IOP Publishing Inverse Problems

Inverse Problems **24** (2008) 025025 (20pp) doi:10.1088/0266-5611/24/2/025025

Efficient computation of the Tikhonov regularization parameter by goal-oriented adaptive discretization

Anke Griesbaum[1], **Barbara Kaltenbacher**[2] **and Boris Vexler**[3]

[1] University of Heidelberg, Heidelberg, Germany
[2] University of Stuttgart, Stuttgart, Germany
[3] RICAM Linz, Austrian Academy of Sciences, Linz, Austria

Received 30 October 2007, in final form 13 February 2008
Published 14 March 2008
Online at stacks.iop.org/IP/24/025025

Abstract

Parameter identification problems for partial differential equations (PDEs) often lead to large-scale inverse problems. For their numerical solution it is necessary to repeatedly solve the forward and even the inverse problem, as it is required for determining the regularization parameter, e.g., according to the discrepancy principle in Tikhonov regularization. To reduce the computational effort, we use adaptive finite-element discretizations based on goal-oriented error estimators. This concept provides an estimate of the error in a so-called quantity of interest, which is a functional of the searched for parameter q and the PDE solution u. Based on this error estimate, the discretizations of q and u are locally refined. The crucial question for parameter identification problems is the choice of an appropriate quantity of interest. A convergence analysis of the Tikhonov regularization with the discrepancy principle on discretized spaces for q and u provides a possible answer: it shows, that in order to determine the correct regularization parameter, one has to guarantee sufficiently high accuracy in the squared residual norm—which is therefore our quantity of interest—whereas q and u themselves need not be computed precisely everywhere. This fact allows for relatively low dimensional adaptive meshes and hence for a considerable reduction of the computational effort. In this paper, we study an efficient inexact Newton algorithm for determining an optimal regularization parameter in Tikhonov regularization according to the discrepancy principle. With the help of error estimators we guide this algorithm and control the accuracy requirements for its convergence. This leads to a highly efficient method for determining the regularization parameter.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In this paper, we consider inverse problems for partial differential equations and develop an efficient algorithm for determining the regularization parameter for Tikhonov regularization. The proposed method is based on the discrepancy principle on the one hand and on exploiting adaptive finite-element discretizations on the other hand.

Driven by efficiency requirements for the solution of increasingly large-scale inverse problems, adaptivity has recently been attracting more and more interest in the inverse problems community. For instance, we point to [1], where refinement and coarsening indicators are extracted from Lagrange multipliers for the misfit functional with constraints incorporating local changes of the discretization. Moreover, we refer to [15, 22], where sloppily speaking the magnitude of gradients is used as a criterion for local refinement. Also, we would like to refer to very interesting ideas on '*a priori*' adaptivity in [8].

Inverse problems for partial differential equations such as parameter identification or inverse boundary value problems can usually be written as operator equations, where the forward operator is the composition

$$F = C \circ S$$

of a parameter-to-solution map for a PDE

$$\begin{aligned} S : Q &\to V \\ q &\mapsto u \end{aligned}$$

with some measurement operator

$$\begin{aligned} C : V &\to G \\ u &\mapsto g, \end{aligned}$$

where Q, V, G are appropriate Hilbert spaces. Throughout the paper we denote by $\|\cdot\|_Q$ the norm and by $(\cdot,\cdot)_Q$ the inner product in Q. Similar notation is used for V and G.

Here, we will write the underlying (possibly nonlinear) PDE in its weak form

$$u \in V : A(q,u)(v) = f(v) \qquad \forall v \in V, \tag{1}$$

where u denotes the solution of the forward (state) equation (1), q some searched for parameter or boundary function and $f \in V^*$ some given right-hand side. We will assume that the forward equation (1) and especially also its linearization at (q,u) is uniquely and stably solvable, i.e.

$$A'_u(q,u)^{-1} \in L(V^*, V). \tag{2}$$

As a matter of fact, parameter identification problems often lead to nonlinear equations

$$F(q) = g \tag{3}$$

where F is a nonlinear operator between Hilbert spaces Q and G. Still, there are also many linear inverse problems for PDEs such as certain inverse source or inverse boundary value problems, thus in this paper we will mainly consider the linear case of (3)

$$Tq = g, \tag{4}$$

and refer to the forthcoming paper [13] for the fully nonlinear case.

Since we are interested in the situation that the solution of (4) does not depend continuously on the data and we are only given noisy data g^δ with noise level δ according to

$$\|g^\delta - g\|_G \leqslant \delta, \tag{5}$$

it is necessary to apply regularization methods for their stable solution. When applying one of the well-known regularization methods such as Tikhonov regularization

$$\text{Minimize} \qquad j_\alpha(q) = \|F(q) - g^\delta\|_G^2 + \alpha\|q\|_Q^2 \text{ over } q \in Q, \tag{6}$$

or with $F = C \circ S$ equivalently

$$\text{Minimize} \qquad J_\alpha(q,u) = \|C(u) - g^\delta\|_G^2 + \alpha\|q\|_Q^2 \text{ over } q \in Q, \qquad u \in V,$$
$$\text{under the constraints} \qquad A(q,u)(v) = f(v) \qquad \forall\, v \in V,$$

it is essential to choose some regularization parameter (here $\alpha > 0$) in an appropriate way. A both theoretically and practically well-established method for doing so *a posteriori* among others (see, e.g., [12, 17, 18] as well as a remark in section 5) is the discrepancy principle: the parameter α_* is determined by

$$\left\| F\left(q_{\alpha_*}^\delta\right) - g^\delta \right\|_G = \tau\delta \tag{7}$$

with some constant $\tau \geqslant 1$, where q_α^δ denotes a (global) minimizer of (6) for given regularization parameter α. We introduce the reciprocal of the regularization parameter $\beta = 1/\alpha$ and define a function $i: \mathbb{R}_+ \to \mathbb{R}$ describing the squared residual norm as a function of β:

$$i(\beta) = \left\| F\left(q_\alpha^\delta\right) - g^\delta \right\|_G^2, \qquad \alpha = \frac{1}{\beta}. \tag{8}$$

Then, an optimal value of regularization parameter has to be computed as a solution to the one-dimensional nonlinear equation

$$i(\beta) = \tau^2\delta^2. \tag{9}$$

Newton's iteration can be applied to (9) and is known to be fast and in the linear case (4) globally convergent (cf, e.g., chapter 9 in [11]). However, this iteration can be numerically very expensive, since the evaluation of $i(\beta)$ for the current iterate β requires the solution of the optimization problem (6). In addition, the derivative $i'(\beta)$ has to be computed in each Newton step. To reduce the computational effort, we will use adaptive finite elements for the discretization of (6) guided by specially designed *a posteriori* error estimators. The underlying meshes should on the one hand be as coarse as possible to save computational effort and on the other hand locally fine enough to preserve global and fast convergence of Newton's method as well as sufficient accuracy in the solution of (9).

For this purpose, we use the concept of goal-oriented error estimators as introduced in [6, 7] for optimization problems with partial differential equations based on the approach from [5]. These *a posteriori* error estimators allow us to assess the discretization error between the solution of the optimization problem (6) and its discrete counterpart obtained by a finite-element discretization. The discretization error can be estimated with respect to a given quantity of interest $I(q,u)$ which may depend on the parameter (control) q as well as the state variable u. On the basis of these error estimators finite-element meshes are locally refined in order to achieve given accuracy requirements on the quantity of interest in an efficient way.

A crucial question that we have to answer beforehand is the choice of an appropriate quantity of interest. Note that in the identification of a distributed parameter one might think of having infinitely many quantities of interest, namely the values of the parameter function in each point, which would obviously be practically useless for the definition of an efficient refinement strategy, though. However, considering Tikhonov regularization with the discrepancy principle as well as Newton's method for solving (7), it is intuitively clear that the value of the Tikhonov functional $j_\alpha(q)$, the squared residual norm $i(\beta)$ and its derivative with respect to the regularization parameter $i'(\beta)$ are important quantities. As a matter of fact, our analysis shows that these quantities are sufficient for guaranteeing convergence and optimal convergence rates of Tikhonov regularization as well as fast convergence of Newton's method for (9). It is an essential result of this paper to provide this link between the analysis in the sense of regularization methods and the requirements on the adaptive discretization. Our main contributions are the derivation of the necessary error estimates, their efficient evaluation as well as the fast computation of $i'(\beta)$ required for the Newton method. With all these ingredients we obtain an efficient algorithm for choosing the regularization parameter and for solving the underlying inverse problem.

The paper is organized as follows: in section 2 we introduce the concept of goal-oriented error estimators and apply it in the context of Tikhonov regularization to estimate the error in $j_\alpha(q)$, $i(\beta)$ and $i'(\beta)$. Moreover, we provide easy to evaluate expressions for $i'(\beta)$ and even the second derivative $i''(\beta)$. The next section deals with computation of the regularization parameter by Newton's method as well as the accuracy requirements for this purpose. Also, convergence and optimal convergence rates for the Tikhonov minimizer with so computed regularization parameter are proven. Section 4 shows the results of numerical experiments based on the proposed methodology that demonstrate its efficiency. In section 5 we summarize and give an outlook.

