Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems

2011 Inverse Problems 27 125008

(http://iopscience.iop.org/0266-5611/27/12/125008)

IOP Publishing | Inverse Problems

Inverse Problems **27** (2011) 125008 (28pp) | doi:10.1088/0266-5611/27/12/125008

Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems

Barbara Kaltenbacher[1], **Alana Kirchner**[2] **and Boris Vexler**[2]

[1] Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria
[2] Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik, Boltzmannstraße 3, 85748 Garching b. München, Germany

E-mail: barbara.kaltenbacher@aau.at, kirchner@ma.tum.de and vexler@ma.tum.de

Received 26 March 2011, in final form 27 October 2011
Published 24 November 2011
Online at stacks.iop.org/IP/27/125008

Abstract
Parameter identification problems for partial differential equations usually lead to nonlinear inverse problems. A typical property of such problems is their instability, which requires regularization techniques, like, e.g., Tikhonov regularization. The main focus of this paper will be on efficient methods for determining a suitable regularization parameter by using adaptive finite element discretizations based on goal-oriented error estimators. A well-established method for the determination of a regularization parameter is the discrepancy principle where the residual norm, considered as a function i of the regularization parameter, should equal an appropriate multiple of the noise level. We suggest to solve the resulting scalar nonlinear equation by an inexact Newton method, where in each iteration step, a regularized problem is solved at a different discretization level. The proposed algorithm is an extension of the method suggested in Griesbaum A *et al* (2008 *Inverse Problems* **24** 025025) for linear inverse problems, where goal-oriented error estimators for i and its derivative are used for adaptive refinement strategies in order to keep the discretization level as coarse as possible to save computational effort but fine enough to guarantee global convergence of the inexact Newton method. This concept leads to a highly efficient method for determining the Tikhonov regularization parameter for nonlinear ill-posed problems. Moreover, we prove that with the so-obtained regularization parameter and an also adaptively discretized Tikhonov minimizer, usual convergence and regularization results from the continuous setting can be recovered. As a matter of fact, it is shown that it suffices to use stationary points of the Tikhonov functional. The efficiency of the proposed method is demonstrated by means of numerical experiments.

(Some figures may appear in colour only in the online journal)

1. Introduction

Consider the nonlinear ill-posed operator equation

$$F(q) = g, \tag{1}$$

where $F : \mathcal{D}(\subseteq Q) \to G$ is a nonlinear operator between Hilbert spaces Q and G. Since we are interested in the situation that the solution of (1) does not depend continuously on the data and we are only given noisy data g^δ with noise level δ according to

$$\|g^\delta - g\|_G \leqslant \delta, \tag{2}$$

it is necessary to apply regularization methods for their solution.

Our study is motivated by inverse problems for partial differential equations such as parameter identification or inverse boundary value problems, where F is the composition of a parameter-to-solution map

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

with some measurement operator

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

where V is an appropriate Hilbert space. Here, we will write the underlying (possibly nonlinear) PDE in its weak form

$$A(q, u)(v) = (f, v) \quad \forall v \in V, \tag{3}$$

where u denotes the PDE solution, q some searched for parameter or boundary function, and $f \in V^*$ is some given right-hand side. We assume that the PDE (3) and especially also its linearization at (q, u) is uniquely and stably solvable. We therefore assume the existence and continuity of S and S', i.e. for all $q \in \mathcal{D}$ there exists $u = S(q) \in V$ such that (3) is fulfilled, where u depends continuously on q and for all $\delta q \in Q$ there exists $\delta u = S'(q)(\delta q) \in V$ such that $A'_q(q, S(q))(\delta q) + A'_u(q, S(q))(\delta u) = 0$, where $S'(q) \in L(Q, V)$.

In the (regularized) numerical solution of (1), it is—for accuracy and efficiency reasons—of interest to use an adaptive discretization not only of the state u but also of the parameter q. For this purpose, in [6] we proposed a goal-oriented approach, based on the ideas in [2, 3]. This enabled us to not only save computational effort when solving a single regularized problem, but yielded an even higher gain in CPU time and storage when used within a Newton method for determining the regularization parameter according to the discrepancy principle, which involves several regularized problems to be solved at different discretization levels. In [6], the analysis and numerical tests were carried out for linear problems

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

in view of possible application for certain inverse source or inverse boundary value problems for PDEs.

As a matter of fact, many interesting parameter identification problems are governed by nonlinear PDEs, and even when dealing with a linear PDE, the inverse problem under consideration is often nonlinear. This motivated us to extend the results from [6] as much as possible to the nonlinear case. In addition to the quadratic convergence of the sequence of regularization parameters produced by the resulting algorithm, we will also consider linear convergence in order to get less restrictive refinement criteria and consequently coarser discretizations, which implies less computational effort.

More precisely, we will consider Tikhonov regularization, which defines q_β^δ as a minimizer (or actually a stationary point) in $\mathcal{D}$ of

$$j_\beta(q) = \|F(q) - g^\delta\|_G^2 + \frac{1}{\beta}\|q - q_0\|_Q^2. \tag{5}$$

As we will formulate the discrepancy principle in terms of β, cf, e.g., proposition 9.8 in [5], already at this point we write the Tikhonov functional in terms of $\beta > 0$ instead of (as usual) $j_\alpha(q) = \|F(q) - g^\delta\|_G^2 + \alpha\|q - q_0\|_Q^2$ with $\alpha > 0$.

For choosing the regularization parameter $\beta > 0$ in an appropriate way, a both theoretically and practically well established method is the discrepancy principle

$$\left\|F\left(q_{\beta_*}^\delta\right) - g^\delta\right\|_G = \tau\delta \tag{6}$$

with some constant $\tau > 1$. The one-dimensional equation (6) gives an implicit definition of an appropriate regularization parameter $\beta_* > 0$, which motivates the definition of

$$\mathrm{i}(\beta) = \left\|F\left(q_\beta^\delta\right) - g^\delta\right\|_G^2.$$

Hence, for finding β_*, we apply Newton's method to the equation

$$\mathrm{i}(\beta_*) = \tau^2\delta^2.$$

In this process, we have to solve (5) in each iteration in order to obtain a minimizer (or rather a stationary point) q_β^δ of j_β for the current value of β, which is required to evaluate $\mathrm{i}(\beta) = \left\|F\left(q_\beta^\delta\right) - g^\delta\right\|_G^2$.

We propose to do so on adaptively refined discretizations of the problem—denoted by the subscript h—which enables us to save a considerable amount of computational effort.

In section 2, we will first of all carry over the convergence analysis of Tikhonov regularization from the continuous setting to an adaptively discretized one. The crucial point here is that we do not impose accuracy in the sense of smallness of operator norms or closeness of Hilbert space elements, but only three scalar-valued quantities have to be computed precisely enough, namely the value of i and its derivative i' at the current iterate β, and finally the value of the Tikhonov functional $j_\beta(q_\beta^\delta)$. In this context, we provide a general rates result based on Jensen's inequality that might be of interest on its own.

Then, we derive refinement criteria in order to guarantee convergence of the sequence of regularization parameters generated by an inexact Newton-like method in order to satisfy the discrepancy principle. As in [6], we use goal-oriented error estimators (see section 3) and i and i' turn out to be appropriate quantities of interest. The main crucial aspect in transferring the results from the linear case [6] to nonlinear inverse problems is the loss of convexity of the function i. This leads to the need for the modification of the Newton step computation as well as for a new theoretical analysis, which we will provide in this paper. Following the ideas of [23], we rely on the existence of a lower and an upper bound to the second derivative of the quantity of interest i in order to guarantee quadratic convergence of the sequence of regularization parameters β. But we will see that for showing linear convergence, the existence of a lower bound suffices.

In section 4, the obtained theoretical results will be tested practically in terms of numerical experiments.

2. Tikhonov regularization

The Tikhonov regularization can be written in the form

$$\text{Minimize}\quad J_\beta(q,u) = I(q,u) + \frac{1}{\beta}\|q - q_0\|_Q^2 = \|C(u) - g^\delta\|_G^2 + \frac{1}{\beta}\|q - q_0\|_Q^2 \tag{7}$$

under the constraints

$$
\begin{aligned}
&A(q, u)(v) = f(v) \quad \forall v \in V, \\
&q \in \mathcal{D} \subseteq Q, \ v \in V.
\end{aligned} \tag{8}
$$

Since for error estimation it will be sufficient to consider stationary points instead of minimizers in (7), we introduce the Lagrange function

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

and replace (7) and (8) by the optimality system

$$
\begin{aligned}
\mathcal{L}'(q, u, z)(\mathrm{d}q,\ \mathrm{d}u,\ \mathrm{d}z) &= 2\langle C(u) - g^\delta, C'(u)(\mathrm{d}u)\rangle_G + 2\frac{1}{\beta}\langle q - q_0,\ \mathrm{d}q\rangle_Q \\
&\quad - (A'_q(q, u)(\mathrm{d}q))(z) - (A'_u(q, u)(\mathrm{d}u))(z) + f(\mathrm{d}z) - A(q, u)(\mathrm{d}z) \\
&= 0
\end{aligned} \tag{9}
$$

for all $(\mathrm{d}q,\ \mathrm{d}u,\ \mathrm{d}z) \in (Q \times V \times V)$.

Thus, solving (9) is equivalent to finding a stationary point of the reduced functional

$$j_\beta(q) = J_\beta(q, F(q)) = \|F(q) - g^\delta\|_G^2 + \frac{1}{\beta}\|q - q_0\|_Q^2 \tag{10}$$

in $\mathcal{D}$. In the following we denote such a continuous stationary point for some fixed β by $q_\beta^\delta \in \mathcal{D}$.

A discretized form of (9) is defined by inserting finite-dimensional spaces $Q_h \subset Q, V_h \subset V$ in place of Q, V, i.e. we search for $(q_h, u_h, z_h) \in (\mathcal{D} \cap Q_h \times V_h \times V_h)$ such that

$$\mathcal{L}'(q_h, u_h, z_h)(\mathrm{d}q_h,\ \mathrm{d}u_h,\ \mathrm{d}z_h) = 0 \tag{11}$$

for all $(\mathrm{d}q_h,\ \mathrm{d}u_h,\ \mathrm{d}z_h) \in (Q_h \times V_h \times V_h)$.

Considering (11) in place of (9) corresponds to using an approximation F_h to F in (1) (or K_h to K in (4)).

Analogously, by solving (11), we get a stationary point of the discrete reduced cost functional

$$j_{h,\beta}(q) = J_\beta(q, F_h(q)) = \|F_h(q) - g^\delta\|_G^2 + \frac{1}{\beta}\|q - q_0\|_Q^2 \tag{12}$$

in $\mathcal{D} \cap Q_h$. We denote such a discrete stationary point for some fixed β by $q_{h,\beta}^\delta \in Q_h$.

Besides we introduce the following notation for the residual norm:

$$\iota(q) = I(q, F(q)) = \|F(q) - g^\delta\|_G^2$$

and its discrete equivalent

$$\iota_h(q) = I(q, F_h(q)) = \|F_h(q) - g^\delta\|_G^2.$$

Our quantity of interest is the squared residual norm at the Tikhonov regularized solution that is required for evaluating the discrepancy principle (6)

$$\mathrm{i}(\beta) = \iota\big(q_\beta^\delta\big) = \big\|F\big(q_\beta^\delta\big) - g^\delta\big\|_G^2, \tag{13}$$

$$i_h(\beta) = \iota_h\big(q_\beta^\delta\big) = \big\|F_h\big(q_\beta^\delta\big) - g^\delta\big\|_G^2. \tag{14}$$

Computation of β_* according to the discrepancy principle (6) amounts to solving the one-dimensional equation (cf [18] for its solvability)

$$\mathrm{i}(\beta_*) = \tau^2\delta^2. \tag{15}$$

This has to be done iteratively, so we repeatedly have to evaluate this quantity of interest corresponding to solutions of (7) and (8) for different values of β. It is obvious that this can be done with lower precision at the beginning of the iteration, which helps to save computational effort by using coarse discretizations.

Moreover, (15) does not need to be solved exactly. Namely, it will be shown that for obtaining convergence and optimal convergence rates, it suffices to have a regularization parameter $\widehat{\beta}$ that satisfies

$$\underline{\tau}\delta \leqslant \left\| F\left(q^\delta_{\widehat{\beta}}\right) - g^\delta \right\|_G \leqslant \overline{\tau}\delta \tag{16}$$

for some constants $0 < \underline{\tau} \leqslant \overline{\tau}$.

The existence of a global minimizer of the Tikhonov functional and therewith of a stationary point (provided that the minimizer is an interior point of the domain $\mathcal{D}$) can be guaranteed if F is weakly sequentially closed:

$$(q_n \rightharpoonup q \wedge F(q_n) \rightharpoonup g) \;\Rightarrow\; (q \in \mathcal{D} \;\wedge\; F(q) = g) \tag{17}$$

for all $(q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$, cf, e.g., the proof of theorem 1 in [26].

Since we work with a stationary point of the Tikhonov functional, it is of interest under which conditions on F the convergence analysis for global minimizers of the Tikhonov functional can be carried over (see also [11, 22]). To conclude the existence of a stationary point from the existence of a global minimizer, we have to assume that the domain $\mathcal{D}$ of F has a nonempty interior (for a possibility of avoiding this assumption under different conditions on F than (19), we refer to proposition 4 in [13]).

