International Journal of Pure and Applied Mathematics
Volume 43 No. 4 2008, 773-794

CONVERGENCE RATES FOR MULTI-PARAMETER REGULARIZATION IN BANACH SPACES

Torsten Hein

Faculty of Mathematics
Technical University of Chemnitz
Chemnitz, 09107, GERMANY
e-mail: torsten.hein@mathematik.tu-chemnitz.de

Abstract: In this paper we present a multi-parameter regularization approach for solving nonlinear ill-posed problems when a finite-dimensional 'vector' of data is given. Based on the the convergence analysis for nonlinear Tikhonov regularization we show stability and convergence of the method in reflexive Banach spaces. Additionally we prove convergence rates results by using Bregman distances and discuss numerical algorithms for solving the underlying minimization problem in an efficient way. The advantage of considering multi-parameter regularization approaches is illustrated by an example arising in mathematical finance.

AMS Subject Classification: 47J06, 49N45, 65J20, 91B28
Key Words: inverse problem, nonlinear ill-posed problem, multi-parameter regularization, Bregman distance, convergence rates, Lagrangian methods

1. Introduction

Classical regularization theory are based on ill-posed operator equations

$$F(x) = y^\delta,$$

with given (single) noisy data y^δ and (single) noise level δ. Thereby $F : \mathcal{D}(F) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ denotes a (nonlinear) mapping between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Usually a regularization strategy is applied in order to find an approximate solution of this equation in a stable way, see e.g. [7] for an overview over different

Received: March 4, 2008

regularization methods for linear and nonlinear ill-posed problems. The balance between stability and approximation can be expressed by the regularization parameter $\alpha > 0$. However, there are often identification problems, where the measurement y^δ can be considered as a 'vector' $y^\delta = (y_1^\delta, \ldots, y_l^\delta)$ of different types of data. Here, the exactness of the different measurements may vary, so it is natural to consider the noise level as vector $\underline{\delta} := (\delta_1, \ldots, \delta_l)^T \in \mathbb{R}^l$ as well. This gives us the motivation for using a vector $\underline{\alpha} = (\alpha_1, \ldots, \alpha_l)^T \in \mathbb{R}^l$ as regularization parameter.

Let $\mathcal{X}$, $\mathcal{Y}_j$, $1 \le j \le l$, now denote reflexive Banach spaces. Moreover, let $F_j : \mathcal{D}(F_j) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}_j$, $1 \le j \le l$, define (nonlinear) operators which describe the relation between the parameter $x \in \mathcal{X}$ which has to to be identified and the observations $y_j \in \mathcal{Y}_j$, $1 \le j \le l$. We introduce $\mathcal{D} := \bigcap \mathcal{D}(F_j) \subset \mathcal{X}$ as the set of all feasible parameters and suppose $\mathcal{D} \neq \emptyset$. For given exact data we have to solve the system of equation

$$F_j(x) = y_j, \qquad 1 \le j \le l, \; x \in \mathcal{D}, \tag{1}$$

whereas in the case of noisy data y_j^δ with $\|y_j - y_j^\delta\| \le \delta_j$, $1 \le j \le l$, we have the perturbed system

$$F_j(x) = y_j^\delta, \qquad 1 \le j \le l, \; x \in \mathcal{D}. \tag{2}$$

In many applications such problems are ill-posed. Even if we suppose unique solvability of the system (1) or (2), a solution $x \in \mathcal{D}$ does not depend continuously on the given data. That means, if $x^\dagger$ denotes the unique solution of (1) and x^δ is the unique solution of (2) then the distance $\|x^\dagger - x^\delta\|$ might become arbitrarily large even the noise levels δ_j, $1 \le j \le l$, are arbitrarily small.

Therefore regularization strategies has to be applied. One standard approach is to introduce a stabilizing functional $J : \mathcal{D} \subseteq \mathcal{X} \longrightarrow \mathbb{R}$ and replace (2) by the constrained minimizing problem

$$\begin{cases} J(x) \to \min \quad \text{subject to} \quad x \in \mathcal{D} \\ \left\| F_j(x) - y_j^\delta \right\|_{\mathcal{Y}_j} \le \delta_j, \quad 1 \le j \le l. \end{cases} \tag{3}$$

Such problems were originally considered in [11, Chapter 4.2] for discrete data, see also [5] for an application in Hilbert spaces. Our aim is to present a stability and convergence theory including convergence rates results for the convergence of solutions of (3) to a solution of the unperturbed system (1) when the noise levels δ_j, $1 \le j \le l$, tend to zero.

Therefore, the paper is organized as follows: based on standard assumptions we show the well-posedness of (3). In fact it will be shown, that the analysis is quite similar as in the case of a single data. In Section 3 we consider convergence rates in terms of Bregman distances which provide a concept for generalizing convergence rates for arbitrary penalty functionals. Section 4 presents two algorithms for solving (3) in Hilbert spaces. The algorithms are closely related to specials variants of the multiplier method for inequality constrained minimization problems. Afterwards we show convergences rates for Tikhonov multi-parameter regularization again in terms of Bregman distances. Finally we shortly present an application which arises in financial mathematics including some numerical results.

2. Stability Results

First of all we want to show the well-posedness of the minimizing problem (3). In particular, we give conditions for the existence of a solution x^δ of (3), which depends stable on the given data y_j^δ (and the noise-level δ_j) for $1 \leq j \leq l$. As we will see, the analysis does not much differ from the one for proving existence, stability and convergence of minimizers x_α^δ of the Tikhonov functional

$$\left\| F(x) - y^\delta \right\|_{\mathcal{Y}}^2 + \alpha J(x) \tag{4}$$

in classical regularization theory, see [6] for $J(x) := \|x - x^*\|^2$, [13] for $J(x) := \|D(x - x^*)\|^2$ with closed linear operator D or [17] for more general penalty terms $J(x)$. Although the basic assumptions slightly differ the main ideas of the proofs are in general the same,see also [2]. Recently, the well-posedness of (4) in more general topologies was shown in [12].

The assumptions under considerations are in detail:

(A1) For exact data there exists a solution $x^\dagger$ of (1), i.e. $F_j(x^\dagger) = y_j$ for all $1 \leq j \leq l$.

(A2) The operators F_j, $1 \leq j \leq l$, are continuous and weakly sequentially closed, i.e. for $x_n \in \mathcal{D}(F_j)$, $x_n \rightharpoonup x$, $F_j(x_n) \rightharpoonup y_j$ we have $x \in \mathcal{D}(F_j)$ and $F_j(x) = y_j$.

(A3) The penalty functional J is nonnegative and weakly sequentially lower semi-continuous, i.e. $x_n \rightharpoonup x$ implies $J(x) \leq \liminf_{n\to\infty} J(x_n)$.

(A4) The sets

$$\mathcal{A}(C) := \left\{ x \in \mathcal{X} \,:\, J(x) + \sum_{j=1}^{l} \|F_j(x)\|_{\mathcal{Y}_j} \leq C \right\}$$

are bounded in $\mathcal{X}$ for all $C \geq 0$.

(A5) The properties $x_n \rightharpoonup x$, $J(x_n) \to J(x)$ imply $x_n \to x$.

Note, that assumption (A1) seems to be natural for identification problems. If we have (exact) observations y_j, $1 \le j \le l$, we can assume a 'cause' $x^\dagger \in \mathcal{D}$ for the given data. This might not remain true for given noisy data y_j^δ, $1 \le j \le l$. Hence, a solution of (2) does not need to exist.