2. Goal-oriented error estimators

We start this section by formulating the optimization problem for Tikhonov regularization and corresponding optimality conditions explicitly taking into account the dependence on the regularization parameter. To make the discrepancy principle, equation (7), less nonlinear, we will use the reciprocal β of the parameter α, cf [11]. The optimization problem for a fixed $\beta \in \mathbb{R}_+$ is formulated as follows:

$$\text{Minimize} \quad J(\beta, q, u) = \|C(u) - g_\delta\|_G^2 + \frac{1}{\beta}\|q\|_Q^2, \qquad q \in Q, u \in V \tag{10}$$

subject to

$$A(q, u)(v) = f(v) \qquad \forall\, v \in V. \tag{11}$$

To formulate necessary optimality conditions we introduce the Lagrange functional:

$$\mathcal{L}: \mathbb{R} \times X \to \mathbb{R}, \qquad \mathcal{L}(\beta, q, u, z) = J(\beta, q, u) + f(z) - A(q, u)(z),$$

where we have used the notation $X = Q \times V \times V$ and $z \in V$ will denote the adjoint state. Using these notation the necessary optimality conditions can be formulated as follows:

$$\mathcal{L}'_x(\beta, x)(\mathrm{d}x) = 0 \qquad \forall\, \mathrm{d}x \in X, \tag{12}$$

where $x = (q, u, z)$.

Remark 1. The solution x depends on the noise level δ and the regularization parameter β. In some cases we will therefore write $x_\beta^\delta = \left(q_\beta^\delta, u_\beta^\delta, z_\beta^\delta\right)$ to stress this dependence. In this section, however, we suppress the notation for this dependence for better readability.

To discretize (10), (11) we use a Galerkin-type discretization based on finite-dimensional subspaces $Q_h \subset Q$, $V_h \subset V_h$. For standard construction of finite elements spaces we refer, e.g., to [9, 10]. The discrete counterpart of (10), (11) is given as

$$\text{Minimize} \quad J(\beta, q_h, u_h), \qquad q_h \in Q_h, u_h \in V_h \tag{13}$$

subject to

$$A(q_h, u_h)(v_h) = f(v_h) \qquad \forall\, v_h \in V_h. \tag{14}$$

The discrete optimality system for fixed $\beta \in \mathbb{R}$ has the form

$$\mathcal{L}'_x(\beta, x_h)(\mathrm{d}x_h) = 0 \qquad \forall\, \mathrm{d}x_h \in X_h, \tag{15}$$

where $x_h = (q_h, u_h, z_h)$ and $X_h = Q_h \times V_h \times V_h$.

2.1. Error estimator for the error in the cost functional

Following [5] we provide an error estimator for the discretization error with respect to the cost functional (for fixed $\beta \in \mathbb{R}$), i.e. for the error

$$J(\beta, q, u) - J(\beta, q_h, u_h),$$

where (q, u) is a solution of (10), (11) and (q_h, u_h) is the solution of the discretized problem (13), (14). There holds the following error representation [5]:

Proposition 1. *Let for fixed $\beta \in \mathbb{R}_+$, (q, u) be a solution of (10), (11) and (q_h, u_h) a solution of (13), (14). Then there holds*

$$J(\beta, q, u) - J(\beta, q_h, u_h) = \tfrac{1}{2}\mathcal{L}'_x(\beta, x_h)(x - \tilde{x}_h) + \mathcal{R},$$

where $\tilde{x}_h \in X_h$ is arbitrary and $\mathcal{R}$ is a third-order remainder term given by

$$\mathcal{R} = \frac{1}{2}\int_0^1 \mathcal{L}'''_{xxx}(\beta, x + se_x)(e_x, e_x, e_x) \cdot s \cdot (s - 1)\,\mathrm{d}s$$

with $e_x = x - x_h$.

In order to turn the above error representation into a computable error estimator, we proceed as follows. First we choose $\tilde{x}_h = i_h x$ with a suitable interpolation operator $i_h: X \to X_h$, then the interpolation error is approximated using an operator $\pi: X_h \to \tilde{X}_h$, with $\tilde{X}_h \neq X_h$, such that $x - \pi x_h$ has a better local asymptotical behavior than $x - i_h x$. Then we approximate

$$J(\beta, q, u) - J(\beta, q_h, u_h) \approx \eta^J = \tfrac{1}{2}\mathcal{L}'_x(\beta, x_h)(\pi_h x_h - x_h).$$

Such an operator can be constructed for example by the interpolation of the computed bilinear finite-element solution in the space of biquadratic finite elements on patches of cells. For this operator the improved approximation property relies on local smoothness of the solution and super-convergence properties of the approximation x_h. The use of such 'local higher-order approximations' is observed to work very successfully in the context of *a posteriori* error estimation, see, e.g., [5, 6].

2.2. Error estimation for the error in the squared residual

In [6, 7] an approach for error estimation with respect to a given quantity of interest is presented. To control the accuracy within the Newton algorithm for solving (9) we first choose the squared residual

$$I(u) = \|C(u) - g_\delta\|_G^2$$

as a quantity of interest. As introduced in (8), $i(\beta)$ denotes the value of $I(u)$ if (q, u) is the solution of (10), (11). On the discrete level we define the function $i_h: \mathbb{R}_+ \to R$ by

$$i_h(\beta) = I(u_h), \qquad \text{where} \qquad (q_h, u_h) \text{ is the solution of (13), (14)}\,. \tag{16}$$

Our aim is now to estimate the error with respect to *I*, i.e.,

$$I(u) - I(u_h) = i(\beta) - i_h(\beta).$$

To this end, we introduce an auxiliary Lagrange functional

$$\mathcal{M}: \mathbb{R} \times X^2 \to \mathbb{R}, \qquad \mathcal{M}(\beta, x, x_1) = I(u) + \mathcal{L}'_x(\beta, x)(x_1).$$

We abbreviate $y = (x, x_1)$ and $x_1 = (q_1, u_1, z_1)$. Then similarly to proposition 1 an error representation for the error in *I* can be formulated using continuous and discrete stationary points of $\mathcal{M}$, cf [6, 7].

Proposition 2. *Let* $y = (x, x_1) \in X^2$ *be stationary point of* $\mathcal{M}$*, i.e.,*

$$\mathcal{M}'_y(\beta, y)(\mathrm{d}y) = 0 \qquad \forall\, \mathrm{d}y \in X^2,$$

and let $y_h = (x_h, x_{1,h}) \in X_h^2$ *be a discrete stationary point, i.e.,*

$$\mathcal{M}'_y(\beta, y_h)(\mathrm{d}y_h) = 0 \qquad \forall\, \mathrm{d}y_h \in X_h^2,$$

then there holds

$$I(u) - I(u_h) = i(\beta) - i_h(\beta) = \tfrac{1}{2}\mathcal{M}'_y(\beta, y_h)(y - \tilde{y}_h) + \mathcal{R}_1,$$

where $\tilde{y}_h \in X_h^2$ *is arbitrary and the remainder term is given as*

$$\mathcal{R}_1 = \frac{1}{2}\int_0^1 \mathcal{M}'''_y(\beta, y + s e_y)(e_y, e_y, e_y) \cdot s \cdot (s-1)\,\mathrm{d}s$$

with $e_y = y - y_h$*.*

Again, in order to turn this error identity into a computable error estimator, we neglect the remainder term $\mathcal{R}_1$ and approximate the interpolation error using a suitable approximation of the interpolation error leading to

$$I(u) - I(u_h) = i(\beta) - i_h(\beta) \approx \eta^I = \tfrac{1}{2}\mathcal{M}'_y(\beta, y_h)(\pi_h y_h - y_h).$$

For a concrete form of this error estimator consisting of some residuals we refer to [6, 7].

A crucial question is of course how to compute the discrete stationary point y_h of $\mathcal{M}$ required for this error estimator. At the first glance it seems that the solution of the stationarity equation for $\mathcal{M}$ leads to coupled system of double size compared with the optimality system for (10), (11). However, solving this stationarity equation can be easily done using the already computed stationary point $x = (q, u, z)$ of $\mathcal{L}$ and exploiting existing structures. The following proposition shows that the computation of auxiliary variables $x_{1,h} = (q_{1,h}, u_{1,h}, z_{1,h})$ is equivalent to one step of an SQP method, which is often applied for solving (10), (11). The corresponding equation can also be solved by a Schur complement technique reducing the problem to the control space, cf, e.g., [19].