2.1. Convergence of adaptively discretized stationary points of the Tikhonov functional

Throughout this paper, we assume that a solution $q^\dagger \in \mathcal{D}$ to (1) exists.

Proposition 1. *Let the reduced forward operator F be continuous and satisfy*

$$(q_n \rightharpoonup q \;\wedge\; F(q_n) \to g) \;\Rightarrow\; (q \in \mathcal{D} \;\wedge\; F(q) = g) \tag{18}$$

for all $(q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ as well as a tangential cone condition (cf, e.g., [5, 25])

$$\|F(q) - F(\bar{q}) - F'(q)(q - \bar{q})\|_G \leqslant c_{tc}\|F(q) - F(\bar{q})\|_G \tag{19}$$

for all $q, \bar{q} \in \mathcal{D}$ and for some $0 < c_{tc} < 1$. Moreover, let $q^\delta_{h,\widehat{\beta}}$ be defined by (11) with $\mathcal{L}$ as in (9), where $\widehat{\beta} = \widehat{\beta}(\delta, g^\delta) > 0$ and Q_h, V_h such that for ι as in (13)

$$\underline{\underline{\tau}}^2\delta^2 \leqslant \iota_h\left(q^\delta_{h,\widehat{\beta}}\right) \leqslant \overline{\overline{\tau}}\delta^2 \tag{20}$$

for constants $\overline{\overline{\tau}} > \tau > \underline{\underline{\tau}} > \frac{1+c_{tc}}{1-c_{ct}}$ as well as

$$\left|\iota_h\left(q^\delta_{h,\widehat{\beta}}\right) - \iota\left(q^\delta_{\widehat{\beta}}\right)\right| \leqslant \tilde{\tau}^2\delta^2 \tag{21}$$

for some constant $\tilde{\tau} > 0$ with

$$\tilde{\tau}^2 < \underline{\underline{\tau}}^2 - \left(\frac{1 + c_{tc}}{1 - c_{tc}}\right)^2. \tag{22}$$

Then, $q^\delta_{\widehat{\beta}}$ converges (weakly) subsequentially to a solution of (1) as $\delta \to 0$ in the sense that it has a weakly convergent subsequence and each weakly convergent subsequence converges strongly to a solution of (1). If the solution $q^\dagger$ to (1) is unique, then $q^\delta_{\widehat{\beta}}$ converges to $q^\dagger$ as $\delta \to 0$.

Note that (18) is less restrictive as compared to (17), since the premiss contains strong convergence of $(F(q_n))_{n\in\mathbb{N}}$.

Proof. From (20) and (21), we can conclude that (16) holds with

$$\underline{\tau}^2 = \underline{\underline{\tau}}^2 - \tilde{\tau}^2, \quad \overline{\tau}^2 = \overline{\overline{\tau}}^2 + \tilde{\tau}^2. \tag{23}$$

Forming the inner product of the equation for a stationary point

$$F'\big(q_\beta^\delta\big)^*\big(F\big(q_\beta^\delta\big) - g^\delta\big) + \beta\big(q_\beta^\delta - q_0\big) = 0 \tag{24}$$

with the error $q_\beta^\delta - q^\dagger$, we get

$$\langle F\big(q_\beta^\delta\big) - g^\delta, F'\big(q_\beta^\delta\big)\big(q_\beta^\delta - q^\dagger\big)\rangle_G + \beta\, \big\|q_\beta^\delta - q^\dagger\big\|_Q^2 = \beta\langle q_0 - q^\dagger, q_\beta^\delta - q^\dagger\rangle_Q;$$

hence, for $\beta = \widehat{\beta}$, by (2), (16), (19), (22)

$$\begin{aligned}\widehat{\beta}\big\langle q_0 - q^\dagger, q_{\widehat{\beta}}^\delta - q^\dagger\big\rangle_Q &= \big\|F\big(q_{\widehat{\beta}}^\delta\big) - g^\delta\big\|_G^2 + \widehat{\beta}\big\|q_{\widehat{\beta}}^\delta - q^\dagger\big\|_Q^2\\ &\quad - \big\langle F\big(q_{\widehat{\beta}}^\delta\big) - g^\delta, F\big(q_{\widehat{\beta}}^\delta\big) - g^\delta - F'\big(q_{\widehat{\beta}}^\delta\big)\big(q_{\widehat{\beta}}^\delta - q^\dagger\big)\big\rangle_G\\ &\geqslant \underbrace{\Big(1 - c_{tc} - \tfrac{1+c_{tc}}{\underline{\tau}}\Big)}_{>0}\big\|F\big(q_{\widehat{\beta}}^\delta\big) - g^\delta\big\|_G^2 + \widehat{\beta}\big\|q_{\widehat{\beta}}^\delta - q^\dagger\big\|_Q^2,\end{aligned}$$

and therewith

$$\big\|q_{\widehat{\beta}}^\delta - q^\dagger\big\|_Q^2 \leqslant \big\langle q_0 - q^\dagger, q_{\widehat{\beta}}^\delta - q^\dagger\big\rangle. \tag{25}$$

By the Cauchy–Schwarz inequality, this implies that

$$\big\|q_{\widehat{\beta}}^\delta - q^\dagger\big\|_Q \leqslant \big\|q_0 - q^\dagger\big\|_Q; \tag{26}$$

hence, a weakly convergent subsequence of $q_{\widehat{\beta}(\delta)}^\delta$ exists. By

$$\big\|F\big(q_{\widehat{\beta}}^\delta\big) - g^\delta\big\|_G \leqslant \overline{\tau}\delta \to 0 \text{ as } \delta \to 0 \tag{27}$$

and the (weak) sequential closedness of F, the weak limit of any weakly convergent subsequence of $q_{\widehat{\beta}(\delta)}^\delta$ is a solution to (1) and therefore can be inserted in place of $q^\dagger$ in (25), which implies even strong convergence by a standard argument (see, e.g, [5]). □

Remark 1. Instead of (21), it suffices to require

$$\big|\iota_h\big(q_{h,\beta}^\delta\big) - \iota\big(q_\beta^\delta\big)\big| \leqslant \max\big\{c \cdot \iota_h\big(q_{h,\beta}^\delta\big), \tilde{\tau}^2\delta^2\big\} \tag{28}$$

for some constant $c \leqslant \frac{\tilde{\tau}^2}{\overline{\overline{\tau}}^2}$, since this less restrictive condition implies (21), once (20) is satisfied.

For proving convergence rates, we make use of the following general result, which might be of interest on its own.

Theorem 1. *Let F satisfy the tangential cone condition (19) for all $q, \bar{q} \in \mathcal{D}$ and for some $0 < c_{tc} < 1$, and let $\tilde{q}$ be a regularized approximation (not necessarily defined by Tikhonov regularization) of a solution $q^\dagger$ of $F(q) = g$ with $\|g - g^\delta\|_G \leqslant \delta$ such that*

$$\|\tilde{q} - q_0\|_Q \leqslant \|q^\dagger - q_0\|_Q \tag{29}$$

$$\|F(\tilde{q}) - g^\delta\|_G \leqslant \hat{\tau}\delta \tag{30}$$

with $\hat{\tau}$ independent of δ. Moreover, let, with some $v \in Q$, the source condition

$$q^\dagger - q_0 = \kappa(F'(q^\dagger)^*F'(q^\dagger))v \tag{31}$$

hold with $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ *such that* κ^2 *is strictly monotonically increasing on* $(0, \|F'(q^\dagger)\|^2]$, ϕ *defined by* $\phi^{-1}(\lambda) = \kappa^2(\lambda)$ *is convex and* ψ *defined by* $\psi(\lambda) = \kappa(\lambda)\sqrt{\lambda}$ *is strictly monotonically increasing on* $(0, \|F'(q^\dagger)\|^2]$.

Then, the rates

$$\|\tilde{q} - q^\dagger\|_Q^2 \leqslant \frac{\bar{C}^2\delta^2}{\psi^{-1}\left(\frac{\bar{C}}{2\|v\|_Q}\delta\right)} = 4\|v\|_Q^2\kappa^2\left(\psi^{-1}\left(\frac{\bar{C}}{2\|v\|_Q}\delta\right)\right) \tag{32}$$

hold with $\bar{C}$ *independent of* δ.

Proof. If $\|\tilde{q} - q^\dagger\|_Q$ vanishes, we are trivially done. So we assume $\|\tilde{q} - q^\dagger\|_Q \neq 0$ for the rest of the proof.

If (31) holds, then we have by Jensen's inequality (cf proposition 1 in [12])

$$\begin{aligned} |\langle q^\dagger - q_0, q - q^\dagger\rangle| &= |\langle v, \kappa(F'(q^\dagger)^*F'(q^\dagger))(q - q^\dagger)\rangle| \\ &\leqslant \|v\|_Q\|q - q^\dagger\|_Q\kappa\left(\frac{\|F'(q^\dagger)(q - q^\dagger)\|_G^2}{\|q - q^\dagger\|_Q^2}\right) \end{aligned} \tag{33}$$

for any $q \in Q$.

Combining (29) and (33), we obtain

$$\begin{aligned} 0 &\geqslant \|\tilde{q} - q_0\|_Q^2 - \|q^\dagger - q_0\|_Q^2 \\ &= \left\|\tilde{q} - q^\dagger\right\|_Q^2 + 2\langle q^\dagger - q_0, \tilde{q} - q^\dagger\rangle \\ &\geqslant \|\tilde{q} - q^\dagger\|_Q^2 - 2\|v\|_Q\|\tilde{q} - q^\dagger\|_Q\kappa\left(\frac{\|F'(q^\dagger)(\tilde{q} - q^\dagger)\|_G^2}{\|\tilde{q} - q^\dagger\|_Q^2}\right); \end{aligned}$$

hence, by monotonicity of κ, (19) and (30)

$$\begin{aligned} \|\tilde{q} - q^\dagger\|_Q^2 &\leqslant 2\|v\|_Q\|\tilde{q} - q^\dagger\|_Q\kappa\left(\frac{\|F'(q^\dagger)(\tilde{q} - q^\dagger)\|_G^2}{\|\tilde{q} - q^\dagger\|_Q^2}\right) \\ &\leqslant 2\|v\|_Q\|\tilde{q} - q^\dagger\|_Q\kappa\left(\frac{((1 + c_{tc})(\|F(\tilde{q}) - g^\delta\|_G + \delta))^2}{\|\tilde{q} - q^\dagger\|_Q^2}\right) \\ &\leqslant 2\|v\|_Q\|\tilde{q} - q^\dagger\|_Q\kappa\left(\frac{\bar{C}^2\delta^2}{\|\tilde{q} - q^\dagger\|_Q^2}\right) \end{aligned} \tag{34}$$

with $\bar{C} = (1 + c_{tc})(\hat{\tau} + 1)$.

As we consider the case $\left\|\tilde{q} - q^\dagger\right\|_Q \neq 0$, we can multiply both sides of (34) by $\frac{\bar{C}\delta}{\|\tilde{q}-q^\dagger\|_Q^2}$ to obtain

$$\bar{C}\delta \leqslant 2\;\|v\|_Q\,\frac{\bar{C}\delta}{\left\|\tilde{q} - q^\dagger\right\|_Q}\kappa\left(\frac{\bar{C}^2\delta^2}{\left\|\tilde{q} - q^\dagger\right\|_Q^2}\right) = 2\;\|v\|_Q\,\psi\left(\frac{\bar{C}^2\delta^2}{\left\|\tilde{q} - q^\dagger\right\|_Q^2}\right).$$

By strict monotonicity of ψ, this yields

$$\psi^{-1}\left(\frac{\bar{C}}{2\;\|v\|_Q}\delta\right) \leqslant \frac{\bar{C}^2\delta^2}{\left\|\tilde{q} - q^\dagger\right\|_Q^2},$$

which proves the first part of (32). The second part of (32) just follows by definition of ψ. □

Proposition 2. *Let the conditions of proposition 1 be fulfilled. Let, moreover, the source condition (31) hold with κ, ψ as in theorem 1.*

Then, the following convergence rates with some $C > 0$ independent of δ are obtained:

$$\left\| q_{\widehat{\beta}}^{\delta} - q^{\dagger} \right\|_Q = \mathcal{O}\left(\frac{C\delta}{\sqrt{\psi^{-1}(C\delta)}} \right). \tag{35}$$

Proof. Follows from estimates (26), (27) and theorem 1. □

Remark 2. With

$$(a)\ \kappa(\lambda) = \lambda^{\nu}, \quad \nu \in (0, \tfrac{1}{2}] \quad \text{or} \quad (b)\ \kappa(\lambda) = (-\ln \lambda)^{-p}, \quad p > 0, \tag{36}$$

we can conclude the usual optimal Hölder type or logarithmic convergence rates, respectively.

For a convergence (rates) proof under different conditions on the forward operator than those used here and in a continuous setting, we refer to [27].

To obtain convergence rates

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

For the discrete version $q_{h,\widehat{\beta}}^{\delta}$ of $q_{\widehat{\beta}}^{\delta}$, we need another quantity of interest to be computed with sufficient precision, namely $j_{h,\widehat{\beta}}(q_{h,\widehat{\beta}}^{\delta})$, which is shown by means of the following proposition.

