Inverse Problems **14** (1998) 1247–1264. Printed in the UK PII: S0266-5611(98)91567-2

Iterative choices of regularization parameters in linear inverse problems

Karl Kunisch†§ and Jun Zou‡‖

† Institut für Mathematik, Karl Franzens Universität Graz, Heinrichstrasse 36, A-8010 Graz, Austria

‡ Department of Mathematics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, People's Republic of China

Received 12 February 1998

Abstract. We investigate possibilities of choosing reasonable regularization parameters for the output least squares formulation of linear inverse problems. Based on the Morozov and damped Morozov discrepancy principles, we propose two iterative methods, a quasi-Newton method and a two-parameter model function method, for finding some reasonable regularization parameters in an efficient manner. These discrepancy principles require knowledge of the error level in the data of the considered inverse problems, which is often inaccessible or very expensive to achieve in real applications. We therefore propose an iterative algorithm to estimate the observation errors for linear inverse problems. Numerical experiments for one- and two-dimensional elliptic boundary value problems and an integral equation are presented to illustrate the efficiency of the proposed algorithms.

1. Introduction

Inverse problems are encountered in many industrial and engineering applications. As the problems are often ill-posed, small perturbations in the observation data can have large effects on the considered solutions. To make a numerical resolution feasible some type of regularization has to be introduced, which entails the necessity of choosing an appropriate regularization parameter. In fact, the effectiveness and success of a regularization method depends strongly on the choice of a good regularization parameter. In practice the regularization parameters are still most frequently chosen heuristically. This is unsatisfactory, both from the practical as well as conceptual points of view. The choice of reasonably good regularization parameters by deterministic numerical methods is one of the most important issues in solving inverse problems.

There exists a significant amount of research in the literature on the development of appropriate strategies for selecting regularization parameters. We refer the readers to [1, 5, 9, 10, 4] and references therein. Much less work has been carried out on the numerical realization of such strategies, and in fact it appears that very few of the strategies are utilized for practical applications. One of the causes may be related to the fact that these methods require knowledge of the noise level of the data which is frequently unavailable in practice.

It is one of the goals of this paper to make a contribution to the subject of practical parameter choice strategies. We will investigate possibilities of choosing reasonable

§ E-mail address: karl.kunisch@kfunigraz.ac.at

‖ E-mail address: zou@math.cuhk.edu.hk

regularization parameters for one of the most frequently used regularization methods, i.e. the output least squares formulation for linear inverse problems. Our basic tool is the well known Morozov discrepancy principle [9, 10, 4] and the damped Morozov discrepancy principle [8]. In [8] the asymptotic behaviour of the damped Morozov principle as the error level in the observation data tends to zero was studied. In this paper, we will propose two iterative methods, a quasi-Newton method and a two-parameter model function method, for finding practically reasonable regularization parameters in an efficient manner. The model function (four-parameter) approach was earlier used in [7] for solving a nonlinear parameter identification problem.

Most parameter-choice strategies and discrepancy principles require knowledge of the observation error of the considered inverse problem, which are often inaccessible or very expensive to achieve in real applications. A second goal of this paper is to estimate the observation error from the available data by an iterative method. The estimated observation error can subsequently be used as a basis for the choice of the regularization parameter. Many numerical experiments for one- and two-dimensional elliptic boundary value problems and an integral equation will be presented to illustrate the efficiency of the proposed algorithms.

Let us now formulate the problem to be discussed in the paper. We consider inverse problems of the form

$$Tf = z \tag{1.1}$$

where T is a bounded operator mapping the parameter space X into the observation space Y. Here $z \in Y$ are the observation data which may be corrupted by error. The noisy data with noise level δ are denoted by z^δ.

The above problem is often ill-posed due to lack of a continuous inverse of T so that small perturbations in the data can result in large changes of the solution f of (1.1). To transform the problem into a well-posed problem and make a numerical resolution feasible, we formulate the inverse problem as the following output least squares problem

$$\min_{f\in X} J(f,\beta) = \frac{1}{2}\|Tf - z^\delta\|_Y^2 + \frac{\beta}{2}\|f\|_X^2 \tag{1.2}$$

where $\beta > 0$ is the regularization parameter, and $\|\cdot\|_Y$ and $\|\cdot\|_X$ denote the norms in the Hilbert spaces Y and X respectively. The corresponding inner products are denoted by $(\cdot,\cdot)_Y$ and $(\cdot,\cdot)_X$.

We end this section with a well known existence result for (1.2). As usual, $T^* : Y \to X$ denotes the adjoint operator of T. In some situations, we shall express the solution of (1.2) as $f(\beta)$ to emphasize its dependence on β.

Lemma 1.1. *For any $\beta > 0$ there exists a unique solution $f(\beta)$ to the minimization problem (1.2). It is characterized as the solution to the system*

$$T^*Tf + \beta f = T^*z^\delta$$

or in variational form

$$(Tf, Tg)_Y + \beta\,(f, g)_X = (z^\delta, Tg)_Y \qquad \textit{for all } g \in X. \tag{1.3}$$

2. Differentiability of $f(\beta)$

In this section we discuss the differentiability of the function $\beta \to f(\beta)$. We first verify the following lemma.

Lemma 2.1. *The function $f(\beta)$ is infinitely differentiable at every $\beta > 0$ and its derivative $f^{(n)}(\beta) \in X$, for each $n \geqslant 1$, is the unique solution w to the following equation:*

$$(Tw, Tg)_Y + \beta\,(w, g)_X = -n\,(f^{(n-1)}(\beta), g)_X \qquad \text{for all } g \in X. \tag{2.1}$$

Proof. For every t we have by (1.3),

$$(Tf(\beta+t), Tg)_Y + (\beta+t)(f(\beta+t), g)_X = (z^\delta, Tg)_Y \qquad \text{for all } g \in X. \tag{2.2}$$

Choosing $g = f(\beta+t)$ the Cauchy–Schwarz inequality implies that

$$\|f(\beta+t)\|_X^2 \leqslant \frac{1}{2\beta}\|z^\delta\|_Y^2 \tag{2.3}$$

for all t with $|t|$ sufficiently small. Subtracting (1.3) from (2.2) yields

$$(T(f(\beta+t) - f(\beta)), Tg)_Y + \beta\,(f(\beta+t) - f(\beta), g)_X = -t\,(f(\beta+t), g)_X \quad \text{for all } g \in X. \tag{2.4}$$

Taking $g = f(\beta+t) - f(\beta)$ in (2.4), we obtain

$$\|T(f(\beta+t) - f(\beta))\|_Y^2 + \beta\|f(\beta+t) - f(\beta)\|_X^2 \leqslant -t(f(\beta+t), f(\beta+t) - f(\beta))_X.$$

Applying Young's inequality leads to

$$\beta\,\|f(\beta+t) - f(\beta)\|_X^2 \leqslant \frac{t^2}{\beta}\|f(\beta+t)\|_X^2$$

which, together with the bound (2.3), proves that $f(\beta)$ is Lipschitz continuous at β.