Proposition 3. *Let* $x = (q, u, z)$ *and* $x_h = (q_h, u_h, z_h)$ *be continuous and discrete stationary points of* $\mathcal{L}$*. Then* $y = (x, x_1)$ *is a stationary point of* $\mathcal{M}$ *if and only if*

$$\mathcal{L}''_{xx}(\beta, x)(\mathrm{d}x, x_1) = -I'(u)(\mathrm{d}u) \qquad \forall\, \mathrm{d}x = (\mathrm{d}q, \mathrm{d}u, \mathrm{d}z) \in X.$$

Moreover, $y_h = (x_h, x_{1,h})$ *is a discrete stationary point of* $\mathcal{M}$ *if and only if*

$$\mathcal{L}''_{xx}(\beta, x_h)(\mathrm{d}x_h, x_{1,h}) = -I'(u_h)(\mathrm{d}u_h) \qquad \forall\, \mathrm{d}x_h = (\mathrm{d}q_h, \mathrm{d}u_h, \mathrm{d}z_h) \in X_h.$$

Proof. There holds for $\mathrm{d}y = (\mathrm{d}x, \mathrm{d}x_1) \in X^2$

$$\mathcal{M}'_y(\beta, y)(\mathrm{d}y) = I'(u)(\mathrm{d}u) + \mathcal{L}''_{xx}(\beta, x)(\mathrm{d}x, x_1) + \mathcal{L}'_x(\beta, x)(\mathrm{d}x_1).$$

The last term vanishes due to the fact that x is a stationary point of $\mathcal{L}$. This completes the proof. □

2.3. Derivative of the squared residual with respect to the regularization parameter

The derivative of the squared residual with respect to the regularization parameter β, i.e. $i'(\beta)$, as well as its discrete counterpart $i_h'(\beta)$ is required for the Newton algorithm for solving (9). In the next proposition we show that once $y_h = (x_h, x_{1,h})$ is computed for the error estimation with respect to i, we can also use these quantities—with almost no additional effort—for evaluation of $i_h'(\beta)$. Similar results for evaluation of sensitivity derivatives of a quantity of interest with respect to some parameters can be found in [7, 14].

Proposition 4. *Let $y = (q, u, z, q_1, u_1, z_1)$ and $y_h = (q_h, u_h, z_h, q_{1,h}, u_{1,h}, z_{1,h})$ be continuous and discrete stationary points of $\mathcal{M}$. Then there holds*

$$i'(\beta) = -\frac{2}{\beta^2}(q, q_1)_Q \qquad \textit{and} \qquad i_h'(\beta) = -\frac{2}{\beta^2}(q_h, q_{1,h})_Q.$$

Proof. Due to the fact that $x = (q, u, z)$ and $x_h = (q_h, u_h, z_h)$ are continuous and discrete stationary points of $\mathcal{L}$ we have by definition of $\mathcal{M}$

$$i(\beta) = I(u) = \mathcal{M}(\beta, y) \qquad \text{and} \qquad i_h(\beta) = I(u_h) = \mathcal{M}(\beta, y_h).$$

Denoting here the dependence $y = y(\beta)$ explicitly we obtain

$$i'(\beta) = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathcal{M}(\beta, y(\beta)) = \mathcal{M}_\beta'(\beta, y(\beta)) + \mathcal{M}_y'(\beta, y(\beta))(y'(\beta)).$$

The last term vanishes due to the stationarity of y. The expression for $i'(\beta)$ is then calculated taking the partial derivative of $\mathcal{M}$ with respect to β leading to

$$i'(\beta) = \mathcal{M}_\beta'(\beta, y) = -\frac{2}{\beta^2}(q, q_1)_Q.$$

The corresponding result on the discrete level is obtained in the same way. □

Remark 2. The above proof uses the existence of directional derivatives of y with respect to β. Sufficient conditions for the existence of this sensitivity derivative can be found in [14].

2.4. Error estimator for the error in the derivative of the squared residual

For the control of the Newton method for solving (9) not only the value of $i(\beta)$ but also the value of its derivative $i'(\beta)$ has to be computed with certain accuracy. Therefore we will estimate the error between $i'(\beta)$ and $i_h'(\beta)$ for fixed value of β. To this end, we introduce a new error functional (quantity of interest) motivated by the expression for $i'(\beta)$ from proposition 4:

$$K\colon \mathbb{R} \times Q^2 \to \mathbb{R}, \qquad K(\beta, q, q_1) = -\frac{2}{\beta^2}(q, q_1)_Q.$$

The aim of this subsection is to derive an error estimator for the error

$$i'(\beta) - i_h'(\beta) = K(\beta, q, q_1) - K(\beta, q_h, q_{1,h}).$$

To this end, we introduce an additional Lagrange functional $\mathcal{N}\colon \mathbb{R} \times X^4 \to \mathbb{R}$ of the same structure as $\mathcal{M}$:

$$\mathcal{N}(\beta, x, x_1, x_2, x_3) = K(\beta, q, q_1) + \mathcal{M}_x'(\beta, x, x_1)(x_2) + \mathcal{M}_{x_1}'(\beta, x, x_1)(x_3),$$

where we have introduced additional variables $x_2 = (q_2, u_2, z_2)$ and $x_3 = (q_3, u_3, z_3)$. Additionally, we introduce a new abbreviation $\hat{y} = (x_2, x_3)$ and can rewrite the definition of $\mathcal{N}$ as

$$\mathcal{N}(\beta, y, \hat{y}) = K(\beta, q, q_1) + \mathcal{M}_y'(\beta, y)(\hat{y}).$$

With this notation we obtain an error representation for the error with respect to K using the same approach as in the previous section.

Proposition 5. *Let* $(y, \hat{y}) = (x, x_1, x_2, x_3) \in X^4$ *be a stationary point of* $\mathcal{N}$, *i.e.,*

$$\mathcal{N}'_y(\beta, y, \hat{y})(\mathrm{d}y) = 0\ \forall\,\mathrm{d}y \in X^2 \qquad \textit{and} \qquad \mathcal{N}'_{\hat{y}}(\beta, y, \hat{y})(\mathrm{d}\hat{y}) = 0\ \forall\,\mathrm{d}\hat{y} \in X^2,$$

and let $(y_h, \hat{y}_h) = (x_h, x_{1,h}, x_{2,h}, x_{3,h}) \in X_h^4$ *be a discrete stationary point of* $\mathcal{N}$, *i.e.,*

$$\mathcal{N}'_y(\beta, y_h, \hat{y}_h)(\mathrm{d}y_h) = 0\ \forall\,\mathrm{d}y_h \in X_h^2 \qquad \textit{and} \qquad \mathcal{N}'_{\hat{y}}(\beta, y_h, \hat{y}_h)(\mathrm{d}\hat{y}_h) = 0\ \forall\,\mathrm{d}\hat{y}_h \in X_h^2,$$

then there holds

$$K(\beta, q, q_1) - K(\beta, q_h, q_{1,h}) = \tfrac{1}{2}\mathcal{N}'_y(\beta, y_h, \hat{y}_h)(y - \tilde{y}_h) + \tfrac{1}{2}\mathcal{N}'_{\hat{y}}(\beta, y_h, \hat{y}_h)(\hat{y} - \bar{y}_h) + \mathcal{R}_2,$$

where $\tilde{y}_h, \bar{y}_h \in X_h^2$ *are arbitrary and* $\mathcal{R}_2$ *is a third-order remainder term.*

This error representation is again turned into a computable error estimate by

$$i'(\beta) - i'_h(\beta) \approx \eta^K = \tfrac{1}{2}\mathcal{N}'_y(\beta, y_h, \hat{y}_h)(\pi_h y - y_h) + \tfrac{1}{2}\mathcal{N}'_{\hat{y}}(\beta, y_h, \hat{y}_h)(\pi_h \hat{y} - \hat{y}_h).$$

As in the previous section the main question here is how to compute auxiliary variables $\hat{y}_h = (x_{2,h}, x_{3,h})$. The system to be solved for $\hat{y}_h$ has double size compared with the system for $x_{1,h}$. However, this system can be decoupled leading to two systems, and each of them can be solved using the existing structure. The numerical effort is equivalent to two steps of an SQP method for the original problem or to two steps of the Newton method reduced to the control space. The required decoupling is given in the following proposition.

Proposition 6. *Let* $y = (x, x_1)$ *and* $y_h = (x_h, x_{1,h})$ *be continuous and discrete stationary points of* $\mathcal{M}$, *cf proposition 3. Then* $(y, \hat{y}) = (x, x_1, x_2, x_3)$ *is a stationary point of* $\mathcal{N}$ *if and only if* $\hat{y} = (x_2, x_3) \in X^2$ *fulfils the following two equations:*

$$\begin{aligned}
&\mathcal{L}''_{xx}(\beta, x)(x_2, \mathrm{d}x_1) = -K'_{q_1}(\beta, q, q_1)(\mathrm{d}q_1) \qquad \forall\,\mathrm{d}x_1 \in X,\\
&\mathcal{L}''_{xx}(\beta, x)(x_3, \mathrm{d}x) = -K'_q(\beta, q, q_1)(\mathrm{d}q) - I''_{uu}(u)(u_2, \mathrm{d}u) - \mathcal{L}'''_{xxx}(\beta, x)(x_1, x_2, \mathrm{d}x)\\
&\qquad\qquad\qquad\qquad \forall\,\mathrm{d}x \in X.
\end{aligned}$$

Moreover, $(y_h, \hat{y}_h) = (x_h, x_{1,h}, x_{2,h}, x_{3,h})$ *is a discrete stationary point of* $\mathcal{N}$ *if and only if the discrete counterparts of the above equations are fulfilled for* $\hat{y}_h = (x_{2,h}, x_{3,h}) \in X_h^2$.