With

$$\mathcal{M}_\delta := \Big\{ x \in \mathcal{X} \,:\, \|F_j(x) - y_j^\delta\| \le \delta_j,\ 1 \le j \le l \Big\} \tag{5}$$

we denote the set of feasible elements. Since $x^\dagger \in \mathcal{M}_\delta$ for all $\delta \ge 0$ the sets are nonempty. By the continuity of the operators F_j, $1 \le j \le l$, the sets $\mathcal{M}_\delta$ are closed.

We show the existence of a solution x^δ of the problem (3).

Lemma 1. *Under the conditions (A1)-(A4) there exists a solution x^δ of (3).*

Proof. Let $\{x_n\} \subset \mathcal{M}_\delta$ be a minimizing sequence, i.e. $J(x_{n+1}) \le J(x_n)$. Hence, because of (A4), the sequence $\{(x_n, F_1(x_n), \ldots, F_l(x_n))\}$ is bounded in $\mathcal{X} \times \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_l$. Consequently, there exists a weak subsequence $\{x_{n_k}\}$ with $x_{n_k} \rightharpoonup x_{n_k}$ and $F_j(x_{n_k}) \rightharpoonup \hat{y}_j$, $1 \le j \le l$. By (A2) and the lower semi-continuity of the norm we have

$$\|F_j(x) - y_j\| \le \liminf_{k\to\infty} \|F_j(x_{n_k}) - y_j\| \le \delta_j$$

which gives $x \in \mathcal{M}_\delta$. Hence, by (A3), $J(x) \le \liminf_{k\to\infty} J(x_{n_k})$, which shows, that x is a minimizer of (3). □

Under certain conditions to perturbations in the side constraints we can prove stable dependence of a solution of (3) from the given data. A similar proof for discrete data can be found in [11, Theorem 4.24].

Lemma 2. *Let (A1)-(A4) hold. Moreover, let be $y_j^{(n)}$ with $\|y_j^{(n)} - y_j^\delta\| \le c_j^{(n)}$ with $c_j^{(n)} \to 0$ for $n \to \infty$. Let $\{x_n^\delta\}$ denotes a sequence of solutions of (3), where y_j^δ and δ_j are replaced by $y_j^{(n)}$ and $c_j^{(n)} + \delta_j$. Then there exists a subsequence $\{x_{n_k}^\delta\}$ converging weakly to a minimizer x^δ of (3). If the minimizer is unique then $x_n^\delta \rightharpoonup x^\delta$.*

Proof. For given n let $\mathcal{M}_\delta^{(n)}$ denote the set of all feasible elements. By construction we have $x^\delta \in \mathcal{M}_\delta^{(n)}$ and $\mathcal{M}_\delta^{(n)} \to \mathcal{M}_\delta$. Hence, $J(x_n^\delta) \le J(x^\delta)$ for all n. Same arguments as above we have a subsequence $\{x_{n_k}^\delta\}$ with $x_{n_k}^\delta \rightharpoonup x \in \mathcal{M}_\delta$.

This implies

$$J(x^\delta) \leq J(x) \leq \liminf_{k\to\infty} J(x^\delta_{n_k}) \leq J(x^\delta).$$

Now, $J(x) = J(x^\delta)$ which shows, that x is a minimizer of (3). The uniqueness of the minimizer says $x = x^\delta$ which trivially shows the weak convergence of the whole sequence $\{x_n\}$. □

The introduction of the variables $c_j^{(n)}$ has technical reasons. They only ensure that for fixed n^* the element $x^\delta_{n^*}$ belongs to $\mathcal{M}_\delta^{(n)}$ for $n \leq n^*$, which is used in the proof.

The additional assumption (A5) clearly gives strong convergence.

Lemma 3. *Let the assumptions of Lemma 2 hold. If additionally (A5) holds, then the convergence is strong.*

Finally we want to prove the convergence of a solution of (3) to a solution $x^\dagger$ of (1) for $\delta_j \to 0$, $1 \leq j \leq l$. Moreover, such a solution hat the property that it minimizes the penalty functional J under all solutions of (1). Therefore we introduce the following notation.

Definition 4. We call $x^\dagger$ a J-minimizing solution of (1) if

$$J(x^\dagger) = \min\left\{J(x) \,:\, F_j(x) = y_j,\ 1 \leq j \leq l\right\}.$$

The previous Lemmas 2 and 3 now immediately show the convergence x^δ to a J-minimizing solution $x^\dagger$ of (1) for $\delta_j \to 0$, $1 \leq j \leq l$. In fact, we have to replace $\delta_j + c_j^{(n)}$ by δ_j and y_j^δ by y for $\leq j \leq l$ and repeat the proof of Lemma 2. The corresponding statement is given below.

Lemma 5. *Let (A1)-(A4) hold and $\delta_j \to 0$, i.e. $y_j^\delta \to y_j$, $1 \leq j \leq l$. If x^δ denote the corresponding solutions of (3), then there exists a subsequence $\{\tilde{x}^\delta\}$ converging weakly to a J-minimizing solution $x^\dagger$ of (1). Moreover, if the J-minimizing element $x^\dagger$ is unique, then $x^\delta \rightharpoonup x^\dagger$. If additionally (A5) holds, the convergence is strong.*

3. Convergence Rates

Now we consider the rate of convergence of $x^\delta \to x^\dagger$ for $\delta_j \to 0$, $1 \leq j \leq l$. Thereby we formulate convergence rates in terms of Bregman distances. This concept was introduced in [4], see also [12] for some newer results. For simplicity we assume Fréchet-differentiability of the functional J. Then we can define

Bregman distances.

Definition 6. Let be J convex and Fréchet-differentiable with Fréchet-derivative $J'(x) \in \mathcal{X}^*$ for $x \in \mathcal{D}$. Then we define the Bregman distance $D(\tilde{x}, x)$ for two elements $\tilde{x}, x \in \mathcal{D}$ as

$$D(\tilde{x}, x) := J(\tilde{x}) - J(x) - \left\langle J'(x), \tilde{x} - x \right\rangle_{\mathcal{X}^*,\mathcal{X}},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{X}^*,\mathcal{X}}$ denotes the duality product in $\mathcal{X}$.

The convexity of J implies $D(\tilde{x}, x) \geq 0$ for all $\tilde{x}, x \in \mathcal{D}$. If we additional assume strict convexity, then $D(\tilde{x}, x) > 0$ for $\tilde{x} \neq x$. The idea of Bregman distances originally was applied to non-smooth functionals J, where $D(\tilde{x}, x)$ denote sets of subgradients of J at the point x, see [4]. Note, that all assertions stated below remain true in that case.

For the proof of convergence rates we suppose the following condition, which was firstly introduced in [12].

(A6) There exist $0 \leq \gamma < 1$ and $\underline{\beta} = (\beta_1, \ldots, \beta_l)^T \in \mathbb{R}^l$ with $\beta_j \geq 0$, $1 \leq j \leq l$, such that

$$|\langle J'(x^\dagger), x - x^\dagger \rangle| \leq \gamma D(x, x^\dagger) + \sum_{j=1}^{l} \beta_j \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j} \tag{6}$$

for all $x \in \mathcal{B}_r(x^\dagger) \cap \mathcal{D}$ with $\mathcal{B}_r(x^\dagger) := \{x \in \mathcal{X} \, : \, \|x - x^\dagger\| < r\}$ and $r > 0$ sufficiently large.

Under this additional assumption we can easily prove convergence rates depending on $\|\underline{\delta}\|_2$ where $\| \cdot \|_2$ denotes the Euclidean norm in $\mathbb{R}^l$.