Proposition 3. *Let the conditions of proposition 1 be fulfilled. Let moreover $\underline{\tau}$, $\tilde{\tau}$ be chosen such that*

$$\sqrt{\underline{\tau}^2 - \tilde{\tau}^2} > \frac{1 + c_{tc} + \sqrt{(1 + c_{tc})^2 + 2(1 - c_{tc})(1 + \tilde{\tau}^2)}}{1 - c_{tc}}$$

is fulfilled (instead of the weaker condition (22)). If for the discretization error with respect to the cost functional

$$\left| j_{\widehat{\beta}}(q_{\widehat{\beta}}^{\delta}) - j_{h,\widehat{\beta}}(q_{h,\widehat{\beta}}^{\delta}) \right| \leqslant \Theta^2 \delta^2 \tag{37}$$

holds, where Θ is sufficiently small so that

$$\sqrt{\underline{\tau}^2 - \tilde{\tau}^2} \geqslant \frac{1 + c_{tc} + \sqrt{(1 + c_{tc})^2 + 2(1 - c_{tc})(1 + \Theta^2 + \tilde{\tau}^2)}}{1 - c_{tc}}, \tag{38}$$

then $q_{h,\widehat{\beta}}^{\delta}$ converges to $q^{\dagger}$ in Q as $\delta \to 0$.

If additionally the source condition (31) holds with κ, ψ as in theorem 1, then the convergence rate (35) is obtained for $q_{h,\widehat{\beta}}^{\delta}$ in place of $q_{\widehat{\beta}}^{\delta}$.

Proof. From the Euler equation (24) and (19), we get

$$\begin{aligned}
j_{\widehat{\beta}}(q_{\widehat{\beta}}^{\delta}) - j_{\widehat{\beta}}(q^{\dagger}) &= \left\| F(q_{\widehat{\beta}}^{\delta}) - g^{\delta} \right\|_G^2 - \|F(q^{\dagger}) - g^{\delta}\|_G^2 \\
&\quad + \frac{1}{\widehat{\beta}} \langle q_{\widehat{\beta}}^{\delta} + q^{\dagger} - 2q_0, q_{\widehat{\beta}}^{\delta} - q^{\dagger} \rangle_Q \\
&= \frac{2}{\widehat{\beta}} \langle q_{\widehat{\beta}}^{\delta} - q_0, q_{\widehat{\beta}}^{\delta} - q^{\dagger} \rangle_Q + 2 \left\| F(q_{\widehat{\beta}}^{\delta}) - g^{\delta} \right\|_G^2 \\
&\quad - \left\| F(q_{\widehat{\beta}}^{\delta}) - g^{\delta} \right\|_G^2 - \frac{1}{\widehat{\beta}} \left\| q_{\widehat{\beta}}^{\delta} - q^{\dagger} \right\|_Q^2 - \|g - g^{\delta}\|_G^2 \\
&= 2 \langle F(q_{\widehat{\beta}}^{\delta}) - g^{\delta}, F(q_{\widehat{\beta}}^{\delta}) - g^{\delta} - F'(q_{\widehat{\beta}}^{\delta})(q_{\widehat{\beta}}^{\delta} - q^{\dagger}) \rangle_G
\end{aligned}$$

$$
\begin{aligned}
&-\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2-\frac{1}{\widehat{\beta}}\left\|q^{\delta}_{\widehat{\beta}}-q^{\dagger}\right\|_Q^2-\|g-g^{\delta}\|_G^2\\
&\leqslant -(1-2c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G\\
&-\frac{1}{\widehat{\beta}}\left\|q^{\delta}_{\widehat{\beta}}-q^{\dagger}\right\|_Q^2.
\end{aligned}
$$

By (37) and (19), we further get

$$
\begin{aligned}
\left\|F_h\left(q^{\delta}_{h,\widehat{\beta}}\right)-g^{\delta}\right\|_G^2+\frac{1}{\widehat{\beta}}\left\|q^{\delta}_{h,\widehat{\beta}}-q_0\right\|_Q^2 &= j_{h,\widehat{\beta}}\left(q^{\delta}_{h,\widehat{\beta}}\right)\leqslant j_{\widehat{\beta}}\left(q^{\delta}_{\widehat{\beta}}\right)+\Theta^2\delta^2\\
&\leqslant j_{\widehat{\beta}}(q^{\dagger})+\Theta^2\delta^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G\\
&-(1-2c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2\\
&\leqslant (1+\Theta^2)\delta^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G\\
&-(1-2c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2+\frac{1}{\widehat{\beta}}\|q^{\dagger}-q_0\|_Q^2,
\end{aligned}
$$

i.e. by (21)

$$
\begin{aligned}
\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2-\tilde{\tau}^2\delta^2+\frac{1}{\widehat{\beta}}\left\|q^{\delta}_{h,\widehat{\beta}}-q_0\right\|_Q^2 &\leqslant (1+\Theta^2)\delta^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G\\
&-(1-2c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2+\frac{1}{\widehat{\beta}}\|q^{\dagger}-q_0\|_Q^2,
\end{aligned}
$$

which implies

$$
\begin{aligned}
\left\|q^{\delta}_{h,\widehat{\beta}}-q_0\right\|_Q^2-\left\|q^{\dagger}-q_0\right\|_Q^2 &\leqslant \widehat{\beta}\Big((1+\Theta^2+\tilde{\tau})\delta^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G\\
&-2(1-c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2\Big).
\end{aligned} \tag{39}
$$

From the fact that (16) holds with (23) (see the beginning of the proof of proposition 1) and by (38), we can conclude that

$$
\begin{aligned}
\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G &\geqslant \sqrt{\underline{\tau}^2\delta^2}=\delta\sqrt{\underline{\underline{\tau}}^2-\tilde{\tau}^2}\\
&\geqslant \delta\,\frac{(1+c_{ct})+\sqrt{(1+c_{ct})^2+2(1-c_{ct})(1+\Theta^2+\tilde{\tau})}}{2(1-c_{tc})},
\end{aligned}
$$

which implies

$$
\widehat{\beta}\Big((1+\Theta^2+\tilde{\tau})\delta^2+2(1+c_{tc})\delta\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G-2(1-c_{tc})\left\|F\left(q^{\delta}_{\widehat{\beta}}\right)-g^{\delta}\right\|_G^2\Big)\leqslant 0,
$$

since

$$
x_{1/2}=\delta\,\frac{(1+c_{ct})\pm\sqrt{(1+c_{ct})^2+2(1-c_{ct})(1+\Theta^2+\tilde{\tau})}}{2(1-c_{tc})}
$$

are the only solutions to the quadratic equation

$$
(1+\Theta^2+\tilde{\tau})\delta^2+2(1+c_{tc})\delta x-2(1-c_{ct})x^2=0,
$$

and $-2(1-c_{ct})<0$. By (39), this finally leads to

$$
\left\|q^{\delta}_{h,\widehat{\beta}}-q_0\right\|_Q^2\leqslant \|q^{\dagger}-q_0\|_Q^2.
$$

The rest of the proof follows the lines of the proof of proposition 1 as well as from theorem 1. □

2.2. Determination of the regularization parameter

While $\beta \mapsto \mathrm{i}(\beta)$ is convex and monotone in the linear case, which implies global monotone convergence of Newton's method for (15), these properties partially get lost in the nonlinear case. Still, monotonicity follows in a very general setting from minimality arguments (see lemma 1). Under a condition similar to (19), we obtain strict monotonicity even with stationary points, as well as lower and upper bounds for i'', provided F is sufficiently smooth (see lemma 2).

Lemma 1. *Assume that for all* $\beta \in [\underline{\beta}, \overline{\beta}]$, *with some* $\overline{\beta} > \underline{\beta} > 0$, *there exists a global minimizer* q_β^δ *of the Tikhonov functional (5).*

Then, the function i defined by $\mathrm{i}(\beta) = \left\| F(q_\beta^\delta) - g^\delta \right\|_G^2$ *with any selection* $q_\beta \in \mathrm{argmin}(j_\beta)$ *being a monotonically decreasing function on* $[\underline{\beta}, \overline{\beta}]$.

Proof. For any $\underline{\beta} \leqslant \beta_1 \leqslant \beta_2 \leqslant \overline{\beta}$ and any $q(\beta_i) = q_{\beta_i}^\delta \in \mathrm{argmin}(j_{\beta_i})$, $i = 1, 2$, we have by minimality

$$\begin{aligned} j_{\beta_2}(q(\beta_2)) &\leqslant j_{\beta_2}(q(\beta_1)) = j_{\beta_1}(q(\beta_1)) + \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right) \|q(\beta_1) - q_0\|_Q^2 \\ &\leqslant j_{\beta_1}(q(\beta_2)) + \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right) \|q(\beta_1) - q_0\|_Q^2, \end{aligned} \tag{40}$$

which implies $\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\left(\|q(\beta_1) - q_0\|_Q^2 - \|q(\beta_2) - q_0\|_Q^2 \right) \geqslant 0$, thus monotone increase of the mapping $\beta \mapsto \|q(\beta) - q_0\|_Q^2$. Monotone decrease of i follows from (40) and the monotonicity of $\beta \mapsto \|q(\beta) - q_0\|_Q^2$. □

Lemma 2. *Assume that for all* $\beta \in [\underline{\beta}, \overline{\beta}]$, *with some* $\overline{\beta} > \underline{\beta} > 0$, *there exists a stationary point* q_β^δ *of the Tikhonov functional (5). Let F be twice continuously differentiable and satisfy (19) as well as the condition*

$$\left\|F(q) - g^\delta\right\|_G \left\|F''(q)[v, v]\right\|_G \leqslant \left\|F'(q)v\right\|_G^2 + \left(\frac{1}{\overline{\beta}} - \eta\right) \|v\|^2 \qquad \forall q \in \mathcal{D}, v \in Q, \tag{41}$$

for some $\eta > 0$. *Moreover, we assume for the signal-to-noise ratio*

$$\left\|F\left(q_\beta^\delta\right) - g^\delta\right\|_G > \frac{1 + c_{tc}}{1 - c_{tc}} \delta. \tag{42}$$

Then, there exists a choice of stationary points $q(\beta) = q_\beta^\delta$ *of* j_β, $\beta \in [\underline{\beta}, \overline{\beta}]$ *such that the function i defined by* $\mathrm{i}(\beta) = \left\|F\left(q_\beta^\delta\right) - g^\delta\right\|_G^2$ *is continuously differentiable and strictly monotonically decreasing on* $[\underline{\beta}, \overline{\beta}]$:

$$i'(\beta) \leqslant -2\underline{\beta}\eta \left\| \frac{\mathrm{d}}{\mathrm{d}\beta} q_\beta^\delta \right\|_Q^2 < 0. \tag{43}$$

If additionally F''' *is continuous, then we obtain continuity of* i'' *on the compact interval* $[\underline{\beta}, \overline{\beta}]$; *hence, there exist* $\underline{\gamma}, \overline{\gamma} \in \mathbb{R}$ *such that* $\underline{\gamma} \leqslant i''(\beta) \leqslant \overline{\gamma}$ *for all* $\beta \in [\underline{\beta}, \overline{\beta}]$.

Proof. A stationary point $q(\beta) := q_\beta^\delta$ has to satisfy the operator equation $\Psi(q, \beta) = 0$ with Ψ defined by

$$\forall v \in Q: \quad \langle \Psi(q, \beta), v \rangle := \beta \langle F(q) - g^\delta, F'(q)v \rangle + \langle q - q_0, v \rangle.$$

Since the derivative of Ψ with respect to q is positive due to (41),

$$\langle \Psi_q'(q,\beta)v, v\rangle = \beta\big\{\|F'(q)v\|_G^2 + \langle F(q) - g^\delta, F''(q)[v,v]\rangle\big\} + \|v\|_Q^2 \geqslant \left(1 - \frac{\beta}{\overline{\beta}} + \beta\eta\right)\|v\|_Q^2 \geqslant \beta\eta\,\|v\|_Q^2$$

for all $v \in Q$ and $\beta \in [\underline{\beta}, \overline{\beta}]$; the implicit function theorem provides the existence of a smooth path $q(\beta)$ defined on the interval $[\underline{\beta}, \overline{\beta}]$ with the claimed differentiability under the respective differentiability assumptions on F. This implies $\mathrm{i} \in C^1([\underline{\beta}, \overline{\beta}])$ if $F \in C^2$ and $\mathrm{i} \in C^2([\underline{\beta}, \overline{\beta}])$ if $F \in C^3$.

