, as mentioned above, the degree of ill-posedness of a nonlinear problem is frequently characterized via the decay of the singular values of its linearization, this is not always appropriate: It is shown in [EKN89] that a nonlinear ill-posed problem may have a well-posed linearization and that well-posed nonlinear problems may have ill-posed linearizations. If one accepts a quantification of ill-posedness via the linearization, then, e.g., inverse scattering is severely ill-posed with $\sigma_n = \mathcal{O}(\frac{1}{n!} (\frac{k}{n}))$, where k denotes the wave number (cf. [CK92]). Since the linearization is used in most numerical algorithms, this certainly makes sense.

As in the linear case, compactness of F , together with, e.g., the (local) injectivity of the operator again serves as a sufficient condition for the ill-posedness of (21), see [EKN89]:

Proposition 1. *Let F be a (nonlinear) compact and continuous operator, and let $\mathcal{D}(F)$ be weakly closed. Furthermore, assume that $F(x^\dagger) = y$ and that there exists an $\varepsilon > 0$ such that $F(x) = \hat{y}$ has a unique solution for all $\hat{y} \in \mathcal{R}(F) \cap U_\varepsilon(y)$. If there exists a sequence $\{x_n\} \subset \mathcal{D}(F)$ satisfying*

$$x_n \rightharpoonup x^\dagger \quad \text{but} \quad x_n \not\rightarrow x^\dagger$$

then F^{-1} - defined on $\mathcal{R}(F) \cap U_\varepsilon(y)$ - is not continuous in y .

Note that if $\mathcal{D}(F)$ happens to be compact then F^{-1} is continuous as soon as it exists due to the Arzela-Ascoli Theorem. This property allows to regularize a nonlinear ill-posed problem by simply restricting the domain of F to a compact set; however, this usually does not yield qualitative stability estimates. In the following, we survey two widely used approaches for solving nonlinear inverse problems in a stable way, namely Tikhonov regularization and iterative regularization methods.

Tikhonov Regularization

In Tikhonov regularization, problem (21) with data satisfying (4) is replaced by the minimization problem

$$\|F(x) - y^\delta\|^2 + \alpha \|x - x^*\|^2 \rightarrow \min, \quad x \in \mathcal{D}(F), \quad (22)$$

where $x^* \in X$ is an initial guess for a solution of (21), motivated from the linear case, see (13). For a positive regularization parameter α , minimizers always exist under the above-mentioned assumptions on F but need not be unique, whence we call *any* global minimizer of (22) a regularized solution x_α^δ . One can show that x_α^δ depends continuously on the data for α fixed and that x_α^δ converges towards a solution of (21) in a set-valued sense with $\alpha(\delta) \rightarrow 0$ and $\delta^2/\alpha(\delta) \rightarrow$ as δ tends to zero, see [EKN89].

The basic result on convergence rates for Tikhonov regularization is

Theorem 3. *Let x^\dagger be an element in the interior of the convex domain $\mathcal{D}(F)$. Furthermore, let F be Fréchet differentiable with*

$$\|F'(x^\dagger) - F'(x)\| \leq C \|x^\dagger - x\|$$

in a neighborhood of x^\dagger . If there exists an element $w \in Y$ satisfying the source condition

$$x^\dagger - x^* = (F'(x^\dagger)^* F'(x^\dagger))^\nu w \quad (23)$$

for some $\nu \in [1/2, 1]$ with $C\|w\| < 1$, then the (a priori) parameter choice

$$\alpha \sim \delta^{\frac{2}{2\nu+1}} \quad (24)$$

yields

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}(\delta^{\frac{2\nu}{2\nu+1}}).$$

In (23), $F'(x^\dagger)^*$ denotes the Hilbert space adjoint of the Fréchet-derivative. Formally, the source condition as well as the obtained rate correspond to (19) and (20). Again, (23) is an abstract smoothness condition on the difference between the true solution x^\dagger and the a-priori guess x^* used in (22), once more supporting the importance of the choice of the latter. Since (23) also plays a crucial role in the convergence analysis of iterative regularization methods, we postpone a further discussion. For the proof and variants of Theorem 3, see [EHN96].

The disadvantage of rule (24) is that the parameter depends on the smoothness index ν of the exact solution x^\dagger which is not known in practice. A slight variant of Tikhonov regularization which allows to prove the (then no longer always optimal) rate $\mathcal{O}(\sqrt{\delta})$ as long as (23) holds with $\nu \geq 1/2$ for the choice $\alpha(\delta) = \mathcal{O}(\delta^2)$ (now independent of the unknown ν) can also be found in [EHN96]. Turning to a-posteriori rules, the use of the discrepancy principle, where $\alpha(\delta, y^\delta)$ is defined as the solution of

$$\|F(x_\alpha^\delta) - y^\delta\| = C\delta, \quad (25)$$

is rather problematic since - in the nonlinear situation - problem (25) only admits a solution under severe additional assumptions, see [KS85]. For a (quite complicated) a posteriori strategy that always leads to optimal rates see [SEK93].

In (22), it is not obligatory to use the norm induced by the inner product in X as penalty term. Other possibilities include *maximum entropy* regularization

$$\|F(x) - y^\delta\|^2 + \alpha \int_{\Omega} x(t) \log \frac{x(t)}{x^*(t)} dt \rightarrow \min,$$

see [EL93], [Egg93], [EL96], [LA96], or *bounded variation* regularization

$$\|F(x) - y^\delta\|^2 + \alpha \int_{\Omega} |\nabla x(t)| dt \rightarrow \min, \quad (26)$$

which enhances sharp features in x as needed in, e.g., image reconstruction, see [Rud94], [NS98], [Sch02].

With respect to the numerical implementation of Tikhonov regularization one can relax the task of exactly solving problem (22) to looking for an element $x_{\alpha,\eta}^\delta$ satisfying

$$\|F(x_{\alpha,\eta}^\delta) - y^\delta\|^2 + \alpha \|x_{\alpha,\eta}^\delta - x^*\|^2 \leq \|F(x) - y^\delta\|^2 + \alpha \|x - x^*\|^2 + \eta$$

for all $x \in \mathcal{D}(F)$ with η a small positive parameter, see [EKN89]. Tikhonov regularization combined with finite dimensional approximation of X (and of F , see also Section 2.2) is discussed e.g. in [Neu89], [NS90].

However, finding a global minimizer (even only approximately) to a nonlinear optimization problem is in general not an easy task. Numerical experience shows that the functional in (22), which is in general not convex (unlike (13) in the linear case) has usually many local minima in which a descent method tends to get stuck if the underlying problem is ill-posed. Since furthermore the determination of an appropriate regularization parameter α can require high computational efforts, iterative regularization methods are an attractive alternative.

Iterative Methods

A first candidate for solving (21) in an iterative way could be Newton's method

$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k)), \quad (27)$$

starting from an initial guess x_0 . Even if the iteration is well-defined and $F'(\cdot)$ is invertible for every $x \in \mathcal{D}(F)$, the inverse is usually unbounded for ill-posed problems (e.g. if F is continuous and compact). Hence, (27) is inappropriate since each iteration means to solve a linear ill-posed problem, and some regularization technique has to be used instead. For instance, Tikhonov regularization applied to the linearization of (21) yields the Levenberg Marquardt method (see [Han97])

$$x_{k+1} = x_k + (F'(x_k)^* F'(x_k) + \alpha_k I)^{-1} F'(x_k)^* (y - F(x_k)), \quad (28)$$

where α_k is a sequence of positive numbers. Augmenting (28) by the term

$$-(\alpha_k I + F'(x_k)^* F'(x_k))^{-1} \alpha_k (x_k - x^*)$$

for additional stabilization gives the *iteratively regularized Gauss-Newton method* (see [Bak92], [BNS97])

$$x_{k+1} = x_k + (F'(x_k)^* F'(x_k) + \alpha_k I)^{-1} [F'(x_k)^* (y - F(x_k)) - \alpha_k (x_k - x^*)]. \quad (29)$$

Usually, x^* is taken as x_0 , but this is not necessary. As (28) and (29), most iterative methods for solving the nonlinear ill-posed problem (21) are based on solving the normal equation

$$F'(x)^* (F(x) - y) = 0 \quad (30)$$

via successive iteration starting from x_0 . Equation (30) is the first-order optimality condition for the nonlinear output least-squares problem

$$\frac{1}{2} \|y - F(x)\|^2 \rightarrow \min, \quad x \in \mathcal{D}(F). \quad (31)$$

Alternatively to Newton type methods like (28) and (29), methods of steepest descent like the *Landweber iteration*

$$x_{k+1} = x_k + F'(x_k)^*(y - F(x_k)), \quad (32)$$

see [HNS95], are used, where the negative gradient of the functional in (31) determines the update direction for the current iterate. From now on, we shall use x_k^δ in our notation of the iterates in order to take possibly perturbed data y^δ with (4) into account.

In the ill-posed case, due to the instability inherent in (21), it is common to all iterative methods that the iteration must not be arbitrarily continued. Instead, an iterative method only then can become a regularization method, if it is stopped “at the right time”, i.e., only for a suitable stopping index k_* , the iterate $x_{k_*}^\delta$ yields a stable approximation to the solution x^\dagger of (21). Due to the ill-posedness, a mere minimization of (31), i.e., an ongoing iteration, leads to unstable results and to a typical error behavior as shown in Figures 2 and 3, compare also to Figure 1. While the error in the output decreases as

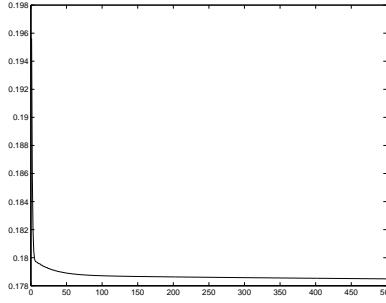


Fig. 2. $\|F(x_k^\delta) - y^\delta\|$ vs. k

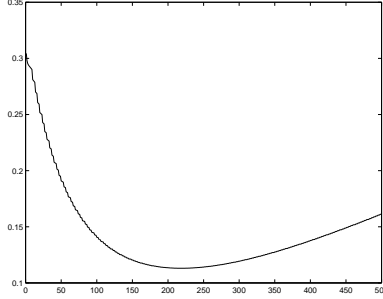


Fig. 3. $\|x_k^\delta - x^\dagger\|$ vs. k

the iteration number increases, the error in the parameter starts to increase after an initial decay.