Theorem 7. *Assume (A1)-(A6) to be hold. Then with* $\underline{\delta} := (\delta_1, \ldots, \delta_l)^T \in \mathbb{R}^l$ *we have*

$$D(x^\delta, x^\dagger) \leq \frac{2}{1-\gamma} \|\underline{\beta}\|_2 \|\underline{\delta}\|_2 \,. \tag{7}$$

Proof. The minimizing property $J(x^\delta) \leq J(x^\dagger)$ implies

$$\begin{aligned}
D(x^\delta, x^\dagger) &\leq \left\langle J'(x^\dagger), x^\dagger - x^\delta \right\rangle_{\mathcal{X}^*,\mathcal{X}} \\
&\leq \gamma D(x^\delta, x^\dagger) + \sum_{j=1}^{l} \beta_j \|F_j(x^\delta) - F_j(x^\dagger)\|_{\mathcal{Y}_j} \\
&\leq \gamma D(x^\delta, x^\dagger) + \sum_{j=1}^{l} \beta_j \left(\|F_j(x^\delta) - y_j^\delta\|_{\mathcal{Y}_j} + \|y_j^\delta - y_j\|_{\mathcal{Y}_j} \right) \\
&\leq \gamma D(x^\delta, x^\dagger) + 2\|\underline{\beta}\|_2 \|\underline{\delta}\|_2 \,,
\end{aligned}$$

which shows the assertion. $\square$

In order to prove convergence rates in the norm of $\mathcal{X}$ we need additional conditions to the operators F_j, $1 \leq j \leq l$, and/or the penalty functional J. If, for example, J strongly convex in $x^\dagger$, then (7) implies

$$\left\| x^\delta - x^\dagger \right\|_{\mathcal{X}}^2 \leq c\, D(x^\delta, x^\dagger) \leq \frac{2c}{1-\gamma} \|\underline{\beta}\|_2 \|\underline{\delta}\|_2$$

for a constant $c > 0$. On the other hand, if $J(x) = \|x - x^*\|^2$ and $\mathcal{X}$ is supposed to be a Hilbert space, then obviously $D(\tilde{x}, x) = \|\tilde{x} - x\|^2$ so that (7) already presents convergence rates in the norm of $\mathcal{X}$ in that case. In the case $J(x) = \|D(x - x^*)\|^2$ conditions for convergence rates can be found in [15] for regularization of linear operator equations and in [13] in the nonlinear case.

We shortly discuss the assumption (A6). It can be fulfilled by assuming a source condition

$$J'(x^\dagger) = \sum_{j=1}^{l} G_j^* \omega_j, \qquad \omega_j \in \mathcal{Y}_j^*,\ 1 \leq j \leq l\,, \tag{8}$$

where $G_j \in L(\mathcal{X}, \mathcal{Y}_j)$ is a linear approximation of F_j with the residual

$$r_j(x, x^\dagger) := F_j(x) - F_j(x^\dagger) - G(x - x^\dagger) \tag{9}$$

satisfying an additional restriction. We give three examples, which show that (A6) is closely related to assumptions which are uses in classical regularization theory.

Lemma 8. *Assume (8) holds with*

$$\|r_j(x, x^\dagger)\|_{\mathcal{Y}_j} \leq C_j D(x, x^\dagger) \quad \text{and} \quad C := \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} C_j < 1\,.$$

Then (A6) hold with $\gamma = C$ and $\beta_j = \|\omega_j\|_{\mathcal{Y}_j^*}$, $1 \leq j \leq l$.

Proof. We have

$$\begin{aligned}
\left\langle J'(x^\dagger), x - x^\dagger \right\rangle_{\mathcal{X}^*, \mathcal{X}} &= \sum_{j=1}^{l} \left\langle \omega_j, G_j(x - x^\dagger) \right\rangle_{\mathcal{Y}_j^*, \mathcal{Y}_j} \\
&= \sum_{j=1}^{l} \left\langle \omega_j, F_j(x) - F_j(x^\dagger) - r_j(x, x^\dagger) \right\rangle_{\mathcal{Y}_j^*, \mathcal{Y}_j} \\
&\leq \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} \left(\|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j} + \|r_j(x, x^\dagger)\|_{\mathcal{Y}_j} \right)
\end{aligned}$$

$$\leq \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j} + \left(\sum_{j=1}^{l} C_j \|\omega_j\|_{\mathcal{Y}_j^*} \right) D(x, x^\dagger). \quad \square$$

Assume that $J(x) = \|x - x^*\|^2$ and that the operators F_j are Fréchet-differentiable. Then the conditions of Lemma 8 are similar to the one for classical Tikhonov regularization (see e.g. [6]), where Lipschitz continuity of the Fréchet-derivative was supposed. We only have to set $G_j = F_j'(x^\dagger)$, $1 \leq j \leq l$. This shows, that condition (6) is a generalization of classical assumptions for proving convergence rates. On the other hand, the Lipschitz continuity of the Fréchet-derivative is not sufficient to prove convergence (rates) for iterative regularization methods. Hence different assumptions to restrict the nonlinearity of the operators were used (see [8] for a brief overview over different iterative regularization methods and their convergence analysis). So we give two more conditions, under which (6) is satisfied.

Lemma 9. *Assume (8) holds with*

$$\|r_j(x, x^\dagger)\|_{\mathcal{Y}_j} \leq C_j \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j}, \qquad 1 \leq j \leq l.$$

Then (A6) hold with $\gamma = 0$ *and* $\beta_j = \|\omega_j\|_{\mathcal{Y}_j^*}(C_j + 1)$, $1 \leq j \leq l$.

Proof. Following the lines of the proof above we conclude

$$\left\langle J'(x^\dagger), x - x^\dagger \right\rangle_{\mathcal{X}^*, \mathcal{X}} \leq \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} \left(\|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j} + \|r_j(x, x^\dagger)\|_{\mathcal{Y}_j} \right)$$

$$\leq \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} (C_j + 1) \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j}. \quad \square$$

For $C_j < \frac{1}{2}$ the condition above is known as so-called η-condition, which was firstly used to prove convergence of the nonlinear Landweber iteration method [9]. Alternatively we can use the operators G_j to restrict the residuals $r_j(x, x^\dagger)$.

Lemma 10. *Assume (8) holds with*

$$\|r_j(x, x^\dagger)\|_{\mathcal{Y}_j} \leq C_j \|G(x - x^\dagger)\|_{\mathcal{Y}_j}, \qquad 1 \leq j \leq l\,,$$

and $C_j < 1$. *Then (A6) hold with* $\gamma = 0$ *and* $\beta_j = \|\omega_j\|_{\mathcal{Y}_j^*}(1 - C_j)^{-1}$, $1 \leq j \leq l$.

Proof. By applying triangle and inverse triangle inequality we have

$$(1 - C_j)\|G_j(x - x^\dagger)\|_{\mathcal{Y}_j} \leq \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j} \leq (1 + C_j)\|G_j(x - x^\dagger)\|_{\mathcal{Y}_j}.$$

Taking the first part we have

$$\left\langle J'(x^\dagger), x - x^\dagger \right\rangle_{\mathcal{X}^*,\mathcal{X}} \leq \sum_{j=1}^{l} \|\omega_j\|_{\mathcal{Y}_j^*} \|G_j(x - x^\dagger)\|_{\mathcal{Y}_j}$$

$$\leq \sum_{j=1}^{l} \frac{\|\omega_j\|_{\mathcal{Y}_j^*}}{1 - C_j} \|F_j(x) - F_j(x^\dagger)\|_{\mathcal{Y}_j}. \quad \square$$

These three examples shows, that condition (6) provides an interesting and reasonable way for presenting convergence rates results.

