BIT manuscript No.
(will be inserted by the editor)

Fractional Tikhonov regularization for linear discrete ill-posed problems

Michiel E. Hochstenbach · **Lothar Reichel**

Received: date / Accepted: date

Abstract Tikhonov regularization is one of the most popular methods for solving linear systems of equations or linear least-squares problems with a severely ill-conditioned matrix A. This method replaces the given problem by a penalized least-squares problem. The present paper discusses measuring the residual error (discrepancy) in Tikhonov regularization with a seminorm that uses a fractional power of the Moore-Penrose pseudoinverse of AA^T as weighting matrix. Properties of this regularization method are discussed. Numerical examples illustrate that the proposed scheme for a suitable fractional power may give approximate solutions of higher quality than standard Tikhonov regularization.

Keywords Ill-posed problem · regularization · fractional Tikhonov · weighted residual norm · filter function · discrepancy principle · solution norm constraint.

Mathematics Subject Classification (2000) 65F10 · 65F22 · 65R30

Fröberg, Björck, Ruhe: A Golden Braid for 50 Years of BIT.

1 Introduction

This paper is concerned with the approximate solution of linear least-squares problems

$$\min_{x\in\mathbb{R}^n} \|Ax-b\| \tag{1.1}$$

Michiel E. Hochstenbach
Department of Mathematics and Computer Science
Eindhoven University of Technology, PO Box 513, 5600 MB, The Netherlands
E-mail: www.win.tue.nl/∼hochsten

Lothar Reichel
Department of Mathematical Sciences
Kent State University, Kent, OH 44242, USA
E-mail: reichel@math.kent.edu

with a matrix $A \in \mathbb{R}^{m\times n}$ of ill-determined rank, i.e., A has many singular values of different orders of magnitude close to the origin. In particular, A is severely ill-conditioned and may be singular. Least-squares problems with a matrix of this kind are often referred to as discrete ill-posed problems. For notational convenience, we assume that $m \geq n$, however, the methods discussed also can be applied when $m < n$. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm.

The vector $b \in \mathbb{R}^m$ represents available data, which is contaminated by an error $e \in \mathbb{R}^m$. The error may stem from measurement inaccuracies or discretization. Thus,

$$b = \hat{b} + e, \tag{1.2}$$

where $\hat{b}$ is the unknown error-free vector associated with b. We will assume the unavailable error-free system

$$Ax = \hat{b} \tag{1.3}$$

to be consistent and denote its solution of minimal Euclidean norm by $\hat{x}$. We would like to determine an approximation of $\hat{x}$ by computing a suitable approximate solution of (1.1). Due to the ill-conditioning of the matrix A and the error e in b, the solution of the least-squares problem (1.1) of minimal Euclidean norm is typically a poor approximation of $\hat{x}$.

Tikhonov regularization is a popular approach to determine an approximation of $\hat{x}$. This method replaces the minimization problem (1.1) by a penalized least-squares problem. We consider penalized least-squares problems of the form

$$\min_{x\in\mathbb{R}^n}\{\|Ax-b\|_W^2 + \mu\,\|x\|^2\}, \tag{1.4}$$

where $\|x\|_W = (x^TWx)^{1/2}$ and W is a symmetric positive semidefinite matrix. The superscript ^{T} denotes transposition. The problem (1.4) has a unique solution x_μ for all positive values of the regularization parameter μ. The value of μ determines how sensitive x_μ is to the error e in b, and how much x_μ differs from the desired solution $\hat{x}$ of (1.3). We propose to let

$$W = (AA^T)^{(\alpha-1)/2} \tag{1.5}$$

for a suitable value of $\alpha > 0$, where A^T denotes the transpose of A. When $\alpha < 1$, we define W with the aid of the Moore-Penrose pseudoinverse of AA^T. The seminorm $\|\cdot\|_W$ allows the parameter α to be chosen to improve the quality of the computed solution $x_{\mu,\alpha}$ of (1.4). We refer to (1.4) with W given by (1.5) as the *weighted Tikhonov method* or as the *fractional Tikhonov method*. Standard Tikhonov regularization based on the Euclidean norm is obtained when $\alpha = 1$. Then W is the identity matrix. Recently, Klann and Ramlau [7] proposed a fractional Tikhonov regularization method different from (1.4)–(1.5). We comment on their approach in Sections 2 and 6.

The normal equations associated with the Tikhonov minimization problem (1.4) with W defined by (1.5) are given by

$$((A^TA)^{(\alpha+1)/2} + \mu I)x = (A^TA)^{(\alpha-1)/2}A^Tb. \tag{1.6}$$

Their solution $x = x_{\mu,\alpha}$ is uniquely determined for any $\mu > 0$ and $\alpha > 0$. When A is of small to moderate size, $x_{\mu,\alpha}$ can be conveniently computed from (1.6) with the aid of the singular value decomposition of A; see Section 3. Large-scale problems can be solved by first projecting them onto a subspace of small dimension, e.g., a low-dimensional Krylov subspace, and then applying the approach of Section 3 to the projected problem. This is described in Section 4.

The present paper is organized as follows. Section 2 discusses properties of filter functions associated with (1.4) and other regularization methods. The determination of μ and α so that the solution of (1.4) satisfies the discrepancy principle is considered in Section 3 for small problems and in Section 4 for large ones. Perturbation bounds are derived in Section 5, and Section 6 reports a few computed results. Section 7 contains concluding remarks.

2 Filter functions

Introduce the singular value decomposition (SVD),

$$A = U\Sigma V^T, \tag{2.1}$$

where $U = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{m\times m}$ and $V = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and

$$\Sigma = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] \in \mathbb{R}^{m\times n}.$$

The singular values are ordered according to

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0,$$

where the index r is the rank of A; see, e.g., [4] for discussions on properties and the computation of the SVD. We first review filter functions for some popular solution methods in Subsections 2.1–2.3. Note that this list is far from complete; for instance, we do not mention the exponential approach of [2]. Some desirable properties of filter functions are summarized in Subsection 2.4, and Subsection 2.5 discusses properties of filter functions associated with (1.6).

2.1 Truncated SVD

Approximate solutions of (1.1) determined by the truncated SVD (TSVD) are of the form

$$x_{\mathrm{tsvd}} = \sum_{j=1}^{k} \frac{1}{\sigma_j} (u_j^T b)\, v_j \tag{2.2}$$

for some cut-off parameter $k \leq r$.