We next show the differentiability of $f(\beta)$. For this purpose, we divide (2.4) by t, subtract (2.1) with $n = 1$ from the resulting equation and obtain

$$\|Tg(t)\|_Y^2 + \beta\|g(t)\|_X^2 = (f(\beta) - f(\beta+t), g(t))_X$$

where $g(t) = t^{-1}(f(\beta+t) - f(\beta)) - w$. Applying Young's inequality to the right-hand side implies

$$\beta\|g(t)\|_X^2 \leqslant \frac{1}{\beta}\|f(\beta) - f(\beta+t)\|_X^2$$

which together with the continuity of $f(\beta)$ shows that $g(t) \to 0$ as $t \to 0$. Thus, the first derivative $f'(\beta)$ exists and is equal to the solution w of (2.1).

The proof of (2.1) follows by induction on n. □

Let $F(\beta)$ denote the minimal value function, i.e.

$$F(\beta) = J(f(\beta), \beta) = \frac{1}{2}\|Tf(\beta) - z^\delta\|_Y^2 + \frac{\beta}{2}\|f(\beta)\|_X^2 \tag{2.5}$$

for $\beta > 0$. We have the following.

Lemma 2.2. *The first and second derivatives of $F(\beta)$ are given by*

$$F'(\beta) = \tfrac{1}{2}\|f(\beta)\|_X^2 \qquad \text{and} \qquad F''(\beta) = (f(\beta), f'(\beta))_X \qquad \text{for all } \beta > 0. \tag{2.6}$$

Proof. The differentiability of $F(\beta)$ follows immediately from its definition and lemma 2.1. To derive the formulae in (2.6), we differentiate both sides of (2.5) with respect to β and obtain

$$F'(\beta) = (Tf(\beta) - z^\delta, Tf'(\beta))_Y + \beta(f(\beta), f'(\beta))_X + \tfrac{1}{2}\|f(\beta)\|_X^2.$$

The desired relation for $F'(\beta)$ follows immediately from (1.3) with $g = f'(\beta)$. The remaining formula for $F'(\beta)$ can be derived directly from the one for $F'(\beta)$. □

The function $F(\beta)$ has some additional nice properties stated in the following lemma.

Lemma 2.3. *Assume that* $z^\delta \notin \ker T^*$. *Then the non-negative function* $F(\beta)$ *is strictly monotonically increasing and strictly concave.*

Proof. From (2.1) with $n = 1$ and $g = f'(\beta)$ we deduce that

$$(Tf'(\beta), Tf'(\beta))_Y + \beta(f'(\beta), f'(\beta))_X = -(f(\beta), f'(\beta))_X.$$

Using lemma 2.2, we obtain

$$F''(\beta) = (f(\beta), f'(\beta))_X = -\|Tf'(\beta)\|_Y^2 - \beta\|f'(\beta)\|^2 \leqslant 0 \qquad \text{for all } \beta > 0.$$

In fact, $F''(\beta) < 0$ for every $\beta > 0$. Otherwise if $F''(\bar{\beta}) = 0$ for some $\bar{\beta}$, then we have $f'(\bar{\beta}) = 0$. By lemma 2.1 this implies $f(\bar{\beta}) = 0$ which by lemma 1.1 contradicts $T^*z^\delta \neq 0$. Thus we have $F'(\beta) > 0$ and $F''(\beta) < 0$ for every positive β. This implies that $F(\beta)$ is strictly monotonically increasing and strictly concave. □

3. Iterative realization of parameter choice strategies

In this section we investigate numerical realizations of some parameter choice strategies. We shall repeatedly use the expressions for the first and second derivatives of the minimal value function $F(\beta)$ given in lemma 2.2 and the fact that these derivatives can be computed in a stable manner if β is sufficiently large. For the most part, our analysis will focus on the Morozov principle.

We first derive an important identity which will be used later. Let $f(\beta)$ be the unique minimizer to problem (1.2) for $\beta > 0$. Then we have by lemma 1.1

$$T^*Tf(\beta) + \beta f(\beta) = T^*z^\delta \tag{3.1}$$

and upon differentiating with respect to β,

$$T^*Tf'(\beta) + f(\beta) + \beta f'(\beta) = 0. \tag{3.2}$$

Taking the inner product with $f(\beta)$ we obtain

$$(Tf'(\beta), Tf(\beta))_Y + (f(\beta), f(\beta))_X + \beta(f'(\beta), f(\beta))_X = 0$$

which by lemma 2.2 can be written as

$$2F'(\beta) + \beta F''(\beta) + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\beta}(Tf(\beta), Tf(\beta))_Y = 0$$

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\{\beta F'(\beta) + F(\beta) + \tfrac{1}{2}(Tf(\beta), Tf(\beta))_Y\} = 0.$$

Integrating with respect to β we find

$$2\beta F'(\beta) + 2F(\beta) + (Tf(\beta), Tf(\beta))_Y = 2C_0 \tag{3.3}$$

where C_0 is an integration constant.

3.1. Morozov's principle

The well known Morozov principle has received a considerable amount of attention in linear inverse problems (cf [1, 5, 9, 10]). The principle states that the regularization parameter β should be chosen such that the error due to the regularization is equal to the error due to the observation data. That is, β is chosen according to

$$\|Tf(\beta) - z^\delta\|_Y^2 = \delta^2 \tag{3.4}$$

where δ is the observation error defined by

$$\delta = \|z - z^\delta\|_Y.$$

Throughout this section we assume that $z^\delta \notin \ker T^*$. We observe that equation (3.4) can be expressed in terms of $F(\beta)$ as

$$F(\beta) - \beta F'(\beta) = \tfrac{1}{2}\delta^2. \tag{3.5}$$

In some applications, the Morozov principle may not be so satisfactory. We therefore consider a more general class of damped Morozov principles [8, 10] given by

$$\|Tf(\beta) - z^\delta\|_Y^2 + \beta^\gamma \|f(\beta)\|_X^2 = \delta^2$$

where $\gamma \in [1, \infty]$, or equivalently,

$$F(\beta) + (\beta^\gamma - \beta)F'(\beta) = \tfrac{1}{2}\delta^2. \tag{3.6}$$

Note that the exact Morozov principle (3.5) is a special case of the damped case with $\gamma = \infty$.

We now discuss the existence and uniqueness of the solutions to the exact Morozov and the damped Morozov equation. We shall make the assumption that $F(0) < \delta^2/2 \leqslant F(1)$, where

$$F(0) = \inf_{f \in X} \tfrac{1}{2}\|Tf - z^\delta\|_Y^2 = \tfrac{1}{2}\|(I - P)z^\delta\|_Y^2 \tag{3.7}$$

with P being the orthorgonal projection of z^δ onto the closure of the range of T. The proof of the following lemma will reveal that the upper bound on δ can be replaced by

$$\tfrac{1}{2}\delta^2 \leqslant F(\infty) = \tfrac{1}{2}\|z^\delta\|_Y^2$$

for $\gamma = \infty$. This is justified by the fact that for $\beta \to \infty^+$ we have $f(\beta) \to 0$ and hence $F(\beta) \to \|z^\delta\|_Y^2/2$. For practical purposes it suffices certainly to restrict β to $(0, 1]$.