Again, there are two classes of methods for choosing the regularization para-

meter, i.e., for the determination of k_* , namely a-priori stopping rules with $k_* = k_*(\delta)$ and a-posteriori rules with $k_* = k_*(\delta, y^\delta)$. Once more, the discrepancy principle, where k_* now is determined by

$$\|y^\delta - F(x_{k_*}^\delta)\| \leq \tau\delta < \|y^\delta - F(x_k^\delta)\|, \quad 0 \leq k < k_*, \quad (33)$$

for some sufficiently large $\tau > 0$ is a widely used representative for the latter. As opposed to (25) for Tikhonov regularization, (33) now is a rule easy to implement, provided that an estimate for the data error as in (4) is available. The discrepancy principle for determining the index k_* is based on stopping as soon as the residual $\|y^\delta - F(x_k^\delta)\|$ is in the order of the data error, which is somehow the best one should expect. For solving (21) when only noisy data y^δ with (4) are given, it would make no sense to ask for an approximate solution \tilde{x} with $\|y^\delta - F(\tilde{x})\| < \delta$, the price to pay would be instability.

Iterative regularization methods are also used for linear inverse problems such as (3). Some of them then can even be analyzed by means of the general theory provided by Theorems 1 and 2. For instance, the “linear version” of Landweber iteration (32), i.e.,

$$x_{k+1}^\delta = x_k^\delta + T^*(y^\delta - Tx_k^\delta),$$

see also (17), is represented by the filter function

$$U(k, \lambda) = \sum_{j=0}^{k-1} (1 - \lambda)^j$$

with k^{-1} playing the role of α . However, for nonlinear problems (21) the investigation of an iterative method is much more complicated and has mostly to be done for each class of methods individually. Some kind of general framework is provided in [ES00].

Theoretical studies of iterative methods for nonlinear ill-posed problems based on a fixed point formulation of the nonlinear problem (21) under contractivity and nonexpansivity assumptions on the fixed point operator can be found in [Vas87], [Vas92], [VA95], [Vas98]. The Mann iteration and its variants, see [Man53], [OG69], [Gro72] are further popular methods for solving fixed point equations. There, the basic principle is to feed the fixed point operator with an weighted average of the previous iterates. For ill-posed problems, it has been analyzed in [EL01], [KL03].

Though it is also possible to formulate the methods (32), (28) and (29) as fixed point operations, the necessary assumptions on the corresponding fixed point operators may become too restrictive for a reasonable theory (see [Sch95], [ES00]) such that also alternative approaches for their investigation are considered. In [HNS95], the convergence analysis of the Landweber iteration (32)

is carried out under the following assumptions: for a ball $\mathcal{B}_\rho(x_0)$ of radius ρ around x_0 with

$$\mathcal{B}_\rho(x_0) \subset \mathcal{D}(F), \quad (34)$$

the Fréchet-differentiable forward operator F is required to satisfy

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq \eta \|F(\tilde{x}) - F(x)\|, \quad x, \tilde{x} \in \mathcal{B}_\rho(x_0) \quad (35)$$

with $\eta < 1/2$. If furthermore, the Fréchet derivative is locally bounded by one, i.e.,

$$\|F'(x)\| \leq 1, \quad x \in \mathcal{B}_\rho(x_0), \quad (36)$$

at least local convergence of the iterates x_k to a solution of (21) in $\mathcal{B}_{\rho/2}(q_0)$ can be guaranteed. Together with (34) these assumptions also guarantee that all iterates x_k remain in $\mathcal{D}(F)$, which makes the iteration well-defined. In case of noisy data y^δ not belonging to the range of F , the iterates x_k^δ cannot converge. Still, condition (35) again forces the iterates x_k^δ to remain in $\mathcal{D}(F)$ and allows a stable approximation $x_{k_*}^\delta$ of a solution to (21), provided that the iteration is terminated after $k_* = k_*(\delta, y^\delta)$ steps according to the discrepancy principle (33) with τ satisfying

$$\tau > 2 \frac{1 + \eta}{1 - 2\eta} > 2. \quad (37)$$

In order to fulfill (36) for a (locally) bounded Fréchet derivative, one eventually has to rescale (21), i.e., instead to consider

$$\lambda F(x) = \lambda y. \quad (38)$$

If λ is chosen appropriately, then (36) holds, while condition (35) is scaling invariant, meaning that the requirement $\eta < 1/2$ cannot be weakened. At the first glance, condition (35) looks like a standard closeness assumption. In [HNS95] it is compared to the weaker Fréchet estimate

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq C \|\tilde{x} - x\|^2$$

for a Lipschitz continuous F' , in [Sch95] even a geometric interpretation is given. Summarizing, (35) can be seen as a nonlinearity condition on F , actually requiring that the nonlinearity of F must not be too strong. Note that a linear operator F clearly would satisfy this condition. However, the conditions (36) and (35) only yield stability, i.e., the regularized solution $x_{k_*}^\delta$ depends continuously on y^δ , and convergence, i.e., $x_{k_*}^\delta \rightarrow x^\dagger$ for $\delta \rightarrow 0$, of the Landweber method, but not a convergence rate.

As already mentioned, for any iterative regularization method, the rate of convergence of $x_k \rightarrow x^\dagger$ for $k \rightarrow \infty$ (in case of exact data) or $x_{k_*}^\delta \rightarrow x^\dagger$ for $\delta \rightarrow 0$ may be arbitrarily slow. As in Tikhonov regularization, convergence rate estimates can only be obtained under a source condition accompanied by additional assumptions. For Landweber iteration, a typical rate result reads as (see [HNS95])

Theorem 4. Assume that x^\dagger is a solution of (21) in $\mathcal{B}_{\rho/2}(x_0)$ and that F satisfies (36), (35) and

$$F'(x) = R_x F'(x^\dagger), \quad x \in \mathcal{B}_\rho(x_0), \quad (39)$$

where $\{R_x \mid x \in \mathcal{B}(x_0)\}$ is a family of bounded linear operators $R_x : Y \rightarrow Y$ with

$$\|R_{\tilde{x}} - I\| \leq C \|\tilde{x} - q^\dagger\|, \quad \tilde{x} \in \mathcal{B}_\rho(x_0) \quad (40)$$

for a positive constant C . If $x^\dagger - x_0$ fulfills the source condition (23) with some $\nu \in (0, 1/2]$ and $\|w\|$ sufficiently small, then

$$\|x^\dagger - x_{k_*}^\delta\| = \mathcal{O}(\delta^{\frac{2\nu}{2\nu+1}}), \quad (41)$$

where $x_{k_*}^\delta$ is defined according to the discrepancy principle (33).

Conditions (39) and (40) mean that the derivative of F at any point $x \in \mathcal{B}_\rho(x_0)$ can be decomposed into $F'(x^\dagger)$ and an operator which is bounded and boundedly invertible on the range of $F'(x^\dagger)$, such that for the linearized problem, the part which changes with the linearization point is well-posed. For a linear operator F we would have $R_x = I$, therefore (39) can be considered as a further restriction on the nonlinearity of F .

The convergence (rate) results for (32) given in [HNS95] are reproven in [DES98], where the assumptions (39) and (40) are replaced by a Newton-Mysovskii condition on F , i.e.,

$$\|(F'(x) - F'(x^\dagger))F'(x^\dagger)^\sharp\| \leq C_{NM} \|x - x^\dagger\|, \quad x \in \mathcal{D}(F). \quad (42)$$

Here $F'(x^\dagger)^\sharp$ denotes a left inverse of $F'(x^\dagger)$. Furthermore, a logarithmic type source condition

$$\exists w \in Y : \quad x^\dagger - x_0 = g_p(F'(x^\dagger)^* F'(x^\dagger))w \quad (43)$$

with

$$g_p(\lambda) := \begin{cases} \left(\ln \frac{\exp(1)}{\lambda} \right)^{-p} & \text{for } 0 < \lambda \leq 1 \\ 0 & \text{else.} \end{cases} \quad (44)$$

is used in [DES98] in order to derive the rate

$$\|x^\dagger - x_{k_*}^\delta\| = \mathcal{O}(-\ln \delta)^{-p}.$$

The motivation for this is that source conditions of the type (23) are too restrictive for problems where the operator $F'(x)$ is strongly smoothing, i.e., for severely ill-posed problems. There, (43) is more appropriate since it gives rise to interpretable conditions and still allows to give a rate estimate, although a slower one. Discussions of (43) and (44) are especially led in [Hoh97] in the context of severely ill-posed inverse scattering problems. Further variants of

Landweber iteration are discussed in [Sch98] and [Sch95].

In the field of Newton type iteration methods, the Levenberg Marquardt iteration (28) has been analyzed in [Han97]. There, convergence and stability of the scheme in combination with the discrepancy principle was proven essentially under the assumption that

$$\|F(x) - F(\tilde{x}) - F'(\tilde{x})(x - \tilde{x})\| \leq \tilde{C} \|F(x) - F(\tilde{x})\| \|x - \tilde{x}\| \quad x, \tilde{x} \in \mathcal{B}(x^\dagger), \quad (45)$$

if the parameter α_k is chosen such that

$$\|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_{k+1}^\delta - x_k^\delta)\| \leq \rho \|y^\delta - F(x_k^\delta)\|$$

is satisfied with some $\rho < 1$. A convergence rate result for (28) is still missing. This is different to the iteratively regularized Gauss-Newton method (29) discussed [BNS97]. There, the source condition (23) already is needed in order to obtain stability and convergence of the iterates with the sequence of regularization parameters chosen as

$$\alpha_k > 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r, \quad \lim_{k \rightarrow \infty} = 0$$

for some $r > 1$. If the iteration is terminated a priori after k_* steps with

$$\delta \sim \alpha_{k_*}^{\nu+1/2},$$

the Lipschitz continuity of the Fréchet derivative F' suffices to prove convergence (rates) with $\nu \in [1/2, 1]$ in (23) (and in (41)). Using the discrepancy principle (33) as stopping rule, convergence (rates) for $\nu \in [0, 1/2]$ are obtained if F satisfies

$$\begin{aligned} F'(\tilde{x}) &= R(\tilde{x}, x)F'(x) + Q(\tilde{x}, x) \\ \|I - R(\tilde{x}, x)\| &\leq C_R \quad \tilde{x}, x \in \mathcal{B}_{2\rho}(x_0) \\ \|Q(\tilde{x}, x)\| &\leq C_Q \|F'(x^\dagger)(\tilde{x} - x)\| \end{aligned} \quad (46)$$

with ρ , C_R and C_Q sufficiently small. Similar to (39) and (40), these conditions guarantee that the linearization is not too far away from the nonlinear operator. At first sight, Newton type methods would be considered to converge much faster than Landweber iteration; this is of course true in the sense that an approximation to a solution of (21) with a given accuracy can be obtained by fewer iteration steps. However, since a single iteration step in (28) or (29) is more expensive than in (32) and also since the instability shows its effect earlier in Newton type methods so that the iteration has to be stopped earlier, it cannot be said that Newton type methods are in general preferable for ill-posed problems to the much simpler Landweber method.