Proof. There holds by the stationarity of y with respect to $\mathcal{M}$:

$$\mathcal{N}'_{\hat{y}}(\beta, y, \hat{y})(\mathrm{d}\hat{y}) = \mathcal{M}'_y(\beta, y)(\mathrm{d}\hat{y}) = 0.$$

For the derivative with respect to y we obtain

$$\mathcal{N}'_y(\beta, y, \hat{y})(\mathrm{d}y) = K'_q(\beta, q, q_1)(\mathrm{d}q) + K'_{q_1}(\beta, q, q_1)(\mathrm{d}q_1) + \mathcal{M}''_{yy}(\beta, y)(\hat{y}, \mathrm{d}y).$$

The last term can be explicitly rewritten as

$$\begin{aligned}
\mathcal{M}''_{yy}(\beta, y)(\hat{y}, \mathrm{d}y) &= I''_{uu}(u)(u_2, \mathrm{d}u) + \mathcal{L}''_{xx}(\beta, x)(x_2, \mathrm{d}x_1)\\
&\quad + \mathcal{L}''_{xx}(\beta, x)(x_3, \mathrm{d}x) + \mathcal{L}'''_{xxx}(\beta, x)(x_1, x_2, \mathrm{d}x).
\end{aligned}$$

Separating the terms with $\mathrm{d}x = (\mathrm{d}q, \mathrm{d}u, \mathrm{d}z)$ and $\mathrm{d}x_1 = (\mathrm{d}q_1, \mathrm{d}u_1, \mathrm{d}z_1)$ we obtain the desired equations for x_2 and x_3. The argumentation for the discrete solutions is analog. □

2.5. Second derivative of the squared residual with respect to the regularization parameter

In section 2.3, we have shown that the quantities $x_{1,h} = (q_{1,h}, u_{1,h}, z_{1,h})$ computed for the error estimation of the error in $I(u)$ can be directly used for the evaluation of $i_h'(\beta)$. Similarly, once the quantities $x_{2,h} = (q_{2,h}, u_{2,h}, z_{2,h})$ and $x_{3,h} = (q_{3,h}, u_{3,h}, z_{3,h})$ are computed for the estimation of the error in $K(\beta, q, q_1)$, one can evaluate the second derivative $i_h''(\beta)$ almost without extra numerical effort. Although this second derivative is not required in Newton's method, it can be useful for other purposes, e.g., one can easily check the correct computation of $x_{2,h}$ and $x_{3,h}$ by comparing $i_h''(\beta)$ with difference quotients. In the next proposition we provide expressions for $i''(\beta)$ and $i_h''(\beta)$.

Proposition 7. *Let $(y, \hat{y}) = (x, x_1, x_2, x_3)$ be a stationary point of $\mathcal{N}$ as in proposition 6. Then the following representation for $i''(\beta)$ holds:*

$$i''(\beta) = \frac{4}{\beta^3}(q, q_1)_Q - \frac{2}{\beta^2}(q_2, q_1)_Q - \frac{2}{\beta^2}(q, q_3)_Q.$$

A similar representation holds on the discrete level for $i_h''(\beta)$.

Proof. Due to the fact that $y = (x, x_1)$ is a stationary point of $\mathcal{M}$ we have

$$i'(\beta) = K(\beta, q, q_1) = \mathcal{N}(\beta, y, \hat{y}).$$

We differentiate totally with respect to β, use the fact that $(y, \hat{y})$ is a stationary point of $\mathcal{N}$ and obtain

$$i''(\beta) = \mathcal{N}_\beta'(\beta, y, \hat{y}).$$

Calculating the partial derivative of $\mathcal{N}$ with respect to β completes the proof. □

3. Determination of the Tikhonov regularization parameter

In this section we will restrict ourselves to the linear case (4), where the minimizer of the Tikhonov functional is given by

$$q_\beta^\delta = \left(T^*T + \frac{1}{\beta} I\right)^{-1} T^* g^\delta. \tag{17}$$

This is the case if the solution operator S of the forward equation and the observation operator C are both linear, i.e. $T = C \circ S$. Throughout this section we assume that a solution to (4) exists and denote by $q^\dagger$ the best approximate solution (i.e., the solution with minimal norm). Due to the linearity of T this solution is unique.

Our aim is to determine the regularization parameter $\beta = \beta(g^\delta, \delta)$ in such a way that the corresponding recovered parameter converges to $q^\dagger$ as δ tends to zero.

3.1. An inexact Newton method

To compute the regularization parameter we would like to apply Newton's method to the one-dimensional equation (9). However, neither $i(\beta)$ nor $i'(\beta)$ required for the Newton method are available. Rather approximations $i_h(\beta)$ and $i_h'(\beta)$ can be evaluated for each fixed discretization with discrete spaces V_h, Q_h, see the previous section. Therefore, we apply an inexact Newton algorithm, where we control and change the accuracy of discretizations in such a way that the algorithm converges globally as well as quadratically to the solution β_* of (9). Moreover, we will derive a stopping criterion in such a way that the iterate β^{k_*} fulfilling this criterion leads to convergence of $q_{\beta^{k_*}}^\delta$ and $q_{h,\beta^{k_*}}^\delta$ to $q^\dagger$ as δ tends to zero. In the following we sketch this multilevel inexact Newton algorithm, where the detailed form of the accuracy requirements and the stopping criterion is given in theorem 1.

Multilevel Inexact Newton Method

1. Choose initial guess $\beta^0 > 0$, initial discretization Q_{h_0}, V_{h_0}, set $k = 0$
2. Solve optimization problem (13)–(14), compute $x_{h_k} = (q_{h_k}, u_{h_k}, z_{h_k})$
3. Evaluate $i_{h_k}(\beta^k)$
4. Compute $x_{1,h_k} = (q_{1,h_k}, u_{1,h_k}, z_{1,h_k})$, see proposition 3
5. Evaluate $i'_{h_k}(\beta^k)$, see proposition 4
6. Evaluate error estimator η^I, see proposition 2
7. Compute $x_{2,h_k} = (q_{2,h_k}, u_{2,h_k}, z_{2,h_k})$ and $x_{3,h_k} = (q_{3,h_k}, u_{3,h_k}, z_{3,h_k})$, see proposition 6
8. Evaluate error estimator η^K, see proposition 5
9. If the accuracy requirements for η^I, η^K are fulfilled, set
$$\beta^{k+1} = \beta_k - \frac{i_{h_k}(\beta^k) - \tau^2\delta^2}{i'_{h_k}(\beta^k)}$$
10. else: refine discretization $h_k \to h_{k+1}$ using local information from η^I, η^K
11. if stopping criterion is fulfilled: break
12. else: Set $k = k + 1$ and go to 2.

For Newton's method with exact evaluation of $i(\beta)$ and $i'(\beta)$, one can show global convergence, see [11], provided g^δ is not in the null space of T^*, i.e. $g^\delta \notin \mathcal{N}(T^*)$. This fact relies on the following lemma.

Lemma 1. *The function $i:\mathbb{R}_+ \to \mathbb{R}$ defined by (8) satisfies for all $\beta \in \mathbb{R}_+$ the following inequalities:*

$$-\frac{2\|g^\delta\|_G}{\beta} \leqslant i'(\beta) \leqslant 0, \qquad \frac{6\|g^\delta\|_G}{\beta^2} \geqslant i''(\beta) \geqslant 0, \qquad i'''(\beta) \leqslant 0.$$

If additionally $g^\delta \notin \mathcal{N}(T^*)$, then strict inequalities hold.

Proof. It is readily checked that

$$\begin{aligned}
i(\beta) &= \|(\beta T T^* + I)^{-1} g^\delta\|_G^2,\\
i'(\beta) &= -2\|(\beta T^* T + I)^{-3/2} T^* g^\delta\|_G^2,\\
i''(\beta) &= 6\|(\beta T T^* + I)^{-2} T T^* g^\delta\|_G^2,\\
i'''(\beta) &= -24\|(\beta T^* T + I)^{-5/2} T^* T T^* g^\delta\|_G^2.
\end{aligned}$$

Using the fact that $\|(TT^*+\beta^{-1}I)^{-1}TT^*\|_G \leqslant 1$, $\|(TT^*+\beta^{-1}I)^{-1}TT^*\|_G \leqslant \beta$, for all $\beta > 0$ (cf, e.g., [11]), we complete the proof. □

In the following theorem we derive the accuracy requirements for the inexact Newton algorithm presented above. For setting up the stopping criterion we exploit the fact that we do not need to reach $\tau^2\delta^2$ in (9) exactly but only up to some accuracy $\tilde{\tau}^2\delta^2$ for some $\tilde{\tau} < \tau$, see subsection 3.2 for details.