Lemma 3.1. *If $F(0) < \frac{1}{2}\delta^2 \leqslant F(1)$, then there exists a unique solution $\beta_* \in (0, 1]$ to the Morozov equation (3.6).*

Proof. Let us define

$$G(\beta) = F(\beta) + (\beta^\gamma - \beta)F'(\beta) - \tfrac{1}{2}\delta^2. \tag{3.8}$$

Due to lemma 2.2 we have

$$G'(\beta) = \gamma\beta^{\gamma-1}F'(\beta) + (\beta^\gamma - \beta)F''(\beta). \tag{3.9}$$

Lemma 2.3 implies that $F''(\beta) < 0$ and hence $G'(\beta) > 0$ for every $\beta \in (0, 1]$. (In the case where $\gamma = \infty$ we have $G'(\beta) = -\beta F''(\beta) > 0$ for every $\beta \in (0, \infty)$.) Therefore the function G is strictly increasing on $(0, 1]$. Continuity of G together with

$$G(0) = F(0) - \tfrac{1}{2}\delta^2 \qquad G(1) = F(1) - \tfrac{1}{2}\delta^2$$

implies the result. □

3.2. Newton's and quasi-Newton's method

We now propose using Newton's method or a quasi-Newton's method to solve the damped nonlinear Morozov equation (3.6), that is,

$$G(\beta) = F(\beta) + (\beta^\gamma - \beta)F'(\beta) - \tfrac{1}{2}\delta^2 = 0.$$

We know from (3.9) that

$$G'(\beta) = \gamma\beta^{\gamma-1}F'(\beta) + (\beta^\gamma - \beta)F''(\beta)$$
$$= \tfrac{1}{2}\gamma\beta^{\gamma-1}(f(\beta), f(\beta))_X + (\beta^\gamma - \beta)(f(\beta), f'(\beta))_X.$$

Thus computing $G'(\beta)$ involves the evaluation of $f'(\beta)$ that solves the equation

$$T^*Tw + \beta w = -f(\beta). \tag{3.10}$$

Newton's method for solving equation (3.6) is formulated as follows.

Newton's method. Given an initial guess β_0, generate the Newton's sequence $\beta_1, \beta_2, \ldots,$ according to

$$\beta_{k+1} = \beta_k - \frac{2G(\beta_k)}{\gamma\beta_k^{\gamma-1}\|f(\beta_k)\|_X^2 + 2(\beta_k^\gamma - \beta_k)(f(\beta_k), f'(\beta_k))_X} \tag{3.11}$$

where $f'(\beta_k)$ is obtained from (3.10). As usual, this Newton's method converges quadratically. But at each iteration we need to solve for both $f(\beta)$ and $f'(\beta)$ and this seems to be a bit too expensive.

To avoid solving equation (3.10) for $f'(\beta)$, we propose replacing $f'(\beta_k)$ in the Newton's method by the finite difference quotient

$$f_k(\beta_k, \beta_{k-1}) \equiv \frac{f(\beta_k) - f(\beta_{k-1})}{\beta_k - \beta_{k-1}}.$$

This leads to the following.

Quasi-Newton's method. Given initial guesses β_0 and β_1. Generate the quasi-Newton's sequence $\beta_2, \beta_3, \ldots,$ according to

$$\beta_{k+1} = \beta_k - \frac{2G(\beta_k)}{\gamma\beta_k^{\gamma-1}\|f(\beta_k)\|_X^2 + 2(\beta_k^\gamma - \beta_k)(f(\beta_k), f_k(\beta_k, \beta_{k-1}))_X}. \tag{3.12}$$

This quasi-Newton's method has the following convergence property.

Theorem 3.1. *Assume that $F(0) < \frac{1}{2}\delta^2 \leqslant F(1)$ and let $\beta_* \in (0, 1]$ be the unique solution of the Morozov equation (3.6). Then there exists a positive constant ε such that whenever the initial guesses β_0 and β_1 belong to the interval $I = [\beta_* - \varepsilon, \beta_* + \varepsilon]$, the whole sequence $\{\beta_k\}_{k=0}^\infty$ generated by the quasi-Newton's method is contained in I and converges to β_* superlinearly.*

Proof. We give the proof for $\gamma = \infty$. The case $\gamma \in [1, \infty)$ can be carried out analogously. By lemma 2.1 the following constant M is finite:

$$\frac{\sqrt{2}}{3}M = \max_{\beta_*/2 \leqslant \beta \leqslant 3\beta_*/2}\{\|f(\beta)\|_X, \|f'(\beta)\|_X, \|f''(\beta)\|_X\}.$$

Together with (3.9) for $\gamma = \infty$ this implies

$$|G''(\beta)| \leqslant M^2 \qquad \text{for } \tfrac{1}{2}\beta_* \leqslant \beta \leqslant \tfrac{3}{2}\beta_*. \tag{3.13}$$

By continuity of $G'(\beta)$ at β_*, there exists a constant $\varepsilon \in (0, \beta_*/2)$ such that

$$|G'(\beta)| \geqslant \tfrac{1}{2}|G'(\beta_*)| \qquad \text{for } \beta_* - \varepsilon \leqslant \beta \leqslant \beta_* + \varepsilon.$$

We can assume that

$$\varepsilon \leqslant \frac{3}{4}\frac{|G'(\beta_*)|}{M^2}.$$

Next we show that each iterate β_k generated by the quasi-Newton's method is contained in I and that β_k converges to β_* superlinearly provided that the start-up values β_0, β_1 are chosen in I.

For β_2 we have by the mean-value theorem

$$\begin{aligned}\beta_2 - \beta_* &= (\beta_1 - \beta_*) + \frac{G(\beta_1) - G(\beta_*)}{\beta_1(f(\beta_1), f_1(\beta_1, \beta_0))_X} \\ &= (\beta_1 - \beta_*)\frac{\beta_1(f(\beta_1), f_1(\beta_1, \beta_0))_X + G'(\eta_1)}{\beta_1(f(\beta_1), f_1(\beta_1, \beta_0))_X} \\ &\equiv: (\beta_1 - \beta_*)\frac{B_1}{A_1}\end{aligned} \tag{3.14}$$

where η_1 lies between β_1 and β_*. We now bound A_1 and B_1. By Taylor expansion we have

$$f_1(\beta_1, \beta_0) = f'(\beta_1) + \tfrac{1}{2}f''(\xi_1)(\beta_0 - \beta_1)$$

with $\xi_1 \in (\beta_1, \beta_0)$. Hence we obtain

$$\begin{aligned}A_1 &= \beta_1(f(\beta_1), f_1(\beta_1, \beta_0) - f'(\beta_1))_X - G'(\beta_1) \\ &= \frac{\beta_1}{2}(\beta_0 - \beta_1)(f(\beta_1), f''(\xi_1))_X - G'(\beta_1).\end{aligned} \tag{3.15}$$

We can bound A_1 as follows

$$|A_1| \geqslant \tfrac{1}{2}|G'(\beta_*)| - \tfrac{2}{9}\beta_* M^2\varepsilon \geqslant \tfrac{1}{4}|G'(\beta_*)|.$$

To estimate B_1, we use (3.15) to obtain

$$B_1 = \frac{\beta_1}{2}(\beta_0 - \beta_1)(f(\beta_1), f''(\xi_1))_X + (G'(\eta_1) - G'(\beta_1)).$$

Thus B_1 can be bounded by using (3.13),

$$|B_1| \leqslant 2\beta_* M^2\varepsilon + M^2\varepsilon \leqslant \tfrac{3}{2}M^2\varepsilon.$$

Combining the bounds for A_1 and B_1, we obtain from (3.14)

$$|\beta_2 - \beta_*| \leqslant |\beta_1 - \beta_*|\frac{6M^2\varepsilon}{|G'(\beta_*)|} \leqslant \frac{1}{2}|\beta_1 - \beta_*|$$

which implies $\beta_2 \in [\beta_* - \varepsilon, \beta_* + \varepsilon]$.