Nearly all assumptions in the theory sketched above as well as the iteration

schemes themselves are formulated in terms of the Fréchet derivative and its adjoint operator. We show how Landweber iteration is realized for our prototype parameter identification problem (2). The ideas presented then also apply in a similar way to other methods as (28) and (29).

In a first step, we translate the problem into a Hilbert space framework and therefore consider the underlying partial differential equation in its weak operator formulation

$$A(q)u = \hat{f} \quad (47)$$

with $A(q) : H_0^1(\Omega) \rightarrow H^{-1}(\Omega)$ and $\hat{f} \in H^{-1}(\Omega)$ defined by

$$(A(q)u, v) = \int_{\Omega} q(x) \nabla u \nabla v \, dx \quad \text{and} \quad (\hat{f}, v) = \int_{\Omega} f v \, dx.$$

For a set $\mathcal{D}(F) \subset X = H^s(\Omega)$ (with $s > d/2$ where d is the dimension of Ω) of admissible parameters q , the direct problem (47) admits a unique solution $A(q)^{-1} \hat{f} \in H_0^1(\Omega)$ which will be denoted by u_q in order to emphasize its dependence on q . If we regard for simplicity the case of distributed L^2 -temperature measurements, the parameter identification problem can put into the form (21) with

$$F : \mathcal{D}(F) \subset X \rightarrow Y = L^2(\Omega), q \rightarrow Eu_q \quad (48)$$

and $y = Eu_{q^\dagger}$, where $E : H_0^1(\Omega) \rightarrow L^2(\Omega)$ is the embedding operator.

For given $q \in \mathcal{D}(F)$, a formal linearization of the direct problem (47) in direction $p \in X$ yields

$$A(q)u'_q p = -A(p)u_q, \quad (49)$$

where the right-hand side is due to the linearity of $A(\cdot)$ with respect to q . Therefore, the Fréchet derivative of (48) is given by

$$F'(q) : X \rightarrow Y, p \rightarrow Eu'_q p,$$

where $u'_q p \in H_0^1(\Omega)$ denotes the solution of (49), i.e.,

$$u'_q p = -A(q)^{-1} A(p)u_q. \quad (50)$$

Hence, if we build the inner product in (32) with an arbitrary test function $p \in X$, the k -th iteration step becomes (where we omit E for reasons of readability)

$$(q_{k+1}^\delta, p) = (q_k^\delta, p) + (F'(q_k^\delta)^*(y^\delta - u_{q_k^\delta}), p) \quad (51)$$

$$(q_{k+1}^\delta, p) = (q_k^\delta, p) + (y^\delta - u_{q_k^\delta}, F'(q_k^\delta)p)$$

$$(q_{k+1}^\delta, p) = (q_k^\delta, p) - (y^\delta - u_{q_k^\delta}, A(q_k^\delta)^{-1} A(p)u_{q_k^\delta}). \quad (52)$$

Because of

$$(y^\delta - u_{q_k^\delta}, A(q_k^\delta)^{-1}A(p)u_{q_k^\delta}) = (A(q_k^\delta)^{-1*}(y^\delta - u_{q_k^\delta}), A(p)u_{q_k^\delta}),$$

the iteration can also be written as

$$\begin{aligned} (q_{k+1}^\delta, p) &= (q_k^\delta, p) - (w_k, A(p)u_{q_k^\delta}) \\ &= (q_k^\delta, p) - \int_{\Omega} p(x) \nabla w_k \nabla u_{q_k^\delta} dx, \end{aligned} \quad (53)$$

where w_k denotes the solution of the linear *adjoint* problem

$$A(q_k^\delta)^* w_k = y^\delta - u_{q_k^\delta}. \quad (54)$$

Hence, each iteration step in (32) requires to solve the direct problem (47) in order to obtain $u_{q_k^\delta}$ and the adjoint problem (54) with the residual $y^\delta - u_{q_k^\delta}$ as right-hand side. Eventually, the update according to (53) can be numerically realized as follows: If $\{p_1, p_2, \dots, p_n\}$ is an n -dimensional basis of the parameter space $X_n \subset X$ with \mathbf{q}_k^δ denoting the vectorial representation of q_k^δ , then (53) means to solve the linear system

$$M \mathbf{s}_k^\delta = \mathbf{r}_k^\delta,$$

where M is the Gramian Matrix

$$M(i, j) = (p_i, p_j)$$

and the vector \mathbf{r}_k is defined via

$$r_k^\delta(j) = \int_{\Omega} p_j(x) \nabla w_k \nabla u_{q_k^\delta} dx,$$

and to update the parameter via

$$\mathbf{q}_{k+1}^\delta = \mathbf{q}_k^\delta + \mathbf{s}_k^\delta.$$

Note that the approach (52) would require to solve n linear problems (49), clearly showing the advantages of (53) which gets by with solving a single problem (54).

Returning to our general discussion, we have already indicated with (15) that some iterative regularization methods can also be derived from certain initial value problems, which are then in turn called *continuous iteration methods*. For nonlinear problems, some of these methods are analyzed and related to their discrete analogues (especially (29)) in [ARS00] and [KNR02]. The asymptotic regularization method

$$u'_\delta(t) = F'(u_\delta(t))^*(y^\delta - F(u_\delta(t))), \quad (55)$$

see (15), is studied in the nonlinear setting in [Tau94]; it is also called inverse scale-space method in the context of imaging problems, see [SG01], [Sch03], [RSW00]. In [LS03], it is shown that (55) applied to (21), where F is the concatenation of a forward operator and a certain projection operator, can in fact be considered as a level set method. Level set methods, see [OF02], [Set99] have been successfully used for shape reconstruction problems e.g. in [San96], [OS01], [Bur03], their role as regularization methods for inverse problems has been analyzed in [Bur01].

As in Tikhonov regularization, a practical realization of an iterative method requires to take into account that in general the forward operator F cannot be exactly evaluated due to the nonlinearity and that only a sequence of approximations F_n with

$$\|F_n(x) - F(x)\| \leq \varepsilon_n, \quad x \in \mathcal{B}_\rho(x^\dagger),$$

is available, where ε_n denotes a (given) locally uniform approximation quality. For instance, F_n could represent an iterative solver for the direct problem with n denoting its iteration index. Then, an obvious question is how to link the inner iteration, i.e., the approximation level ε_n , to the index k of the inverse iteration in an efficient way. Such issues are explicitly addressed by multilevel techniques, see [Sch98b], [DES98] and [Ram99]. The basic principle is to start the iteration scheme on a rather rough approximation level in order to initially keep the computational efforts low. Of course - for actually approximating a solution of (21) - the quality of F_n has to be gradually increased, and in order to guarantee the desired regularization properties, the choice of ε_n has to be closely coupled to the outer iteration index k and the noise level of the data. A methodically different approach that is also based on a hierarchy of different approximation levels is to solve (21) (directly) via multigrid methods, see [Kin92], [Kal01].

Not only the evaluation of F but also that of F' and of its adjoint operator may cause computational problems and hence call for modifications of standard iteration methods. Especially in case of an already nonlinear direct problem, the basic assumptions of the convergence theory, which all involve F' , may become too restrictive, or the adjoint problem gets too complicated though it is linear. Hence, one might think of replacing F' in the iteration process by another linear operator which is easier to handle. Based on this idea, the convergence analysis of an iterative method related to the Landweber method (32) has been performed in [Sch95] with a modification of the nonlinearity condition (35). In the context of distributed parameter identification, an iteration operator is constructed in [Küg03] such that the desired regularization properties already follow from the unique solvability of the direct problem and the differentiability assumptions on F get redundant. This also might serve as a basis for the development of iterative routines for the

identification of parameters that appear in variational inequalities, which is a prime example for a nonlinear and non-differentiable inverse problem.

3 Some Industrial Applications

In this section, we report about some inverse problems posed by industrial partners; the companies involved in the first two projects, which are connected with iron and steel making, were VOEST Alpine Stahl and VAI, both global players based in Linz. We close with a current example from quantitative finance. All problems are nonlinear inverse problems and have been attacked with regularization methods as described in this paper. Although usually, practical problems are too complicated for an application of a mathematical theory in a way where all the assumptions can be checked, a sound theory is indispensable in order to know what effects are to be expected and how to tackle them. This is especially true for the instabilities associated with inverse problems: only a sound theory can act as a guideline on how to solve inverse problems in a stable and efficient way.

3.1 Determining the Inside Lining of an Iron Making Furnace

Our first example is concerned with the determination of the inside lining of a furnace from temperature measurements in its wall (cf. [RW98]). The blast furnace process is still the most important technology for producing iron from ore, although more recent alternative like the COREX process are available (cf. [Sch00] for a discussion of a mathematical model for this process, which is an adaptation of the kinetic model of the blast furnace process discussed in See [DZSKFES98], see also PATENT PCT/EP00/01463).

Due to variations in the working conditions of the furnace, slag may deposit on its sides and on its bottom, increasing the thickness of the lining. At the same time, the wall of the blast furnace is subject to physical and chemical wear, that the resulting erosion of the bricks causes the thickness of the lining to decrease. Therefore, the inner contour of the furnace has to be observed in order to avoid a breakthrough which would cause extensive damage and would also be highly dangerous. On the other hand, one should not stop the process before it is actually necessary for obvious economic reasons.

A direct observation of the wall from the inside of the furnace is obviously impossible, the temperatures inside being up to 1500°C . Hence, one has to reliably calculate the thickness of the wall in order to stop the process at the right time. For that purpose, temperature and heat flux measurement devices are placed inside the wall when the lining is bricked. Based on these measurements, one wants to calculate the shape of the wall on the inner, inaccessible side of the furnace.