To obtain an estimate of i', we differentiate the Euler equation

$$\forall v \in Q: \quad \beta\langle F(q(\beta)) - g^\delta, F'(q(\beta))v\rangle + \langle q(\beta) - q_0, v\rangle = 0$$

with respect to β, which yields

$$\begin{aligned}\forall v \in Q: \quad &\langle F(q(\beta)) - g^\delta, F'(q(\beta))v\rangle + \beta\langle F'(q(\beta))q'(\beta), F'(q(\beta))v\rangle\\ &+ \beta\langle F(q(\beta)) - g^\delta, F''(q(\beta))[q'(\beta), v]\rangle + \langle q'(\beta), v\rangle \;=\; 0.\end{aligned} \tag{44}$$

Therewith, setting $v = q'(\beta)$, we can compute

$$\begin{aligned} i'(\beta) &= \frac{\mathrm{d}}{\mathrm{d}\beta}\,\big\|F(q(\beta)) - g^\delta\big\|_G^2\\ &= 2\langle F(q(\beta)) - g^\delta, F'(q(\beta))q'(\beta)\rangle\\ &= -2\beta\big\{\|F'(q(\beta))q'(\beta)\|_G^2 + \langle F(q(\beta)) - g^\delta, F''(q(\beta))[q'(\beta), q'(\beta)]\rangle\big\}\\ &\quad -2\|q'(\beta)\|_Q^2. \end{aligned} \tag{45}$$

The term in braces is bounded from below by $(\eta - \frac{1}{\beta})\|q'(\beta)\|_Q$ due to the Cauchy–Schwarz inequality, condition (41), and $\underline{\beta} \leqslant \beta \leqslant \overline{\beta}$. This implies the first estimate in (43). The fact that $q'(\beta) \neq 0$ can be seen from (44) as well: if $q'(\beta)$ would vanish, then all terms except for the first one in (44) would be zero. Setting $v := q(\beta) - q^\dagger$ and using (19), we would then arrive at

$$\begin{aligned} 0 &= \langle F(q(\beta)) - g^\delta, F'(q(\beta))(q(\beta) - q^\dagger)\rangle\\ &\geqslant \|F(q(\beta)) - g^\delta\|_G((1 - c_{tc})\|F(q(\beta)) - g^\delta\|_G - (1 + c_{tc})\delta), \end{aligned}$$

which gives a contradiction to (42). □

Remark 3. As can be seen from the proof, instead of (41), we only need

$$j_\beta''(q_\beta^\delta)[v,v] = 2\big\|F'\big(q_\beta^\delta\big)v\big\|_G^2 + 2\big\langle F\big(q_\beta^\delta\big) - g^\delta, F''\big(q_\beta^\delta\big)[v,v]\big\rangle + \frac{2}{\beta}\|v\|_Q^2 \geqslant 2\eta\|v\|_Q^2$$

for all $v \in Q$ to hold at stationary points q_β^δ of the Tikhonov functional $j_\beta(q)$, i.e. these stationary points should satisfy a second order sufficient optimality condition.

Therewith we are led to make use of globally and superlinearly convergent monotone modifications of Newton's method, cf, e.g. [23], where the solution is approached from above and below by two simultaneously computed sequences of iterates. In the simple one-dimensional situation we deal with here, we can make use of the fact that quadratic equations can be solved explicitly (see (58)) to get rid of the necessity of computing two sequences. For this purpose like in [23], we need a lower (and partially also an upper) bound on the second derivative of i. An approximation to the second derivative can be computed very efficiently in the context of goal-oriented error estimation, see section 3 and section 2 in [6].

For the following proposition we will temporarily repudiate the explicit definition of $\mathrm{i}(\beta) = \big\|F\big(q_\beta^\delta\big) - g^\delta\big\|_G^2$, since the result holds for any function i fulfilling the assumptions imposed at the beginning of the proposition.

Proposition 4. *Let* $\mathrm{i} \in C^2(\mathbb{R}^+)$,

$$i'(\beta) < 0, \qquad \underline{\gamma} \leqslant i''(\beta)$$

for all $\beta > 0$ *with some constant* $-\underline{\gamma} < \overline{\gamma}$, $\overline{\gamma} \geqslant 0$ *independent of* β, *and let* β_* *solve* $i(\beta_*) = \tau^2\delta^2$ *with* $\tau, \delta > 0$. *Choose* β^0 *so that* $\mathrm{i}(\beta^0) \geqslant \tau^2\delta^2$, *and define*

$$k_* = \inf\{k \in \mathbb{N} \mid i_h^k - \overline{\overline{\tau}}^2\delta^2 \leqslant 0\} \in \mathbb{N} \cup \{\infty\} \tag{46}$$

for some $\overline{\overline{\tau}} > \tau$:

$$s_N^{k+1} = -\frac{i_h^k - \tau^2\delta^2}{i'^k_h}, \qquad \sigma^{k+1} = \frac{2}{1 + \sqrt{1 - 2\gamma s_N^{k+1}/i'^k_h}}$$

$$s^{k+1} = \sigma^{k+1} s_N^{k+1}, \qquad \beta^{k+1} = \beta^k + s^{k+1} \tag{47}$$

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

$$\left|\mathrm{i}(\beta^k) - i_h^k\right| \leqslant c_1 \left|i_h^k - \tau^2\delta^2\right| \tag{48}$$

$$\left|i'(\beta^k) - i'^k_h\right| \leqslant c_2 \left|i'^k_h\right| \tag{49}$$

$$\left|\mathrm{i}(\beta^k) - i_h^k\right| + \left|i'(\beta^k) - i'^k_h\right| |s^{k+1}| \leqslant \frac{\gamma + \underline{\gamma}}{2} |s^{k+1}|^2 \tag{50}$$

$$\left|\mathrm{i}(\beta^{k_*}) - i_h^{k_*}\right| \leqslant (\tau^2 - \underline{\underline{\tau}}^2)\delta^2 \tag{51}$$

for some constants $c_1 \in (0, 1)$, $c_2 \in \left(0, \frac{1}{2}\right)$, $\underline{\underline{\tau}} < \tau$ *independent of* k, k_*, *as well as for* $k \geqslant k_*$ *additionally:*

$$\left|\mathrm{i}(\beta^k) - i_h^k\right| + \left|\mathrm{i}(\beta^{k-1}) - i_h^{k-1}\right| + \left|i'(\beta^{k-1}) - i'^{k-1}_h\right| |s^k| < \frac{\gamma + \underline{\gamma}}{2} |s^k|^2. \tag{52}$$

Then, k_* *is finite, and for all* $k \in \mathbb{N}$ *we have well-defined, monotonicity*

$$\beta^{k-1} \leqslant \beta^k \leqslant \beta_*, \tag{53}$$

$$s_N^k \geqslant 0, \qquad \sigma^k \in (0, 1], \tag{54}$$

and

$$\underline{\underline{\tau}}^2\delta^2 \leqslant i_h^{k_*} \leqslant \overline{\overline{\tau}}^2\delta^2, \tag{55}$$

as well as boundedness of i'^k_h, *i.e. there exist constants* $a, b > 0$, *such that for all* $k \in \mathbb{N}$

$$a \leqslant \left|i'^k_h\right| \leqslant b. \tag{56}$$

Moreover, we have convergence

$$\beta^k \to \beta_* \quad \text{and} \quad \sigma^{k+1} \to 1 \quad \text{as} \quad k \to \infty. \tag{57}$$

Proof. Monotonicity (53) and (54) can be shown by induction. We only do the induction step here; the case $k = 0$ goes analogously or follows directly from the assumptions.

Observe that s^{k+1} solves the quadratic equation

$$i_h^k - \tau^2\delta^2 + i'^k_h s - \frac{\gamma}{2} s^2 = 0. \tag{58}$$

Therefore, we can write

$$\begin{aligned} \mathrm{i}(\beta^{k+1}) - \tau^2\delta^2 &= \mathrm{i}(\beta^{k+1}) - i_h^k - i_h'^k s^{k+1} + \frac{\gamma}{2}(s^{k+1})^2 \\ &= \left(\int_0^1 (1-t) i''(\beta^k + t s^{k+1})\,\mathrm{d}t + \frac{\gamma}{2}\right)(s^{k+1})^2 \\ &\quad + \mathrm{i}(\beta^k) - i_h^k + (i'(\beta^k) - i_h'^k) s^{k+1} \\ &\geqslant \frac{\underline{\gamma} + \gamma}{2}|s^{k+1}|^2 - \left|\mathrm{i}(\beta^k) - i_h^k\right| - \left|i'(\beta^k) - i_h'^k\right||s^{k+1}| \geqslant 0, \end{aligned} \tag{59}$$

due to (50), which by $i' < 0$ and $\tau^2\delta^2 = \mathrm{i}(\beta_*)$ yields the right inequality in (53).

In the case $k \geqslant k_*$, due to (52), the estimate (59) with k replaced by $k-1$ also yields

$$i_h^k - \tau^2\delta^2 > 0, \tag{60}$$

and in the case $k \leqslant k_* - 1$ positivity of the residual (60) just follows from the definition of k_*:

$$i_h^k - \tau^2\delta^2 \geqslant i_h^k - \overline{\overline{\tau}}^2\delta^2 > 0.$$

For the approximate derivative, we have by (49)

$$i_h'^k \leqslant i'(\beta^k) + \left|i'(\beta^k) - i_h'^k\right| \leqslant i'(\beta^k) + \frac{c_2}{1-c_2}\underbrace{|i'(\beta^k)|}_{=-i'(\beta^k)} = \frac{1-2c_2}{1-c_2} i'(\beta^k) < 0. \tag{61}$$

A combination of (60) and (61) yields

$$s_N^{k+1} = \frac{i_h^k - \tau^2\delta^2}{-i_h'^k} > 0.$$

Hence, since the factor $\sigma^{k+1} = \frac{2}{1+\sqrt{1-2\gamma s_N^{k+1}/i_h'^k}}$ is therewith obviously well defined and contained in the interval $(0, 1]$, we get the left inequality in (53).

So $(\beta^k)_{k\in\mathbb{N}}$ is a monotonically increasing sequence which is bounded from above by β_* and consequently convergent. In the following, we show that its limit is in fact β_*.

In order to do so, we first prove the boundedness of $|i_h'^k|$, i.e. (56). To this end, we use the fact that by our assumption $i'(\beta) < 0$ we have

$$0 < \underline{c} := \min_{\beta\in[\beta^0,\beta_*]} -i'(\beta) \leqslant -i'(\beta^k) \leqslant \max_{\beta\in[\beta^0,\beta_*]} -i'(\beta) =: \overline{c}$$

and therewith, by (61)

$$|i_h'^k| = -i_h'^k \geqslant \frac{1-2c_2}{1-c_2}\underline{c}$$

as well as by (49)

$$\begin{aligned} |i_h'^k| = -i_h'^k &\leqslant -i'(\beta^k) + |i'(\beta^k) - i_h'^k| \leqslant -i'(\beta^k) + \frac{c_2}{1-c_2}|i'(\beta^k)| \\ &= \frac{1}{1-c_2}|i'(\beta^k)| \leqslant \frac{1}{1-c_2}\overline{c}. \end{aligned}$$

By (47) and the convergence of the sequence $(\beta^k)_{k\in\mathbb{N}}$, there holds

$$|\sigma^{k+1} s_N^{k+1}| = |\beta^{k+1} - \beta^k| \to 0 \tag{62}$$

for $k \to \infty$. This leads to two cases.

(i) Case: $s_N^{k+1} \nrightarrow 0$. Hence, there exists a subsequence $(s_N^{k_l+1})_{l\in\mathbb{N}}$ which is bounded away from zero, which by (62) implies $\sigma^{k_l+1} \to 0$. This by the definition of σ^{k_l+1} and (61) implies $\frac{s_N^{k_l+1}}{|i'^{k_l}_h|} \to \infty$, which by (56) implies $s_N^{k_l+1} \to \infty$, so that $\left|\sigma^{k_l+1} s_N^{k_l+1}\right| \geqslant \left|\frac{2 s_N^{k_l+1}}{1+\sqrt{1+2\gamma\frac{s_N^{k_l+1}}{a}}}\right| \to \infty$, which is a contradiction to (62).

(ii) Case: $s_N^{k+1} \to 0$. Then, $\frac{|i_h^k-\tau^2\delta^2|}{b} \leqslant \frac{|i_h^k-\tau^2\delta^2|}{|i'^k_h|} = s_N^{k+1} \to 0$; hence, $\left|i_h^k - \tau^2\delta^2\right| \to 0$.

As the first case led to a contradiction, we can conclude that

$$s_N^{k+1} \to 0, \qquad \left|i_h^k - \tau^2\delta^2\right| \to 0, \tag{63}$$

which also implies that k_* is finite.

The convergence $\beta^k \to \beta_*$ can now be shown by using (48) and (63):

$$\begin{aligned}\underline{c}|\beta^k - \beta_*| &\leqslant |\mathrm{i}(\beta^k) - \mathrm{i}(\beta_*)| \leqslant |\mathrm{i}(\beta^k) - i_h^k| + |i_h^k - \mathrm{i}(\beta_*)| \\ &\leqslant (c_1+1)|i_h^k - \tau^2\delta^2| \ \to 0 \quad \text{as} \quad k \to \infty.\end{aligned}$$

To show (55), we use monotonicity and (51) to conclude the lower bound

$$i_h^{k_*} \geqslant \mathrm{i}(\beta^{k_*}) - |i_h^{k_*} - \mathrm{i}(\beta^{k_*})| \geqslant \mathrm{i}(\beta_*) - |i_h^{k_*} - \mathrm{i}(\beta^{k_*})| \geqslant \tau^2\delta^2 - (\tau^2 - \underline{\underline{\tau}})\delta^2,$$

while the upper bound follows directly from the definition of k_*.

Finally, the second limit in (57) follows from (56) and (63). □

Note that by lemma 2, the assumptions of this proposition are satisfied for the function i defined by $\mathrm{i}(\beta) = \left\|F(q_\beta^\delta) - g^\delta\right\|^2$ on the interval $[\underline{\beta}, \overline{\beta}]$ and extended smoothly to all of $\mathbb{R}^+$ by a strictly monotonically decreasing C^2 function with a uniformly bounded second derivative, in such a way that C^2 smoothness over the endpoints $\underline{\beta}, \overline{\beta}$ is preserved. This extension beyond the interval $[\underline{\beta}, \overline{\beta}]$ is only a theoretical construction, since when choosing the interval $[\underline{\beta}, \overline{\beta}]$ sufficiently large so that it contains β^0 and β_*, the actual iterates will stay in this interval according to lemma 2. Concerning the existence of an exact solution β_* to (15) with this choice of i, we refer to [18].