By induction one can show that $\beta_k \in [\beta_* - \varepsilon, \beta_* + \varepsilon]$ and

$$|\beta_k - \beta_*| \leqslant \tfrac{1}{2}|\beta_{k-1} - \beta_*| \qquad \text{for } k = 2, 3, \ldots.$$

Therefore we have $\beta_k \to \beta_*$ as $k \to \infty$.

Finally in the same way as (3.14), we have

$$\beta_{k+1} - \beta_* = (\beta_k - \beta_*)\frac{\beta_k(f(\beta_k), f_k(\beta_k, \beta_{k-1}))_X + G'(\eta_k)}{\beta_k(f(\beta_k), f_k(\beta_k, \beta_{k-1}))_X}.$$

Taking the limit as $k \to \infty$, we obtain

$$\lim_{k\to\infty}\frac{|\beta_{k+1} - \beta_*|}{|\beta_k - \beta_*|} = \frac{|\beta_*(f(\beta_*), f'(\beta_*))_X + G'(\beta_*)|}{|\beta_*(f(\beta_*), f'(\beta_*))_X|} = 0$$

and superlinear convergence of the sequence $\{\beta_k\}$ follows. □

3.3. Two-parameter model functions

In this section we discuss a model function approach for solving the general Morozov equation (3.6) approximately. By a model function we mean a parametrized function which preserves the major properties of the non-negative function $F(\beta)$ and which approximates or interpolates $F(\beta)$ in a manner to be specified below. Some results for nonlinear inverse problems using model function approaches can be found in [7], where a four-parameter model function approach was investigated. We are now going to derive a two-parameter model function approach which will be seen to perform well for linear inverse problems. Moreover, we shall demonstrate in section 4 that a hybrid scheme based on model functions during the start-up phase and on the quasi-Newton method locally provides a very efficient method to solve the Morozov equation.

To derive the model function, we make the following approximation in the equation (3.3):

$$(Tf(\beta), Tf(\beta))_Y \approx T_1(f(\beta), f(\beta))_X$$

where T_1 is a positive constant to be determined. Then equation (3.3) reduces to

$$\beta m'(\beta) + m(\beta) + T_1 m'(\beta) = C_0. \tag{3.16}$$

Solving the ordinary differential equation (3.16) we obtain

$$m(\beta) = C_0 + \frac{C_1}{T_1 + \beta} \tag{3.17}$$

where C_1 is an integration constant. We know that $F(\beta)$ is an increasing and concave function, and obviously the model function $m(\beta)$ preserves these properties when $C_1 < 0$, and is non-negative if $C_0 + C_1/T_1 \geqslant 0$.

Note that when the linear operator T has a dense range in $L^2(\Omega)$, then we have

$$\inf_{f\in X} \|Tf - z^\delta\|_Y^2 = 0$$

and therefore

$$F(0) = 0.$$

So to further simplifiy the model function $m(\beta)$ in (3.17), we require $m(0) = 0$. Then we can write the model function as

$$m(\beta) = C\left\{1 - \frac{T}{T + \beta}\right\} \tag{3.18}$$

which has only two parameters involved, and we refer to it as the *two-parameter model function*.

To update the two parameters C and T in the above model function and so solve the general Morozov equation (3.6) approximately, we suggest the following algorithm.

Two-parameter algorithm. Set $k = 0$ and choose $\beta_0 > 0$.

(1) Compute $F'(\beta_k)$ and $F(\beta_k)$ using (2.6). Compute T_k and C_k from

$$m(\beta_k) = C_k\left\{1 - \frac{T_k}{T_k + \beta_k}\right\} = F(\beta_k) \tag{3.19}$$

$$m'(\beta_k) = \frac{C_k T_k}{(T_k + \beta_k)^2} = F'(\beta_k). \tag{3.20}$$

(2) Set

$$m(\beta) = C_k \left\{ 1 - \frac{T_k}{T_k + \beta} \right\}.$$

(3) Solve for β_{k+1} the approximate Morozov's equation

$$m(\beta) + (\beta^\gamma - \beta) m'(\beta) = \tfrac{1}{2}\delta^2. \tag{3.21}$$

(4) If $|\beta_{k+1} - \beta_k| \leqslant tolerance$, STOP; otherwise set $k := k + 1$, GOTO (1).

In step (1) of the above two-parameter algorithm, one needs to compute T_k and C_k from (3.19) and (3.20). Combining (3.19) and (3.20), we can easily find the following direct formulae for computing T_k and C_k:

$$T_k = \frac{\beta_k^2 F'(\beta_k)}{F(\beta_k) - \beta_k F'(\beta_k)} \qquad C_k = \frac{F^2(\beta_k)}{F(\beta_k) - \beta_k F'(\beta_k)}. \tag{3.22}$$

Note that by (2.5) and lemma 2.2 the two-parameter model function in step (2) can be determined from one evaluation of (1.2) at β_k. The denominators in (3.22) do not vanish as we always have

$$F(\beta_k) - \beta_k F'(\beta_k) = \tfrac{1}{2}\|Tf(\beta_k) - z^\delta\|_Y^2 > 0$$

if $z^\delta \notin \ker T^*$.

Numerically we use the Newton's method to solve the approximate Morozov equation (3.21) for β if $\gamma \neq \infty$. But for $\gamma = \infty$ equation (3.21) is quadratic in β and can be solved directly.

Note that equation (3.21) always has a solution for $\gamma \in (2, \infty)$, while the existence is guaranteed for $\gamma \in [1, 2]$ and $\gamma = \infty$ if $0 = m(0) \leqslant \delta^2/2 < m(\infty) = C_k$. If $2m(1) \geqslant \delta^2$ then (3.21) has a unique solution in $(0, 1]$.

As a safeguard we had used in our implementation of the algorithm the additional stopping criterion 'if $m(\beta_k) + (\beta_k^\gamma - \beta_k)m'(\beta) \leqslant \frac{1}{2}\delta^2$, STOP'. In our examples it was never activated.

3.4. Predictions of observation errors

Most parameter-choice strategies, including in particular the Morozov principle, require knowledge of the observation error level δ. In practice, δ is often inaccessible, expensive to achieve or itself error-prone. In such situations it can be helpful to utilize some heuristic approach to estimate δ.

In this section, we propose to use the model function $m(\beta)$ to obtain an estimate for δ. If the unperturbed data z are attainable by some $f^* \in X$, then

$$F(0) \leqslant \tfrac{1}{2}\|Tf^* - z^\delta\|_Y^2 = \tfrac{1}{2}\delta^2$$

and thus $\sqrt{2F(0)}$ gives a lower bound for the error δ. In numerical implementations, we will enlarge this lower bound $\sqrt{2F(0)}$ and use $2\sqrt{m(0)}$ to predict the observation error. We propose the following algorithm.