It is convenient to express approximate solutions $\widetilde{x}$ of (1.1) with the aid of filter functions φ, i.e.,

$$\widetilde{x} = \sum_{j=1}^{n} \varphi(\sigma_j)(u_j^T b) v_j; \tag{2.3}$$

see, e.g., [5] for a discussion on filter functions. For instance, the approximate solution (2.2) can be written as

$$x_{\text{tsvd}} = \sum_{j=1}^{n} \varphi(\sigma_j)\,(u_j^T b)\,v_j,$$

where

$$\varphi(\sigma) = \varphi_{\text{tsvd}}(\sigma) = \begin{cases} 1/\sigma & \text{if } \sigma \geq \tau, \\ 0 & \text{otherwise,} \end{cases}$$

and $\sigma_{k+1} < \tau \leq \sigma_k$ is arbitrary.

2.2 Tikhonov

Standard Tikhonov regularization (1.4) (with $W = I$) corresponds to the filter function

$$\varphi_{\text{tikh}}(\sigma) = \frac{\sigma}{\sigma^2 + \mu},$$

where $\mu > 0$ is the regularization parameter. The asymptotics of this function are

$$\begin{aligned} \varphi_{\text{tikh}}(\sigma) &= \frac{\sigma}{\mu} + \mathscr{O}(\sigma^3) \quad (\sigma \searrow 0), \\ \varphi_{\text{tikh}}(\sigma) &= \sigma^{-1} + \mathscr{O}(\sigma^{-3}) \quad (\sigma \to \infty). \end{aligned}$$

Figure 2.1 displays functions φ_{tikh} for several values of μ.

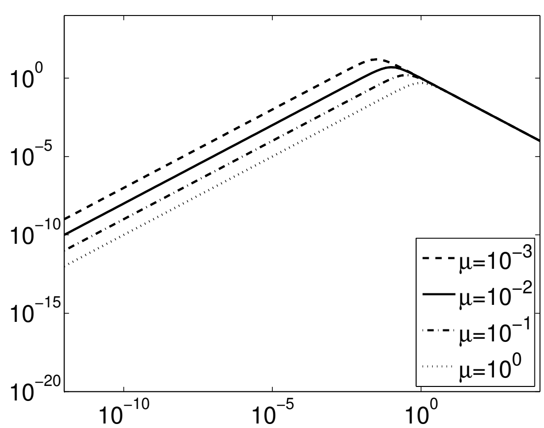

Fig. 2.1 Filter functions $\varphi_{\text{tikh}}(\sigma) = \frac{\sigma}{\sigma^2+\mu}$ for $\mu = 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}$, and $10^{-12} \leq \sigma \leq 10^{4}$. Note the logarithmic scales.

2.3 Klann and Ramlau's filter functions

Klann and Ramlau [7] consider the family of filter functions

$$\varphi_{\mathrm{KR}}(\sigma) = \frac{\sigma^{2\gamma-1}}{(\sigma^2+\mu)^\gamma} \tag{2.4}$$

with parameter $\gamma > 1/2$. Its asymptotics are

$$\varphi_{\mathrm{KR}}(\sigma) = \frac{\sigma^{2\gamma-1}}{\mu^\gamma} + \mathscr{O}(\sigma^{2\gamma+1}) \quad (\sigma \searrow 0), \qquad \varphi_{\mathrm{KR}}(\sigma) = \sigma^{-1} + \mathscr{O}(\sigma^{-3}) \quad (\sigma \to \infty).$$

Standard Tikhonov regularization is recovered for $\gamma = 1$. The analogue of the normal equations (1.6) associated with the filter function (2.4) is given by

$$(A^TA+\mu I)^\gamma x = (A^TA)^{(\gamma-1)}A^Tb. \tag{2.5}$$

For $\gamma \neq 1$, this equation is not related to a simple minimization problem of the form (1.4).

2.4 Desirable properties of filter functions

We now discuss some desirable properties of filter functions. Equation (2.3) yields

$$\begin{aligned} \|\widetilde{x}\|^2 &= \sum_{j=1}^{n} (\varphi(\sigma_j))^2\,(u_j^Tb)^2, \\ \|b-A\widetilde{x}\|^2 &= \sum_{j=1}^{n} (1-\sigma_j\varphi(\sigma_j))^2\,(u_j^Tb)^2 + \sum_{j=n+1}^{m} (u_j^Tb)^2. \end{aligned} \tag{2.6}$$

To get a small residual norm for matrices with large singular values, we require in view of (2.6) that

$$\varphi(\sigma) = \sigma^{-1} + o(\sigma^{-1}) \quad (\sigma \to \infty). \tag{2.7}$$

Moreover, we would like the filter function to satisfy

$$\varphi(\sigma) = o(1) \quad (\sigma \searrow 0). \tag{2.8}$$

This ensures that the computed approximate solution $\widetilde{x}$ of (1.1) only contains small multiples of singular vectors associated with small singular values. These singular vectors usually represent high-frequency oscillations.

The above filter functions differ in how quickly they converge to zero when σ decreases to zero. Fast convergence implies significant smoothing of the computed approximate solution (2.3). The interest of Klann and Ramlau [7] in the filter functions (2.4) stems from the fact that they provide less smoothing for $1/2 < \gamma < 1$ than φ_{tikh}.

2.5 Fractional Tikhonov

We turn to the family of filter functions associated with weighted Tikhonov regularization (1.4) with W given by (1.5). The following properties are easy to show and illustrate that these filter functions satisfy the desirable properties of filter functions stated in the previous subsection.

Proposition 2.1 *The filter function for weighted Tikhonov regularization* (1.4) *with* W *defined by* (1.5) *for some* $\alpha > 0$ *is given by*

$$\varphi_{\mathrm{tikh},W}(\sigma) = \frac{\sigma^{\alpha}}{\sigma^{\alpha+1}+\mu}. \tag{2.9}$$

It has the asymptotics

$$\varphi_{\mathrm{tikh},W}(\sigma) = \sigma^{-1} + \mathscr{O}(\sigma^{-(\alpha+2)}) \quad (\sigma \to \infty),$$
$$\varphi_{\mathrm{tikh},W}(\sigma) = \frac{\sigma^{\alpha}}{\mu} + \mathscr{O}(\sigma^{2\alpha+1}) \quad (\sigma \searrow 0).$$

In particular, $\varphi_{\mathrm{tikh},W}$ *satisfies* (2.7) *and* (2.8).