4. Multi-Parameter Regularization in Hilbert Spaces

In this section we present two approaches for solving the minimizing problem (3). Therefore, let us now assume for simplicity, that $\mathcal{X}$ and $\mathcal{Y}_j$, $1 \leq j \leq l$, are Hilbert spaces. A generalization of the idea to arbitrary (reflexive) Banach spaces is in principle possible. Moreover, we suppose that the stabilizing functional $J(x)$ is twice (continuously) differentiable.

We apply Lagrange techniques for solving (3). By reformulation of the inequality constraints we can introduce the Lagrangian functional

$$L(x, \underline{\lambda}) := J(x) + \sum_{j=1}^{l} \lambda_j \left(\left\| F_j(x) - y_j^\delta \right\|_{\mathcal{Y}_j}^2 - \delta_j^2 \right) \tag{10}$$

as well as the Tikhonov-like functional

$$\tilde{T}_\lambda(x) := J(x) + \sum_{j=1}^{l} \lambda_j \left\| F_j(x) - y_j^\delta \right\|_{\mathcal{Y}_j}^2 . \tag{11}$$

For $\lambda_j > 0$, the Lagrange multiplier λ_j can be considered as inverse of regularization parameter $\alpha_j := \lambda_j^{-1}$, so that we have up to l regularization parameter instead of single parameter as common in classical regularization theory.

For finding a solution x^δ of (3), the following algorithm was suggested in [11, Algorithm 4.26]:

- For given $x_k^\delta \in \mathcal{D}$ and $\underline{\lambda}^{(k)} = (\lambda_1^{(k)}, \ldots, \lambda_l^{(k)}) \geq 0$ find a minimizer $x_{k+1}^\delta \in \mathcal{D}$ of (11) and update the Lagrange multipliers via

$$\lambda_j^{(k+1)} := \lambda_j^{(k)} \max\left(\frac{\|F_j(x_{k+1}^\delta) - y_j^\delta\|^\nu}{\delta_j^\nu}, \varepsilon \right) \tag{12}$$

  for $0 < \varepsilon \ll 1$ and arbitrary $\nu > 0$.

It can be shown, that if we have convergence $x_k^\delta \to x^\delta$ and $\underline{\lambda}^{(k)} \to \underline{\lambda}$ for

$k \to \infty$, then the x^δ is in fact a solution of (3) ([11, Theorem 4.24], see also [5, Theorem 8]).

Theorem 11. *As* $k \to \infty$ *let* $\underline{\lambda}^{(k)} \to \underline{\lambda} \geq 0$ *and* $x_k^\delta \to x^\delta \in \mathcal{D} \subset \mathcal{X}$. *Then the pair* $(x^\delta, \underline{\lambda}) \in \mathcal{X} \times \mathbb{R}^l$ *forms a saddle point of the Lagrangian functional (10) and hence* x^δ *is an optimal solution of the problem (3).*

The approach considered above has two disadvantages in numerical realization. First, we have to solve a nonlinear optimization problem in each iteration, which can become expensive. The second point is the slow convergence of the iteration (12) for updating the Lagrange multipliers. In fact, formula (12) is based on a fixed point equation approach so we can expect only linear speed of convergence. This matter leads to an increasing numbers of iterations which raise the numerical costs again.

An improvement of the speed can be obtained by realizing, that the algorithm above is closely related to classical multiplier methods for constrained minimization problems for some ν which is shown in Appendix. Therefore we can apply, what is known as *diagonalized method of multipliers* (see [3, pp.240]) in combination with a Gauß-Newton approximation of the second derivatives of the constraints. The idea is the following: instead of finding a minimizer of (11) for fixed $\underline{\lambda}$ we apply a Newton iteration for x and update the Lagrange multipliers after each Newton-step by some rule, e.g. by (12). Therefore, for given x_k^δ and $x \in \mathcal{D}$ we linearize

$$F_j(x) \approx F_j(x_k^\delta) + F_j'(x_k^\delta)(x - x_k^\delta), \qquad 1 \leq j \leq l\,.$$

Applying one single Newton step for minimizing the functional $T_{\lambda^{(k)}}(x)$ by using this linearization we have to solve the equation

$$\left[J''(x_k^\delta) + \sum_{j=1}^{l} \lambda_j F_j'(x_k^\delta)^* F_j'(x_k^\delta) \right] d = \sum_{j=1}^{l} \lambda_j F_j'(x_k^\delta)^* \left(y_j^\delta - F_j(x_k^\delta) \right) - J'(x_k^\delta), \quad (13)$$

where $F_j'(x_k^\delta)^*$ denote the (Hilbert space-)adjoint operators of $F_j'(x_k^\delta)$, $1 \leq j \leq l$, and $d := x - x_k^\delta$. Now we can reformulate the algorithm, which gives in fact a reduction of the numerical costs. We present the algorithm in detail.

Algorithm 1

1. give initial guess $x_0^\delta \in \mathcal{X}$, $\underline{\lambda}^{(0)} := (\lambda_1^{(0)}, \ldots, \lambda_l^{(0)})^T \geq 0$, set $k := 0$

2. solve (13) and set $x_{k+1}^\delta := x_k^\delta + d$
3. update the Lagrange multipliers by (12)
4. if $\|\underline{\lambda}^{(k+1)} - \underline{\lambda}^{(k)}\|_2 < TOL1$ and $\|\Delta x\|_{\mathcal{X}} < TOL2$, then STOP, else $k := k+1$ and return to 2.

The forth step sometimes has to be modified to ensure $x_k^\delta \in \mathcal{D}$. If we suppose for example upper and/or lower bounds which restrict the domain $\mathcal{D}$, then the correction can be realized by a simple projection of $x_k^\delta + d$ onto these bounds.

The algorithm can be basically applied in general reflexive Banach spaces. The Hilbert space properties are only used for establishing the normal equation (13). So we only have replace (13) by an appropriate condition for calculating Δx in Banach spaces. This can be done for example by applying duality maps, see e.g. [18, Chapter 47.12].

Alternatively, we can apply a Lagrange-Newton technique for solving (3) in combination with an active-set strategy. Let us therefore introduce the notation

$$g_j(x) := \frac{1}{2}\left(\|F_j(x) - y_j^\delta\|^2 - \delta_j^2\right), \qquad 1 \le j \le l.$$

For given $x \in \mathcal{D} \subset \mathcal{X}$ and $\underline{\lambda} = (\lambda_1, \ldots, \lambda_l)^T \in \mathbb{R}^l$ with $\lambda_j \ge 0$, $1 \le j \le l$, we define the Lagrangian functional as

$$\tilde{L}(x\underline{\lambda}) := J(x) + \sum_{j \in \mathcal{A}_c(x)} \frac{\lambda_j}{2}\left(\|F_j(x) - y_j^\delta\|^2 - \delta_j^2\right)$$

where

$$\mathcal{A}_c(x) := \left\{ j \,:\, \|F_j(x) - y_j^\delta\|^2 - \delta_j^2 > -\frac{\lambda_j}{c} \right\}$$

denotes the sets of constraints predicted of being active at the solution x^δ of (3). Here, $c > 0$ is a given constant.