Observation error prediction algorithm. Given a ratio $\sigma \in [0.5, 1]$ and $\beta_0 > 0$,

(1) compute $F(\beta_0)$ and $F'(\beta_0)$. Then update C_1 and T_1 in (3.18) using

$$m(\beta_0) = F(\beta_0) \qquad m'(\beta_0) = F'(\beta_0).$$

Find y_0, the y-intercept of the tangent to the curve $y = m(\beta)$ at $(\beta_0, m(\beta_0))$, and β_1 satisfying

$$m(\beta_1) = \sigma y_0.$$

(2) Compute $F(\beta_1)$, $F'(\beta_1)$, and C_0 (use (3.3)). Then update C_1 and T_1 in (3.17) using

$$m(\beta_1) = F(\beta_1) \qquad m'(\beta_1) = F'(\beta_1).$$

Return $2\sqrt{m(0)}$ as the observation error δ.

Remark 3.1. The parameter σ can always be taken to be 1. But one may obtain better results when it varies in the interval $[0.5, 1)$.

In the first step, we need to compute y_0 and β_1. Both y_0 and β_1 exist uniquely and are positive as the two-parameter model function $m(\beta)$ in (3.18) is a concave and strictly non-decreasing function and $m(0) = 0$.

The first step can be run a few times, that is, when β_1 is found, set $\beta_0 = \beta_1$ and re-run the step. But in all our numerical implementations, at most two iterations of step (1) gave good results. In any case, β_0 is chosen significantly larger than the expected optimal β so that the corresponding evaluation of (1.2) is numerically stable. The reduction in β from β_0 to β_1 depends primarily on the concavity of $m(\beta)$ and thus on the product $C_1 T_1$.

The information $F(\beta_0)$ and $F'(\beta_0)$ can be saved for use in the two-parameter algorithm or the quasi-Newton's algorithm for solving the damped Morozov equation.

4. Numerical results for solving the Morozov equation

We now present some numerical experiments to show the effectiveness of the quasi-Newton's method, the two-parameter model function approach (two-parameter algorithm) and the resulting hybrid method.

In all tables of this section, β_{opt} stands for the optimal β value which achieves the minimum for $\|f(\beta) - f^*\|_{L^2(\Omega)}$, it is computed as follows. We first compute the L^2-norm error for 200 uniformly distributed β-values in the interval $[10^{-7}, 10^{-3}]$ to find an approximate optimal β, denoted by $\tilde{\beta}$, then a much smaller interval including $\tilde{\beta}$ is chosen to compute an accurate β_{opt}. β_M stands for the solution of the general Morozov equation (3.6). It can be determined by means of a bisection algorithm, for example. In the tables Iter denotes the required number of iterations of the specified algorithm to achieve the listed β values.

Example 1. Consider the following two point boundary value problem

$$-(q(x)u_x)_x = f(x) \qquad \text{in } (0, 1) \text{ with } u(0) = u(1) = 0. \tag{4.1}$$

We take the coefficient function $q(x)$ and the observation data z of u as

$$q(x) = \mathrm{e}^{1+x^2} \qquad z = u(f^*) = \mathrm{e}^{-x}\sin(\pi x)$$

and then the source term $f(x)$ which is to be recovered can be obtained from the above differential equation,

$$f^* = -q_x \mathrm{e}^{-x}\{\pi\cos(\pi x) - \sin(\pi x)\} + q\mathrm{e}^{-x}\{2\pi\cos(\pi x) + (\pi^2 - 1)\sin(\pi x)\}.$$

We assume that the available observed data are the superposition of the error free data z and the sinusoidal noise:

$$z^\delta(x) = z(x) + \hat{\delta}\sin(1.5\pi(2x - 1)).$$

Table 1. Optimal β values and the β's obtained by Morozov principle.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.277×10^{-5}	0.990×10^{-5}	0.172×10^{-4}	0.260×10^{-4}	0.434×10^{-4}
β_M	0.362×10^{-5}	0.125×10^{-4}	0.223×10^{-4}	0.324×10^{-4}	0.474×10^{-4}

Table 2. Convergence of the quasi-Newton's method with $\beta_0 = 10^{-3}$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_i	0.362×10^{-5}	0.125×10^{-4}	0.223×10^{-4}	0.324×10^{-4}	0.474×10^{-4}
Iter	7	7	6	5	7

Table 3. Convergence of the two-parameter algorithm with $\beta_0 = 0.1$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
Iter(2)	0.575×10^{-5}	0.206×10^{-4}	0.365×10^{-4}	0.527×10^{-4}	0.773×10^{-4}
β_i	0.362×10^{-5}	0.125×10^{-4}	0.223×10^{-4}	0.324×10^{-4}	0.474×10^{-4}
Iter	9	11	14	15	16

Table 4. Quasi-Newton's convergence with initial guesses from the two-parameter algorithm.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_i	0.362×10^{-5}	0.125×10^{-4}	0.223×10^{-4}	0.324×10^{-4}	0.474×10^{-4}
Iter	2	2	3	3	3

In our implementations, we use piecewise-linear finite element method to solve the elliptic problem (4.1) and the variational equation (1.3) satisfied by the optimal $f(\beta)$. For this purpose, we first partition the domain $\Omega = (0, 1)$ into N equally distributed subintervals and then define V^h to be the continuous piecewise-linear finite element space associated with the partition, and $h = 1/N$. Let V_0^h be a subspace of V^h with functions vanishing at two endpoints $x = 0, 1$. Then the finite element approximation $f_h(\beta)$ of $f(\beta)$ is formulated as follows. Find $f_h(\beta) \in V^h$ such that

$$(u_h(f_h(\beta)), u_h(g)) + \beta(f_h(\beta), g) = (z^\delta, u_h(g)) \qquad \text{for all } g \in V^h$$

where $u_h \equiv u_h(f_h(\beta)) \in V_0^h$ satisfies

$$(q(x)(u_h)_x, v_x) = (f_h(\beta), v) \qquad \text{for all } v \in V_0^h.$$

In tables 1–4, we present some of the numerical results with different noise parameters $\hat{\delta}$ and $N = 20$.

Table 1 gives the optimal β values as well as the β values computed from the exact Morozov equation ($\gamma = \infty$):

$$F(\beta) - \beta F'(\beta) = \tfrac{1}{2}\delta^2.$$

We can see that the Morozov principle gives very accurate approximations to the optimal β values for the considered example.

Table 2 shows the numbers of iterations of the quasi-Newton's method with the initial guess $\beta_0 = 10^{-3}$. It takes about 5–7 iterations for the method to converge to the exact solution of the Morozov equation. The stopping criterion for the quasi-Newton iteration and also for the two-parameter iteration is chosen as $|\beta_{k+1} - \beta_k|/\beta_{k+1} \leqslant 10^{-2}$. If the initial guess β_0 is too rough, say 0.1, the iteration may diverge. This is consistent with the local convergence of the quasi-Newton's method.