The asymptotic behavior of $\varphi_{\mathrm{tikh},W}(\sigma)$ as $\sigma \searrow 0$ shows this function to provide less smoothing than φ_{tikh} for $0 < \alpha < 1$. Figure 2.2 displays the behavior of the functions $\varphi_{\mathrm{tikh},W}$ for $\mu = 10^{-2}$ and several values of α. A comparison with Figure 2.1 shows that components of the solution (2.3) associated with "tiny" singular values are damped less by the function $\varphi_{\mathrm{tikh,W}}$ than by φ_{tikh}. This often yields computed approximate solutions of (1.1) of higher quality than with standard Tikhonov regularization.

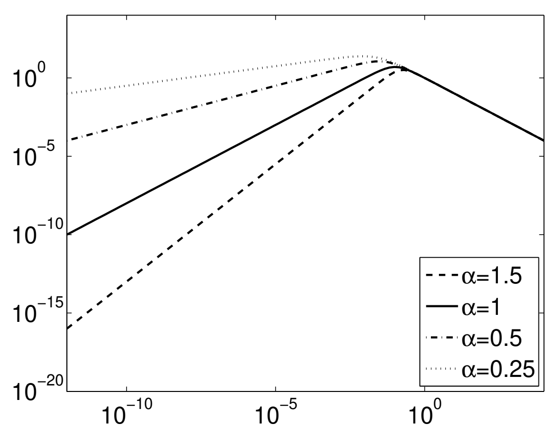

Fig. 2.2 Filter functions $\varphi_{\mathrm{tikh},W}(\sigma) = \frac{\sigma^{\alpha}}{\sigma^{\alpha+1}+\mu}$ for $\alpha = 0.25, 0.5, 1, 1.5$, $\mu = 10^{-2}$, and $10^{-12} \le \sigma \le 10^{4}$.

3 Choosing μ and α

We first investigate the dependence of the solution $x_{\mu,\alpha}$ of (1.6) on the parameters μ and α. This is conveniently carried out with the help of the SVD of A. Subsequently, we determine μ with the discrepancy principle and study how the computed solutions vary with α. The situation when $x_{\mu,\alpha}$ is required to be of specified norm is also considered.

Substituting the SVD (2.1) into (1.6) yields

$$((\Sigma^T\Sigma)^{(\alpha+1)/2}+\mu I)y=(\Sigma^T)^{\alpha}U^Tb.$$

Denote the solution by $y_{\mu,\alpha}$. Then $x_{\mu,\alpha}=Vy_{\mu,\alpha}$ solves (1.6), and

$$\|x_{\mu,\alpha}\|^2=\|y_{\mu,\alpha}\|^2=\sum_{j=1}^{r}\frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1}+\mu)^2}\,(u_j^Tb)^2, \tag{3.1}$$

where r is the rank of A. Thus,

$$\frac{\partial}{\partial\mu}\,\|x_{\mu,\alpha}\|^2=-2\sum_{j=1}^{r}\frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1}+\mu)^3}\,(u_j^Tb)^2. \tag{3.2}$$

Clearly, $\mu\mapsto\|x_{\mu,\alpha}\|^2$ is a monotonically decreasing function. Similarly,

$$\frac{\partial}{\partial\alpha}\,\|x_{\mu,\alpha}\|^2=2\mu\sum_{j=1}^{r}\frac{\log(\sigma_j)\sigma_j^{-\alpha}}{(\sigma_j+\mu\sigma_j^{-\alpha})^3}\,(u_j^Tb)^2.$$

We may rescale the problem (1.1) so that $\|A\|<1$. Then $\log(\sigma_j)<0$ and it follows that $\alpha\mapsto\|x_{\mu,\alpha}\|^2$ is monotonically decreasing. We assume this scaling in the present section.

The choice of the regularization parameter μ depends on the amount of error e in b. Consider for the moment standard Tikhonov regularization, i.e., the situation when $\alpha=1$. Generally, the larger $\|e\|$, the larger μ should be; see, e.g., Proposition 3.1 below. However, it follows from (3.2) that increasing μ decreases the norm of the computed solution $x_{\mu,1}$. Therefore, the computed solution may be of significantly smaller norm than the desired solution $\hat{x}$. This difficulty can be remedied by choosing $\alpha<1$, because this increases the norm of the computed solution. Computed examples in Section 6 illustrate that, indeed, $\alpha<1$ typically yields more accurate approximations of $\hat{x}$ than $\alpha=1$.

We turn to the situation when a fairly accurate bound for the error in b,

$$\|e\|\leq\varepsilon,$$

is available. Then we can apply the discrepancy principle to determine a suitable value of the regularization parameter μ. Let $\alpha>0$ be fixed and define

$$\delta=\eta\varepsilon, \tag{3.3}$$

where $\eta > 1$ is a user-supplied constant independent of ε. We would like to determine $\mu > 0$, so that the solution $x_{\mu,\alpha}$ of (1.4) satisfies

$$\|b - Ax_{\mu,\alpha}\| = \delta. \tag{3.4}$$

Then the vector $x_{\mu,\alpha}$ is said to satisfy the discrepancy principle; see, e.g., [5] for discussions on this choice of regularization parameter.

The change of variable $\lambda = \mu^{-1}$ gives a simple expression for the coefficients in the leftmost sum in (2.6) with $\varphi = \varphi_{\mathrm{tikh},W}$,

$$1 - \sigma_j \, \frac{\sigma_j^{\alpha}}{\sigma_j^{\alpha+1} + \mu} = 1 - \frac{\lambda\, \sigma_j^{\alpha+1}}{\lambda\, \sigma_j^{\alpha+1} + 1} = \frac{1}{\lambda\, \sigma_j^{\alpha+1} + 1}.$$

The solution of (3.4) for $\mu > 0$ is equivalent to the computation of the positive zero of the function

$$F_\alpha(\lambda) = \sum_{j=1}^{r} (\lambda\,\sigma_j^{\alpha+1} + 1)^{-2} \, (u_j^T b)^2 + \sum_{j=r+1}^{m} (u_j^T b)^2 - \delta^2. \tag{3.5}$$

We are in a position to show how μ, such that $x_{\mu,\alpha}$ satisfies (3.4) for fixed $\alpha > 0$, depends on δ.