We now consider derivatives. For the constraints we have

$$g_j'(x) = F_j'(x)^*\left(F_j(x) - y_j^\delta\right) \in \mathcal{X}, \qquad 1 \le j \le l,$$

whereas the partial derivative of the Lagrangian functional with respect to x is given as

$$\begin{aligned}\tilde{L}_x(x, \underline{\lambda}) = J'(x) + &\sum_{j \in \mathcal{A}_c(x)} \lambda_j g_j'(x) \\ &= J'(x) + \sum_{j \in \mathcal{A}_c(x)} \lambda_j F_j'(x)^*\left(F_j(x) - y_j^\delta\right) \in L(\mathcal{X}).\end{aligned}$$

Here we applied the Riesz Representation Theorem in order to identify the derivatives, which belong to the dual space $\mathcal{X}^*$, with elements of the Hilbert space $\mathcal{X}$ itself. Moreover the second partial derivative $\tilde{L}_{xx}(x,\underline{\lambda})$ we approximate by

$$\tilde{L}_{xx}(x,\underline{\lambda}) \approx J''(x) + \sum_{j\in\mathcal{A}_c(x)} \lambda_j F_j'(x)^* F_j'(x) =: H(x,\underline{\lambda}).$$

Then we can formulate the following quadratic problem

$$\begin{cases} \frac{1}{2}\langle H(x,\underline{\lambda})\, d, d\rangle_{\mathcal{X}} + \langle J'(x), d\rangle_{\mathcal{X}} \to \min, \quad \text{subject to } \ d \in \mathcal{X}, \\ g_j(x) + \langle g_j'(x), d\rangle_{\mathcal{X}} = 0, \quad J \in \mathcal{A}_c(x). \end{cases} \tag{14}$$

Now we can present a fast and efficient algorithm for solving (3).

Algorithm 2

1. give initial guess $x_0^\delta \in \mathcal{X}$, $\underline{\lambda}^{(0)} := (\lambda_1^{(0)},\ldots,\lambda_l^{(0)})^T \geq 0$, $c > 0$, set $k := 0$
2. determine the active set $\mathcal{A}_c(x_k^\delta)$
3. for $(x,\underline{\lambda}) = (x_k^\delta, \underline{\lambda}^{(k)})$ find a solution $d \in \mathcal{X}$ of (14) with corresponding Lagrange multipliers $\tilde{\lambda}_j$, $j \in \mathcal{A}_c(x_k)$.
4. set $x_{k+1}^\delta := x_k^\delta + d$ and update the Lagrange multipliers via
$$\lambda_j^{(k+1)} := \begin{cases} \max(\tilde{\lambda}_j, 0), & j \in \mathcal{A}_c(x_k), \\ 0, & \text{else.} \end{cases}$$
4. if $\|d\|_{\mathcal{X}} < TOL$, then STOP, else $k := k+1$ and return to 2.

For the local convergence of the algorithm we refer to standard literature, see e.g. [1]. We also want to mention [3] for further discussions of the finite-dimensional case and to [16] for aspects of the numerical realization.

5. Tikhonov-Multi-Parameter Regularization

We consider $\underline{\lambda} := (\lambda_1,\ldots,\lambda_l)^T \in \mathbb{R}^l$ with $\|\underline{\lambda}\|_\infty = 1$ and assume $\lambda_j > 0$ for all $1 \leq j \leq l$. An alternative regularization approach to (3) can be achieved by applying a Tikhonov-like regularization approach, i.e. we search a solution x_α^δ of the minimizing problem

$$T_\alpha(x) := \sum_{j=1}^{l} \lambda_j \left\| F_j(x) - y_j^\delta \right\|_{\mathcal{Y}_j}^2 + \alpha J(x) \to \min\,. \tag{15}$$

We have two interpretations of (15) concerning convergence studies. First we can assume $\lambda_1, \ldots, \lambda_l$ to be fixed and suppose $\alpha > 0$ to be the (single) regularization parameter. Then, by introducing the Banach space $\mathcal{Y} := \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_l$ with weighted norm

$$\|\underline{y}\|_{\mathcal{Y}} := \left(\sum_{j=1}^{l} \lambda_j \|y_j\|_{\mathcal{Y}_j}^2 \right)^{\frac{1}{2}}, \quad \underline{y} = (y_1, \ldots, y_l),\ y_j \in \mathcal{Y}_j,\ 1 \le j \le l,$$

and the operator $F : \mathcal{D} \subset \mathcal{X} \longrightarrow \mathcal{Y}$ as $F(x) := (F_1(x), \ldots, F_l(x))$ we can apply stability and convergence results of classical Tikhonov regularization [17]. Here we have also one noise level δ. On the other hand, the additional parameter α in (15) can be considered as scaling parameter of $\tilde{T}_\lambda(x)$, so we in fact consider $\tilde{T}_\lambda(x) \to \min$ and $\alpha := \|\underline{\lambda}\|_\infty^{-1}$ is a parameter which is only introduced to present convergence analysis in a proper way. The vector $\underline{\alpha} := (\alpha_1, \ldots, \alpha_l)^T \in \mathbb{R}^l$ with $\alpha_j := \frac{\alpha}{\lambda_j}$, $1 \le j \le l$, play the role of the regularization parameter. The convergence rates results below will show, that it makes sense to have the second idea in mind.

By standard arguments it can be easily shown, that under the conditions (A1)-(A5) the problem (15) has at least one solution x_α^δ, which depends stable on the given data. For the sake of completeness we show the prove of convergence of x_α^δ to a solution $x^\dagger$ of (1) for $\delta \to 0$. Moreover we consider the case of incomplete minimization, i.e. we have $x_\alpha^{\delta,\eta}$ which satisfies

$$T_\alpha(x_\alpha^{\delta,\eta}) \le T_\alpha(x_\alpha^\delta) + \eta \tag{16}$$

for given $\eta \ge 0$.

Theorem 12. *Assume (A1)-(A5) to be hold. Let be $y_j^\delta \in \mathcal{Y}_j$ with $\|y_j - y_j^\delta\|_{\mathcal{Y}_j} \le \delta_j$, $\underline{\delta} := (\delta_1, \ldots, \delta_l)^T$ and let $\alpha := \alpha(\underline{\delta}, \eta)$ be such that*

$$\alpha(\underline{\delta}, \eta) \to 0,\ \frac{\|\underline{\delta}\|_2^2}{\alpha(\underline{\delta}, \eta)} \to 0 \ \text{ and } \ \frac{\eta}{\alpha(\underline{\delta}, \eta)} \to 0$$

as $\underline{\delta} \to 0$ and $\eta \to 0$. Then every sequences $\{x_{\alpha_k}^{\delta_k,\eta_k}\}$ with $\underline{\delta}_k \to 0$ and $\eta_k \to 0$, $\alpha_k = \alpha(\delta_k, \eta_k)$ with $x_{\alpha_k}^{\delta_k,\eta_k}$ satisfies (16) has a convergent subsequence. Each limit $\hat{x}$ of such subsequence is a J-minimizing solution of (1). If – in addition – the solution $x^\dagger$ of (1) is unique, then

$$\lim_{k\to\infty} x_{\alpha_k}^{\delta_k,\eta_k} = x^\dagger.$$

Proof. We set $x_k := x_{\alpha_k}^{\delta_k,\eta_k}$. By definition of x_k we have

$$\begin{aligned}\sum_{j=1}^{l}\lambda_j \left\|F_j(x_k)-y_j^{\delta_{j,k}}\right\|_{\mathcal{Y}_j}^2 &+ \alpha_k J(x_k)\\ &\le \sum_{j=1}^{l}\lambda_j \left\|F_j(x^\dagger)-y_j^{\delta_{j,k}}\right\|_{\mathcal{Y}_j}^2 + \alpha_k J(x^\dagger)+\eta_k\\ &\le \|\underline{\delta}_k\|_2^2 + \alpha_k J(x^\dagger)+\eta_k.\end{aligned}$$