Table 3 shows the convergence of the two-parameter algorithm discussed in section 3.3, with a very rough initial guess $\beta_0 = 0.1$. The second row contains the β values obtained at the second iterations. We observe that the algorithm gives very good approximations to the solutions of the Morozov equation after only two iterations. But afterwards the convergence of the algorithm becomes much slower. The last two rows of table 3 give the numbers of iterations required for the algorithm to converge to the exact solutions of the Morozov equation. The corresponding β-values are shown in the third row.

Many additional numerical experiments have confirmed the convergence phenomena we have seen above about the quasi-Newton's algorithm and the two-parameter model function algorithm. The former converges faster than the latter but only locally, i.e. one has to start at a very good initial guess. The two-parameter model function algorithm converges very fast during the first few iterations and then it slows down. How about combining the advantages of these two algorithms? What will happen if we take the approximate β values obtained from the first or second iterations of the two-parameter algorithm as the initial guesses of the quasi-Newton's method? The results are very positive. Table 4 gives the numbers of iterations for the quasi-Newton's algorithm to converge to the exact solutions of the Morozov equation, when the second iterates from the two-parameter algorithm have been taken as initial guesses. One can see that it then needs only two or three iterations.

Example 2. We consider the following two-dimensional elliptic problem

$$-\nabla \cdot (q(x, y)\nabla u) + c(x, y)u = f(x, y) \qquad \text{in } \Omega \tag{4.2}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial\Omega. \tag{4.3}$$

We use the piecewise-linear finite element method to discretize the Neumann boundary value problem (4.2), (4.3) with triangular elements of uniform mesh size $h = \frac{1}{20}$. Let V^h be the piecewise-linear finite element space with the standard nodal basis functions $\{\phi_j\}_j^N$, $N = 21 \times 21$ being the number of nodal points. The finite element problem approximating (4.2), (4.3) is: find $u_h(f) \in V^h$ such that

$$(q\nabla u_h(f), v_h) + (cu_h(f), v_h) = (f, v_h) \qquad \text{for all } v_h \in V^h. \tag{4.4}$$

The integrals involved on each element for computing the coefficient matrix were done by the quadrature rule which takes the average of three midpoint values on three sides of the element. The resulting stiffness matrix is denoted as A.

To approximate the optimal $f(\beta)$ which minimizes the real function $F(\beta)$ defined in (2.5) for a fixed β we use

$$(u_h(f_h), u_h(g)) + \beta(f_h, g) = (z^\delta, u_h(g)) \qquad \text{for all } g \in V^h \tag{4.5}$$

where $f_h = f_h(\beta)$. Let M be the N by N mass matrix, i.e. $\mathsf{M} = (m_{ij})$, $m_{ij} = (\phi_i, \phi_j)$. Then equations (4.4) and (4.5) can be written algebraically as follows

$$\boldsymbol{u}(\boldsymbol{g})^\top \mathsf{M}\boldsymbol{u}(\boldsymbol{f}) + \beta \boldsymbol{g}^\top \mathsf{M}\boldsymbol{f} = \boldsymbol{u}(\boldsymbol{g})^\top \mathsf{M}\boldsymbol{z}^\delta \qquad \mathsf{A}\boldsymbol{u}(\boldsymbol{g}) = \mathsf{M}\boldsymbol{g} \qquad \text{for all } \boldsymbol{g} \in R^N \tag{4.6}$$

where $\boldsymbol{f}$ and $\boldsymbol{u}(\boldsymbol{f})$ are vectors consisting of the nodal values of $f_h(\beta)$ and $u_h(f_h)$ respectively. Similarly $\boldsymbol{z}^\delta$, $\boldsymbol{g}$ and $\boldsymbol{u}(\boldsymbol{g})$ represent the nodal values of z^δ, g and $u(g)$.

Table 5. Optimal β values and the β's obtained by Morozov principle.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.121×10^{-5}	0.381×10^{-5}	0.685×10^{-5}	0.983×10^{-5}	0.139×10^{-4}
β_M	0.103×10^{-5}	0.328×10^{-5}	0.556×10^{-5}	0.782×10^{-5}	0.110×10^{-4}

Table 6. Convergence of the quasi-Newton's method with $\beta_0 = 10^{-4}$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_i	0.103×10^{-5}	0.328×10^{-5}	0.556×10^{-5}	0.782×10^{-5}	0.111×10^{-4}
Iter	5	6	6	7	6

By substituting the second relation into (4.6) into the first one and making some simple re-arrangements, we arrive at

$$(\mathbf{M}\mathbf{A}^{-1}\mathbf{M} + \beta\mathbf{A})\boldsymbol{f} = \mathbf{M}z^{\delta}.$$

Solving this equation is usually very expensive as it involves the inverse of $\mathbf{A}$ which is ill-conditioned. Instead we multiply both sides of the equation by $\mathbf{A}\mathbf{M}^{-1}$ and obtain another equivalent form:

$$(\mathbf{M} + \beta\mathbf{A}\mathbf{M}^{-1}\mathbf{A})\boldsymbol{f} = \mathbf{A}\boldsymbol{z}^{\delta}. \tag{4.7}$$

This equation is easier to solve than the previous one as one can show that $\mathbf{M}^{-1}$ is well-conditioned. However, it is still expensive to obtain this inverse. To make the computation more efficient while keeping the same finite element accuracy, we compute the mass matrix using the lumped mass approximation, namely its entries m_{ij} are evaluated as follows

$$m_{ij} = \sum_K \int_K \phi_i \phi_j \, \mathrm{d}x \, \mathrm{d}y \approx \sum_K \int_K \Pi_h(\phi_i \phi_j) \, \mathrm{d}x \, \mathrm{d}y$$

where Π_h is the standard finite element interpolant associated with V^h and the summation is done over at most two elements K's on which the product $\phi_i \phi_j$ does not vanish. This results in a diagonal mass matrix $\mathbf{M}$ so that solving equation (4.7) becomes much cheaper. In real implementations, the algebraic system (4.7) can be solved very efficiently by using domain-decomposition-based or multilevel-method-based preconditioned iterative methods (cf [3, 2]). In our computations we solved the algebraic system by the conjugate gradient method. Tables 5–8 show the numerical experiments related to example 2, where we have taken the coefficint functions $q(x, y)$, $c(x, y)$ and the unperturbed observation data as

$$q(x, y) = \mathrm{e}^{x+y} \qquad c(x, y) = \mathrm{e}^{1+x^2+y^2} \qquad u(f^*) = \cos(\pi x)\cos(\pi y).$$

The noisy data were assumed to be of the form

$$z^{\delta}(x, y) = u(x, y) + \hat{\delta}\sin(1.5\pi(2x - 1))\sin(1.5\pi(2y - 1)).$$

The exact source term $f(x, y)$ to be recovered is the right-hand side function of equation (4.2) using the given coefficients $q(x, y)$, $c(x, y)$ and the exact observation $u(x, y)$.