The problem is an inverse heat conduction problem and is severely ill-posed. The first step is the development of a mathematical model that allows the numerical simulation of the process, i.e., a model for the direct problem. Since the furnace is (essentially) rotationally symmetric, the wall of the furnace in the region considered can be modeled in cylindrical coordinates. The (practically justified) assumptions of stationarity (due to the long time scales of changes in the shape of the inner lining) and rotational symmetry lead to the following nonlinear elliptic equation in the radial and height variables (the temperature u does not depend on the angular variable):

$$\frac{\partial}{\partial r}(\lambda \frac{\partial u}{\partial r}) + \frac{\lambda}{r} \cdot \frac{\partial u}{\partial r} + \frac{\partial}{\partial z}(\lambda \frac{\partial u}{\partial z}) = 0 \quad \text{in } \Omega, \quad (56)$$

where Ω denotes a radial cross-section of the wall (see Figure 3.1) with the boundary $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$. The heat conductivity depends on the

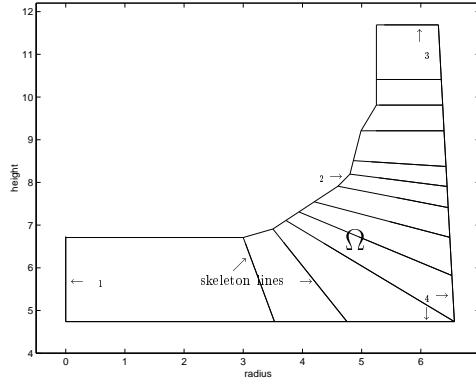


Fig. 4. Sketch of the furnace: two-dimensional, rotationally symmetric

material, which changes over Ω (since Ω is covered by different types of brick and possibly slag), and hence on position, and also on the temperature:

$$\lambda = \lambda(r, z, u).$$

This temperature-dependence cannot be neglected and makes even the direct problem nonlinear. Due to rotational symmetry and the fact that the lining continues past Γ_3 , we have the boundary conditions

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \Gamma_1 \quad \text{and on } \Gamma_3.$$

The outer surface is cooled by water with temperature T_w , which leads to the boundary condition

$$-\lambda \frac{\partial u}{\partial n} = \alpha_0 \cdot (u - T_w) \quad \text{on } \Gamma_4$$

with a (measurable) heat transfer coefficient α_0 . At the inner surface of the wall, an analogous condition

$$-\lambda \frac{\partial u}{\partial n} = \alpha \cdot (u - T_i) \quad \text{on } \Gamma_2 \quad (57)$$

holds, where the heat transfer coefficient α depends on the actual material present at Γ_2 and T_i is the interior temperature of the furnace (i.e., of molten iron).

Now, if Γ_2 , i.e., the shape of the inside lining, and hence Ω were known, we could solve the direct problem (56)-(57) in order to compute the temperature field u in Ω . Especially, we could predict the temperatures

$$\tilde{u}_j := u(x_j), j \in \{1, \dots, m\} \quad (58)$$

measured by thermo couples located at finitely many points $x_j \in \Omega$. In other words, given Γ_2 , we could evaluate the nonlinear forward operator

$$F : \Gamma_2 \rightarrow \tilde{u}. \quad (59)$$

The inverse problem now is to determine the inner contour Γ_2 and hence Ω from measurements of \tilde{u}_j .

For this problem, which could also be viewed as a shape reconstruction problem, the issue of uniqueness is of relevance since one wants to find a unique inner lining which gives information about the current status of the furnace, and not several (or even infinitely many) possible inner contours. In the setup based on (59) uniqueness cannot be expected, since only finitely many data $\tilde{u}_1, \dots, \tilde{u}_m$ are available for the determination of the curve Γ_2 . One could consider the corresponding infinite-dimensional problem of determining Γ_2 from temperature values at a whole curve inside Ω or, which is more practical, describe Γ_2 by finitely many parameters. In [RW98], Γ_2 has been described by the distances p_i from the outer contour Γ_4 along the skeleton lines (see Figure 3.1). If we denote, for $p = (p_1, \dots, p_n)$, by u_p the temperature field according to (56)-(57) with Γ_2 (and hence Ω) determined by p , then our inverse problem can be reformulated as the least squares problem

$$\Phi(p) := \sum_{j=1}^m (u_p(x_j) - \tilde{u}_j)^2 \rightarrow \min., \quad p \in C, \quad (60)$$

where C symbolizes constraints on p , see [RW98]. But even in this formulation the lack of uniqueness remains: Figure 5 shows two domains Ω which are markedly different, but nearly yield the same value for Φ . This example

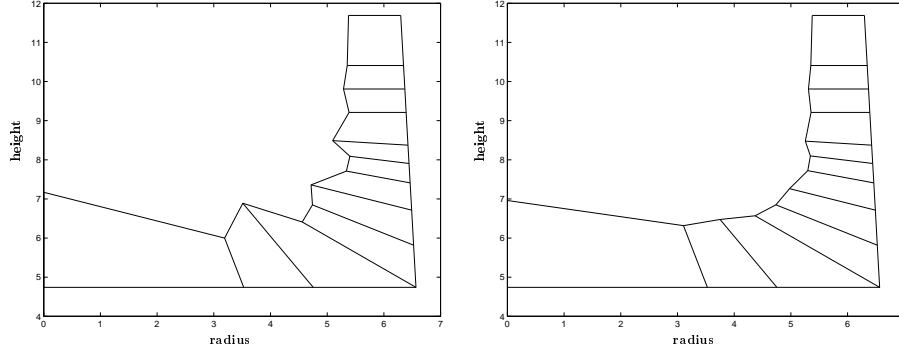


Fig. 5. Two possible solutions, the first one being not very realistic, while the second one is physically reasonable

indicates not only non-uniqueness, but also the inherent instability of the problem. In order to deal with this situation, the optimization problem (60) has to be regularized and - based on physical considerations - a-priori smoothness requirements for Γ_2 have to be incorporated. In [RW98], Tikhonov regularization has been applied for stably solving the inverse problem and the least squares minimization problem (60) has been replaced by

$$\sum_{j=1}^m (u_p(x_j) - \tilde{u}_j)^2 + \alpha \sum_{j=1}^{n+1} (\psi_j(p) - \psi_{j-1}(p))^2 \rightarrow \min., p \in C.$$

Thereby, $\psi_0(p) = \psi_{n+1}(p) = \frac{\pi}{2}$ and $\psi_j(p) (j = 1, \dots, n)$ denotes the angle between the j -th skeleton line and Γ_2 (see Figure 3.1). This regularization term penalizes oscillations in Γ_2 as those shown in the left picture of Figure 5 and enforces uniqueness. Figure 6 illustrates the results obtained by this regularization method for simulated data. In each picture, the boxes show the measurement points, the dashed line is the starting contour, the continuous line denotes the regularized solution and the third line (---) is the “true solution”. The latter was used to generate the data before they were corrupted by 10 % noise. These calculations, which were done for tuning the regularization parameter by experiment, show that without regularization, no useful solution can be obtained, while regularization yields quite good results (in consideration of the high noise level). The value $\alpha = 1225$ has then been used for solving the inverse problem with the real world data from the industrial company supposed to be of the same noise level. Of course, this shows the discrepancy between theory and practice: In the preceding sections, we wrote about strategies for determining the regularization parameter, and now we say that we used a value of 1225. The fact is that in a real practical situation like here, one frequently has too little information (especially about the

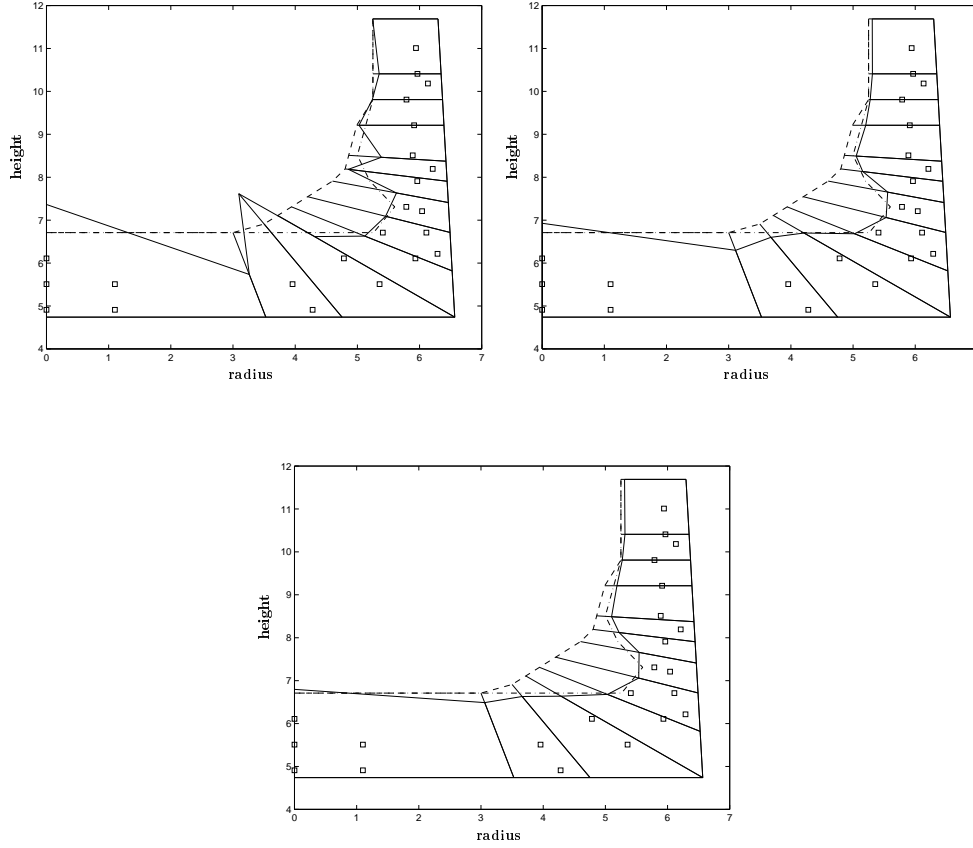


Fig. 6. Solutions with simulated data and different regularization parameters, 10% noise

noise level) to actually use the nice methods provided by the theory. As said above in connection with "error free strategies", one should not try to solve an inverse problem without any idea about the noise level. In the problem just described, one might have and will then use a rough upper bound for the noise level; but with such a rough bound, a regularization parameter based on an a-posteriori strategy will probably not be better than one determined by numerical experiments with simulated data, which is what is frequently used in practice. In our opinion this does not make the mathematical theory superfluous, since the theory indicates what is important and in which way e.g. a tuning of the regularization parameter by numerical experiment should be done.