Proposition 3.1 *Let $\mu = \mu(\delta) > 0$ be such that $x_{\mu,\alpha}$ satisfies* (3.4) *for fixed $\alpha > 0$. Then $d\mu/d\delta > 0$.*

Proof Consider $\lambda(\delta) = 1/\mu(\delta)$. It follows from (3.5) that the inverse function satisfies

$$\delta(\lambda)^2 = \sum_{j=1}^{r} (\lambda\,\sigma_j^{\alpha+1} + 1)^{-2} \, (u_j^T b)^2 + \sum_{j=r+1}^{m} (u_j^T b)^2.$$

Differentiating with respect to λ yields

$$2\,\delta(\lambda)\,\delta'(\lambda) = -2 \sum_{j=1}^{r} \frac{\sigma_j^{\alpha+1}}{(\lambda\,\sigma_j^{\alpha+1} + 1)^3} \, (u_j^T b)^2.$$

It follows that $\delta'(\lambda) < 0$. Consequently, $\lambda'(\delta) < 0$ and $\mu'(\delta) > 0$.

We consider properties of Newton's method when applied to the computation of the positive zero of the function (3.5). However, other zero-finders also can be used. A discussion of Newton's method and other zero-finders for the situation when $\alpha = 1$ is provided in [9].

Proposition 3.2 *Newton's method applied to the computation of the positive zero of F_α with initial iterate $\lambda_0 = 0$ converges quadratically and monotonically.*

Proof The quadratic convergence is a consequence of the analyticity of $F_\alpha(\lambda)$ in a neighborhood of the positive real axis in the complex plane. The monotonic convergence follows from the fact that for every fixed $\alpha > 0$ and $\lambda \geq 0$, the function F_α satisfies $F_\alpha'(\lambda) < 0$ and $F_\alpha''(\lambda) > 0$.

Let $\alpha > 0$ and let $\mu = \mu(\alpha)$ be determined so that $x_{\mu,\alpha}$ satisfies the discrepancy principle. The following result shows how $x_{\mu,\alpha}$ depends on $\alpha > 0$.

Proposition 3.3 *Let for $\alpha > 0$ the regularization parameter $\mu = \mu(\alpha)$ be such that $x_{\mu,\alpha}$ satisfies* (3.4)*. Then there is an open real interval Ω containing unity such that* $\underset{\alpha\in\Omega}{argmin}\,\|x_{\mu(\alpha),\alpha}\| = 1$.

Proof The equation $F_\alpha(\lambda) = 0$ can be expressed as

$$\sum_{j=1}^{r} \frac{\mu^2}{(\sigma_j^{\alpha+1}+\mu)^2}\,(u_j^T b)^2 = \delta^2 - \sum_{j=r+1}^{m} (u_j^T b)^2. \tag{3.6}$$

We may consider $\mu = \mu(\alpha)$ a function of α. Implicit differentiation of (3.6) with respect to α yields

$$2\mu \sum_{j=1}^{r} \frac{\sigma_j^{\alpha+1}(\mu' - \mu\log(\sigma_j))}{(\sigma_j^{\alpha+1}+\mu)^3}\,(u_j^T b)^2 = 0, \tag{3.7}$$

which, since $\mu > 0$, implies that

$$\sum_{j=1}^{r} \xi_j\,(\mu' - \mu\log(\sigma_j)) = 0, \tag{3.8}$$

where

$$\xi_j = \frac{\sigma_j^{\alpha+1}}{(\sigma_j^{\alpha+1}+\mu)^3}\,(u_j^T b)^2. \tag{3.9}$$

Introduce the function

$$G(\alpha) = \|x_{\mu(\alpha),\alpha}\|^2 = \sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1}+\mu)^2}\,(u_j^T b)^2.$$

Then

$$\begin{aligned} G'(\alpha) &= \sum_{j=1}^{r} \frac{2\sigma_j^{2\alpha}\log(\sigma_j)(\sigma_j^{\alpha+1}+\mu) - 2\sigma_j^{2\alpha}(\sigma_j^{\alpha+1}\log(\sigma_j)+\mu')}{(\sigma_j^{\alpha+1}+\mu)^3}\,(u_j^T b)^2 \\ &= \sum_{j=1}^{r} \frac{2\sigma_j^{2\alpha}(\log(\sigma_j)\mu - \mu')}{(\sigma_j^{\alpha+1}+\mu)^3}\,(u_j^T b)^2 \\ &= 2\sum_{j=1}^{r} \xi_j \sigma_j^{\alpha-1}(\mu\log(\sigma_j) - \mu'). \end{aligned}$$

It follows from (3.8) that $G'(1) = 0$. Moreover, differentiating (3.8) yields

$$\sum_{j=1}^{r} \left\{ \xi_j'(\mu' - \mu\log(\sigma_j)) + \xi_j(\mu'' - \mu'\log(\sigma_j)) \right\} = 0. \tag{3.10}$$

Since

$$G''(\alpha) = 2 \sum_{j=1}^{r} \sigma_j^{\alpha-1} \left\{ (\xi_j' + \xi_j \log(\sigma_j))(\mu \log(\sigma_j) - \mu') + \xi_j (\mu' \log(\sigma_j) - \mu'') \right\},$$

we obtain, in view of (3.10),

$$G''(1) = 2 \sum_{j=1}^{r} \xi_j \log(\sigma_j)(\mu \log(\sigma_j) - \mu').$$

The above sum is obtained by multiplying the terms in (3.8) by the positive weights $-\log(\sigma_j)$; the largest weights multiply the largest terms. Therefore, $G''(1) > 0$. By continuity, $G''(\alpha)$ is positive in a neighborhood Ω of $\alpha = 1$. Thus, $G(\alpha)$ has a local minimum at $\alpha = 1$.

For some linear discrete ill-posed problems (1.1) an estimate Δ of the norm of the desired solution $\hat{x}$ may be known. Then it may be desirable to require the computed solution $x_{\mu,\alpha}$ to be of the same norm, i.e.,

$$\Delta = \|x_{\mu,\alpha}\|. \tag{3.11}$$

This type of problems is discussed in [3, 8, 10]. The following result sheds light on how $\|Ax_{\mu,\alpha} - b\|$ depends on α for solutions that satisfy (3.11).