Taking the limit, we have $F_j(x_k)\to y_j$ as $k\to\infty$ and

$$J(x_k) \le \frac{\|\underline{\delta}_k\|_2^2}{\alpha_k} + J(x^\dagger) + \frac{\eta_k}{\alpha_k}. \tag{17}$$

In particular, the sequence $\{J(x_k), F_1(x_k),\dots,F_l(x_k)\}$ remain bounded. Consequently there exists a weak convergent subsequence $x_{k_i}\rightharpoonup \hat{x}\in\mathcal{D}(F)$. Moreover, by (17) with $\hat{x}$ instead of $x^\dagger$ we obtain

$$J(\hat{x}) \le \liminf_{i\to\infty} J(x_{k_i}) \le \limsup_{i\to\infty} J(x_{k_i}) \le \lim_{i\to\infty}\frac{\|\underline{\delta}_k\|_2^2}{\alpha_k} + J(\hat{x}) + \frac{\eta_k}{\alpha_k} = J(\hat{x})$$

and hence $J(x_{k_i})\to J(\hat{x})$. Condition (A5) now implies $x_{k_i}\to\hat{x}$. Moreover, by (17) we conclude

$$J(\hat{x}) \le \limsup_{i\to\infty} J(x_{k_i}) \le J(x^\dagger) \le J(\hat{x}).$$

In particular, $\hat{x}$ is a J-minimizing solution (1). If the solution $x^\dagger$ is unique, i.e. $\hat{x}=x^\dagger$, the convergence $x_k\to x^\dagger$ follows from the fact, that each weak convergent subsequence of $\{x_k\}$ converges to $x^\dagger$. □

We also want to prove convergence rates.

Theorem 13. *Assume (A1)-(A6) to be hold. Then – for a choice* $\alpha\sim\|\underline{\delta}\|_2$ *– we obtain a convergence rate*

$$D(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(\|\underline{\delta}\|_2\right).$$

Proof. Starting with $T_\alpha(x_\alpha^\delta)\le T_\alpha(x^\dagger)$ we obtain

$$\begin{aligned}\sum_{j=1}^{l}\lambda_j \left\|F_j(x_\alpha^\delta)-y_j^\delta\right\|_{\mathcal{Y}_j}^2 &+ \alpha D(x_\alpha^\delta,x^\dagger)\\ &\le \sum_{j=1}^{l}\lambda_j \left\|F_j(x^\dagger)-y_j^\delta\right\|_{\mathcal{Y}_j}^2 + \alpha\left\langle J'(x^\dagger), x_\alpha^\delta - x^\dagger\right\rangle_{\mathcal{X}^*,\mathcal{X}}\\ &\le \sum_{j=1}^{l}\lambda_j\delta_j^2 + \alpha\gamma D(x_\alpha^\delta,x^\dagger) + \alpha\sum_{j=1}^{l}\beta_j\left(\left\|F_j(x_\alpha^\delta)-y_j^\delta\right\|_{\mathcal{Y}_j}+\delta_j\right)\end{aligned}$$

$$\leq \alpha\gamma D(x_\alpha^\delta, x^\dagger) + \sum_{j=1}^{l} \left(\lambda_j \delta_j^2 + \alpha\beta_j\delta_j + \frac{\alpha^2\beta_j^2}{4\lambda_j} + \lambda_j \left\| F_j(x_\alpha^\delta) - y_j^\delta \right\|_{\mathcal{Y}_j}^2 \right).$$

With the choice $\alpha := c\|\underline{\delta}\|_2$ we have

$$D(x_\alpha^\delta, x^\dagger) \leq \frac{\|\underline{\delta}\|_2}{1-\gamma} \left(\frac{1}{c} + \|\underline{\beta}\|_2 + c\sum_{j=1}^{l} \frac{\beta_j^2}{\lambda_j} \right),$$

which proves the desired convergence rate. □

Following the lines of the proof we have

$$D(x_\alpha^\delta, x^\dagger) \leq \frac{1}{1-\gamma} \sum_{j=1}^{l} \left(\frac{\sqrt{\lambda_j}\delta_j}{\sqrt{\alpha}} + \sqrt{\alpha}\frac{\beta_j}{2\sqrt{\lambda_j}} \right)^2. \tag{18}$$

The error bound becomes minimal if we set

$$\alpha_j^{-1} = \frac{\lambda_j}{\alpha} := \frac{\beta_j}{2\delta_j}, \qquad 1 \leq j \leq l.$$

Then from (18) we derive the error bound

$$D(x_\alpha^\delta, x^\dagger) \leq \frac{2}{1-\gamma} \|\underline{\delta}\|_2 \|\underline{\beta}\|_2,$$

which coincides with the error bound (7) in Theorem 7. Consequently, by solving (3) with Lagrangian methods we obtain Lagrange multipliers λ_j, $1 \leq j \leq l$, which provide an 'optimal' error bound for $D(x_\alpha^\delta, x^\dagger)$ under the assumptions under consideration.

Another point seems to be quite remarkable: the 'optimal' choice of the regularization parameter $\alpha_j = \frac{\alpha}{\lambda_j}$ does not depend only on the noise level δ_j but on the ratio $\frac{\delta_j}{\beta_j}$. Hence it seems to be difficult to find an optimal choice of $\underline{\lambda}$ a priori when considering (15) with fixed $\underline{\lambda}$ and α as the only regularization parameter. Hence Algorithm 1 can be interpreted as multi-parameter Tikhonov regularization approach in combination with an a-posteriori parameter choice strategy. The parameter choice strategy can be considered as a generalization of the discrepancy principle of Morozov, see e.g. [7, Chapter 4.3].

6. A Numerical Example

As example we consider the inverse problem of determining the time-dependent volatility from given prices of European call options. At present time $t^* = 0$ let the price $X^* > 0$ of the asset and the risk-less interest rate $r \geq 0$ be given and fixed. Following the Black-Scholes analysis (see e.g. [14, pp. 71]) the price

$y(K,t)$ for call option with strike $K>0$ and maturity $t\in[0,T]$ is given by

$$y(K,t) := U_{BS}\left(X^*, K, r, t, \int_0^t \sigma^2(\tau)\,d\tau\right),$$

where the so-called Black-Scholes function is given by

$$U_{BS}(X,K,r,t,s) := \begin{cases} X\,\Phi(d_1) - Ke^{-r\,t}\Phi(d_2), & s>0,\\ \max\left(X - Ke^{-r\,t}, 0\right) & s=0,\end{cases}$$

with

$$d_1 := \frac{\ln\frac{X}{K} + r\,t + \frac{s}{2}}{\sqrt{s}}, \qquad d_2 := d_1 - \sqrt{s}$$

and the cumulative density function of the standard normal distribution

$$\Phi(\xi) := \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\xi} e^{-\frac{\tau^2}{2}}\,d\tau.$$

Thereby the volatility function $\sigma(t)$, $t\in[0,T]$, is an additional market parameter, which has to be determined from given option price data.