We finally remark that the same practical problem was also treated in

[TMO98] for the simple case of a constant heat conductivity; instead of Tikhonov regularization, Kalman filtering was used there.

3.2 An Inverse Solidification Problem in Continuous Casting of Steel

Our example deals with the continuous casting process for the production of steel, where liquid steel is solidified in a continuous way before the hot rolling process. This process poses a lot of mathematical modeling and numerical simulation problems, also inverse problems, about two of which we shortly report here (cf. [GBEM95]). At the top end of a continuous caster, liquid steel is cooled in a water-cooled mould to form a solidified shell which can support the liquid pool at the mould exit. Typical temperatures at the end of the mould are 1100 °C at the strand surface and 1550 °C in the center of the strand. Since steel does not solidify at a fixed temperature, but over a temperature interval, there is a mushy region where the steel is neither completely solid nor completely liquid. After the mould, the cooling process is continued in secondary cooling zones by spraying water onto the strand. Near the point of final solidification, called the crater end, segregation and thermal stresses may increasingly occur because of strongly different temperature gradients in the interior and on the boundary of the strand. These effects would then drastically decrease the quality of the final product. In order to counteract the segregation, the strand thickness at the end of the crater is slightly reduced by compression. For a successful application of this technique named soft-reduction one has to ensure by an appropriate secondary cooling strategy that the final solidification takes place within the soft-reduction zone.

Hence, in this inverse problem we have to find the cause, namely the secondary cooling of the strand by cooling water, for a desired effect, namely the crater end to remain within the soft-reduction zone all the time or as long as possible. Here, a unique solution is not necessary since it may even be of advantage to choose among several cooling strategies. However, the main difficulty lies in the fact that in practice the casting speed is not always constant; e.g., when the width is changed, the process is slowed down and then accelerated again. If the speed changes, no cross-section of the steel strand can be considered on its own, which significantly increases the complexity of the problem even though only the location of the crater end has to be controlled. For a constant casting speed, the more general inverse problem of controlling the whole solidification front by secondary cooling was considered already much earlier by several industries and their research partners, see e.g. [EL88] for our contribution. As mentioned, the complexity of the direct (and hence also the inverse) problem is much lower, since for a constant casting speed, only one cross section needs to be considered as it runs through the caster.

Again, the first step is to derive a mathematical model for the direct problem. In [GBEM95], it is described in Lagrangian coordinates: The strand of thickness d moves in the casting direction z with the casting speed $v(t)$, y is the direction of width and x the direction of thickness of the strand in which the spray cooling takes place. The amount of cooling water with temperature U_w sprayed in the secondary cooling region enters into the boundary condition via a heat transfer function $g(z, t)$, furthermore the cooling due to radiation is considered with U_a denoting the temperature of the surrounding air. The material parameters of steel primarily relevant for the problem are the thermal conductivity $k(u)$, the density $\rho(u)$ and the specific heat $c(u)$, which all depend on the temperature $u = u(x, z, t)$. σ is the Stefan–Boltzmann constant and ϵ a further material parameter that depends on the steel grade. Under symmetry assumptions on the cooling and the initial temperature f and under the (physically justified) neglect of the heat conduction in casting direction, the temperature field is described by the following spatially two-dimensional problem

$$[k(u)u_x]_x = \rho(u)c(u)u_t \quad (61)$$

$$u(x, z = -\int_0^t v(\tau) d\tau, t) = f(x, t)$$

$$u_x(d/2, z, t) = 0 \quad (62)$$

$$k(u(0, z, t))u_x(0, z, t) = g(z + \int_0^t v(\tau) d\tau, t)(u(0, z, t) - U_w) + \sigma\epsilon(u^4(0, z, t) - U_a^4). \quad (63)$$

Here, the Eulerian coordinate 0 describes the lower end of the mould which is passed at time t by the cross-section with Lagrangian coordinate $z = -\int_0^t v(\tau) d\tau$. The temperature $f(x, t)$ at the end of the mould in (62) for this cross-section can be computed by calculating the temperature in the mould. It is assumed that at the beginning of the mould there is a constant initial temperature. The function f depends on time since the casting speed is not constant, so that different cross-sections remain in the mould for different amounts of time and thus cool down in a different way. $z + \int_0^t v(\tau) d\tau$ is the Eulerian coordinate of the cross-section with Lagrangian coordinate z at time t . This is used to compute the cooling for the cross-section z in (63).

If the material parameters k , ρ and c , the casting speed v and the initial temperature f are known, the direct problem is to calculate the temperature field u satisfying (61)–(63) for a given heat transfer function $g(z, t)$. In fact, the temperature and - as a consequence - the point of complete solidification can be determined for each cross section z separately. For mathematical questions like existence and uniqueness of a solution of (61)–(63) we refer to [Gre98].

Given a prescribed soft-reduction zone, the inverse problem now is to find a heat transfer function g such that with this cooling, the resulting solution of the direct problem is such that the crater end of each cross-section of the

strand remains in the soft-reduction zone. The amount of water sprayed onto the strand in the secondary cooling zones has to be constant in each cooling zone (there are typically six cooling zones) for technical reasons but can vary in time. Therefore, one has to admit piecewise constant, i.e., non-smooth heat transfer functions g , which makes the mathematical theory even for the direct problem a bit complicated (see [Gre98]). On the other hand, as a function of time, g should not jump too often, since each jump reflects a change in the cooling setup. Furthermore, upper and lower bounds for the total amount of water sprayed onto the strand have to be taken into account.

Though in the direct problem each cross-section z can be considered separately, they have to be treated simultaneously in the inverse problem due to the varying casting speed. The basic idea now is to penalize cross-sections that are not in the soft-reduction zone when they become solid, to integrate the penalties (which are a function of z) over all relevant cross-sections and then to minimize the resulting functional (which is a function of the heat transfer function g) over all admissible heat transfer functions g . To make things more tractable by approximating this non-differentiable functional by a differentiable one, a differentiable function P with

$$P(s) \begin{cases} = 0 & : s \leq 0 \\ > 0 & : s > 0 \end{cases}$$

is introduced; with $t(z)$ denoting the time when the cross-section z passes the end of the mould and $t(z, g)$ the time of its complete solidification for a given heat transfer function g , the functional

$$J(g) = \int_{z_1}^{z_2} \left\{ P \left(L_b - \int_{t(z)}^{t(z, g)} v(\tau) d\tau \right) + P \left(\int_{t(z)}^{t(z, g)} v(\tau) d\tau - L_e \right) \right\} dz \quad (64)$$

is minimized over all admissible heat transfer functions g . Here, L_b is the beginning of the soft-reduction zone, L_e is its end, z_1 is the cross-section that enters the mould at the time after which the casting speed remains constant again, and z_2 is the cross-section that passes the end of the secondary cooling zone at the time when the casting speed starts to change. The term $P(L_b - \int_{t(z)}^{t(z, g)} v(\tau) d\tau)$ penalizes the cross-section z if it solidifies before the soft-reduction zone, the second term penalizes a solidification after L_e . Note that computing J involves solving the direct problem and is hence time-consuming; a single evaluation of the functional (64) requires to solve many direct problems (61)–(63).

Since the practically relevant heat transfer functions are piecewise constant both in space and time (there are only finitely many prescribed times when a change in cooling can be made), the inverse problem finally leads a finite-dimensional nonlinear optimization problem with bounds for the variables,

which also serves as some regularization of the problem. However, if the number of degrees of freedom is more than just a few, this regularization alone is not enough to remove the instability. In [GBEM95], the optimization problem was solved with a Quasi-Newton method, i.e., with an iterative method considered as a regularization method with an appropriate stopping rule, where for the efficient computation of the search direction the adjoint method was used. Figure 7 shows an example where the casting speed is increased from

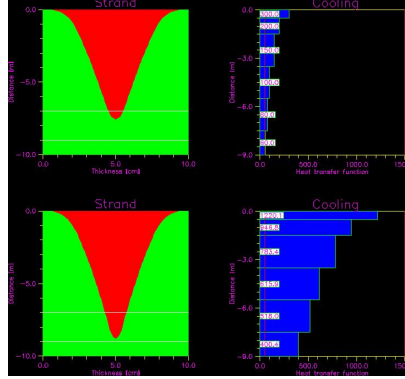


Fig. 7. The strand and computed cooling strategies

1.8 to 3.2 m/min within about 10 minutes. At the beginning of the process a rather low cooling rate suffices to keep the crater end in the soft-reduction zone. After the acceleration of the strand speed, only a significantly higher cooling rate allows to hold the crater end in the desired region.

The next question that arose in the practical problem was how the heat transfer function g actually depends on the water flow rate - in the following denoted by q - which is the quantity which can be used to control the process in practice, while g is only a mathematical quantity not directly available in practice. While in the first phase of the project, g was considered as $g = g(z + \int_0^t v(\tau) d\tau, t)$ and values from the literature were used, this turned out not to be accurate enough for practical purposes. Hence, we were asked to consider the problem of actually determining the connection between g and q , again an inverse problem about which we report in more detail in [CE00]: The heat transfer function g in the boundary condition (63) is set up as

$$g = g(q) \quad (65)$$

with the water flow rate having the form

$$q = q(z + \int_0^t v(\tau) d\tau, t),$$

and the main focus is laid on the determination of the dependence of g on q in (65). As data, temperature measurements on the strand surface at different cross-sections are available. Since the goal is to estimate a function that appears in the boundary condition of a nonlinear parabolic equation from boundary measurements at a later time, this inverse problem can be understood as a combination of a sideways and backwards heat equation and hence has to be expected to be severely ill-posed.

If we denote by u_i^* the temperature measured at the i -th cross-section and by $u_i(g)$ those predicted by (61)–(63) (with the boundary condition modified according to (65)) for given g at the measurement point, our problem can be formulated as finding $g \in V$ with

$$\|F(g)\|^2 \rightarrow \min \text{ over } V, \quad (66)$$

where F is defined by

$$F(g) := \begin{pmatrix} u_1(g) - u_1^* \\ \vdots \\ u_N(g) - u_N^* \end{pmatrix}$$

and V is a set of functions one considers admissible for g .

Now, the traditional approach in the engineering literature is to represent g as

$$g(q) := a \cdot q^b$$

such that only two positive parameters a and b are left to be determined. However, for real temperature data, i.e., in the presence of data noise, this exponential ansatz is not appropriate, since it turns out that the parameters a and b obtained in (66) are highly sensitive to noise and the data set used. Even the use of Tikhonov regularization does not improve the results significantly, there are just too few parameters to match real data both in a reasonably accurate and stable way. This instability is in fact not primarily associated with the ill-posedness of the inverse problems, but results from the attempt to match experimental data with just two parameters.