Proposition 3.4 *Let, for $\alpha > 0$, the regularization parameter $\mu = \mu(\alpha)$ be such that $x_{\mu,\alpha}$ satisfies* (3.11)*. Then there is an open real interval Ω containing unity, such that* $\underset{\alpha\in\Omega}{\mathit{argmin}} \|b - Ax_{\mu(\alpha),\alpha}\| = 1$.

Proof This result is shown in a similar fashion as Proposition 3.3. Differentiating the right-hand side and left-hand side of (3.11) with respect to α, keeping in mind that $\mu = \mu(\alpha)$, gives analogously to (3.8) the equation

$$\sum_{j=1}^{r} \zeta_j (\mu \log(\sigma_j) - \mu') = 0, \qquad \zeta_j = \xi_j \sigma_j^{\alpha-1}, \tag{3.12}$$

where ξ_j is defined by (3.9). Introduce the function

$$H(\alpha) = \|b - Ax_{\mu(\alpha),\alpha}\|^2 = \sum_{j=1}^{r} \frac{\mu^2}{(\sigma_j^{\alpha+1} + \mu)^2} (u_j^T b)^2 + \sum_{j=r+1}^{m} (u_j^T b)^2, \tag{3.13}$$

where the right-hand side is obtained by substituting (2.9) into (2.6). Then (cf. (3.7))

$$H'(\alpha) = 2\mu \sum_{j=1}^{r} \zeta_j \sigma_j^{1-\alpha} (\mu' - \mu \log(\sigma_j))$$

and it follows from (3.12) that $H'(1) = 0$.

The representation

$$H'(\alpha) = 2\mu \sum_{j=1}^{r} \xi_j (\mu' - \mu \log(\sigma_j))$$

conveniently can be differentiated to give

$$\begin{aligned} H''(\alpha) = {} & 2\mu' \sum_{j=1}^{r} \xi_j \left(\mu' - \mu \log(\sigma_j)\right) \\ & + 2\mu \sum_{j=1}^{r} \left\{ \xi_j' \left(\mu' - \mu \log(\sigma_j)\right) + \xi_j \left(\mu'' - \mu' \log(\sigma_j)\right)\right\}. \end{aligned} \tag{3.14}$$

Differentiating (3.12) yields

$$\sum_{j=1}^{r} \zeta_j' \left(\mu \log(\sigma_j) - \mu'\right) + \zeta_j \left(\mu' \log(\sigma_j) - \mu''\right) = 0. \tag{3.15}$$

Let $\alpha = 1$. Then $\zeta_j = \xi_j$ for all j. Using this property when substituting (3.15) into (3.14) gives, in view of (3.12),

$$H''(1) = 2\mu \sum_{j=1}^{r} (\xi_j' - \zeta_j')(\mu' - \mu \log(\sigma_j)). \tag{3.16}$$

It follows from $\xi_j = \zeta_j \sigma_j^{1-\alpha}$ that, for $\alpha = 1$, $\xi_j' = \zeta_j' - \zeta_j \log(\sigma_j)$. Substituting the latter expression into (3.16) yields

$$H''(1) = -2\mu \sum_{j=1}^{r} \zeta_j (\mu' - \mu \log(\sigma_j)) \log(\sigma_j).$$

Comparing this sum with (3.12) shows that $H''(1) > 0$, similarly as the analogous result for $G''(1)$ in the proof of Proposition 3.3. By continuity, H is convex in a neighborhood Ω of $\alpha = 1$.

Propositions 3.3 and 3.4 show the choice $\alpha = 1$, which corresponds to standard Tikhonov regularization, to be quite natural; by Proposition 3.3 this choice minimizes $\|x_{\mu(\alpha),\alpha}\|$ locally when the residual norm $\|b - Ax_{\mu(\alpha),\alpha}\|$ is specified and by Propositions 3.4 the residual norm has a local minimum for $\alpha = 1$ when $\|x_{\mu(\alpha),\alpha}\|$ is specified. We remark that the value of δ used in Proposition 3.3 does not have to be defined by (3.3) and, similarly, the value of Δ in Proposition 3.4 does not have to be close to $\|\hat{x}\|$. However, despite these properties of standard Tikhonov regularization, numerical examples of Section 6 illustrate that $\alpha < 1$ may yield more accurate approximations of $\hat{x}$.

4 Large-scale problems

The solution method described in the previous section, based on first computing the SVD of A, is too expensive to be applied to large problems. We therefore propose to project large-scale problems onto a Krylov subspace of small dimension and then apply the solution method of Section 3 to the small problem so obtained. For instance, application of ℓ steps of Lanczos bidiagonalization to A with initial vector $b/\|b\|$ yields the decompositions

$$AV_\ell = U_{\ell+1}\bar{C}_\ell, \qquad A^T U_\ell = V_\ell C_\ell^T, \qquad U_{\ell+1} e_1 = b/\|b\|, \tag{4.1}$$

where the matrices $U_{\ell+1} \in \mathbb{R}^{m\times(\ell+1)}$ and $V_\ell \in \mathbb{R}^{n\times\ell}$ have orthonormal columns, and the lower bidiagonal matrix $\bar{C}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ has positive subdiagonal entries. Moreover, $U_\ell \in \mathbb{R}^{m\times\ell}$ is made up of the ℓ first columns of $U_{\ell+1}$, $C_\ell \in \mathbb{R}^{\ell\times\ell}$ consists of the first ℓ rows of $\bar{C}_\ell$, and $e_1 = [1,0,\ldots,0]^T$ denotes the first axis vector. The columns of V_ℓ span the Krylov subspace

$$\mathscr{K}_\ell(A^TA, A^Tb) = \mathrm{span}\{A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{\ell-1}A^Tb\}; \tag{4.2}$$

see, e.g., [1] for a discussion. The number of bidiagonalization steps, ℓ, is generally chosen quite small; we assume ℓ to be small enough so that the decompositions (4.1) with the stated properties exist.

It follows from (4.1) that

$$\min_{x\in\mathscr{K}_\ell(A^TA,A^Tb)} \|Ax-b\| = \min_{y\in\mathbb{R}^\ell} \|\bar{C}_\ell y - e_1\|b\|\,\|. \tag{4.3}$$

Thus, application of ℓ steps of Lanczos bidiagonalization reduces the large minimization problem (1.1) to the small minimization problem in the right-hand of (4.3). We apply the fractional Tikhonov method to the latter problem as described in Section 3. A numerical illustration can be found in Section 6.