Table 5 gives the optimal β values and the β values computed from the exact Morozov equation ($\gamma = \infty$):

$$F(\beta) - \beta F'(\beta) = \tfrac{1}{2}\delta^2.$$

Table 7. Convergence of the two-parameter algorithm with $\beta_0 = 0.1$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
Iter(2)	0.122×10^{-5}	0.428×10^{-5}	0.776×10^{-5}	0.115×10^{-4}	0.176×10^{-4}
β_i	0.103×10^{-5}	0.328×10^{-5}	0.556×10^{-5}	0.782×10^{-5}	0.111×10^{-4}
Iter	5	6	7	8	8

Table 8. Quasi-Newton's convergence with initial guesses from the two-parameter algorithm.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_i	0.103×10^{-5}	0.328×10^{-5}	0.556×10^{-5}	0.783×10^{-5}	0.111×10^{-4}
Iter	3	3	4	3	3

We see that the Morozov principle again gives very accurate approximations to the optimal β values for the considered example. Table 6 gives the number of iterations of the quasi-Newton's method with a very close initial guess $\beta_0 = 10^{-4}$ when it converges to the given β values. For this example the quasi-Newton algorithm still converges with $\beta_0 = 0.1$ but the number of iterations is more than doubled compared with $\beta_0 = 10^{-4}$.

Table 7 gives the convergence of the two-parameter algorithm discussed in section 3.3, with a very rough initial guess $\beta_0 = 0.1$. The second row contains the β values obtained at the second iteration. We observe that the algorithm gives already very good approximations to the solutions of the Morozov equation after only two iterations. However, then the convergence slows down during the subsequent iterations. The last two rows of table 7 give the converged β-values and the number of iterations that are required to reach them.

Just as in example 1, the current example again demonstrates the local convergence of the quasi-Newton algorithm and the global convergence of the two-parameter model function algorithm. When we combine the advantages of the two algorithms, we can speed up the whole iterative process. Table 8 gives the numbers of iterations for the quasi-Newton algorithm to converge to the exact solutions of the Morozov equation, when the second iterates from the two-parameter algorithm were taken as initial guesses. One needs only three or four iterations to reach the stopping criterion. In practical implementations, we do not need such accurate results. Only one or two quasi-Newton's iterations will give very satisfactory results.

Example 3 (cf [6]). Traditional agricultural fields are often watered from elevated irrigation canals by removing a solid gate from a weir notch. Suppose that the depth of water in the canal is h and the notch is symmetric about a vertical centre line (cf figure 1).

By Torricelli's law (cf [6]), the volume of flow per unit time through the notch is

$$2\int_0^h \sqrt{2g(h-y)}\,f(y)\,\mathrm{d}y$$

where $g = 9.80$ m s^{-2} is the gravitaional constant and $x = f(y)$ specifies the shape of the notch. Suppose that one wishes to design a notch so that this quantity is a given function $z(h)$ of the water depth in the canal (or equivalently suppose one wants to determine the shape f from observations of the flow rate $z(h)$). Then one is led to solve the following integral equation

$$z(h) = 2\int_0^h \sqrt{2g(h-y)}\,f(y)\,\mathrm{d}y \tag{4.8}$$

for f.

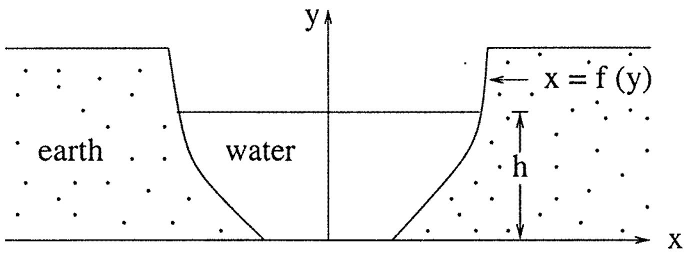

Figure 1. A weir notch.

Table 9. Optimal β's and β's obtained by the exact Morozov principle ($\gamma = \infty$).

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.430×10^{-4}	0.134×10^{-3}	0.230×10^{-3}	0.333×10^{-3}	0.502×10^{-3}
β_M	0.656×10^{-3}	0.317×10^{-2}	0.697×10^{-2}	0.114×10^{-1}	0.184×10^{-1}

Table 10. Optimal β's and β's obtained by the Morozov's principle with $\gamma = 1$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.430×10^{-4}	0.134×10^{-3}	0.230×10^{-3}	0.333×10^{-3}	0.502×10^{-3}
β_M	0.154×10^{-5}	0.138×10^{-4}	0.382×10^{-4}	0.748×10^{-4}	0.154×10^{-3}

Table 11. Optimal β's and the L^2-norm errors.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.430×10^{-4}	0.134×10^{-3}	0.230×10^{-3}	0.333×10^{-3}	0.502×10^{-3}
L^2-norm error	0.501×10^{-2}	0.148×10^{-1}	0.245×10^{-1}	0.342×10^{-1}	0.486×10^{-1}

In our numerical implementation, we take the following true parameter function

$$f(y) = \mathrm{e}^{-y}(2\pi \cos(\pi y) + (\pi^2 - 1)\sin(\pi y))$$

and the observation function $z(h)$, $h \in [0, 1]$, was computed using formula (4.8). Then we add noise to the observation data as follows

$$z^{\delta}(h) = z(h) + \hat{\delta}\sin(3\pi h).$$

To evaluate the integrals involved, we divide the intergal $[0, 1]$ into $n = 20$ subintervals, and on each subinterval the trapezoidal rule is used.

For this example, we use the damped Morozov principle:

$$F(\beta) + (\beta^{\gamma} - \beta)F'(\beta) = \tfrac{1}{2}\delta^2$$

since the exact Morozov principle ($\gamma = \infty$) behaves very disappointingly, compared with the previous boundary value inverse problems (examples 1 and 2). It overestimates the optimal β values about 15–40 times when the noise level ranges from 1% to 10%, see table 9.

When we take $\gamma \in [1, 2]$, all the results are acceptable, except for smaller noise level ($\leqslant$ 1%), see table 10 for $\gamma = 1$. The optimal γ seems to be around 1.4. Table 12 gives

Table 12. β's obtained by Morozov's equation ($\gamma = 1.5$) and the L^2-norm errors.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_M	0.128×10^{-3}	0.543×10^{-3}	0.107×10^{-2}	0.167×10^{-2}	0.269×10^{-2}
L^2-norm error	0.676×10^{-2}	0.227×10^{-1}	0.374×10^{-1}	0.497×10^{-1}	0.654×10^{-1}

Table 13. Optimal β's and β's obtained from Morozov's equation with $\gamma = 1.3$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.430×10^{-4}	0.134×10^{-3}	0.230×10^{-3}	0.333×10^{-3}	0.502×10^{-3}
β_M	0.336×10^{-4}	0.181×10^{-3}	0.395×10^{-3}	0.652×10^{-3}	0.114×10^{-2}

Table 14. Convergence of the quasi-Newton's method with $\beta_0 = 0.01$.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_i	0.128×10^{-3}	0.543×10^{-3}	0.107×10^{-2}	0.167×10^{-2}	0.269×10^{-2}
Iter	5	4	4	3	3

Table 15. Iteration numbers of the two-parameter algorithm for example 3.