Also, we wish to point out that the monotonicity (53) implies that we approach the exact parameter β_* from the stable side in the sense that smaller β corresponds to a more stable Tikhonov problem.

Moreover, we mention in passing that if we only aim at achieving (55) (as it is the case in our application of this result to the computation of the regularization parameter), then we will stop the iteration at k_* and set $\widehat{\beta} := \beta_{k_*}$. Hence, (52) does not get active but is only a theoretical bound for proving finiteness of the stopping index k_*.

By means of the results of proposition 4, we will prove that the shown convergence $\beta^k \to \beta_*$ is in fact linear.

Proposition 5. *Let the assumptions of proposition 4 hold.*

If additionally $2c_1 + c_2 < 1$, *we get linear convergence, i.e. for all* $C \in (\frac{c_1+c_2}{1-c_1}, 1)$ *there exists* $k_0 \in \mathbb{N}$ *such that for all* $k \geqslant k_0$

$$|\beta^{k+1} - \beta_*| \leqslant C|\beta^k - \beta_*|. \tag{64}$$

Note that we here also recover the superlinear convergence stated in the case of exact evaluation $c_1 = c_2 = 0$ in [23] with only a lower bound on i''.

Proof. With the aid of (57), we can show the linear convergence result (64) as follows:

$$\begin{aligned}|\beta^{k+1}-\beta_*| &= \beta_*-\beta^{k+1}\\ &= \frac{1}{-i'^k_h}\left(\sigma^{k+1}\left(\tau^2\delta^2-i^k_h-i'^k_h(\beta_*-\beta^k)\right)+(1-\sigma^{k+1})\left(-i'^k_h\right)(\beta_*-\beta^k)\right)\\ &= \frac{\sigma^{k+1}}{|i'^k_h|}\left(\left(\left(i'\left(\overline{\beta}^k_1\right)-i'(\beta^k)\right)(\beta_*-\beta^k)\right.\right.\\ &\quad \left.+\left(i(\beta_k)-i^k_h\right)+\left(i'(\beta_k)-i'^k_h\right)\right)(\beta_*-\beta^k))\\ &\quad +(1-\sigma^{k+1})(\beta_*-\beta^k)\end{aligned} \tag{65}$$

for some $\overline{\beta}^k_1\in[\beta^k,\beta_*]$, where we have used (61).

With (48), (49), (56) and $\sigma^{k+1}\in(0,1]$ we can further deduce

$$\frac{|\beta^{k+1}-\beta_*|}{|\beta^k-\beta_*|}\leqslant\frac{c_1+c_2}{1-c_1}+\frac{1}{1-c_1}\frac{|i'(\overline{\beta}^k_1)-i'(\beta^k)|}{a}+|1-\sigma^{k+1}|, \tag{66}$$

where we have used the estimate

$$\left|\mathrm{i}(\beta_k)-i^k_h\right|\leqslant\frac{c_1}{1-c_1}\left(\left|i'^k_h\right||\beta_*-\beta^k|+\left|i'\left(\overline{\beta}^k_1\right)-i'(\beta^k)\right||\beta_*-\beta^k|+\left|i'(\beta_k)-i'^k_h\right)\right||\beta_*-\beta^k|)$$

following from

$$\begin{aligned}\left|\mathrm{i}(\beta_k)-i^k_h\right| &\leqslant c_1\left(\left|i'^k_h\right||\beta_*-\beta^k|+\left|\tau^2\delta^2-i^k_h-i'^k_h(\beta_*-\beta^k)\right|\right)\\ &\leqslant c_1\left(\left|i'^k_h\right||\beta_*-\beta^k|+\left|i'\left(\overline{\beta}^k_1\right)-i'(\beta^k)\right||\beta_*-\beta^k|+\left|\mathrm{i}(\beta_k)-i^k_h\right|\right)\\ &\quad +c_1\left(\left|i'(\beta_k)-i'^k_h\right||\beta_*-\beta^k|\right).\end{aligned}$$

As $|i'(\beta^k)-i'(\overline{\beta}^k_1)|\to 0$ and $(1-\sigma^{k+1})\to 0$ for $k\to\infty$, for any $C\in(\frac{c_1+c_2}{1-c_1},1)$, there exists $k_0\in\mathbb{N}$, such that the right-hand side of (66) is smaller than C for all $k\geqslant k_0$, which shows the linear convergence result (64). □

Under additional quadratic conditions on the errors in i and i', as well as the assumption that i'' is bounded from above, there even holds quadratic convergence $\beta^k\to\beta_*$, which we will show in terms of the following proposition.

Proposition 6. *Let the assumptions of proposition 4 be satisfied.*

If additionally $i''(\beta)\leqslant\overline{\gamma}$ for some $\overline{\gamma}\in\mathbb{R}$ independent of β, and

$$\left|\mathrm{i}(\beta^k)-i^k_h\right|\leqslant C_1\frac{\left|i^k_h-\tau^2\delta^2\right|^2}{\left|i'^k_h\right|^2} \tag{67}$$

$$\left|i'(\beta^k)-i'^k_h\right|\leqslant C_2\frac{\left|i^k_h-\tau^2\delta^2\right|}{\left|i'^k_h\right|} \tag{68}$$

hold, then we obtain the quadratic convergence estimate

$$|\beta^{k+1}-\beta_*|\leqslant\frac{1}{\left|i'^k_h\right|}\left(\frac{\max\{0,\overline{\gamma}\}}{2}+\frac{C_1}{(1-c_1)^2}\frac{|i'(\overline{\beta}^k)|^2}{|i'^k_h|^2}+\frac{2C_2+\gamma}{2(1-c_1)}\frac{|i'(\overline{\beta}^k)|}{|i'^k_h|}\right)|\beta^k-\beta_*|^2 \tag{69}$$

for some $\overline{\beta}^k\in[\beta^k,\beta_*]$.

Proof. The quadratic convergence estimate (69) can be concluded by continuing the estimate (65) as follows:

$$
\begin{aligned}
|\beta^{k+1} - \beta_*| = \frac{\sigma^{k+1}}{|i'^k_h|} \bigg(\int_0^1 (1-t) i''(\beta_* + t(\beta_k - \beta_*))(\beta_* - \beta^k)\,\mathrm{d}t + \big(\mathrm{i}(\beta_k) - i^k_h\big) \\
+ \big(i'(\beta_k) - i'^k_h\big))(\beta_* - \beta^k) \bigg) + \left(1 - \frac{2}{1 + \sqrt{1 - 2\gamma s_N^{k+1}/i'^k_h}} \right) (\beta_* - \beta^k) \\
\leqslant \frac{1}{|i'^k_h|} \bigg(\frac{\max\{0, \overline{\gamma}\}}{2} |\beta^k - \beta_*|^2 + |\mathrm{i}(\beta^k) - i^k_h| + |i'(\beta^k) - i'^k_h| |\beta^k - \beta_*| \\
+ \frac{\gamma}{2} \frac{|i^k_h - \tau^2\delta^2|}{|i'^k_h|} |\beta^k - \beta_*| \bigg) \\
\leqslant \frac{1}{|i'^k_h|} \bigg(\frac{\max\{0, \overline{\gamma}\}}{2} |\beta^k - \beta_*|^2 + C_1 \frac{|i^k_h - \tau^2\delta^2|^2}{|i'^k_h|^2} \\
+ \left(C_2 + \frac{\gamma}{2} \right) \frac{|i^k_h - \tau^2\delta^2|}{|i'^k_h|} |\beta^k - \beta_*| \bigg),
\end{aligned}
$$

where we have used (67) and (68). Using the estimate

$$
|i^k_h - \tau^2\delta^2| \leqslant \frac{1}{1-c_1} |\mathrm{i}(\beta^k) - \tau^2\delta^2| \leqslant \frac{1}{1-c_1} |i'(\overline{\beta}^k)| |\beta^k - \beta_*|,
$$

we arrive at (69). □

Note that a discrete version $i''_h(\beta)$ of $i''(\beta)$ and therewith *a posteriori* estimates for the lower and upper bounds on $i''_h \approx i''(\beta)$ can be computed with low numerical effort, once the error estimator for i' has been evaluated (see section 3 and [6]).

In case the upper and lower bounds on i'' are not known (e.g. if the problem is too nonlinear to satisfy (41) and therefore lemma 2 cannot be applied,) bisection still provides a globally and *R*-linearly convergent method for obtaining β such that (55) holds.

To summarize the results from this section, for obtaining convergence of q^δ_β according to propositions 1 and 3 and linear (and quadratic) convergence of the sequence of β's produced by the proposed inexact Newton method according to proposition 5 (and proposition 6), we check if (28), (37), (48), (49), (50), (51) (and (67) and (68)) hold in our search for the correct regularization parameter and refine the discretization according to the corresponding error estimator if one of these conditions is violated. Note that these bounds allow for a rather rough approximation as long as the regularization parameter β^k under consideration is still 'far away' from the actual one in the sense that $\mathrm{i}(\beta^k) - \tau^2\delta^2$ is not small yet.

We also wish to point out that conditions (28), (37), (48), (49), (50), (51), (67) and (68) are tailored to the use with goal-oriented error estimators (see the following section). Indeed, adaptive refinement according to these estimators allows us to enforce error bounds on terms of the type $I(q, F(q)) - I(q_h, F_h(q_h))$, whereas terms of the type $I(q, F(q)) - I(q_h, F(q_h))$ or $I(q, F(q)) - I(q, F_h(q))$ would not be tractable with this technique.

3. Goal-oriented error estimators

For evaluating the error estimates needed to determine β according to section 2, we employ the goal-oriented error estimators proposed in [2] and [3] and make use of the results from [6], where concrete error estimators for j, i and i' and computation formulas for i' and i'' are derived. To simplify exposition, we set $\mathcal{D} = Q$ here; the general case $\mathcal{D} \subset Q$ follows in a straightforward manner.

We begin with an error estimator for j, which is needed to evaluate (37). According to proposition 1 in [6] for continuous and discrete stationary points $(q, u) \in Q \times V$ and $(q_h.u_h) \in Q_h \times V_h$ of $\mathcal{L}$ and $\mathcal{L}_h$, respectively, an error representation is given by

$$j(q) - j_h(q_h) = J(q, u) - J(q_h, u_h) = \tfrac{1}{2}\mathcal{L}'(q_h, u_h, z_h)(q - \tilde{q}_h, u - \tilde{u}_h, z - \tilde{z}_h) + R_1 \tag{70}$$

for arbitrary $(\tilde{q}_h, \tilde{u}_h, \tilde{z}_h) \in Q_h \times V_h \times V_h$, where R_1 is a third-order remainder term. We skipped the index β from (10) and (12), since the representation holds for all β.

For the purpose of formulating an error estimator for i (needed in (21), (48), (50), (51), (67) and (28)), we define the auxiliary functional

$$\mathcal{M} : X^2 \to \mathbb{R}, \quad \mathcal{M}(x, x_1) = I(q, u) + \mathcal{L}'(x)(x_1),$$

where

$$X = Q \times V \times V, \qquad x = (q, u, z), \qquad x_1 = (q_1, u_1, z_1).$$

Let $x = (q, u, z)$ and $x_h = (q_h, u_h, z_h)$ solve (9) and (11), respectively. According to proposition 3 in [6], we get continuous and stationary points (x, x_1) and $(x_h, x_{h,1})$ of $\mathcal{M}$ by solving

$$\begin{aligned}
&\mathcal{L}''(x)(\mathrm{d}x, x_1) = -I'(u)(\mathrm{d}u) \qquad \forall\, \mathrm{d}x = (\mathrm{d}q,\ \mathrm{d}u,\ \mathrm{d}z) \in X,\\
&\mathcal{L}''(x_h)(\mathrm{d}x_h, x_{h,1}) = -I'(u_h)(\mathrm{d}u_h) \qquad \forall\, \mathrm{d}x_h = (\mathrm{d}q_h,\ \mathrm{d}u_h,\ \mathrm{d}z_h) \in X_h.
\end{aligned} \tag{71}$$

For such stationary points, the error representation

$$\rho = I(q, u) - I(q_h, u_h) = \mathrm{i}(\beta) - i_h(\beta) = \tfrac{1}{2}\mathcal{M}'(x_h)(x - \tilde{x}_h) + R_2 \tag{72}$$

holds for arbitrary $\tilde{x}_h \in X_h$, where R_2 is a third order remainder term.

By means of (x, x_1) and $(x_h, x_{h,1})$ the continuous and discrete version of the first derivative i' can be evaluated by

$$i'(\beta) = \frac{2}{\beta^2}(q, q_1)_Q \qquad \text{and} \qquad i_h'(\beta) = \frac{2}{\beta^2}(q_h, q_{h,1})_Q, \tag{73}$$

see proposition 4 in [6].