Hence, in the project reported about in [CE00], we increased the number of parameters by subdividing the q -interval into M subintervals and to model g as a piecewise cubic spline in q . But this increase in the number of parameters now gives rise to possible instability due to the ill-posed nature of the inverse problem. Because of this and also since the data were very unevenly distributed over the water flow rates, i.e., in the q -space, a mere minimization of the functional in (66) turned out to be problematic. In order to compensate resulting negative effects, more smoothness is enforced by solving the regularized problem

$$\|F(g)\|^2 + \alpha \|g''\|^2 \rightarrow \min, \quad g \in \{\text{cubic splines}\}.$$

Figure 8, taken from [CE00], indicates that satisfactory results can be obtained

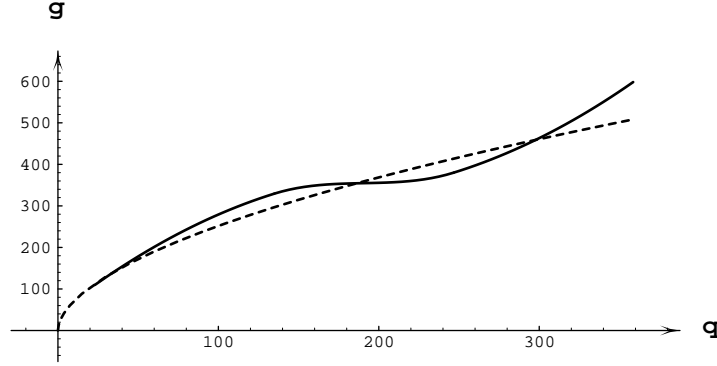


Fig. 8. exact (dotted line) and computed heat transfer functions

by this method. For this test using simulated data with simulated noise (independent pseudo-random numbers normally distributed in the interval $[-5, 5]$ were added to the simulated data), $M = 4$, $N = 40$ and $\alpha = 14.1$ were chosen. The same remarks as in the first example apply concerning the relation between mathematical theory and practice: Reasonable values for the regularization parameter were determined by numerical experiments on carefully (i.e., theory-based!) chosen examples. The method also worked reasonably when applied to actual data, meaning that the results were reproducible for different data sets.

We close this section by mentioning that similar considerations concerning the proper modeling of functions are also of relevance for the material parameters of steel used in the direct problem (61)–(63). Usually, the thermal conductivity $k(u)$, the density $\rho(u)$ and the specific heat $c(u)$ are either interpolated between given discrete temperature values or adjusted to them via tuning of some parameters in physical laws taken from literature. Alternatively, one could make the full functional dependence of the material parameters on the temperature u subject of a separate inverse problem. For recent work on the identification of nonlinearities we refer to [KE02], [Küg03].

3.3 Inverse Problems in Quantitative Finance

One of the fastest growing fields in applied mathematics is computational finance. The modeling of the fair price of financial derivatives can by now be

considered classic (see [Wil98]). All these by now very refined models contain parameters like the volatility of the underlying assets, and in recent years, the problem of identifying such parameters from observed data (e.g., prices of some derivative products observed in the market)) became

To make this presentation reasonably self-contained, we review some basic facts about mathematical finance: A financial derivative is a contract where payment is derived from some underlying benchmark like a stock, bond, interest rate or exchange rate. For instance, a European call option on a stock gives its owner the right (but not the obligation) to buy a specified amount of the underlying stock at a given date (the maturity T) for a given price (the strike K). Similarly, a European put option gives the right to sell, while for American options, the strike cannot be done only at, but also before a specified date. The correct pricing of such financial instruments requires mathematical models which reasonably describe the stochastic processes of the underlying, efficient methods for the calculation of fair prices, which are usually described by a partial differential equation model (see [BES03] for one approach), and robust estimates for the coefficients of these PDEs. The prices in the financial markets, where the derivatives are traded, then serve as data for the identification of these coefficients.

In the classical Black Scholes model, see [BS77], the spot price S of an equity follows a random walk

$$dS(t) = \mu S(t)dt + \sigma S dW,$$

where dW denotes the increment of a standard Wiener process and σ is the so-called volatility. For the pricing of options on this underlying, the Black Scholes trick constructs a portfolio whose value evolves risk-free such that the same return r as that of a risk-free cash account has to be expected. Based on these considerations, the value V of a European option can then be shown to satisfy the (by now famous) Black Scholes equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0. \quad (67)$$

In order to completely describe the (direct) problem of calculating V for a given σ , the parabolic differential equation (67) backwards in time is augmented by an end condition and boundary conditions at zero and at infinity. For the easiest case of European call or put (together called "vanilla")) options, characterized by the terminal conditions

$$V(T) = \max(S - K, 0) \quad \text{or} \quad V(T) = \max(K - S, 0),$$

analytic solutions of (67) are available. Once the volatility is known, a wide range of derivatives with the same underlying S can be priced by the use of equivalents to (67). Hence it is important to find the proper parameter σ .

Given the contract rules of a European vanilla option and its price, the volatility can even be directly calculated from the analytical solution formula of (67). Using this implied volatility in (67) then yields exactly the given option price. However, the market typically observes that the such determined implied volatility depends on the exercise price K of the option, which is unreasonable, and therefore cannot be constant. Based on a duality argument from [Dup94], where this functional correlation of σ can also be expressed as a dependence on the spot price S , one possible generalization of the Black Scholes model is to put up the volatility in (67) as a deterministic function of S , i.e., $\sigma = \sigma(S)$, then denoted as local volatility. Other approaches are based on modeling σ also as a stochastic process ("stochastic volatility"), see [Shr97].

Since the volatility is not directly observable in the markets, one has to face the inverse problem of identifying σ from available data such as the prices of liquidly traded derivatives. Option prices can be observed for different maturities T and strike prices K , i.e., the data take the form $V_{\text{market}}(T_i, K_j)$. Such inverse problems have e.g. been considered in [AFHS97], [CCE00], [Cre03a], [Cre03b], [BI99], [JSH99], [LO97], [LY01]. In [Egg01], the case of data available only for single maturity T_0 has been studied. With $V(T_0, K_j)$ denoting the solution of the direct problem (67) for different K_j and given local volatility, the output least squares problem is formulated as

$$\sum_j^n (V(T_0, K_j) - V_{\text{market}}(T_0, K_j))^2 \rightarrow \min \text{ over } \Sigma. \quad (68)$$

Here, Σ denotes the set of piecewise cubic interpolating splines used for the representation of the volatility. As long as there is no noise in the data, the approach (68) works quite well as indicated by Figure 9: The blue dotted line

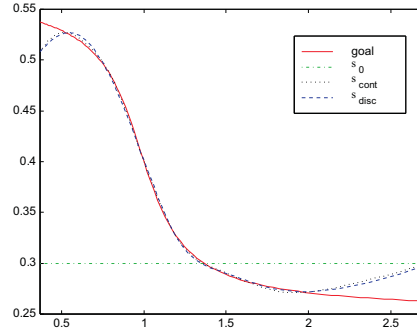


Fig. 9.

shows the volatility as a function of K recovered from $n = 10$ market prices of a European vanilla option (simulated by using the red volatility function). As expected, the identification is satisfactory as long as we are not too deep in the money or too deep out of the money (for a call option, these terms refer to an asset price above or below the strike price, respectively). At the extreme ends, the option prices do not contain much information and therefore the solution is mainly determined by the initial guess (green). For comparison, also continuous market data have been used. The violet line demonstrates that the error introduced by only using discrete data points (and interpolating between them) is negligible.

However, in practice, there will always be noise in the data, at least in the order of magnitude of the bid-offer spreads in the option prices. Then, the situation becomes unstable without regularization as shown in Figure 10: For

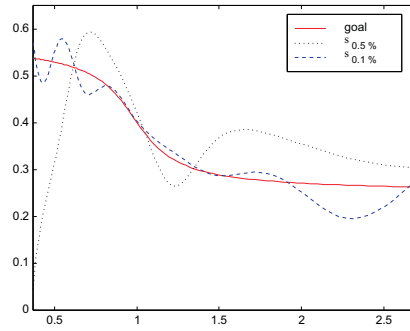
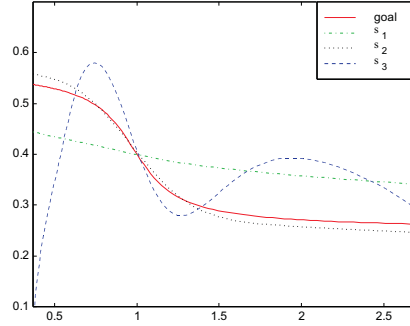


Fig. 10.

(quite small) noise levels of 0.1 and 0.5 percent, the pure minimization of the functional in (68) leads to oscillating results though the degrees of freedom are kept low in the spline representation. In order to reduce these oscillations, Tikhonov regularization is used in [Egg01], where the penalty term

$$\alpha(\|\sigma'\|^2 + \|\sigma''\|^2),$$

with $\|\cdot\|$ denoting the L^2 -norm, is added to the objective functional. Figure 11 now illustrates the influence of the regularization parameter: If α is chosen too small, the main emphasis is still laid on the approximation of the noisy data only possible by means of an oscillating volatility (blue). On the other hand, for a regularization parameter which is too large, variations in the volatility are penalized too much (green), too little information is used from the data. However, if parameter choice strategies like the ones described in the preceding

**Fig. 11.**

sections, based on the noise level obtained e.g. from the bid-offer spread, are used, a stable and accurate solution can be obtained (violet). In [Egg01], the theory of Tikhonov regularization including results about convergence rates is actually applied to this inverse problem, so that the gap between theory and practice is not as big as in the two problems from iron and steel industry described above.