Theorem 1. *Let $i \in C^3(\mathbb{R}^+)$, $i'(\beta) < 0$, $i''(\beta) > 0$, $i'''(\beta) \leqslant 0$ for all $\beta > 0$, and β_* solve (9). Let moreover a sequence $\{\beta^k\}$ be defined by*

$$\beta^{k+1} = \beta^k - \frac{i_h^k - \tau^2\delta^2}{i_h'^k}, \qquad 0 < \beta^0 \leqslant \beta_*, \tag{18}$$

with i_h^k, i'^k_h *satisfying*

$$\left|i(\beta^k) - i_h^k\right| \leqslant \min\left\{c_1\left|i_h^k - \tau^2\delta^2\right|, \frac{C_2\|g^\delta\|_G^2}{\left|i'^k_h\right|^2(\beta^k)^2}\left|i_h^k - \tau^2\delta^2\right|^2\right\} \tag{19}$$

$$\left|i'(\beta^k) - i'^k_h\right| \leqslant \min\left\{C_3\left|i'^k_h\right|, \frac{C_2\|g^\delta\|_G^2}{\left|i'^k_h\right|(\beta^k)^2}\left|i_h^k - \tau^2\delta^2\right|\right\} \tag{20}$$

for some constants $c_1, C_2, C_3 > 0, c_1 < 1$ *independent of* k*. Let moreover* k_* *be given as*

$$k_* = \min\left\{k \in \mathbb{N}\middle| i_h^k - (\tau^2 + \tilde{\tau}^2/2)\delta^2 \leqslant 0\right\} \tag{21}$$

and the following conditions be fulfilled:

$$i'^k_h < 0 \qquad \textit{for all} \qquad k \leqslant k_* - 1, \tag{22}$$

$$\left|i(\beta^{k_*-1}) - i_h^{k_*-1}\right| + \left|\frac{i_h^{k_*-1} - \tau^2\delta^2}{i'^{k_*-1}_h}\right| \left|i'(\beta^{k_*-1}) - i'^{k_*-1}_h\right| \leqslant \tilde{\tau}^2\delta^2, \tag{23}$$

$$\left|i(\beta^{k_*}) - i_h^{k_*}\right| \leqslant \frac{\tilde{\tau}^2}{2}\delta^2 \tag{24}$$

for some $\tilde{\tau} < \tau$*.*

Then k_* *is finite and there holds*

$$\beta^{k+1} \geqslant \beta^k \wedge \beta^k \leqslant \beta_* \qquad \forall\, k \leqslant k_* - 1, \tag{25}$$

β^k *satisfies the local quadratic convergence estimate*

$$|\beta^{k+1} - \beta_*| \leqslant \frac{C\|g^\delta\|_G^2}{i'(\beta^k)(\beta^k)^2}(\beta^k - \beta_*)^2 + \mathcal{O}((\beta^k - \beta_*)^4) \qquad \forall\, k \leqslant k_* - 1 \tag{26}$$

for some $C > 0$ *independent of* β^k *and* k*, and*

$$(\tau^2 - \tilde{\tau}^2)\delta^2 \leqslant i(\beta^{k_*}) \leqslant (\tau^2 + \tilde{\tau}^2)\delta^2. \tag{27}$$

Proof. To show monotonicity up to k_*, note that the definition of k_* and (19) imply that for all $k \leqslant k_* - 1$

$$i(\beta^k) - \tau^2\delta^2 \geqslant i_h^k - \tau^2\delta^2 - \left|i(\beta^k) - i_h^k\right| \geqslant (1 - c_1)\left(i_h^k - \tau^2\delta^2\right) > 0,$$

hence, by the strict monotonicity of $i(\beta)$, $\beta^k \leqslant \beta_*$. Moreover,

$$\beta^{k+1} - \beta^k = \frac{i_h^k - \tau^2\delta^2}{-i'^k_h} \geqslant 0$$

by (22), hence we have shown (25).

By Taylor expansion we obtain the following error decomposition:

$$\beta^{k+1} - \beta_* = \frac{1}{i'(\beta^k)}\left(\frac{1}{2}i''(\bar{\beta}^k)(\beta^k - \beta_*)^2 + i(\beta^k) - i_h^k - \frac{i_h^k - \tau^2\delta^2}{i'^k_h}\left(i'(\beta^k) - i'^k_h\right)\right), \tag{28}$$

where $\bar{\beta}^k \in [\beta^k, \beta_*]$. Hence, by lemma 1, relation (25), and the fact that $i''(\beta)$ is monotonically decreasing

$$i''(\bar{\beta}^k) \leqslant i''(\beta^k) \leqslant \frac{6\|g^\delta\|_G^2}{(\beta^k)^2}, \tag{29}$$

the above error decomposition (28) implies (26) provided

$$\left| i(\beta^k) - i_h^k - \frac{i_h^k - \tau^2\delta^2}{i'^k_h}\left(i'(\beta^k) - i'^k_h\right) \right| \leqslant \frac{\tilde{C}\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2 + \mathcal{O}((\beta^k - \beta_*)^4) \tag{30}$$

for some constant $\tilde{C}$ can be guaranteed. The latter can be concluded from (19), (20), using the fact that

$$i_h^k - \tau^2\delta^2 = i_h^k - i(\beta^k) + i'^k_h(\beta^k - \beta_*) + (i'(\beta^k) - i'^k_h)(\beta^k - \beta_*) - \tfrac{1}{2}i''(\bar{\beta}^k)(\beta^k - \beta_*)^2.$$

and therewith

$$r \leqslant e + \left|i'^k_h\right|(\beta^k - \beta_*) + e'(\beta^k - \beta_*) + \frac{3\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2 \tag{31}$$

for

$$r = \left|i_h^k - \tau^2\delta^2\right|, \qquad e = \left|i(\beta^k) - i_h^k\right|, \qquad \text{and} \qquad e' = \left|i'(\beta^k) - i'^k_h\right|.$$

Namely, with (19), (20), the estimate (31) implies

$$\begin{aligned}(1 - c_1)e &\leqslant c_1\left(\left|i'^k_h\right|(\beta^k - \beta_*) + e'(\beta^k - \beta_*) + \frac{3\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2\right)\\ &\leqslant c_1(1 + C_3)\left|i'^k_h\right|(\beta^k - \beta_*) + c_1\frac{3\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2.\end{aligned}$$

Inserting this and (20) into (31) yields

$$r \leqslant \frac{1}{1 - c_1}\left((1 + C_3)\left|i'^k_h\right|(\beta^k - \beta_*) + \frac{3\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2\right)$$

which by (19), (20) implies

$$\begin{aligned}\max\left\{e, \frac{r}{\left|i'^k_h\right|}e'\right\} &\leqslant \frac{C_2\|g^\delta\|_G^2}{\left|i'^k_h\right|^2(\beta^k)^2}r^2\\ &\leqslant \frac{2C_2(1 + C_3)^2}{(1 - c_1)^2}\frac{\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2 + \frac{2C_2}{(1 - c_1)^2}\frac{9\|g^\delta\|_G^6}{\left|i'^k_h\right|^2(\beta^k)^6}(\beta^k - \beta_*)^4\\ &\leqslant \frac{2C_2(1 + C_3)^2}{(1 - c_1)^2}\left(\frac{\|g^\delta\|_G^2}{(\beta^k)^2}(\beta^k - \beta_*)^2 + \frac{9\|g^\delta\|_G^6}{|i'(\beta^k)|^2(\beta^k)^6}(\beta^k - \beta_*)^4\right),\end{aligned}$$

where we have used $|i'(\beta^k)| \leqslant (1 + C_3)\left|i'^k_h\right|$, and therewith (30).

Existence of $k_* < \infty$ now follows from convergence of β_k to a solution of $i(\beta) = \tau^2\delta^2$ if (19), (20), (22) hold for all $k \in \mathbb{N}$.

To show the lower estimate in (27), we use (23), which implies

$$\begin{aligned}i(\beta^{k_*}) - \tau^2\delta^2 &= i\left(\beta^{k_*-1} - \frac{i_h^{k_*-1} - \tau^2\delta^2}{i'^{k_*-1}_h}\right) - \tau^2\delta^2\\ &= i(\beta^{k_*-1}) - \tau^2\delta^2 + \underbrace{i'(\tilde{\beta}^{k_*-1})}_{\geqslant i'(\beta^{k_*-1})}\underbrace{\frac{i_h^{k_*-1} - \tau^2\delta^2}{-i'^{k_*-1}_h}}_{\geqslant 0}\\ &\geqslant i(\beta^{k_*-1}) - i_h^{k_*-1} + \frac{i_h^{k_*-1} - \tau^2\delta^2}{-i'^{k_*-1}_h}\left(i'(\beta^{k_*-1}) - i'^{k_*-1}_h\right)\\ &\geqslant -\tilde{\tau}^2\delta^2\end{aligned}$$

where $\tilde{\beta}^{k_*-1} \in [\beta_{k_*-1}, \beta_{k_*}]$. The upper estimate in (27) directly follows from the definition of k_* and (24). □

Remark 3. Setting $i_h^k = i_h(\beta^k)$ and $i'^k_h = i'_h(\beta^k)$ in theorem 1, we obtain the accuracy requirements and the stopping criterion for the inexact Newton algorithm described above. The requirement (22) is fulfilled due to the discrete analog of lemma 1. Condition $\beta^0 \leqslant \beta_*$ on the starting value is natural and easy to satisfy: it means that we start with a sufficiently strongly regularized problem such that the residual norm is still larger than $\tau\delta$. (To see the latter note that i is strictly monotone and hence $\beta^0 \leqslant \beta_*$ equivalent to $i(\beta^0) \geqslant \tau^2\delta^2$.) Since no closeness assumption to β_* is made on β^0, theorem 1 describes a globally convergent iteration.