5 Sensitivity analysis

This section studies the sensitivity of the regularization parameter μ in (1.6) to perturbations in the discrepancy $\delta = \eta\varepsilon$ in (3.4) and to changes in

$$\Delta = \|\hat{x}\|. \tag{5.1}$$

Our analysis is motivated by the fact that only approximations of ε and Δ may be available. In this section $\|A\|$ is arbitrarily large.

It is convenient to let μ_{d} denote the solution of (3.4) and to let $x_{\mathrm{d}} = x_{\mu_{\mathrm{d}},\alpha}$ be the associated solution of (1.6). Since we keep the parameter α fixed in this section, we will not explicitly indicate the dependence of μ_{d} and x_{d} on α. Similarly, let μ_{n} denote the value of the regularization parameter such that $\|x_{\mu_{\mathrm{n}}}\| = \Delta$, where Δ is given by (5.1), and define $x_{\mathrm{n}} = x_{\mu_{\mathrm{n}}}$. We will also need the residual error $r_{\mathrm{d}} = b - Ax_{\mathrm{d}}$.

It can be shown that for δ sufficiently large,

$$\mu_{\mathrm{n}} < \mu_{\mathrm{d}}, \qquad \|x_{\mathrm{d}}\| < \|x_{\mathrm{n}}\|.$$

The following bounds shed some light on the sensitivity of $\mu_{\mathrm{n}} = \mu_{\mathrm{n}}(\Delta)$ and $\mu_{\mathrm{d}} = \mu_{\mathrm{d}}(\delta)$ to perturbations in Δ and δ, respectively. The lower bound involves the constant

$$\delta_-^2 = \sum_{j=1}^{r} \frac{\mu_{\mathrm{d}}^2}{(\sigma_j^{\alpha+1} + \mu_{\mathrm{d}})^2}\,(u_j^Tb)^2,$$

which can also be expressed as

$$\delta_-^2 = \delta^2 - \sum_{j=r+1}^{m} (u_j^Tb)^2;$$

cf. (3.13). In particular, $\delta_-^2 = \delta^2$ for consistent least-squares problems (1.1). When A is square, the discrete ill-posed least-squares problems considered are typically consistent with a severely ill-conditioned matrix.

Proposition 5.1 *The following bounds hold,*

$$\frac{\mu_{\rm n}}{\Delta} \leq |\mu_{\rm n}'(\Delta)| \leq \frac{\|A\|^{\alpha+1} + \mu_{\rm n}}{\Delta} \tag{5.2}$$

and

$$\max\left\{ \frac{\delta}{\|A\|^{1-\alpha}\|x_{\rm d}\|^2},\ \frac{\delta\,\mu_{\rm d}^2}{\delta_-^2} \right\} \leq\ \mu_{\rm d}'(\delta). \tag{5.3}$$

Proof To show the inequalities (5.2), we express the constraint $\|x_{\rm n}\|^2 = \Delta^2$ in terms of the singular value decomposition (2.1),

$$\sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1} + \mu_{\rm n})^2}\,(u_j^T b)^2 = \Delta^2; \tag{5.4}$$

cf. (3.1). Considering $\mu_{\rm n}$ a function of Δ and differentiating (5.4) with respect to Δ gives

$$\mu_{\rm n}'(\Delta) = -\Delta\left(\sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1} + \mu_{\rm n})^3}\,(u_j^T b)^2 \right)^{-1}.$$

Therefore, $\mu_{\rm n}'(\Delta) < 0$ and

$$\left|\mu_{\rm n}'(\Delta)\right| \leq \Delta\,(\sigma_1^{\alpha+1} + \mu_{\rm n}) \left(\sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1} + \mu_{\rm n})^2}\,(u_j^T b)^2 \right)^{-1} = \frac{\|A\|^{\alpha+1} + \mu_{\rm n}}{\Delta}.$$

Moreover,

$$\left|\mu_{\rm n}'(\Delta)\right| \geq \Delta\,\mu_{\rm n} \left(\sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1} + \mu_{\rm n})^2}\,(u_j^T b)^2 \right)^{-1} = \frac{\mu_{\rm n}}{\Delta}.$$

We turn to the lower bounds (5.3). The discrepancy principle determines the regularization parameter $\mu_{\rm d} = \mu_{\rm d}(\delta)$ so that $\|r_{\rm d}\|^2 = \delta^2$, which can be written as

$$\sum_{j=1}^{r} \frac{\mu_{\rm d}^2}{(\sigma_j^{\alpha+1} + \mu_{\rm d})^2}\,(u_j^T b)^2 + \sum_{j=r+1}^{m} (u_j^T b)^2 = \delta^2.$$

Differentiating this expression with respect to δ yields

$$\mu_{\rm d}'(\delta) = \mu_{\rm d}^{-1}\delta \left(\sum_{j=1}^{r} \frac{\sigma_j^{\alpha+1}}{(\sigma_j^{\alpha+1} + \mu_{\rm d})^3}\,(u_j^T b)^2 \right)^{-1}.$$

It follows from

$$\sum_{j=1}^{r} \frac{\sigma_j^{\alpha+1}}{(\sigma_j^{\alpha+1}+\mu_{\mathrm{d}})^3}\,(u_j^T b)^2 = \frac{1}{\mu_{\mathrm{d}}^2} \sum_{j=1}^{r} \frac{\sigma_j^{\alpha+1}}{\sigma_j^{\alpha+1}+\mu_{\mathrm{d}}}\,\frac{\mu_{\mathrm{d}}^2}{(\sigma_j^{\alpha+1}+\mu_{\mathrm{d}})^2}\,(u_j^T b)^2$$

$$\leq \frac{1}{\mu_{\mathrm{d}}^2} \sum_{j=1}^{r} \frac{\mu_{\mathrm{d}}^2}{(\sigma_j^{\alpha+1}+\mu_{\mathrm{d}})^2}\,(u_j^T b)^2 = \frac{\delta_-^2}{\mu_{\mathrm{d}}^2}$$