Similar techniques apply also to interest rates models. E.g., for a one-factor model, the underlying differential equation looks like

$$\frac{\partial V}{\partial t} + \frac{1}{2}W^2 \frac{\partial^2 V}{\partial r^2} + (u - \lambda W) \frac{\partial V}{\partial r} - rV = 0,$$

see e.g. [Wil98], [Reb98]. A special case is the Hull-White interest rate model, where the "short rate" r behaves according to

$$dr(t) = (a(t) - b(t)r(t))dt + \sigma(t)dW.$$

For such models, the term "model calibration" is used for the inverse problem of identifying some or all of the parameters $a(t)$, $b(t)$ and $\sigma(t)$ from market prices of swaps, caps, floors, and swaptions. Given the fast-moving field of mathematical finance, it is not surprising that these methods have already found their way into practice: e.g. the pricing software UnRisk (see [Unr]) uses discrete versions of bounded variation regularization, compare to (26) to obtain robust interest rate model parameters. These parameters can then be used to price more complex instruments like callable bonds or callable reverse floaters.

References

- [ARS00] R.G. Airapetyan, A.G. Ramm, A.B. Smirnova, Continuous methods for solving nonlinear ill-posed problems, *Fields Institute Communications* **25**, 111–137 (2000)
- [And86] R.S. Anderssen, The linear functional strategy for improperly posed problems, in: J. Cannon, U. Hornung, eds., *Inverse Problems*, Birkhäuser, Basel, 11–30 (1986)
- [AFHS97] M. Avellaneda, C. Friedman, R. Holmes, D. Samperi, Calibrating Volatility Surfaces Via Relative-Entropy Minimization, *Applied Mathematical Finance* **4**, 37–64 (1997)
- [BG67] G. Backus, F. Gilbert, Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc.* **13**, 247–276 (1967)
- [Bak92] A.B. Bakushinskii, The problem of the convergence of the iteratively regularized Gauss-Newton method, *Comput. Math. Math. Phys.* **32**, 1353–1359 (1992)
- [BB01] T. Banks, K.L. Bihari, Modeling and estimating uncertainty in parameter estimation, *Inverse Problems* **17**, 1–17 (2001)
- [BK89] H. Banks, K. Kunisch, *Parameter Estimation Techniques for Distributed Systems*, Birkhäuser, Boston (1989)
- [BBC85] J. Beck, B. Blackwell, C.S. Clair, *Inverse Heat Conduction: Ill-Posed Problems*; Wiley, New York (1985)
- [BB98] M. Bertero, P. Boccacci, *Introduction to Inverse Problems in Imaging*, Inst. of Physics Publ., Bristol (1998)
- [BEGNS94] A. Binder, H. W. Engl, C. W. Groetsch, A. Neubauer and O. Scherzer, Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization, *Appl. Anal.* **55**, 215–234 (1994)
- [BES03] A. Binder, H.W. Engl, A. Schatz, Advanced numerical techniques for pricing financial derivatives, *Derivatives Week Vol. XII* **17**, 6–7 (2003)
- [BS77] F. Black, M. Scholes, The pricing of options and corporate liabilities, *J. of Political Economy* **81**, 637–659 (1973)
- [BNS97] B. Blaschke, A. Neubauer, O. Scherzer, On convergence rates for the iteratively regularized Gauss-Newton method, *IMA Journal of Numerical Analysis* **17**, 421–436 (1997)
- [BI99] I. Bouchouev, V. Isakov, Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets, *Inverse Problems* **15**, R95 - R116 (1999)
- [Bra87] H. Brakhage, On ill-posed problems and the method of conjugate gradients, in [EG87], 165–175
- [BU97] R.M. Brown, G.A. Uhlmann, Uniqueness in the inverse conductivity problem for non-smooth conductivities in two dimensions, *Commun. Part. Diff. Equ.* **22**, 1009–27 (1997)
- [Bur03] M. Burger, A framework for the construction of level set methods for shape optimization and reconstruction, *Interfaces and Free Boundaries* **5**, 301–329 (2003)
- [Bur01] M. Burger, A level set method for inverse problems, *Inverse Problems* **17**, 1327–1356 (2001)

- [BE00] M. Burger, H.W. Engl, Training neural networks with noisy data as an ill-posed problem, *Adv. Comp. Math.* **13**, 335-354 (2000)
- [CE00] C. Carthel, H.W. Engl, Identification of heat transfer functions in continuous casting of steel by regularization, *J. Inv. Ill-Posed Problems* **8**, 677-693 (2000)
- [CCE00] C. Chiarella, M. Craddock, N. El-Hassan, The Calibration of Stock Option Pricing Models Using Inverse Problem Methodology, *QFRQ Research Papers*, UTS Sydney (2000)
- [CER90] D. Colton, R. Ewing, and W. Rundell, *Inverse Problems in Partial Differential Equations*, SIAM, Philadelphia (1990)
- [CK92] D. Colton, R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, Berlin (1992)
- [Cre03a] S. Crepey, Calibration of the local volatility in a trinomial tree using Tikhonov regularization, *Inverse Problems* **19**, 91-127 (2003)
- [Cre03b] S. Crepey, Calibration of the local volatility in a generalized Black-Scholes model using Tikhonov regularization, *SIAM J. Math. Anal.* **354**, 1183-1206 (2003)
- [DES98] P. Deuffhard, H.W. Engl, O. Scherzer, A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions, *Inverse Problems* **14**, 1081-1106 (1998)
- [DZSKFES98] H. Druckenthauer, H. Zeisel, A. Schieffermüller, G. Kolb, A. Ferstl, H. Engl, A. Schatz, Online simulation of the blast furnace, *Advanced Steel*, 58-61 (1997-1998)
- [Dup94] B. Dupire, Pricing with a smile, *RISK* **7**, 18-20 (1994)
- [Egg01] H. Egger, Identification of Volatility Smiles in the Black Scholes Equation via Tikhonov Regularization, Diploma Thesis, University of Linz, Austria (2001)
- [Egg93] , P.P.B. Eggermont, Maximum entropy regularization for Fredholm integral equations of the first kind, *SIAM J. Math. Anal.* **24**, 1557-1576 (1993)
- [Eng82] H.W. Engl, On least squares collocation for solving linear integral equations of the first kind with noisy right-hand side, *Boll. Geodesia Sc. Aff.* **41**, 291-313 (1982)
- [Eng93] H.W. Engl, Regularization methods for the stable solution of inverse problems, *Surveys on Mathematics for Industry* **3** 71-143 (1993)
- [EG88] H.W. Engl, H. Gfrerer, A posteriori parameter choice for general regularization methods for solving linear ill-posed problems, *Appl. Numer. Math.* **4**, 395-417 (1988)
- [EG94] H.W. Engl, W. Grever, Using the L -curve for determining optimal regularization parameters, *Numerische Mathematik* **69**, 25-31 (1994)
- [EG87] H.W. Engl, C. W. Groetsch, *Inverse and Ill-Posed Problems*, Academic Press, Orlando, (1987)
- [EHN96] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht (1996) (Paperback edition 2000)
- [EKN89] H.W. Engl, K. Kunisch and A. Neubauer, Convergence rates for Tikhonov regularization of nonlinear ill-posed problems, *Inverse Problems* **5**, 523-540 (1989)
- [EL93] H.W. Engl, G. Landl, Convergence rates for maximum entropy regularization, *SIAM J. Numer. Anal.* **30**, 1509-1536 (1993)

- [EL96] H.W. Engl, T. Langthaler, Maximum entropy regularization of nonlinear ill-posed problems, in: V. Lakshmikantham, ed., *Proceedings of the First World Congress of Nonlinear Analysts*, Vol. I, de Gruyter, Berlin, 513–525 (1996)
- [EL88] H.W. Engl, T. Langthaler, Control of the solidification front by secondary cooling in continuous casting of steel, in: H.W. Engl, H. Wacker and W. Zulehner, eds. *Case Studies in Industrial Mathematics* Teubner, Stuttgart, 51–77 (1988)
- [EL01] H.W. Engl, A. Leitao, A Mann iterative regularization method for elliptic Cauchy problems, *Numer. Func. Anal. Optim.* **22**, 861–884 (2001)
- [ELR96a] H. W. Engl, A. K. Louis and W. Rundell (eds.), *Inverse Problems in Geophysics*, SIAM, Philadelphia, (1996)
- [ELR96b] H.W. Engl, A.K. Louis, W. Rundell (eds.), *Inverse Problems in Medical Imaging and Non-Destructive Testing*, Springer, Vienna, New York (1996)
- [EN88] H.W. Engl, A. Neubauer, Convergence rates for Tikhonov regularization in finite dimensional subspaces of Hilbert scales, *Proc. Amer. Math. Soc.* **102**, 587–592 (1988)
- [ER95] H. W. Engl, W. Rundell (eds.), *Inverse Problems in Diffusion Processes*, SIAM, Philadelphia (1995)
- [ES00] H.W. Engl, O. Scherzer, Convergence rate results for iterative methods for solving nonlinear ill-posed problems, in: D. Colton, H.W. Engl, A.K. Louis, J. McLaughlin, W.F. Rundell (eds.), *Surveys on Solution Methods for Inverse Problems*, Springer, Vienna/New York, 7–34 (2000)
- [EW85] H.W. Engl, A. Wakolbinger, Continuity properties of the extension of a locally Lipschitz continuous map to the space of probability measures, *Monatshefte f. Math.* **100**, 85–103 (1985)
- [Gre98] W. Grever, A nonlinear parabolic initial boundary value problem modeling the continuous casting of steel with variable casting speed, *ZAMM* **78**, 109–119 (1998)
- [GBEM95] W. Grever, A. Binder, H.W. Engl, K. Mörwald, Optimal cooling strategies in continuous casting of steel with variable casting speed, *Inverse Problems in Engineering* **2**, 289–300 (1995)
- [Gro72] C.W. Groetsch, A note on segmenting Mann iterates, *J. Math. Anal. Appl.* **40**, 369–372 (1972)
- [Gro77] C.W. Groetsch, *Generalized Inverses of Linear Operators: Representation and Approximation*, Dekker, New York (1977)
- [Gro93] C.W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg, Braunschweig (1993)
- [GN88] C.W. Groetsch, A. Neubauer, Convergence of a general projection method for an operator equation of the first kind, *Houston J. Math.* **14**, 201–208 (1988)
- [Han97] M. Hanke, A regularizing Levenberg-Marquardt scheme with applications to inverse groundwater filtration problems, *Inverse Problems* **13**, 79–95 (1997)
- [Han95] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, Harlow, Essex, (1995)
- [HH93] M. Hanke, P.C. Hansen, Regularization methods for large-scale problems, *Surveys Math. Indust.* **3**, 253–315 (1993)