Remark 4. A similar strategy for choosing accuracy requirements as in theorem 1 can be obtained for a secant method of the following type:

$$\beta^{k+1} = \beta^k - \frac{i_h^k - \tau^2\delta^2}{\frac{i_h^k - i_h^{k-1}}{\beta^k - \beta^{k-1}}}.$$

3.2. *Convergence of the discrete and the continuous Tikhonov minimizer*

The stopping rule from theorem 1 for the multilevel inexact Newton method described in the previous subsection leads to an approximation $\hat{\beta} = \beta^{k_*}$ of the solution β_* of (9), such that the condition (27) is fulfilled. Following the lines of theorem 4.17 in [11], we will show that this condition allows for convergence and optimal convergence rates for the corresponding Tikhonov minimizer $q_{\hat{\beta}}^\delta$ to $q^\dagger$ as δ tends to zero. This result is given in the following proposition.

Proposition 8. *Let $q^\dagger$ be the minimal norm solution of (4) and $u^\dagger$ the corresponding state. Let moreover $\left(q_{\hat{\beta}}^\delta, u_{\hat{\beta}}^\delta\right)$ be the minimizer of the Tikhonov functional with regularization parameter $\hat{\beta} = \hat{\beta}(\delta, g^\delta)$ chosen in such a way that (27) is fulfilled with $\tau > 1, 0 < \tilde{\tau}^2 < \tau^2 - 1$. Then $q_{\hat{\beta}}^\delta$ converges to $q^\dagger$ in Q as δ tends to zero.*

Proof. The lower inequality of (27) and the optimality of $\left(q_{\hat{\beta}}^\delta, u_{\hat{\beta}}^\delta\right)$ implies

$$(\tau^2 - \tilde{\tau}^2)\delta^2 + \hat{\beta}^{-1}\left\|q_{\hat{\beta}}^\delta\right\|_Q^2 \leqslant J\left(\hat{\beta}, q_{\hat{\beta}}^\delta, u_{\hat{\beta}}^\delta\right) \leqslant J(\hat{\beta}, q^\dagger, u^\dagger) \leqslant \delta^2 + \hat{\beta}^{-1}\|q^\dagger\|_Q^2.$$

Hence, by the conditions for $\tau, \tilde{\tau}$ we have

$$\left\|q_{\hat{\beta}}^\delta\right\|_Q^2 \leqslant \hat{\beta}(1 - \tau^2 + \tilde{\tau}^2) + \|q^\dagger\|_Q^2 \leqslant \|q^\dagger\|_Q^2. \tag{32}$$

Considering $q(\delta) = q_{\hat{\beta}}^\delta$ as a sequence in δ, the boundedness (32) implies existence of a weakly convergent subsequence $q(\delta_k)$. For an arbitrary weakly convergent subsequence $q(\delta_k)$ with weak limit q_* it follows from the upper bound in (27) and weak continuity of the bounded linear operator T that q_* solves $Tq_* = g$ and $\|q_*\| \leqslant \|q^\dagger\|$. Since $q^\dagger$ has minimal norm among all solutions to $Tq = g$ and is uniquely determined by this property, we can conclude $q_* = q^\dagger$, which by a subsequence–subsequence argument implies weak convergence of the whole sequence $q(\delta) = q_{\hat{\beta}}^\delta$ to $q^\dagger$ as $\delta \to 0$. Strong convergence follows as usual by the following argument:

$$\left\|q_{\hat{\beta}}^\delta - q^\dagger\right\|_Q^2 = \left\|q_{\hat{\beta}}^\delta\right\|_Q^2 + \|q^\dagger\|_Q^2 - 2\left(q_{\hat{\beta}}^\delta, q^\dagger\right)_Q \leqslant 2\|q^\dagger\|_Q^2 - 2\left(q_{\hat{\beta}}^\delta, q^\dagger\right)_Q \to 0,$$

where we have used (32) and the weak convergence of $q_{\hat{\beta}}^\delta$. □

The proposed choice of $\hat{\beta}$ leads not only to convergence to $q^\dagger$ but also to optimal convergence rates provided that a corresponding source condition is fulfilled.

Proposition 9. *Let the conditions of proposition 8 be fulfilled. Let moreover, the following source condition hold:*

$$q^\dagger \in \mathcal{R}(f(T^*T)), \tag{33}$$

with f such that f^2 is strictly monotonically increasing on $(0, \|T\|^2]$, ϕ defined by $\phi^{-1}(\lambda) = f^2(\lambda)$ is convex and ψ defined by $\psi(\lambda) = f(\lambda)\sqrt{\lambda}$ is strictly monotonically increasing on $(0, \|T\|^2]$. Then the following convergence rates with some $C > 0$ independent of δ are obtained:

$$\|q_{\hat{\beta}}^\delta - q^\dagger\|_Q = \mathcal{O}\left(\frac{C\delta}{\sqrt{\psi^{-1}(C\delta)}}\right).$$

Note that using (27) we could directly conclude the assertion from [21], see also [20] for the more general situation of regularization in Hilbert scales. Nevertheless, we provide a proof here which allows for an easy generalization to the nonlinear case, see [13]; for a convergence rate proof in the nonlinear case under different conditions on the forward operator than those used in [13] we refer to [23].

Proof. From the source condition we obtain the existence of $v \in Q$ such that

$$q^\dagger = f(T^*T)v.$$

Using the notation $e = q_{\hat{\beta}}^\delta - q^\dagger$ and Jensen's inequality, that implies

$$\|f(T^*T)e\|_Q \leqslant \|e\|_Q f\left(\frac{\|Te\|_G^2}{\|e\|_Q^2}\right)$$

(cf, e.g., [16]) we obtain by (32)

$$\begin{aligned}\|e\|_Q^2 &\leqslant 2\|q^\dagger\|_Q^2 - 2\left(q_{\alpha_*}^\delta, q^\dagger\right)_Q = -2(e, f(T^*T)v)_Q \\ &= -2(v, f(T^*T)e)_Q \leqslant 2\|v\|_Q\|e\|_Q f\left(\frac{\|Te\|_G^2}{\|e\|_Q^2}\right).\end{aligned}$$

This implies

$$\frac{\sqrt{\tau^2 - \tilde{\tau}^2} - 1}{2\|v\|_Q}\delta \leqslant \frac{\|Te\|_G}{2\|v\|_Q} \leqslant \psi\left(\frac{\|Te\|_G^2}{\|e\|_Q^2}\right) \leqslant \psi\left(\frac{(\sqrt{\tau^2 + \tilde{\tau}^2} + 1)^2\delta^2}{\|e\|_Q^2}\right),$$

hence

$$\psi^{-1}\left(\frac{\sqrt{\tau^2 - \tilde{\tau}^2} - 1}{2\|v\|_Q}\delta\right) \leqslant \frac{(\sqrt{\tau^2 + \tilde{\tau}^2} + 1)^2\delta^2}{\|e\|_Q^2},$$

from which the proposed assertion follows with $C := \frac{\sqrt{\tau^2 - \tilde{\tau}^2} - 1}{2\|v\|_Q}$. □

The following corollary provides convergence rates for typical source conditions:

Corollary 1. *Let the assumptions of proposition 9 be satisfied and let the source condition (33) hold with*

(*a*) $f(\lambda) = \lambda^\nu$ *for some* $\nu \in \left(0, \frac{1}{2}\right]$

or

(*b*) $f(\lambda) = (-\ln\lambda)^{-p}$ *for some* $p > 0$, *where in case (b) we additionally assume (without loss of generality, by appropriate scaling) that* $\|T\|^2 \leqslant \frac{1}{e}$. *Then optimal convergence rates*

$$\|q^\delta_{\hat\beta} - q^\dagger\|_Q = \mathcal{O}\big(\delta^{\frac{2\nu}{2\nu+1}}\big) \text{ in case } (a), \qquad \|q^\delta_{\hat\beta} - q^\dagger\|_Q = \mathcal{O}((-\ln\delta)^{-p}) \text{ in case } (b)$$

are obtained.

In proposition 8, proposition 9 and in the above corollary, the convergence behavior of $q^\delta_{\hat\beta}$ is studied for $\delta \to 0$. Although $\hat\beta = \beta^{k_*}$ is directly computed by the inexact Newton algorithm presented in the previous section, the solution of the continuous Tikhonov problem $q^\delta_{\hat\beta}$ is not available. Rather the solution of the discrete Tikhonov problem $q^\delta_{h,\hat\beta}$ can be computed. In the next proposition we give an additional accuracy criterion which allows for convergence of $q^\delta_{h,\hat\beta}$ to $q^\dagger$ as $\delta \to 0$.