We define an additional auxiliary functional in order to express an error estimator for i' (needed in (49), (50), (68)):

$$\mathcal{N} : X^4 \to \mathbb{R}, \qquad \mathcal{N}(x, x_1, x_2, x_3) = K(q, q_1) + \mathcal{M}_x'(x, x_1)(x_2) + \mathcal{M}_{x_1}'(x, x_1)(x_3),$$

where

$$x_1 = (q_1, u_1, z_1), \qquad x_2 = (q_2, u_2, z_2), \qquad x_3 = (q_3, u_3, z_3)$$

and

$$K(q, q_1) = i'(\beta) = -\frac{2}{\beta^2}(q, q_1)_Q.$$

According to proposition 6 in [6], we obtain continuous and stationary points $(x, x_1, x_2, x_3) \in X^4$ and $(x_h, x_{h,1}, x_{h,2}, x_{h,3}) \in X_h^4$ of $\mathcal{N}$ by solving

$$\begin{aligned}
&\mathcal{L}''(x)(x_2,\ \mathrm{d}x_1) = -K_{q_1}'(q, q_1)(\mathrm{d}q_1) \qquad \forall\, \mathrm{d}x_1 \in X,\\
&\mathcal{L}''(x)(x_3,\ \mathrm{d}x) = -K_q'(q, q_1)(\mathrm{d}q) - I_{uu}''(q, u)(u_2,\ \mathrm{d}u) - \mathcal{L}'''(x)(x_1, x_2,\ \mathrm{d}x) \qquad \forall\, \mathrm{d}x \in X
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{L}''(x_h)(x_{h,2},\ \mathrm{d}x_{h,1}) = -K_{q_{h,1}}'(q_h, q_{h,1})(\mathrm{d}q_{h,1}) \qquad \forall\, \mathrm{d}x_{h,1} \in X_h,\\
&\mathcal{L}''(x_h)(x_{h,3},\ \mathrm{d}x_h) = -K_q'(q_h, q_{h,1})(\mathrm{d}q_h) - I_{uu}''(q_h, u_h)(u_{h.2},\ \mathrm{d}u_h) - \mathcal{L}'''(x_h)(x_{h,1}, x_{h,2},\ \mathrm{d}x_h)\\
&\qquad \forall\, \mathrm{d}x_h \in X_h.
\end{aligned} \tag{74}$$

For these stationary points, there holds the error representation

$$K(q, q_1) - K(q_h, q_{h,1}) = i'(\beta) - i'_h(\beta) = \tfrac{1}{2}\mathcal{N}'_y(y_h, \widehat{y}_h)(y - \tilde{y}_h) + \tfrac{1}{2}\mathcal{N}'_{\widehat{y}}(y_h, \widehat{y}_h)(\widehat{y} - \overline{y}_h) + R_3 \tag{75}$$

for arbitrary $\tilde{y}_h, \overline{y}_h \in X_h^2$, where R_3 is a third order remainder term, cf proposition 5 in [6].

Finally, the second derivative i'' and its discrete equivalent can be evaluated by

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

and

$$i''_h(\beta) = \frac{4}{\beta^3}(q_h, q_{h,1})_Q - \frac{2}{\beta^2}(q_{h,2}, q_{h,1})_Q - \frac{2}{\beta^2}(q_h, q_{h,3})_Q, \tag{76}$$

see proposition 7 in [6].

When using spaces Q_h and V_h with locally supported basis functions, the error estimators

$$\rho = \tfrac{1}{2}\mathcal{M}'(x_h)(e_h) \qquad \text{and} \qquad \eta = \tfrac{1}{2}\mathcal{N}'_y(y_h, \widehat{y}_h)(\widehat{e}_h) + \tfrac{1}{2}\mathcal{N}'_{\widehat{y}}(y_h, \widehat{y}_h)(\overline{e}_h)$$

(where e_h, $\widehat{e}_h$ and $\overline{e}_h$ are approximations of the interpolation errors $x - \tilde{x}_h$, $y - \tilde{y}_h$ and $\widehat{y} - \overline{y}_h$ obtained by local averaging or higher order approximations, cf, e.g., [20]) can be written as a sum of its local contributions, which enables us to implement a local refinement strategy based on the estimator, cf [2]. Additionally to that, each local error can be decomposed into its components due to the discretization of Q on the one hand and of V on the other hand. Therewith, the proposed method for determining β could also be applied when using different discretizations of Q and V. Then, in each iteration step, it can be decided whether to refine Q_h or V_h (or both), if necessary, cf [20].

For further details on the evaluation of the error estimators, we refer to [6].

Summarizing the results from this and the previous section, the concrete algorithm for determining the regularization parameter β according to propositions 4 and 5 appears as follows [6].

At this point, we assume that $\underline{\gamma}$ and γ are given such that $\underline{\gamma} \leqslant i''(\beta)$ for all $\beta > 0$ and $\gamma > \underline{\gamma}$. In section 4, we will propose a specific alternative for the choice of these parameters.

Algorithm 1. *Inexact Newton method for determining a regularization parameter for nonlinear inverse problems.*

1: *Choose* $c_{tc} < 1$, $\overline{\overline{\tau}} > \tau > \underline{\underline{\tau}} > 1$, $\tilde{\tau} < \underline{\underline{\tau}}$
such that (38) *is fulfilled*, $c_1 \in (0, 1)$, $c_2 \in (0, \frac{1}{2})$ *such that* $2c_1 + c_2 < 1$ *and* $c \leqslant \frac{\tilde{\tau}^2}{\overline{\overline{\tau}}^2}$.
2: *Choose initial guess* $\beta^0 > 0$, *initial discretizations* Q_{h_0}, V_{h_0}, *set* $k = 0$.
3: ***repeat***
4: *Compute* $x_{h_k} = (q_{h_k}, u_{h_k}, z_{h_k})$ *by solving* (11).
5: *Evaluate* $i_{h_k}(\beta^k)$ *by* (14).
6: *Compute* $x_{h_k,1} = (q_{h_k,1}, u_{h_k,1}, z_{h_k,1})$ *by solving* (71).
7: *Evaluate* $i'_{h_k}(\beta^k)$ *by* (73).
8: *Evaluate the error estimator* ρ (*for* $i(\beta^k)$) *according to* (72).
9: *Compute* $x_{h_k,2}$ *and* $x_{h_k,3}$ *by solving* (74).
10: *Evaluate the error estimator* η (*for* $i'(\beta^k)$) *according to* (75).
11: ***if*** (48), (49), (50), (51) or (28) *is violated* ***then***
12: *Refine with respect to the corresponding error estimator.*
13: ***else***
14: *Set* $\beta^{k+1} = \beta^k + s^{k+1}$ *according to* (47) *and* $k = k + 1$.
15: ***until*** $i_h^k \leqslant \overline{\overline{\tau}}^2\delta^2$

16: Set $\widehat{\beta} := \beta_k$.
17: Choose $\Theta > 0$ *such that* (38) *is fulfilled.*
18: Compute $x_{h_k} = (q_{h_k}, u_{h_k}, z_{h_k})$ *by solving* (11).
19: ***while*** (37) *is violated* ***do***
20: *Refine with respect to the error estimator for j (see* (37) and (70)*).*
21: *Compute* $x_{h_k} = (q_{h_k}, u_{h_k}, z_{h_k})$ *by solving (11).*

4. Numerical experiments

For illustrating the performance of the presented method, we consider the following parameter identification problem: identify the source term q in the nonlinear elliptic PDE

$$\begin{aligned} -\Delta u + \zeta u^3 &= q \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{77}$$

from measurements of the state u in Ω, where Ω is a smooth or polygonal and convex domain in $\mathbb{R}^d$, $d \in \{1, 2, 3\}$ and $\zeta > 0$ is a constant.

Due to the monotonicity of the nonlinear term, there exists a unique solution $u \in H_0^1(\Omega)$ for all $q \in L^2(\Omega)$. From this and the energy estimate

$$\|\nabla u\|^2_{L_2(\Omega)} + \zeta \|u\|^4_{L_4(\Omega)} \leqslant \|q\|_{H^{-1}(\Omega)} \|u\|_{H_0^1(\Omega)} \tag{78}$$

following by the use of u as a test function in the weak form of (77), it gets obvious that the forward operator $F : q \mapsto u$ is well defined and bounded as a mapping from $L_2(\Omega)$ to $H_0^1(\Omega)$. Taking L_2 norms and using the Cauchy–Schwarz inequality in the strong form of the PDE (77), we even get

$$\begin{aligned} \|\Delta u\|_{L_2(\Omega)} \leqslant \zeta \|u\|^3_{L_6(\Omega)} + \|q\|_{L_2(\Omega)} \leqslant \zeta C^3 \|\nabla u\|^3_{L_2(\Omega)} \\ + \|q\|_{L_2(\Omega)} \leqslant \zeta C^3 \|q\|^3_{H^{-1}(\Omega)} + \|q\|_{L_2(\Omega)}, \end{aligned}$$

where we have applied Sobolev's embedding theorem $\|y\|_{L_6(\Omega)} \leqslant C\|\nabla y\|_{L_2(\Omega)}$ and used the norm $\|u\|_{H_0^1(\Omega)} = \|\nabla u\|_{L_2(\Omega)}$, which is justified by Friedrich's inequality $\|y\|_{L_2(\Omega)} \leqslant C_F \|\nabla y\|_{L_2(\Omega)}$. By the assumed regularity of the domain we then obtain that F is even bounded as a mapping from $L_2(\Omega)$ to $H^2(\Omega) \cap H_0^1(\Omega)$. However, since only function values and not derivative values can be measured, and in order to remain in a Hilbert space setting, we use the pre-image and image spaces

$$Q = L_2(\Omega), \qquad G = L_2(\Omega).$$

The considerations above also show that for this particular example we can use $\mathcal{D} = Q$, whereas in general coefficient inverse problems for PDEs, the domain of definition will typically be a strict subset of the underlying linear space.

To verify conditions (17), (18), (19) and (41) on F in the case of measurements in all of Ω (see (ii) below), first of all observe that by the above-mentioned boundedness of $F : L_2(\Omega) \to H^2(\Omega) \cap H_0^1(\Omega)$ and the Rellich–Kondrachov theorem, $F : L_2(\Omega) \to H_0^1(\Omega)$ is in fact compact, which implies (17) and (18) as follows: for any sequence $(q_n)_{n\in\mathbb{N}} \subseteq \mathcal{D}$ with $q_n \rightharpoonup q$ in $L_2(\Omega)$ and $u_n := F(q_n) \rightharpoonup g$ in $L_2(\Omega)$, we have that even $u_n \to g$ in $H^1(\Omega)$ and $u_n^3 \to g^3$ in $L_2(\Omega)$. Therewith, $g \in H_0^1(\Omega)$ and for any $v \in H_0^1(\Omega)$ we have

$$\int_\Omega \{\nabla g \cdot \nabla v + \zeta g^3 v - qv\}\,\mathrm{d}x = \lim_{n\to\infty} \int_\Omega \{\nabla u_n \cdot \nabla v + \zeta u_n^3 v - q_n v\} = 0,$$

i.e. $F(q) = g$.

With the abbreviations $u = F(q)$, $\tilde{u} = F(\bar{q})$, $w = F'(q)(\bar{q} - q)$, we obtain condition (19) from the fact that w can be written as the solution of

$$-\Delta w + 3\zeta u^2 w = \bar{q} - q \quad \text{in } \Omega$$
$$w = 0 \quad \text{on } \partial\Omega$$

and $z = F(\bar{q}) - F(q) - F'(q)(\bar{q} - q) = \tilde{u} - u - w$ solves

$$-\Delta z + 3\zeta u^2 z = -\zeta(\tilde{u} + 2u)(\tilde{u} - u)^2 \quad \text{in } \Omega$$
$$w = 0 \quad \text{on } \partial\Omega.$$

Hence, by testing with z we get by Hölder's inequality and Sobolev's embedding theorem

$$\begin{aligned}\|\nabla z\|^2_{L_2(\Omega)} + 3\zeta\|uz\|^2_{L_2(\Omega)} &= -\zeta\int_\Omega(\tilde{u} + 2u)(\tilde{u} - u)^2 z\,\mathrm{d}x\\ &\leqslant \zeta\|\tilde{u} + 2u\|_{L_6(\Omega)}\|\tilde{u} - u\|_{L_6(\Omega)}\|z\|_{L_6(\Omega)}\|\tilde{u} - u\|_{L_2(\Omega)}\\ &\leqslant \zeta C^3\|\nabla(\tilde{u} + 2u)\|_{L_2(\Omega)}\|\nabla(\tilde{u} - u)\|_{L_2(\Omega)}\|\nabla z\|_{L_2(\Omega)}\|\tilde{u} - u\|_{L_2(\Omega)}.\end{aligned} \tag{79}$$

To estimate the term $\|\nabla(\tilde{u} - u)\|_{L_2(\Omega)}$ on the right-hand side, we use the fact that $\tilde{u} - u$ solves

$$-\Delta(\tilde{u} - u) + \tfrac{1}{2}\zeta\{\tilde{u}^2 + u^2 + (\tilde{u} + u)^2\}(\tilde{u} - u) = \bar{q} - q \quad \text{in } \Omega$$
$$\tilde{u} - u = 0 \quad \text{on } \partial\Omega;$$

hence,

$$\|\nabla(\tilde{u} - u)\|^2_{L_2(\Omega)} + \tfrac{1}{2}\zeta\|\sqrt{\tilde{u}^2 + u^2 + (\tilde{u} + u)^2}(\tilde{u} - u)\|^2_{L_2(\Omega)} \leqslant \|\nabla(\tilde{u} - u)\|_{L_2(\Omega)}\|\bar{q} - q\|_{H^{-1}(\Omega)}.$$

Inserting this into (79) and using (78), we obtain

$$\begin{aligned}&\|F(\bar{q}) - F(q) - F'(q)(\bar{q} - q)\|_G\\ &\quad\leqslant \zeta\tilde{C}(\|\bar{q}\|_{H^{-1}(\Omega)} + 2\|q\|_{H^{-1}(\Omega)})\|\bar{q} - q\|_{H^{-1}(\Omega)}\|F(\bar{q}) - F(q)\|_G\end{aligned}$$

which implies (19) with small c_{tc} upon restriction of the domain $\mathcal{D}$ to a sufficiently small H^{-1} neighborhood of a solution $q^\dagger$.