- [HNS95] M. Hanke, A. Neubauer, O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numerische Mathematik* **72**, 21–37 (1995)
- [HO93] P.C. Hansen, D.P. O’Leary, The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.* **14**, 1487–1503 (1993)
- [HN88] J. Haslinger, P. Neittaanmäki, *Finite Element Approximation for Optimal Shape Design: Theory and Applications*, John Wiley & Sons Ltd., Chichester (1988)
- [Hof99] B. Hofmann, *Mathematik inverser Probleme*, Teubner, Stuttgart (1999)
- [Hoh97] T. Hohage, Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem, *Inverse Problems* **13**, 1279 – 1299 (1997)
- [IN99] D. Isaacson, J.C. Newell, Electrical Impedance Tomography, *SIAM Review* **41**, 85–101 (1999)
- [Isa98] V. Isakov, *Inverse Problems in Partial Differential Equations*, Springer, Berlin, New York (1998)
- [JSH99] N. Jackson, E. Süli, S. Howison, Computation of Deterministic Volatility Surfaces, *Journal of Computational Finance* **2**, (1998/99)
- [Kal01] B. Kaltenbacher, On the regularizing properties of a full multigrid method for ill-posed problems, *Inverse Problems* **17**, 767–788 (2001)
- [KNR02] B. Kaltenbacher, A. Neubauer, A.G. Ramm, Convergence rates of the continuous regularized Gauss-Newton method, *J. Inv. Ill-Posed Problems* **10**, 261–280 (2002)
- [Kin92] J. King, Multilevel algorithms for ill-posed problems, *Numerische Mathematik* **61**, 311–334 (1992)
- [Kir96] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, New York (1996)
- [KSB88] A. Kirsch, B. Schomburg and G. Berendt, The Backus-Gilbert method, *Inverse Problems* **4**, 771–783 (1988)
- [KS85] C. Kravaris, J.H. Seinfeld, Identification of parameters in distributed parameter systems by regularization, *SIAM J. Control Optim.* **23**, 217–241 (1985)
- [Küg03] P. Kügler, A Derivative Free Landweber Iteration for Parameter Identification in Elliptic Partial Differential Equations with Application to the Manufacture of Car Windshields, PhD-Thesis, Johannes Kepler Universität Linz, Austria, (2003)
- [Küg03] P. Kügler, Identification of a temperature dependent heat conductivity from single boundary measurements, *SIAM Journal on Numerical Analysis*, accepted
- [KE02] P. Kügler, H.W. Engl, Identification of a temperature dependent heat conductivity by Tikhonov regularization, *Journal of Inverse and Ill-posed Problems* **10**, 67–90 (2002)
- [KL03] P. Kügler, A. Leitao, Mean value iterations for nonlinear elliptic Cauchy problems, *Numerische Mathematik*, accepted for publication
- [LO97] R. Lagnado, S. Osher, A Technique for Calibrating Derivative Security Pricing Models: Numerical Solution of an Inverse Problem, *JCF* **1**, (1997)
- [LA96] , G. Landl, R.S. Anderssen, Non-negative differentially constrained entropy-like regularization, *Inverse Problems* **12**, 35–53 (1996)

- [LS03] A. Leitao, O. Scherzer, On the relation between constraint regularization, level sets and shape optimization, *Inverse Problems* **19**, L1–L11 (2003)
- [LY98] A. Leonov, A. Yagola, Special regularizing methods for ill-posed problems with sourcewise represented solutions, *Inverse Problems* **14**, 1539–1550 (1998)
- [LY01] J. Lishang, T. Youshan, Identifying the volatility of underlying assets from option prices, *Inverse Problems* **17**, 137–155 (2001)
- [Lou87] A.K. Louis, Convergence of the conjugate gradient method for compact operators, in [EG87], 177–183
- [Lou89] A.K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart (1989)
- [LM90] A.K. Louis, P. Maass, A mollifier method for linear operator equations of the first kind, *Inverse Problems* **6** 427–440 (1990)
- [Man53] W.R. Mann, Mean value methods in iteration, *Proc. Am. Math. Soc.* **4**, 506–510 (1953)
- [Nas76] M. Z. Nashed, *Generalized Inverses and Applications*, Academic Press, New York (1976)
- [NS98] M.Z. Nashed, O. Scherzer, Least squares and bounded variation regularization with nondifferentiable functional, *Num. Funct. Anal. and Optimiz.* **19**, 873–901 (1998)
- [Nat77] F. Natterer, Regularisierung schlecht gestellter Probleme durch Projektionsverfahren, *Numer. Math.* **28**, 329–341 (1977)
- [Nat86] F. Natterer, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart (1986)
- [Neu89] A. Neubauer, Tikhonov regularization for non-linear ill-posed problems: optimal convergence rates and finite-dimensional approximation, *Inverse Problems* **5**, 541–557 (1989)
- [NS90] A. Neubauer, O. Scherzer, Finite-dimensional approximation of Tikhonov regularized solutions of non-linear ill-posed problems, *Numer. Funct. Anal. and Optimiz.* **11**, 85–99 (1990)
- [OF02] S. Osher, R.P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer, New York (2002)
- [OS01] S. Osher, F. Santosa, Level set methods for optimization problems involving geometry and constraints I. Frequencies of a two-density inhomogeneous drum, *J. Comp. Phys.* **171**, 272–288 (2001)
- [OG69] C. Outlaw, C.W. Groetsch, Averaging iteration in a Banach space, *Bull. Am. Math. Soc.* **75**, 430–432 (1969)
- [PV90] R. Plato, G. Vainikko, On the regularization of projection methods for solving ill-posed problems, *Numer. Math.* **57**, 63–79 (1990)
- [RSW00] E. Radmoser, O. Scherzer, J. Weickert, Scale-space properties of non-stationary iterative regularization methods, *Journal of Visual Communication and Image Representation* **11**, 96–114 (2000)
- [RW98] E. Radmoser, R. Wincor, Determining the inner contour of a furnace from temperature measurements, *Industrial Mathematics Institute, Johannes Kepler Universität Linz, Technical Report* **12** (1998)
- [Ram99] R. Ramlau, A modified Landweber-method for inverse problems, *Num. Funct. Anal. Opt.* **20**, 79–98 (1999)
- [Ram86] A.G. Ramm, *Scattering by Obstacles*, Reidel, Dordrecht, (1986)

- [Rau84] T. Raus, Residue principle for ill-posed problems, *Acta et Comment. Univers. Tartuensis* **672**, 16–26 (1984)
- [Reb98] R. Rebonato, *Interest-Rate Option Models*, Wiley 2nd edition (1998)
- [Rud94] L.I. Rudin, S. Osher, Total variation based restoration with free local constraints, *Proc. of IEEE Int. Conf. on Image Processing*, Austin, Tx., 31–35 (1994)
- [San96] F. Santosa, A level-set approach for inverse problems involving obstacles, *ESAIM: Control, Optimization and Calculus of Variations* **1**, 17–33 (1996)
- [Sch00] A. Schatz, Lump Ore, Pellets, and Dead Men - Mathematical Modelling and Numerical Simulation of the COREX Reduction Shaft, PhD-Thesis, University of Linz, Austria (2000)
- [Shr97] S. Shreve, *Stochastic Calculus and Finance*, www.cs.cmu.edu/~chal/shreve.html (1997)
- [Sch98] O. Scherzer, A modified Landweber iteration for solving parameter estimation problems, *Applied Mathematics and Optimization* **38**, 45–68 (1998)
- [Sch98b] O. Scherzer, An iterative multi level algorithm for solving nonlinear ill-posed problems, *Numerische Mathematik* **80**, 579–600 (1998)
- [Sch95] O. Scherzer, Convergence criteria of iterative methods based on Landweber iteration for solving nonlinear problems, *Journal of Mathematical Analysis and Applications* **194**, 911–933 (1995)
- [Sch02] O. Scherzer, Explicit versus implicate relative error regularization on the space of functions of bounded variation, *Contemporary Mathematics* **313**, 171–198 (2002)
- [Sch03] O. Scherzer, Scale space methods for denoising and inverse problems, *Advances in Imaging and Electron Physics*, accepted for publication
- [SEK93] O. Scherzer, H.W. Engl, K. Kunisch, Optimal a-posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems, *SIAM J. Numer. Anal.* **30**, 1796–1838 (1993)
- [SG01] O. Scherzer, C.W. Groetsch, *Scale-Space and Morphology in Computer Vision*, Springer Lecture Notes in Computer Science **2106**, Springer, Berlin, 317 – 325 (2001)
- [Sch85] E. Sock, Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence, in: G. Hämmerlin and K. Hoffmann, eds., *Constructive Methods for the Practical Treatment of Integral Equations*, Birkhäuser, Basel, 234–243 (1985)
- [Set99] J.A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, 2nd edition, Cambridge (1999)
- [TMO98] M. Tanaka, T. Matsumoto, S. Oida, Identification of unknown boundary shape of rotationally symmetric body in steady heat conduction via BEM and filter theories, in: M. Tanaka and G.S. Dulikravich eds., *Inverse Problems in Engineering Mechanics*, Elsevier Science B.V., Tokyo, 121–130 (1998)
- [Tau94] U. Tautenhahn, On the asymptotic regularization of nonlinear ill-posed problems, *Inverse Problems* **10**, 1405–1418 (1994)
- [Unr] UnRisk Documentation for the UnRisk PRICING ENGINE for Mathematica, www.unriskderivatives.com

- [Vas87] V.V. Vasin, Iterative methods for the approximate solution of ill-posed problems with a priori informations and their applications, in: [EG87], 211–229 (1987)
- [Vas92] V.V. Vasin, Ill-posed problems and iterative approximation of fixed points of pseudo-contractive mappings, in: A.N. Tikhonov, A.S. Leonov, A.I. Prilepko, I.A. Vasin, V.A. Vatutin, A.G. Yagola, eds., Ill-Posed Problems in Natural Sciences, VSP, Utrecht, 1992, 214–223, (1992)
- [Vas98] V.V. Vasin, On the convergence of gradient-type methods for nonlinear equations, *Doklady Mathematics* **57**, 173–175 (1998) translated from *Doklady Akademii Nauk* **359**, 7–9 (1998)
- [VA95] V.V. Vasin, A.L. Ageev, Ill-Posed Problems with A-Priori Information, VSP, Utrecht (1995)
- [Vog96] C.R. Vogel, Non-convergence of the L-curve regularization parameter selection method, *Inverse Problems* **12**, 535–547 (1996)
- [Wah90] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia (1990)
- [Wil98] P. Wilmott, *Derivatives - The Theory and Practice of Financial Engineering*, Wiley (1998)