Proposition 10. *Let the conditions of proposition 8 and (24) be fulfilled. Let moreover for the discretization error with respect to the cost functional hold:*

$$\big|J\big(\hat\beta, q^\delta_{\hat\beta}, u^\delta_{\hat\beta}\big) - J\big(\hat\beta, q^\delta_{h,\hat\beta}, u^\delta_{h,\hat\beta}\big)\big| \leqslant \sigma^2\delta^2, \tag{34}$$

where $0 \leqslant \sigma^2 \leqslant \tau^2 - \frac{3}{2}\tilde\tau^2 - 1$. *Then* $q^\delta_{h,\hat\beta}$ *converges to* $q^\dagger$ *in* Q *as* $\delta \to 0$.

Proof. Similar as in the proof of proposition 8 we have, by (24) and (27),

$$i_h(\hat\beta) \geqslant i(\hat\beta) - \frac{\tilde\tau^2}{2}\delta^2 \geqslant \left(\tau^2 - \frac{3\tilde\tau^2}{2}\right)\delta^2.$$

Therefore

$$\left(\tau^2 - \frac{3\tilde\tau^2}{2}\right)\delta^2 + \hat\beta^{-1}\|q^\delta_{h,\hat\beta}\|^2_Q \leqslant J\big(\hat\beta, q^\delta_{h,\hat\beta}, u^\delta_{h,\hat\beta}\big) \leqslant J\big(\hat\beta, q^\delta_{\hat\beta}, u^\delta_{\hat\beta}\big) + \sigma^2\delta^2$$
$$\leqslant J(\hat\beta, q^\dagger, u^\dagger) + \sigma^2\delta^2 = (1+\sigma^2)\delta^2 + \hat\beta^{-1}\|q^\dagger\|^2_Q.$$

Hence,

$$\|q^\delta_{h,\hat\beta}\|^2_Q \leqslant \hat\beta\left(1 + \sigma^2 - \tau^2 + \frac{3\tilde\tau^2}{2}\right)\delta^2 + \|q^\dagger\|^2_Q \leqslant \|q^\dagger\|^2_Q.$$

The rest of the proof follows the lines of the proof of proposition 8. □

Once the regularization parameter $\hat\beta = \beta^{k_*}$ is computed, we check if the condition (34) is fulfilled on the current mesh. If it is not the case, the mesh is refined using the information from the corresponding error estimator η^J, see proposition 1. Then on this new mesh a Tikhonov problem for the already computed regularization parameter $\hat\beta$ is solved. This procedure is repeated until the condition (34) is satisfied. Once, it is fulfilled, the accuracy of $q^\delta_{h,\hat\beta}$ is satisfactory.

4. Numerical results

For illustrating the performance of the proposed method, we apply it to the problem of identifying the source term q in the elliptic boundary value problem

$$\begin{aligned} -\Delta u &= q \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

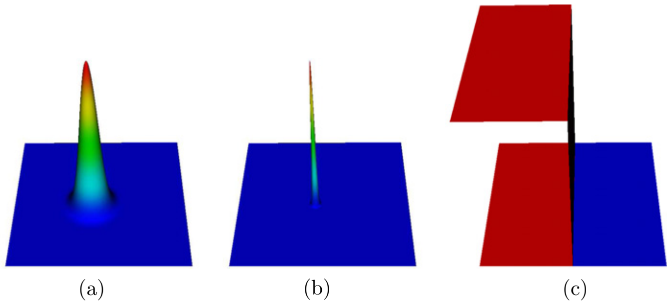


Figure 1. Exact q for examples (a) Gaussian distribution, $s = 0.05$, (b) Gaussian distribution, $s = 0.01$ and (c) step function.

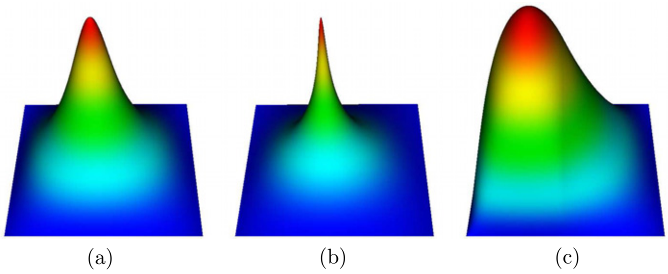


Figure 2. Exact u for examples (a) Gaussian distribution, $s = 0.05$, (b) Gaussian distribution, $s = 0.01$ and (c) step function.

on the unit square $\Omega = (0, 1)^2 \subset \mathbb{R}^2$. We consider three configurations with different exact sources $q^\dagger$: two types of Gaussian distributions and a step function

$$(a), (b) \quad q^\dagger(x, y) = \frac{1}{2\pi s^2} \exp\left(-\frac{\left(x - \frac{5}{11}\right)^2 + \left(y - \frac{5}{11}\right)^2}{2s^2}\right)$$

$$(c) \quad q^\dagger(x, y) = \begin{cases} 1 & x \leqslant \frac{1}{2} \\ 0 & x > \frac{1}{2} \end{cases}$$

where we set

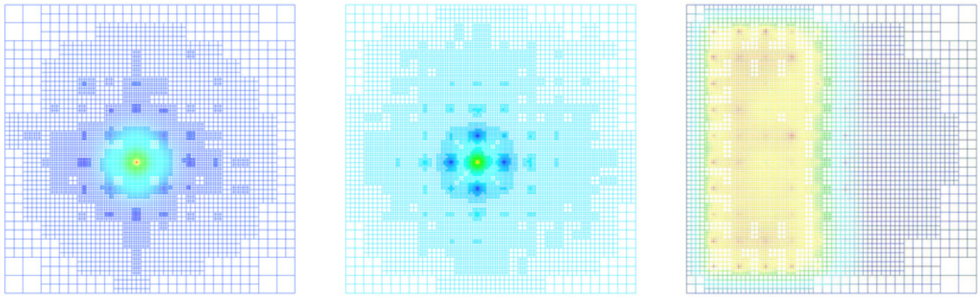
$$(a) \quad s = 0.05 \qquad \text{and} \qquad (b) \quad s = 0.01.$$

Figures 1 and 2 show the exact parameters and states (computed on a very fine grid with more than 10^6 nodes to avoid an inverse crime), respectively.

The measurements were defined by point functionals in $n_m = 100$ uniformly distributed points $\{\xi_i\} \subset \Omega$ and perturbed by uniformly distributed random noise at different percentages. The functional spaces are chosen as $Q = L^2(\Omega)$, $V = H_0^1(\Omega) \cap \mathcal{C}(\bar{\Omega})$ and $G = \mathbb{R}^{n_m}$. The observation operator $C: V \to G$ is defined for our examples by

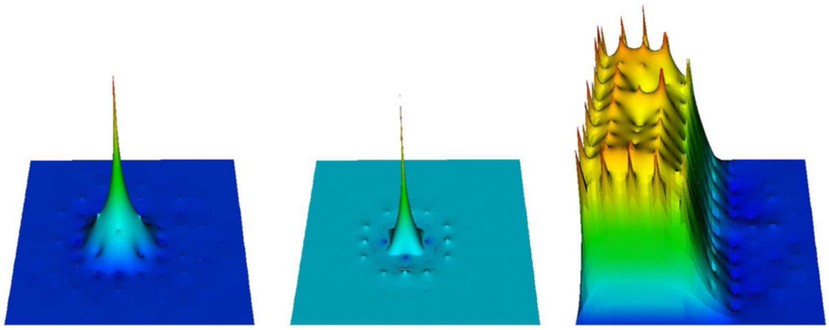
$$(C(v))_i = v(\xi_i), \qquad i = 1, \ldots, n_m.$$

The algorithm consists of two main steps. First the computation of $\hat{\beta} = \beta^{k_*}$ within the inexact Newton algorithm described in section 3 according to theorem 1. Second, taking into account



(a) (b) (c)

Figure 3. Adaptively refined meshes for examples (a) Gaussian distribution, $s = 0.05$, (b) Gaussian distribution, $s = 0.01$ and (c) step function, with noise level $\delta = 1\%$.



(a) (b) (c)

Figure 4. Reconstructed q for examples (a) Gaussian distribution, $s = 0.05$, (b) Gaussian distribution, $s = 0.01$ and (c) step function, with noise level $\delta = 1\%$.

proposition 10, the accordant $q^\delta_{h,\hat\beta}$ is computed. Further parameters used in the computations were

$$\tau = 1.8, \qquad \tilde\tau = 1, \qquad \sigma^2 = 0.24, \qquad \beta^0 = 10.$$

All computations were carried out on a AMD Athlon64 3500+ (2 GB central memory) using the package RoDoBo [4] for treating optimization problems governed by partial differential equations and the finite-element library Gascoigne [2]. For visualization we used the visualization tool VisuSimple [3].

In tables 1–3 we show a comparison of the proposed adaptive refinement strategy versus a uniform refinement for our three example configurations. For the uniform refinement we also used the error estimators derived in section 2 to control the error according to the requirements given in theorem 1 and proposition 10. The noise level for these computations was chosen $\delta = 1\%$. The corresponding reconstructed sources $q^\delta_{h,\hat\beta}$ are shown in figure 4 and the finest adaptive meshes in figure 3. Neither of the examples tested here required further refinement to guarantee condition (34) in proposition 10. In these examples we observe considerable saving in degrees of freedom required for computation of the regularization parameter, if the proposed adaptive algorithm is used. Note that the corresponding algorithm on uniformly refined meshes can only be realized if the proposed error estimators are used.