Similarly, one obtains (41) from the fact that $y = F''(q)(v, v)$ solves

$$-\Delta y + 3\zeta u^2 y = -6\zeta u w^2 \quad \text{in } \Omega$$
$$w = 0 \quad \text{on } \partial\Omega,$$

where again we use abbreviations $u = F(q)$, $w = F'(q)v$. Testing with $\hat{y} = (-\Delta + 3\zeta u^2\mathrm{id})^{-1}y$ (with homogeneous Dirichlet boundary conditions), we get

$$\|y\|^2_{L_2(\Omega)} = -6\zeta\int_\Omega uw^2\hat{y}\,\mathrm{d}x,$$

where analogously to the H_0^1 and the H^2 estimate for u we get

$$\|\nabla\hat{y}\|^2_{L_2(\Omega)} + 3\zeta\|u\hat{y}\|^2_{L_2(\Omega)} \leqslant \|y\|_{H^{-1}(\Omega)}\|\hat{y}\|_{H_0^1(\Omega)} \leqslant C_F\|y\|_{L_2(\Omega)}\|\hat{y}\|_{H_0^1(\Omega)}$$

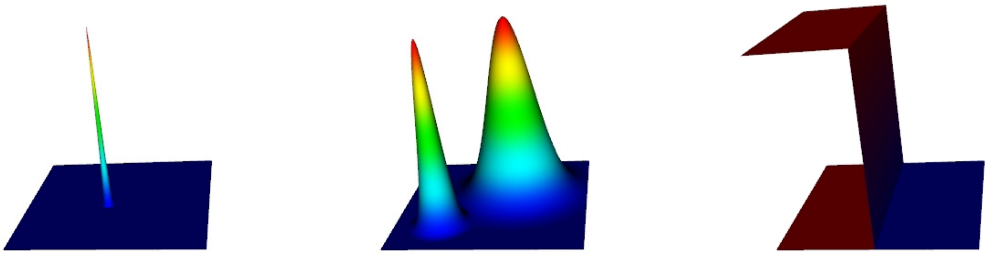
$$\begin{aligned}\|\Delta\hat{y}\|_{L_2(\Omega)} &\leqslant 3\zeta\|u^2\hat{y}\|_{L_2(\Omega)} + \|y\|_{L_2(\Omega)} \leqslant 3\zeta C^3\|\nabla u\|^2_{L_2(\Omega)}\|\nabla\hat{y}\|_{L_2(\Omega)} + \|y\|_{L_2(\Omega)}\\ &\leqslant 3\zeta C^3\|q\|^2_{H^{-1}(\Omega)}\|y\|_{H^{-1}(\Omega)} + \|y\|_{L_2(\Omega)}.\end{aligned}$$

Hence, with $\hat{C}$ the norm of $\Delta^{-1} : L_2(\Omega) \to L_\infty(\Omega)$ altogether:

$$\begin{aligned}\|y\|^2_{L_2(\Omega)} &\leqslant 6\zeta\hat{C}^3\|\Delta u\|^2_{L_2(\Omega)}\|\Delta\hat{y}\|_{L_2(\Omega)}\|w\|^2_{L_2(\Omega)}\\ &\leqslant 6\zeta\hat{C}^3(\zeta C^3\|q\|^3_{H^{-1}(\Omega)} + \|q\|_{L_2(\Omega)})^2(3\zeta C^3\|q\|^2_{H^{-1}(\Omega)}C_F + 1)\|y\|_{L_2(\Omega)}\|w\|^2_{L_2(\Omega)},\end{aligned}$$

i.e. (41), provided q is sufficiently close to $q^\dagger$ so that

$$\left\|F(q) - g^\delta\right\| 6\zeta\hat{C}^2(\zeta C^3\|q\|^3_{H^{-1}(\Omega)} + \|q\|_{L_2(\Omega)})^2(3\zeta C^3\|q\|^2_{H^{-1}(\Omega)}C_F + 1) \leqslant 1.$$



(a) (b) (c)

Figure 1. Exact control $q^\dagger$ for examples (a), (b) and (c).

In our numerical computations, we use the unit square $\Omega = (0, 1)^2 \subset \mathbb{R}^2$ as a domain, and in order to amplify the nonlinearity of this example, we set $\zeta = 1000$.

We consider configurations with three different exact sources $q^\dagger$.

(a) A Gaussian distribution

$$q^\dagger = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\left(\frac{sx-\mu}{\sigma}\right)^2 + \left(\frac{sy-\mu}{\sigma}\right)^2\right)\right),$$

where $\sigma = 0.01$, $\mu = \frac{5}{11}$, $s = 1$.

(b) Two Gaussian distributions added up to one distribution

$$q^\dagger = q_1 + q_2,$$

where

$$q_1 = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\left(\frac{s_1x-\mu}{\sigma}\right)^2 + \left(\frac{s_1y-\mu}{\sigma}\right)^2\right)\right),$$

$$q_2 = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\left(\frac{s_2x-\mu}{\sigma}\right)^2 + \left(\frac{s_2y-\mu}{\sigma}\right)^2\right)\right)$$

with $\sigma = 0.1$, $\mu = 0.5$, $s_1 = 2$, $s_2 = 0.8$.

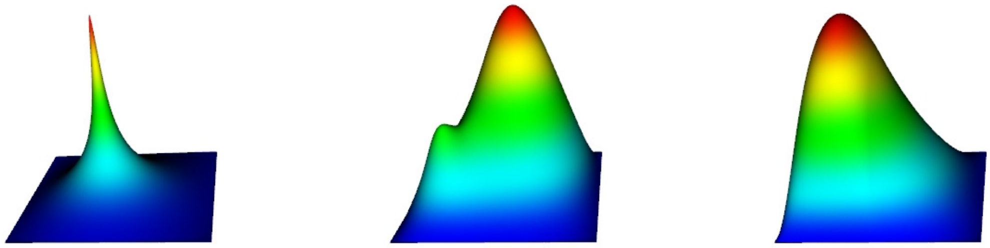
(c) The step function

$$q^\dagger = \begin{cases} 0 & \text{for} \quad x \geqslant \frac{1}{2} \\ 1 & \text{for} \quad x < \frac{1}{2}. \end{cases}$$

Figures 1 and 2 show the exact source distribution $q^\dagger$ and the corresponding state $u^\dagger = S(q^\dagger)$ computed on a very fine grid with 1050 625 nodes and equally sized quadratic cells.

The measurements were simulated in two different ways.

(i) As in [6] via point functionals in $n_m = 100$ uniformly distributed points $\{\xi_i\} \subset \Omega$ and perturbed by uniformly distributed random noise at different percentages. The observation space and the observation operator are chosen as $G = \mathbb{R}^{n_m}$ and $C : V \to G$, $(C(v))_i = v(\xi_i)$ for $i = 1, \ldots, n_m$.



(a) (b) (c)

Figure 2. Exact state $u^\dagger$ for examples (a), (b) and (c).

(ii) Via L^2-projection: then $G = L^2(\Omega)$, $C = \mathrm{i}d$ and

$$g^\delta = g + \delta \frac{r}{\|r\|_{L^2(\Omega)}} = g + p\,\|g\|_{L^2(\Omega)} \frac{r}{\|r\|_{L^2(\Omega)}},$$

where r denotes some uniformly distributed random noise and p the percentage of perturbation. $F(q^\dagger) = S(q^\dagger) = u^\dagger$ and r are generated on a very fine grid with 1050 625 nodes and we denote the corresponding finite element space by V_L. In order to evaluate $\left\|F(q) - g^\delta\right\|^2_{L^2(\Omega)}$ on coarser grids and the corresponding finite element spaces V_l with $l = 0, 1, \ldots, L$ during the optimization algorithm, g^δ has to be transferred from V_L to the current grid V_l. As usual in the finite element context, this is done by the L^2 projection as the restriction operator.

The concrete choice of the parameters for the numerical tests is as follows: $\tilde{\tau} = 0.1$, $\underline{\underline{\tau}} = 3.1$, $\tau = 3.2$, $\overline{\overline{\tau}} = 3.3$, $c_1 = 0.3$, $c_2 = 0.3$. $c_{tc} = 10^{-7}$ and

$$\Theta^2 := \frac{\left(\sqrt{\underline{\underline{\tau}}^2 - \tilde{\tau}^2}(1 - c_{tc}) - 1 - c_{tc}\right)^2 - (1 + c_{tc})^2}{2(1 - c_{tc})} - 1 - \tilde{\tau}.$$

Note that in order to fulfill (38), $\underline{\underline{\tau}}$ and $\tilde{\tau}$ have to be chosen such that $\sqrt{\underline{\underline{\tau}}^2 - \tilde{\tau}^2} > 1 + \sqrt{3}$.

As the theoretical assumption $\underline{\gamma} \leqslant i''(\beta)$ for all $\beta > 0$ is not exactly transferable to practice, we decided to choose $\underline{\gamma}$ as follows:

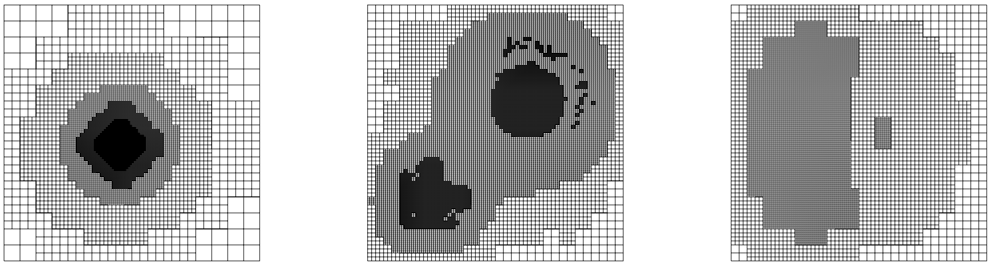
$$\underline{\gamma} = \underline{\gamma}_k := \min\left\{0, i''^k_h\right\}, \qquad \gamma = \gamma_k := \tfrac{1}{3}\left|i''^k_h\right|,$$

where i''^k_h is a discrete version of $i''(\beta^k)$ that can be computed by (76).

Note that only the existence of an upper bound $\overline{\gamma}$ but not its explicit value is needed in order to compute the regularization parameter according to section 2.

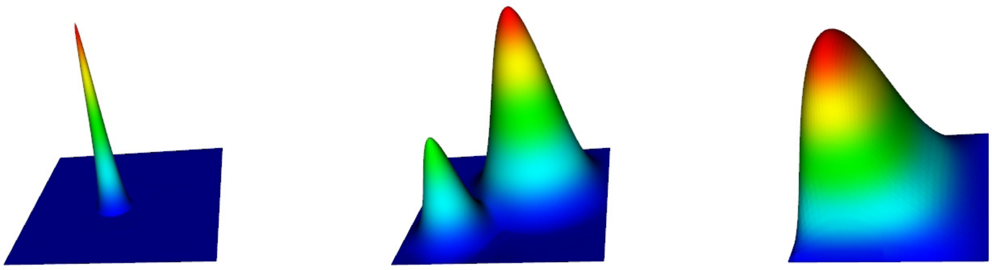
As for the tested examples, the numerical results of the version of the algorithm for quadratic convergence (cf proposition 6) did not differ significantly from the one given above for linear convergence, we will restrict ourselves to the latter case. That means all presented numerical results refer to algorithm 1, which guarantees (only) linear convergence of the produced sequence of β's, but uses less restrictive accuracy requirements, thus allowing for coarser grids than the quadratically convergent version.

In figure 3, one can see the adaptively refined meshes for examples (a), (b) and (c). It can be clearly seen that the algorithm automatically concentrates refinement to regions of large changes. The corresponding reconstructed source distributions and states for $\delta = 1\%\|g\|_G$ are shown in figures 4 and 5, where we used the approach (ii). In order to be able to compare the numerical results of this paper with those from [6], we decided to consider example (c) with



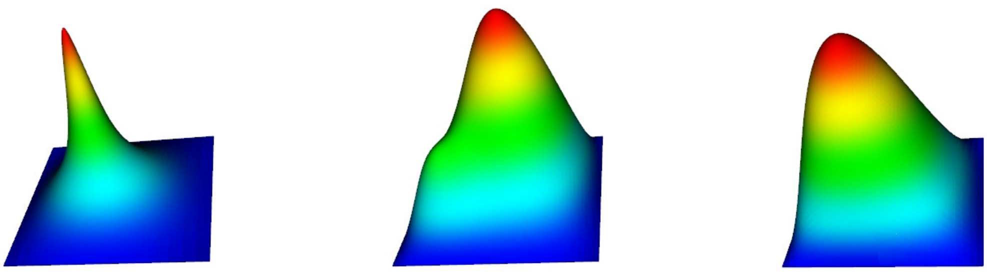
(a) (b) (c)

Figure 3. Adaptively refined meshes for examples (a), (b) and (c), 1% noise, via L^2 projection.



(a) (b) (c)

Figure 4. Reconstructed q for examples (a), (b) and (c), 1% noise, via L^2 projection.