In [10] the following inverse problem was analyzed: for fixed strike price $K=K^*>0$ and (noisy) data $y^*(t)=y(K^*,t)$, $t\in[0,T]$, a positive function $x\in L^2(0,T)$ (with $x=\sigma^2$) as approximate solution of the nonlinear equation

$$U_{BS}\left(X^*, K^*, r, t, \int_0^t x(\tau)\,d\tau\right) = y^*(t), \qquad t\in[0,T],$$

has to be estimated. In particular, only one strike price K^* was considered for the identification of x. On the other hand, data usually is given in form of several price functions $y_j(t)=y_j(K_j,t)$, $t\in[0,T]$, for different strikes $K_j>0$, $1\le j\le l$. So there no reason to pick out a single data for the identification problem. Moreover, by using all given data we can expect stabilizing effects which promise better reconstruction chances. Hence, the multi-parameter approach can be naturally formulated as follows: we set $X=\mathcal{Y}_1=\ldots=\mathcal{Y}_l=L^2(0,T)$, introduce the operators $F_j:\mathcal{D}\subset L^2(0,T)\longrightarrow L^2(0,T)$ as

$$[F_j(x)](t) := U_{BS}\left(X^*, K_j, r, t, \int_0^t x(\tau)\,d\tau\right), \qquad t\in[0,T],$$

whereas the domain $\mathcal{D}$ contains all function in $L^2(0,T)$ which are positive a.e. on $[0,T]$. In fact, here it makes sens to deal with different noise levels. It is well-known, that option prices with $K>X^*$ or $K<X^*$ are less sensible with respect to perturbation of the volatility function that options with $K\approx X^*$. So a higher noise level δ_j can be assumed as larger or smaller the ratio $\frac{X_j}{K^*}$ for $1\le j\le l$ becomes.

	$K = 0.7$		$K = 0.8$		$K = 0.95$	
δ_{rel}	#it.	$\frac{\|x_\alpha^\delta - x^\dagger\|}{\|x^\dagger\|}$	#it.	$\frac{\|x_\alpha^\delta - x^\dagger\|}{\|x^\dagger\|}$	#it.	$\frac{\|x_\alpha^\delta - x^\dagger\|}{\|x^\dagger\|}$
10^{-4}	9	0.0207	9	0.0097	7	0.0050
10^{-3}	8	0.0386	8	0.0251	8	0.0229
10^{-2}	7	0.0758	7	0.0644	7	0.0691
	$K = 1.1$		$K = 1.2$			
δ_{rel}	#it.	$\frac{\|x_\alpha^\delta - x^\dagger\|}{\|x^\dagger\|}$	#it.	$\frac{\|x_\alpha^\delta - x^\dagger\|}{\|x^\dagger\|}$		
10^{-4}	7	0.0081	8	0.0140		
10^{-3}	8	0.0229	8	0.0320		
10^{-2}	7	0.0691	7	0.0916		

Table 1: Regularized solutions of single measurements for different strike prices K and noise levels

We present some numerical results. We choose $X^* = 1$, $T = 1$ and $r = 0.05$. Moreover, the exact solution is given as

$$x^\dagger(t) := (t - 0.5)^2 + 0.1, \quad t \in [0, 1],$$

and the penalty functional

$$J(x) := \frac{1}{2}\|x - x^*\|^2 \quad \text{with} \quad x^* \equiv 0.35 \;\text{ on }\; [0, 1],$$

is applied. The a priori guess x^* was chosen such that $x^\dagger - x^* \in \mathcal{R}(F'(x^\dagger)^*)$, see e.g. [10]. For different strike prices $K > 0$ the option prices $y(K_j, t_i)$ were calculated at $n = 50$ times $t_i = i/n$, $1 \le i \le n$, and perturbed with a random vector. Note, that the noise level δ_{rel} here describes the relative size of this perturbation which differs from the notation in the previous sections.

The first calculations deal with single measurements, i.e. we have $l = 1$. Solving the problem (3) is a numerical variant of applying Tikhonov regularization using the discrepancy principle. The results are presented in Table 1 for three different noise levels. The number #it. denotes the number of iterations needed for solving the problem (3) by Algorithm 2. In order to improve the results obtained above we now use a higher number of measurements for identifying the unknown function x. First we set $l = 3$ and use

$$\underline{K}_{(3)} := (K_1,\, K_2,\, K_3)^T = (0.8,\, 0.95,\, 1.2)^T$$

as vector of the strike prices, for which the (noisy) data y_j^δ, $j = 1, 2, 3$, is given. Then the number of data is increased to $l = 5$ with data y_j^δ, $1 \le j \le 5$, for the

	$\underline{K}_{(3)}$		$\underline{K}_{(5)}$	
δ_{rel}	#it.	$\frac{\lVert x_\alpha^\delta - x^\dagger \rVert}{\lVert x^\dagger \rVert}$	#it.	$\frac{\lVert x_\alpha^\delta - x^\dagger \rVert}{\lVert x^\dagger \rVert}$
0	21	$< 5 \cdot 10^{-7}$	22	$< 6 \cdot 10^{-7}$
10^{-4}	15	0.0026	16	0.0019
10^{-3}	12	0.0153	13	0.0139
10^{-2}	9	0.0487	17	0.0486

Table 2: Multi-parameter regularization of 3/5 measurements for different noise levels

$\underline{\delta}_{rel}$	#it.	$\frac{\lVert x_\alpha^\delta - x^\dagger \rVert}{\lVert x^\dagger \rVert}$
$(10^{-2}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4})$	15	0.0028
$(10^{-4}, 10^{-4}, 10^{-4}, 10^{-2}, 10^{-2})$	17	0.0020
$(10^{-2}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-4})$	13	0.0076
$(10^{-4}, 10^{-4}, 10^{-2}, 0, 10^{-4})$	27	$< 5 \cdot 10^{-7}$
$(10^{-2}, 10^{-3}, 10^{-4}, 10^{-3}, 10^{-2})$	10	0.0051

Table 3: Multi-parameter regularization for variable noise levels

strike prices

$$\underline{K}_{(5)} := (K_1,\, \ldots,\, K_5)^T = (0.7,\, 0.8,\, 0.95,\, 1.1,\, 1.2)^T\,.$$

The numerical results are given in Table 2. Here the same relative noise level δ_{rel} for all l measurements is assumed. In the noiseless case $\delta_{rel} = 0$ the unknown function $x^\dagger$ could be identified exactly up to a numerical error. Comparing the results for $\delta_{rel} > 0$ with the one of Table 1 we observe, that the relative errors in Table 2 are always better than the best result using a single measurement for the identification problem. We also notice, that for $l = 5$ measurements the approximations are still a bit better than the solutions for $l = 3$. The numbers #it. of required iterations also show that the numerical costs do not increase too much. A final calculation should show the effect of data y_j^δ having different noise levels δ_j, $1 \le j \le l = 5$. Here again $\underline{K}_{(5)}$ as vector of the corresponding strike prices is used. The numerical results can be found in Table 3. Here, the vectors $\underline{\delta}_{rel} := (\delta_1, \ldots, \delta_5)$ describe the chosen relative noise levels for each data. Again, having a look at the strike price K_j with the lowest noise level δ_j, the results here are better than the corresponding approximation error using a single measurement in Table 1.

7. Conclusions

We have shown that the stability and convergence theory for nonlinear Tikhonov regularization can be easily generalized to a multi-parameter approach when we consider the given data as a 'vector' of l different measurements. The convergence rates results indicate, that is advantageously not to fix the parameter $\underline{\lambda}$ in the Tikhonov functional $T_\alpha(x)$ a-priori. There is a good chance to obtain better approximate solutions by using a vector of regularization parameter in combination with an a-posteriori parameter choice strategy for choosing the optimal regularization parameter. In particular, solving the minimization problem (3) by Lagrangian methods as suggested in Section 4 can be interpreted as one possibility for a multi-parameter Tikhonov regularization strategy with an a-posteriori choice of the regularization parameter based on the discrepancy principle of Morozov.