that

$$\mu_{\mathrm{d}}'(\delta) \geq \frac{\delta\,\mu_{\mathrm{d}}}{\delta_-^2}.$$

Alternatively, we may substitute the bound

$$\frac{\sigma_j^{1-\alpha}}{\sigma_j^{\alpha+1}+\mu_{\mathrm{d}}} \leq \frac{\sigma_j^{1-\alpha}}{\mu_{\mathrm{d}}}$$

into

$$\mu_{\mathrm{d}}'(\delta) = \mu_{\mathrm{d}}^{-1}\delta \left(\sum_{j=1}^{r} \frac{\sigma_j^{1-\alpha}}{\sigma_j^{\alpha+1}+\mu_{\mathrm{d}}}\,\frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1}+\mu_{\mathrm{d}})^2}\,(u_j^T b)^2 \right)^{-1}$$

to obtain

$$\mu_{\mathrm{d}}'(\delta) \geq \frac{\delta}{\mu_{\mathrm{d}}} \left(\frac{\sigma_1^{1-\alpha}}{\mu_{\mathrm{d}}} \sum_{j=1}^{r} \frac{\sigma_j^{2\alpha}}{(\sigma_j^{\alpha+1}+\mu_{\mathrm{d}})^2}\,(u_j^T b)^2 \right)^{-1} = \frac{\delta}{\|A\|^{1-\alpha}\|x_d\|^2}.$$

Using elementary computations, we can also bound the sensitivity of the solution and residual norms to perturbations in μ to first order.

Corollary 5.1 *We have*

$$\frac{\Delta}{\|A\|^{\alpha+1}+\mu_{\mathrm{n}}} \leq |\Delta'(\mu_{\mathrm{n}})| \leq \frac{\Delta}{\mu_{\mathrm{n}}}$$

and

$$\delta'(\mu_{\mathrm{d}}) \leq \min\left\{ \frac{\|A\|^{1-\alpha}\|x_{\mathrm{d}}\|^2}{\delta},\ \frac{\delta_-^2}{\delta\,\mu_{\mathrm{d}}} \right\}.$$

6 Computed examples

We show numerical experiments carried out for ten linear discrete ill-posed problems from Regularization Tools [6]. These problems are discretized Fredholm integral equations of the first kind. The matrices A for all problems are square. The small problems solved by the method of Section 3 are of order 100; the large-scale problems solved as described in Section 4 are of order 5000. MATLAB codes in [6] determine these matrices and the solutions $\hat{x}$ from which we compute the error-free right-hand side (assumed unknown) of (1.3) by $\hat{b} = A\hat{x}$. The vector b in (1.1) is determined from (1.2), where the entries of the "error vector" e are normally distributed random numbers with zero mean, scaled to correspond to a desired error-level $\|e\|/\|\hat{b}\|$. In the experiments, we consider the error-levels 1%, 5%, and 10%.

Experiment 6.1 We show the performance of the method described in Section 3 when applied to problems (1.1) with matrices of order 100. Tables 6.1–6.3 display the relative errors (qualities) $\|x_{\mu,\alpha} - \hat{x}\| / \|\hat{x}\|$ for standard Tikhonov ($\alpha = 1$, labeled "Tikh"), for fractional Tikhonov for the α-values 0.8, 0.6, and 0.4 (labeled "Frac"), as well as the ratios "Frac"/"Tikh". The regularization parameter μ is determined by the discrepancy principle, i.e., so that $x_{\mu,\alpha}$ satisfies (3.4) with δ given by (3.3), where $\varepsilon = \|e\|$ and $\eta = 1.1$.

The tables show improved accuracy of the computed solutions for the vast majority of the problems. The choice $\alpha = 0.8$ gives better results than $\alpha = 1$ for almost all examples. Smaller values of α, such as $\alpha = 0.5$ work even better for some examples, at the cost of yielding worse results for others. A general rule-of-thumb is that the larger the error-level, the more advantageous it is to let $\alpha < 1$; for smaller error-levels, fractional Tikhonov can be seen to perform best for α-values close to unity.

Table 6.1 Qualities of Tikhonov, fractional Tikhonov, and their ratios for various $n = 100$ examples for different error-levels (1%, 5%, 10%) and $\alpha = 0.8$.

Error	1%			5%			10%		
Problem	Tikh	Frac	Ratio	Tikh	Frac	Ratio	Tikh	Frac	Ratio
baart	2.1e-1	2.0e-1	9.7e-1	3.3e-1	3.1e-1	9.6e-1	3.6e-1	3.5e-1	9.8e-1
deriv2-1	2.8e-1	2.7e-1	9.8e-1	3.8e-1	3.7e-1	9.7e-1	4.4e-1	4.3e-1	9.7e-1
deriv2-2	2.7e-1	2.6e-1	9.7e-1	3.8e-1	3.7e-1	9.7e-1	4.4e-1	4.3e-1	9.6e-1
deriv2-3	3.8e-2	3.4e-2	9.0e-1	7.0e-2	6.0e-2	8.5e-1	9.3e-2	8.2e-2	8.9e-1
foxgood	4.3e-2	4.2e-2	9.6e-1	1.4e-1	1.2e-1	8.7e-1	2.0e-1	1.8e-1	8.8e-1
gravity	3.7e-2	2.9e-2	8.0e-1	7.4e-2	6.5e-2	8.7e-1	1.1e-1	1.0e-1	9.0e-1
heat	1.5e-1	1.5e-1	1.0e-0	3.1e-1	3.1e-1	9.9e-1	4.4e-1	4.3e-1	9.7e-1
ilaplace	1.6e-1	1.5e-1	9.6e-1	2.0e-1	1.9e-1	9.6e-1	2.2e-1	2.1e-1	9.7e-1
phillips	2.8e-2	3.1e-2	1.1e-0	6.5e-2	6.2e-2	9.6e-1	1.1e-1	1.0e-1	9.5e-1
shaw	1.5e-1	1.5e-1	9.4e-1	1.8e-1	1.8e-1	9.8e-1	2.0e-1	2.0e-1	9.7e-1

To shed some more light on the significance of the parameter α, Figure 6.1 provides plots that show the optimal α as a function of the error-level (1%, 2%, ..., 10%) for each of the test problems in Tables 6.1–6.3. For a given random error vector, we determine the best value for α from the discrete set 0.01, 0.02, ..., 1; that is, the α for which the corresponding solution has the smallest relative error compared to $\hat{x}$. The graphs show the averages of the optimal α-values over 100 runs with different random error vectors. The figures suggest that for many of the test problems, the value of the optimal α does not drastically vary with the error level. Also, it is clear that the optimal value of α is smaller than unity for all problems, often even much smaller.