$\gamma\backslash\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
1.0	2	2	2	2	2
1.3	2	2	2	2	2
1.5	2	3	3	3	3
2.0	5	5	5	5	5
∞	7	7	7	6	7

the β values obtained with $\gamma = 1.5$. Compared with the optimal β values in table 9, they are about 2–5 times larger than the optimal β's. While this is quite satisfactory, one can achieve much more accurate results with $\gamma = 1.2$ or $\gamma = 1.3$, see table 13 for $\gamma = 1.3$.

Table 11 shows the optimal β values and the corresponding minimum L^2-norm distance between $f(\beta)$ and the true parameter f. Table 12 contains the β values given by the Morozov principle and their corresponding relative L^2-norm errors. Comparing table 11 with table 12, we can see that the β's obtained by the Morozov principle are about 2–5 times larger than optimal ones but the corresponding L^2-norm errors are quite close, just about 1.3–1.5 times larger than optimal L^2-norm errors.

The convergence of the Newton's method for finding the solutions of the Morozov equations with $\gamma = 1.5$ is shown in table 14. One can see that it takes usually just 3–5 iterations to obtain the Morozov's solutions with very good accuracy.

Finally, table 15 shows the iteration numbers for the convergence of the two-parameter algorithm to the solution of the corresponding damped Morozov equation for different parameters γ (cf tables 9–13), with a very rough initial guess $\beta_0 = 0.1$. Note that the algorithm performs very well for a very large range of γ.

Table 16. Exact and predicted observation errors for example 1.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
δ	7.07×10^{-3}	2.12×10^{-2}	3.53×10^{-2}	4.95×10^{-2}	7.07×10^{-2}
$2\sqrt{m(0)}$	9.31×10^{-3}	2.76×10^{-2}	4.55×10^{-2}	6.34×10^{-2}	9.03×10^{-2}

Table 17. Exact and predicted observation errors for example 2.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
δ	0.50×10^{-2}	1.50×10^{-2}	2.50×10^{-2}	3.50×10^{-2}	5.00×10^{-2}
$2\sqrt{m(0)}$	1.69×10^{-2}	2.01×10^{-2}	2.53×10^{-2}	3.19×10^{-2}	4.37×10^{-2}

Table 18. Exact and predicted observation errors for example 3.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
δ	0.71×10^{-2}	2.12×10^{-2}	3.53×10^{-2}	4.95×10^{-2}	7.07×10^{-2}
$2\sqrt{m(0)}$	1.80×10^{-2}	1.89×10^{-2}	1.99×10^{-2}	2.09×10^{-2}	2.25×10^{-2}

Table 19. β-values given by Morozov equation ($\gamma = \infty$) with estimated δ for example 1.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
β_{opt}	0.277×10^{-5}	0.990×10^{-5}	0.172×10^{-4}	0.260×10^{-4}	0.434×10^{-4}
β_M	0.678×10^{-5}	0.273×10^{-4}	0.501×10^{-4}	0.736×10^{-4}	1.120×10^{-4}

5. Numerical predictions of observation errors

All the numerical results of the last section assumed the availability of exact observation errors δ in the exact or damped Morozov equation. However, in practical applications, the noise level may not be accessible. We now report some numerical results on the performance of the observation error prediction algorithm proposed in section 3.4.

The examples considered here are the same three examples as those chosen in section 4. Tables 16–18 give the exact observation errors and the predicted ones using the observation error prediction algorithm. The numerical methods and the quadrature rules are the same as those used in section 4. The parameter σ required in the algorithm is taken to be 1.0 in all three examples, and the initial β_0 is chosen to be 10^{-3} for example 1, 10^{-4} for example 2, and 0.1 for example 3 respectively. For example 2, we have iterated step (1) of the algorithm twice (see remark 3.1) while only once for examples 1 and 3. From these three examples, we can see that our observation error prediction algorithm appears to work well. The predicted observation errors are all of the same magnitudes as the exact ones.

One may adjust the parameter σ and even the factor 2 in the predicted observation error formula $2\sqrt{m(0)}$ to obtain much better results. However, these choices must be based on programmers' experience for a concrete applied problem.

Tables 19–21 give the β-values obtained by the Morozov principle with the exact observation error δ replaced by the estimated error $2\sqrt{m(0)}$. We observe that the estimated β-values are all of the same order as the optimal ones. We have also run the quasi-Newton

Table 20. β-values given by Morozov equation ($\gamma = \infty$) with estimated δ for example 2.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
$\beta_{\rm opt}$	0.121×10^{-5}	0.381×10^{-5}	0.685×10^{-5}	0.983×10^{-5}	0.139×10^{-4}
β_M	0.250×10^{-5}	0.285×10^{-5}	0.335×10^{-5}	0.394×10^{-5}	0.510×10^{-5}

Table 21. β-values given by Morozov equation ($\gamma = 1.3$) with estimated δ for example 3.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
$\beta_{\rm opt}$	0.430×10^{-4}	0.134×10^{-3}	0.230×10^{-3}	0.333×10^{-3}	0.502×10^{-3}
β_M	1.409×10^{-4}	0.151×10^{-3}	0.164×10^{-3}	0.176×10^{-3}	0.196×10^{-3}

Table 22. Optimal errors and the errors obtained using estimated β-values from Morozov equation.

$\hat{\delta}$	0.01	0.03	0.05	0.07	0.1
Optimal error	5.012×10^{-3}	1.476×10^{-2}	2.452×10^{-2}	3.422×10^{-2}	4.864×10^{-2}
Morozov error	7.145×10^{-3}	1.479×10^{-2}	2.476×10^{-2}	3.509×10^{-2}	5.073×10^{-2}

iteration and two-parameter algorithm when δ is replaced by $2\sqrt{m(0)}$, the convergence behaviours of the two algorithms are exactly the same as in the case with the exact δ, so we do not present those numerical results here. From tables 19–21, we find that the results seem a bit worse for example 3. However, if we compare their corresponding L^2-norm errors, there are no essential differences, see table 22 for the L^2-norm errors obtained by using optimal β-values and the estimated β-values from Morozov equation ($\gamma = 1.3$) with the exact observation error δ replaced by $2\sqrt{m(0)}$.

Acknowledgment

JZ was partially supported by Hong Kong RGC grant no CUHK 338/96E.

References

[1] Baumeister J 1986 *Stable Solutions of Inverse Problems* (Braunschweig: Vieweg)
[2] Chan T, Go S and Zou J 1998 Boundary treatments for multilevel methods on unstructured meshes *SIAM J. Sci. Comput.* accepted
[3] Chan T, Smith B and Zou J 1996 Overlapping schwarz methods on unstructured meshes using non-matching coarse grids *Numer. Math.* **73** 149–67
[4] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer Academic)
[5] Groetsch C W 1983 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Boston, MA: Pitman)
[6] Groetsch C W 1993 *Inverse Problems in the Mathematical Sciences* (Braunschweig: Vieweg)
[7] Ito K and Kunisch K 1992 On the choice of the regularization parameters in nonlinear inverse problems *SIAM J. Optim.* **2** 376–404
[8] Kunisch K 1993 On a class of damped Morozov principles *Computing* **50** 185–98
[9] Louis A K 1989 *Inverse und Schlechtgestellte Problems* (Stuttgart: Teubner)
[10] Morozov V A 1984 *Methods for Solving Incorrectly Posed Problems* (New York: Springer)