References

[1] W. Alt, The Lagrange-Newton method for infinite-dimensional optimization problems, *Numer. Funct. Anal. and Optimiz.*, **11**, No-s: 3, 4 (1990), 201-224.

[2] R. Acar, C.R. Vogel, Analysis of bounded variation penalty methods for ill-posed problems, *Inverse Problems*, **10** (1994), 1217-1229.

[3] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Athena Scientific (1996).

[4] M. Burger, S. Osher, Convergence rates of convex variational regularization, *Inverse Problems*, **20** (2004), 1411-1421.

[5] D. Düvelmeyer, B. Hofmann, A multi-parameter regularization approach for estimating parameters in jump diffusion processes, *J.Inverse Ill-Posed Probl.*, **14**, No. 9 (2006), 861-880.

[6] H.W. Engl, K. Kunisch, A. Neubauer, Convergence rates for Tikhonov regularization of nonlinear ill-posed problems, *Inverse Problems*, **5** (1989), 523-540.

[7] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dortrecht (1996).

[8] H.W. Engl, O.Scherzer, Convergence rate results for iterative methods for solving nonlinear ill-posed problems, In: *Surveys on Solution Methods for Inverse Problems* (Ed. David Colton), Springer Wien (2000), 7-34.

[9] M. Hanke, A. Neubauer, O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numer. Math.*, **72** (1995), 21-27.

[10] T. Hein, B. Hofmann, On the nature of ill-posedness of an inverse problem arising in option pricing, *Inverse Problems*, **19** (2003), 1319-1338.

[11] B. Hofmann, *Regularization for Applied Inverse and Ill-Posed Problems*, Teubner Verlag Leipzig (1986).

[12] B. Hofmann, B. Kaltenbacher, C. Pöschl, O. Scherzer, A convergence rates result in Banach spaces with non-smooth operators, *Inverse Problems*, **23** (2007), 987-1010.

[13] K. Kunisch, W. Ring, Regularization of nonlinear illposed problems with closed operators, *Numer. Funct. Anal. Optimization*, **14**, No-s: 3-4 (1993), 389-404.

[14] Y.K. Kwok, *Mathematical Models of Financial Derivatives*, Springer, Singapore (1998).

[15] J. Locker, P.M.Prenter, Regularization with differential operators. I: General theory, *J. Math. Anal. Appl.*, **74**, 504-529 (1980).

[16] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer-Verlag New York (1999).

[17] T.I. Seidman, C.R. Vogel, Well posedness and convergence of some regularisation methods for non-linear ill-posed problems, *Inverse Problems*, **5** (1989), 227-238.

[18] E. Zeidler, *Nonlinear Functional Analysis and its Applications*, Springer-Verlag, New York (1985).

Appendix. Multi-Parameter Regularization and Multiplier Methods

We show that Algorithm 1 is closely related to multiplier methods for inequality constrained optimization problems. Let again $\mathcal{X}$ be a Banach space. We consider the problem

$$\begin{cases} J(x) \to \min\,, \\ g_j(x) \le 0, \quad 1 \le j \le l\,, \end{cases} \tag{19}$$

with objective functional $J : \mathcal{X} \longrightarrow \mathbb{R}$ and functionals $g_j : \mathcal{X} \longrightarrow \mathbb{R}$, $1 \le j \le l$, which describe the side constraints. For given Lagrange multiplier $\underline{\lambda} = (\lambda_1, \dots, \lambda_l)^T \in \mathbb{R}^l$ with $\lambda_j \ge 0$ and constant $c > 0$, the augmented Lagrangian is given as

$$M(x, \underline{\lambda}, c) := J(x) + \frac{1}{2c} \sum_{j=1}^{l} \left(\max\{0, \lambda_j + c\, g_j(x)\}^2 - \lambda_j^2 \right). \tag{20}$$

The Multiplier method is now given as follows: for given $\underline{\lambda}_k \ge 0$, $c_k > 0$ we estimate x_k as solution of

$$M(x, \underline{\lambda}_k, c_k) \to \min \quad \text{subject to } x \in \mathcal{X}, \tag{21}$$

and update the Lagrange multiplier via

$$\lambda_j^{(k+1)} := \max\left\{0, \lambda_j^{(k)} + c_k g_j(x_k)\right\}, \quad 1 \le j \le l. \tag{22}$$

The parameter c_{k+1} is chosen by some rule (usually $c_{k+1} \ge c_k$ recommended). We introduce a slight modification, which seems to be quite natural. Instead of a single parameter $c \in \mathbb{R}$ we choose a vector $\underline{c} = (c_1, \dots, c_l)^T \in \mathbb{R}^l$ with $c_j > 0$ and replace (20) by

$$M(x, \underline{\lambda}, \underline{c}) := J(x) + \sum_{j=1}^{l} \frac{1}{2c_j} \left(\max\{0, \lambda_j + c_j g_j(x)\}^2 - \lambda_j^2 \right). \tag{23}$$

Now, for $\nu > 0$, we set

$$g_j(x) := 2\delta_j^\nu \left(\left\| F_j(x) - y_j^\delta \right\|_{\mathcal{Y}_j}^\nu - \delta_j^\nu \right).$$

and

$$c_j := \frac{\lambda_j}{2\delta_j^{2\nu}}.$$

Hence

$$
\begin{aligned}
&M(x,\underline{\lambda},\underline{c})\\
&= J(x)+\sum_{j=1}^{l}\tfrac{1}{2c_j}\left(\max\left\{0,\lambda_j+2c_j\delta_j^{\nu}\left(\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}-\delta_j^{\nu}\right)\right\}^2-\lambda_j^2\right)\\
&\quad= J(x)+\sum_{j=1}^{l}\tfrac{\delta_j^{2\nu}}{\lambda_j}\left(\max\left\{0,\lambda_j+\tfrac{\lambda_j}{\delta_j^{\nu}}\left(\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}-\delta_j^{\nu}\right)\right\}^2-\lambda_j^2\right)\\
&\quad= J(x)+\sum_{j=1}^{l}\delta_j^{2\nu}\lambda_j\left(\max\left\{0,\tfrac{\delta_j^{\nu}+\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}-\delta_j^{\nu}}{\delta_j^{\nu}}\right\}^2-1\right)\\
&= J(x)+\sum_{j=1}^{l}\lambda_j\left(\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{2\nu}-\delta_j^{2\nu}\right),
\end{aligned}
$$

which coincides with the Lagrangian (10) for $\nu=1$. Moreover, for the parameter update (22) we conclude

$$
\begin{aligned}
\lambda_j^{(k+1)} &:= \max\left\{0,\lambda_j^{(k)}+2c\delta_j^{\nu}\left(\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}-\delta_j^{\nu}\right)\right\}\\
&= \max\left\{0,\lambda_j^{(k)}\frac{\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}}{\delta^{\nu}}\right\}=\lambda_j^{(k)}\max\left\{0,\frac{\left\|F_j(x)-y_j^{\delta}\right\|_{\mathcal{Y}_j}^{\nu}}{\delta^{\nu}}\right\},
\end{aligned}
$$

which is equal to (12) with $\varepsilon=0$ for the case $\nu=1$.

Remark 14. The functional (20) can be derived after some calculations from the (classical) augmented Lagrangian with quadratic penalty term. By a similar calculation it can be shown, that the choice $\nu=2$ in (12) can be interpreted as a variant of the multiplier method with exponential penalty functional (see e.g. [3, Chapter 5] for theoretical aspects of multiplier methods with nonquadratic penalty terms).