(a) (b) (c)

Figure 5. Reconstructed u for examples (a), (b) and (c), 1% noise, via L^2 projection.

point measurements as well, see (i). The reconstructed control and state and the adaptively refined mesh are shown in figure 6. It can be seen very well that the refinement is concentrated to the location of the measurement points.

As well as in [6], none of the tested examples needed further refinement after determination of the regularization parameter, so steps 19 to 21 did not take effect in the tested examples.

Tables 1, 2 and 3 show the proposed adaptive refinement strategy versus uniform refinement for $\delta = 1\% \, \|g\|_G$ with measurements according to (ii) via L^2 projection. For both cases, we used the same error estimators (see section 3 and [6]). By applying adaptive refinement, in example (a) we save more than 94% of nodes, which yields about 92% reduction of computation time with respect to global refinement. The relative error $\frac{\|q^\delta_{h,\widehat{\beta}}-q^\dagger\|_Q}{\|q^\dagger\|_Q}$ (where $\widehat{\beta}$

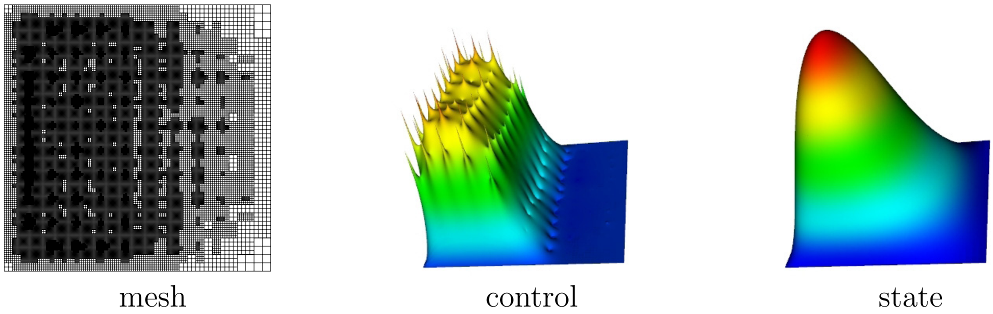


Figure 6. Adaptively refined mesh, reconstructed q and u for example (c), 1% noise, via point functional.

Table 1. Adaptive versus uniform refinement for example (a) with noise level $\delta = 1\% \, \|g\|_G$ (via L^2 projection).

	Adaptive refinement		Uniform refinement	
k	Number of nodes	β^k	Number of nodes	β^k
0	25	10	25	10
0	289	10	289	10
0	709	10	1089	10
1	709	207	1089	207
1	2117	207	4225	207
2	2117	618	4225	618
3	2117	1615	4225	1615
4	2117	4332	4225	4332
5	2117	12 325	4225	12 320
5	3657	12 325	16 641	12 320
6	3657	18 699	16 641	18 694
7	3657	47 750	16 641	47 749
8	3657	117 097	16 641	117 099
8	7601	117 097	66 049	117 099
9	7601	180 693	66 049	180 707
10	7601	420 067	66 049	420 990
11	7601	939 002	66 049	941 214
11	14 157	939 002	263 169	941 214
12	14 157	1435 629	263 169	1439 818
13	14 157	3392 480	263 169	3407 509
14	14 157	6571 883	263 169	6623 004
15	14 157	14 920 457	263 169	15 109 983
16	14 157	43 764 743	263 169	44 718 542
17	14 157	197 697 223	263 169	205 575 293

is produced by algorithm 1) for this example is about 0.116. In example (b), we save 72% of nodes, so that we save 53% of computation time, where the relative error is about 0.233. Unfortunately in example (c), we get about double the number of nodes, which leads to an augmentation of computation time of about 10%. This, however, is not surprising, since adaptivity in general is more effective for more complicated structures than the step function considered here. Moreover, an L_2 regularization term is known not to perform well for the identification of discontinuities, so the adaptive algorithm seems to try to compensate for

Table 2. Adaptive versus uniform refinement for example (b) with noise level $\delta = 1\% \, \|g\|_G$ (via L^2 projection).

	Adaptive refinement		Uniform refinement	
k	Number of nodes	β^k	Number of nodes	β^k
0	25	10	25	10
0	289	10	289	10
0	797	10	1089	10
1	797	187	1089	187
1	2749	187	4225	187
2	2749	544	4225	544
3	2749	1293	4225	1293
4	2749	2773	4225	2773
5	2749	5564	4225	5664
6	2749	10 600	4225	10 597
7	2749	19 253	4225	19 252
8	2749	33 364	4225	33 359
8	9305	33 364	16 641	33 359
9	9305	48 508	16 641	48 503
10	9305	77 676	16 641	77 675
11	9305	116 370	16 641	116 384
12	9305	160 075	16 641	160 134
12	18 035	160 075	66 049	160 134
13	18 035	191 531	66 049	192 061

Table 3. Adaptive versus uniform refinement for example (c) with noise level $\delta = 1\% \, \|g\|_G$, via L^2 projection.

	Adaptive refinement		Uniform refinement	
k	Number of nodes	β^k	Number of nodes	β^k
0	25	10	25	10
0	189	10	289	10
0	2553	10	4225	10
0	26 191	10	16 641	10
1	26 191	19	16 641	19
2	26 191	33	16 641	33
3	26 191	54	16 641	54
4	26 191	88	16 641	88
5	26 191	137	16 641	137
5	–	–	66 049	137
6	26 191	205	66 049	190
7	26 191	295	66 049	276
8	26 191	393	66 049	374
8	26 191	393	–	–
9	56 409	463	66 049	457
10	56 409	498	66 049	496

this weakness. The corresponding relative error amounts to 0.244. In table 4, we listed the produced sequence of betas for example (c) this time with point measurements according to (i). As we deal only with a small number of measurement points, expectedly, the relative error in q is larger than for the case of the L^2-projection, namely 0.421, but we save 14% of nodes and 24% of computation time with respect to global refinement. Note that the step function example is also particularly challenging due to the fact that because of the different values

Table 4. Adaptive versus uniform refinement for example (c) with noise level $\delta = 1\% \, \|g\|_G$, via point functional.

	Adaptive refinement		Uniform refinement	
k	Number of nodes	β^k	Number of nodes	β^k
0	25	10	25	10
0	81	10	81	10
0	239	10	289	10
1	239	182	289	182
1	725	182	1089	182
1	2409	182	–	–
2	2409	187	1089	187
2	–	–	4225	187
3	2409	437	4225	192
4	2409	778	4225	444
5	2409	1165	4225	787
6	2409	1466	4225	1173
6	7705	1466	–	–
7	7705	1575	4225	1470
9	–	–	4225	1579

Table 5. Adaptive versus uniform refinement for example (a) with different noise levels $\delta = p \, \|g\|_G$ (via L^2 projection).

p	$\frac{\left\|q^\delta_{h,\widehat{\beta}} - q^\dagger\right\|_Q}{\|q^\dagger\|_Q}$	$\widehat{\beta}$
8%	0.240	34 514
4%	0.228	607 056
1%	0.116	43 764 743

of initial guess and exact solution on large parts of the boundary, only a very weak source condition (or possibly none at all) holds.

In order to further measure the quality of the presented algorithm, we tested example (a) with different percentages of noise. In table 5, we can observe the same behavior as in the linear case (cf [6]): the larger the noise, the larger the relative error with respect to $q^\dagger$ and the smaller the $\widehat{\beta}$ (corresponding to a larger regularization parameter $\widehat{\alpha}$) produced by algorithm 1.

5. Conclusions and remarks

We presented an extension of the adaptive discretization strategy for parameter identification from [6] to the case of nonlinear ill-posed problems. For this purpose, we used the discrepancy principle and applied an inexact Newton method to the resulting quantity of interest. Imposing conditions on the quantity of interest and the reduced forward operator F, we provided a general rates result based on Jensen's inequality, which is not restricted to the context of this paper. In order to guarantee convergence of the produced sequence of regularization parameters, we derived refinement criteria, which we combined with the goal-oriented error estimators from [6]. In this process, the main difference to [6] consisted in compensating the lack of convexity of the cost functional and the lack of knowledge about lower and upper bounds on the first, second and third derivatives of the quantity of interest with respect to the regularization parameter. We considered not only quadratic convergence (as in [6]) but also linear convergence conditions, where we extended the idea of [23] regarding bounds on the second derivative.

The strength of the proposed algorithm lies in the ability to save a considerable amount of degrees of freedom at the beginning of the Newton-type iteration for determining the regularization parameter. Indeed, our numerical experiments show an up to 90% gain in computation time with our strategy as compared to uniform refinement.

We expect that goal-oriented error estimators can also be used in the context of different regularization parameter choice strategies such as the balancing principle [19], which is particularly of interest in the case $\nu > \frac{1}{2}$ in the Hölder-type source condition (36), where the discrepancy principle fails to provide optimal rates. Moreover, we hope that the use of the balancing principle with a sufficiently strong source condition will allow us to avoid conditions on F like the tangential cone condition, so that we can also take strongly nonlinear problems into consideration.

Moreover, we consider a combination of adaptivity with regularization by projection onto finite-dimensional spaces in place of or additionally to Tikhonov regularization, see the forthcoming paper [16].

Since iterative methods are an attractive alternative to Tikhonov regularization, also the use of adaptivity within Newton-type regularization methods is the subject of this research [15].

Acknowledgments

The authors would like to thank the German Science Foundation (DFG) for financial support within the grant KA 1778/5-1 and VE 368/2-1 'Adaptive Discretization Methods for the Regularization of Inverse Problems'. Moreover, AK gratefully acknowledges support by the TUM Graduate School's Thematic Graduate Center/Faculty Graduate Center at Technische Universität München, Germany. Finally, we thank both referees for their fruitful suggestions and remarks.

References

[1] Bakushinskii A B 1992 The problem of the convergence of the iteratively regularized Gauss–Newton method *Comput. Math. Math. Phys.* **32** 1353–9

[2] Becker R and Vexler B 2004 *A posteriori* error estimation for finite element discretizations of parameter identification problems *Numer. Math.* **96** 435–59

[3] Becker R and Vexler B 2005 Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations *J. Comp. Phys.* **206** 95–110

[4] Deuflhard P, Engl H W and Scherzer O 1998 A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions *Inverse Problems* **14** 1081–106

[5] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)

[6] Griesbaum A, Kaltenbacher B and Vexler B 2008 Efficient computation of the Tikhonov regularization parameter by goal oriented adaptive discretization *Inverse Problems* **24** 025025

[7] Hanke M 1997 Regularizing properties of a truncated Newton–cG algorithm for nonlinear inverse problems *Numer. Funct. Anal. Optim.* **18** 971–93

[8] Hanke M 1997 A regularization Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* **13** 79–95

[9] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration for nonlinear ill-posed problems *Numer. Math.* **72** 21–37

[10] Hohage T 1997 Logarithmic convergence rates of the iteratively regularized Gauss–Newton method for an inverse potential and an inverse scattering problem *Inverse Problems* **13** 1279–99

[11] Hofmann B and Scherzer O 1994 Factors influencing the ill-posedness of nonlinear problems *Inverse Problems* **10** 1277–98

[12] Hohage T 2000 Regularization of exponentially ill-posed problems *Numer. Funct. Anal. Optim.* **21** 439–64

[13] Kaltenbacher B 2008 A note on logarithmic convergence rates for nonlinear Tikhonov regularization *J. Inverse Ill-Posed Problems* **16** 79–88

[14] Blaschke B, Neubauer A and Scherzer O 1997 On convergence rates for the iteratively regularized Gauss–Newton method *IMA J. Numer. Anal.* **17** 421–36

[15] Kaltenbacher B, Kirchner A, Veljovic S and Vexler B 2011 Goal-oriented adaptivity for Newton type regularization methods (in preparation)

[16] Kaltenbacher B, Kirchner A and Vexler B 2011 Regularization by projection with adaptive discretization based on goal oriented error estimators (in preparation)

[17] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Berlin: de Gruyter)

[18] Kravaris C and Seinfeld J H 1985 Identification of parameters in distributed parameter systems by regularization *SIAM J. Control Optim.* **23** 217–41

[19] Lu S, Pereverzev S V and Ramlau R 2007 An analysis of Tikhonov regularization for nonlinear ill-posed problems under a general smoothness assumption *Inverse Problems* **23** 217–30

[20] Meidner D and Vexler B 2007 Adaptive space-time finite element methods for parabolic optimization problems *SIAM J. Control Optim.* **46** 116–42

[21] Nair M T, Schock E and Tautenhahn U 2003 Morozov's discrepancy principle under general source conditions *Zeitschr. Anal. Anw.* **22** 199–214

[22] Neubauer A and Scherzer O 1990 Finite-dimensional approximation of Tikhonov regularized solutions of nonlinear ill-posed problems *Numer. Funct. Anal. Optim.* **11** 85–99

[23] Potra F 1987 On a monotone Newton-like method *Computing* **39** 233–46

[24] Rieder A 1999 On the regularization of nonlinear ill-posed problems via inexact Newton methods *Inverse Problems* **15** 309–27

[25] Scherzer O 1995 Convergence criteria of iterative methods based on Landweber iteration for nonlinear problems *J. Math. Anal. Appl.* **194** 911–33

[26] Seidman T I and Vogel C R 1989 Well-posedness and convergence of some regularization methods for nonlinear ill-posed problems *Inverse Problems* **5** 227–38

[27] Tautenhahn U and Jin Q-n 2003 Tikhonov regularization and *a posteriori* rules for solving nonlinear ill-posed problems *Inverse Problems* **19** 1–21