Table 1. Adaptive versus uniform refinement for example (a) with noise level $\delta = 1\%$.

k	No. of nodes	β^k	k	No. of nodes	β^k
(a) Adaptive refinement			(b) Uniform refinement		
0	25	10	0	25	10
0	81	10	0	81	10
0	189	10	0	289	10
0	561	10	0	1089	10
0	1485	10	0	4225	10
0	4485	10	0	16641	10
1	4485	22.77	1	16641	22.77
2	4485	57.65	2	16641	57.63
3	4485	143.47	3	16641	143.37
4	4485	324.20	4	16641	323.96
5	4485	684.73	5	16641	684.19
6	4485	1352.15	6	16641	1351.17
7	4485	2456.35	7	16641	2454.93
8	4485	4007.33	8	16641	4005.77
9	4485	5670.76	9	16641	5669.44
10	4485	6761.25	10	16641	6760.35
11	4485	7047.47	11	16641	7046.77
11	9261	7047.47	11	66049	7046.77

Table 2. Adaptive versus uniform refinement for example (b) with noise level $\delta = 1\%$.

k	No. of nodes	β^k	k	No. of nodes	β^k
(a) Adaptive refinement			(b) Uniform refinement		
0	25	10	0	25	10
0	25	10	0	81	10
0	81	10	0	289	10
0	165	10	0	1089	10
0	525	10	0	4225	10
0	1373	10	0	16641	10
0	4041	10	1	16641	31.29
1	4041	31.29	2	16641	125.41
1	12413	31.29	3	16641	477.19
2	12413	125.40	3	66049	477.19
3	12413	476.80	4	66049	1592.59
4	12413	1591.70	5	66049	4084.30
5	12413	4083.29	6	66049	8365.45
6	12413	8365.15	7	66049	14888.70
7	12413	14890.09	8	66049	24046.19
8	12413	24050.74	9	66049	35451.46
9	12413	35461.69	10	66049	46650.69
10	12413	46670.03	11	66049	53403.93
11	12413	53433.11	12	66049	54983.99
12	12413	55017.36	12	263169	54983.99
12	18021	55017.36			

The convergence behavior of $q^{\delta}_{h,\hat{\beta}}$ for $\delta \to 0$ for configurations (a) and (b) is demonstrated in table 4. Although in both examples the solutions are sufficiently smooth to satisfy a source condition with $\nu \geqslant \frac{1}{2}$ in the case of continuous measurements (up to zero boundary conditions

Table 3. Adaptive versus uniform refinement for example (c) with noise level $\delta = 1\%$.

k	No. of nodes	β^k	k	No. of nodes	β^k
	(a) Adaptive refinement			(b) Uniform refinement	
0	25	10	0	25	10
0	55	10	0	81	10
0	189	10	0	289	10
0	649	10	0	1089	10
0	2139	10	0	4225	10
0	7033	10	0	16641	10
1	7033	20.15	1	16641	20.15
2	7033	38.17	2	16641	38.16
3	7033	68.69	3	16641	68.68
4	7033	119.42	4	16641	119.40
5	7033	202.84	5	16641	202.77
6	7033	336.88	6	16641	336.67
7	7033	538.37	7	16641	537.79
8	7033	792.94	8	16641	791.59
9	7033	1013.43	9	16641	1011.01
10	7033	1105.83	10	16641	1102.70

Table 4. Convergence as $\delta \to 0$ for example (a) with $\|q^\dagger\|_Q = 5.64$ and (b) with $\|q^\dagger\|_Q = 28.21$.

δ	$\frac{\|q^\delta_{h,\hat\beta} - q^\dagger\|_Q}{\|q^\dagger\|_Q}$	$\hat\beta$	δ	$\frac{\|q^\delta_{h,\hat\beta} - q^\dagger\|_Q}{\|q^\dagger\|_Q}$	$\hat\beta$
8%	0.761	156.390	8%	0.869	2396.281
4%	0.592	660.930	4%	0.776	9044.374
2%	0.414	2426.109	2%	0.744	24364.894
1%	0.288	7047.472	1%	0.734	55017.364
0.5%	0.229	17042.825	0.5%	0.731	117560.866

on $q^\dagger$, that are 'almost' fulfilled due to the exponential decay of $q^\dagger$), the convergence behavior is quite different in both examples. The reason for that is that we assume to have only point measurements at a finite number of locations. In example (b) basically only five measurement points lie within the region where $q^\dagger$ differs from zero significantly enough, so the measurements obviously provide less information that in example (a).

5. Conclusions and remarks

In this paper, we introduced an adaptive discretization strategy for parameter identification in PDEs that is based on the concept of goal-oriented error estimators. The key idea is that if the discrepancy principle is used for regularization parameter choice, the squared residual norm can be considered as an essential quantity of interest in which the discretization error must be controlled. A complete convergence analysis is carried out in the context of Tikhonov regularization and linear inverse problems. However, our idea can be transferred to nonlinear problems and different regularization methods, as will be shown in the forthcoming paper [13].

We expect that goal-oriented error estimators can also be used when using different regularization parameter choice strategies such as the balancing (or Lepskii) principle [12] as well as generalization cross-validation (gcv) and robust GCV [17, 18].

Acknowledgments

The research work leading to this paper was initiated during a stay of the second author at the Radon Institute for Computational and Applied Mathematics (RICAM) in Linz; hence the authors wish to express their thanks to the Austrian Academy of Sciences. Moreover, we wish to thank the referees for valuable comments. The first author is supported by the DFG priority program 1251 'Optimization with Partial Differential Equations'. The third author has been partially supported by the Austrian Science Fund FWF project P18971-N18 'Numerical analysis and discretization strategies for optimal control problems with singularities'.

References

[1] Ameur H B, Chavent G and Jaffré J 2002 Refinement and coarsening indicators for adaptive parametrization to the estimation of hydraulic transmissivities *Inverse Problems* **18** 775–94
[2] Becker R, Braack M, Meidner D, Richter T, Schmich M and Vexler B 2005 *The Finite Element Toolkit* Gascoigne
[3] Becker R, Dunne T and Meidner D 2005 VisuSimple: An interactive VTK-based visualization and graphics/mpeg-generation program
[4] Becker R, Meidner D and Vexler B 2005 RoDoBo: A C++ library for optimization with stationary and nonstationary PDEs with interface to Gascoigne [2]
[5] Becker R and Rannacher R 2001 *An Optimal Control Approach to a-Posteriori Error Estimation in Acta Numerica 2001* ed A Iserles (Cambridge: Cambridge University Press) pp 1–102
[6] Becker R and Vexler B 2004 *A posteriori* error estimation for finite element discretizations of parameter identification problems *Numer. Math.* **96** 435–59
[7] Becker R and Vexler B 2005 Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations *J. Comput. Phys.* **206** 95–110
[8] Borcea L and Druskin V 2002 Optimal finite difference grids for direct and inverse sturm liouville problems *Inverse Problems* **18** 979–1001
[9] Braess D 2007 *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics* (Cambridge: Cambridge University Press)
[10] Brenner S and Scott R 2002 *The Mathematical Theory of Finite Element Methods* (Berlin: Springer)
[11] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[12] Goldenshluger A and Pereverzev S 2000 Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations *Probab. Theory Relat. Fields* **118** 169–86
[13] Griesbaum A, Kaltenbacher B and Vexler B Goal oriented adaptive discretization for nonlinear ill-posed problems, in preparation
[14] Griesse R and Vexler B 2007 Numerical sensitivity analysis for the quantity of interest in PDE-constrained optimization *SIAM J. Sci. Comput.* **29** 22–48
[15] Haber E, Heldmann S and Ascher U 2007 Adaptive finite volume methods for distributed non-smooth parameter identification *Inverse Problems* **23** 1659–76
[16] Hohage T 2000 Regularization of exponentially ill-posed problems *Numer. Funct. Anal. Optim.* **21** 439–64
[17] Lukas M A 1998 Comparisons of parameter choice methods for regularization with discrete noisy data *Inverse Problems* **14** 161–84
[18] Lukas M A 2006 Robust generalized cross-validation for choosing the regularization parameter *Inverse Problems* **22** 1883–902
[19] Meidner D and Vexler B 2007 Adaptive space-time finite element methods for parabolic optimization problems *SIAM J. Control Optim.* **46** 116–42
[20] Nair M, Pereverzev S and Tautenhahn U 2005 Regularization in Hilbert scales under general smoothing conditions *Inverse Problems* **21** 1851–69
[21] Nair M, Schock E and Tautenhahn U 2003 Morozov's discrepancy principle under general source conditions Zeitschr *Anal. Anw.* **22** 199–214
[22] Neubauer A 2007 Solution of ill-posed problems via adaptive grid regularization: convergence analysis *Numer. Funct. Anal. Optim.* **28** 405–23
[23] Tautenhahn U and Nian Jin Q 2003 Tikhonov regularization and *a posteriori* rules for solving nonlinear ill-posed problems *Inverse Problems* **19** 1–21