Next, we compare our fractional Tikhonov method with the approach in [7]. We compare (2.9) with parameter $\alpha = 0.6$ with (2.4) for $\gamma = \frac{1}{2}(\alpha + 1) = 0.8$. These values of α and γ yield the same power of $A^T A$ in the right-hand sides of (1.6) and (2.5), respectively, and render the filter functions (2.4) and (2.9) the same asymptotic be-

Table 6.2 Qualities of Tikhonov, fractional Tikhonov, and their ratios for various $n = 100$ examples for different error-levels (1%, 5%, 10%) and $\alpha = 0.6$.

Error	1%			5%			10%		
Problem	Tikh	Frac	Ratio	Tikh	Frac	Ratio	Tikh	Frac	Ratio
baart	2.1e-1	1.9e-1	9.1e-1	3.3e-1	3.0e-1	9.0e-1	3.6e-1	3.4e-1	9.4e-1
deriv2-1	2.8e-1	2.7e-1	9.6e-1	3.8e-1	3.6e-1	9.3e-1	4.4e-1	4.1e-1	9.3e-1
deriv2-2	2.7e-1	2.5e-1	9.5e-1	3.8e-1	3.5e-1	9.2e-1	4.4e-1	4.1e-1	9.2e-1
deriv2-3	3.8e-2	4.6e-2	1.2e-0	7.0e-2	6.3e-2	8.9e-1	9.3e-2	8.1e-2	8.7e-1
foxgood	4.3e-2	4.0e-2	9.3e-1	1.4e-1	1.0e-1	7.3e-1	2.0e-1	1.5e-1	7.4e-1
gravity	3.7e-2	2.3e-2	6.4e-1	7.4e-2	5.6e-2	7.6e-1	1.1e-1	9.0e-2	8.2e-1
heat	1.5e-1	1.6e-1	1.1e-0	3.1e-1	3.1e-1	9.9e-1	4.4e-1	4.2e-1	9.5e-1
ilaplace	1.6e-1	1.5e-1	9.2e-1	2.0e-1	1.8e-1	9.2e-1	2.2e-1	2.0e-1	9.4e-1
phillips	2.8e-2	4.8e-2	1.7e-0	6.5e-2	7.3e-2	1.1e-0	1.1e-1	1.0e-1	9.8e-1
shaw	1.5e-1	1.3e-1	8.6e-1	1.8e-1	1.7e-1	9.5e-1	2.0e-1	1.9e-1	9.3e-1

Table 6.3 Qualities of Tikhonov, fractional Tikhonov, and their ratios for various $n = 100$ examples for different error-levels (1%, 5%, 10%) and $\alpha = 0.4$.

Error	1%			5%			10%		
Problem	Tikh	Frac	Ratio	Tikh	Frac	Ratio	Tikh	Frac	Ratio
baart	2.1e-1	1.7e-1	8.1e-1	3.3e-1	2.7e-1	8.2e-1	3.6e-1	3.2e-1	8.8e-1
deriv2-1	2.8e-1	2.8e-1	1.0e-0	3.8e-1	3.6e-1	9.4e-1	4.4e-1	4.0e-1	9.1e-1
deriv2-2	2.7e-1	2.6e-1	9.8e-1	3.8e-1	3.4e-1	9.0e-1	4.4e-1	3.9e-1	8.9e-1
deriv2-3	3.8e-2	1.1e-1	2.9e-0	7.0e-2	1.3e-1	1.9e-0	9.3e-2	1.4e-1	1.5e-0
foxgood	4.3e-2	4.8e-2	1.1e-0	1.4e-1	8.6e-2	6.3e-1	2.0e-1	1.3e-1	6.2e-1
gravity	3.7e-2	4.3e-2	1.2e-0	7.4e-2	6.6e-2	8.9e-1	1.1e-1	9.3e-2	8.4e-1
heat	1.5e-1	1.9e-1	1.2e-0	3.1e-1	3.3e-1	1.1e-0	4.4e-1	4.3e-1	9.7e-1
ilaplace	1.6e-1	1.4e-1	9.0e-1	2.0e-1	1.8e-1	9.0e-1	2.2e-1	2.0e-1	9.2e-1
phillips	2.8e-2	1.1e-1	3.7e-0	6.5e-2	1.3e-1	2.0e-0	1.1e-1	1.5e-1	1.4e-0
shaw	1.5e-1	1.2e-1	7.9e-1	1.8e-1	1.7e-1	9.3e-1	2.0e-1	1.9e-1	9.2e-1

Table 6.4 Qualities of the Klann–Ramlau approach of [7], fractional Tikhonov, and their ratios for various $n = 100$ examples for different error-levels (1%, 5%, 10%) and $\alpha = 0.6$.

Error	1%			5%			10%		
Problem	KR	Frac	Ratio	KR	Frac	Ratio	KR	Frac	Ratio
baart	2.1e-1	1.9e-1	9.2e-1	3.2e-1	3.0e-1	9.3e-1	3.5e-1	3.4e-1	9.6e-1
deriv2-1	2.8e-1	2.7e-1	9.7e-1	3.8e-1	3.6e-1	9.5e-1	4.4e-1	4.1e-1	9.5e-1
deriv2-2	2.6e-1	2.5e-1	9.6e-1	3.7e-1	3.5e-1	9.4e-1	4.3e-1	4.1e-1	9.4e-1
deriv2-3	3.9e-2	4.6e-2	1.2e-0	6.8e-2	6.3e-2	9.2e-1	8.8e-2	8.1e-2	9.2e-1
foxgood	4.4e-2	4.0e-2	9.2e-1	1.3e-1	1.0e-1	7.5e-1	1.9e-1	1.5e-1	7.8e-1
gravity	3.4e-2	2.3e-2	6.9e-1	7.0e-2	5.6e-2	8.0e-1	1.1e-1	9.0e-2	8.5e-1
heat	1.6e-1	1.6e-1	1.0e-0	3.2e-1	3.1e-1	9.8e-1	4.4e-1	4.2e-1	9.5e-1
ilaplace	1.6e-1	1.5e-1	9.3e-1	1.9e-1	1.8e-1	9.4e-1	2.1e-1	2.0e-1	9.5e-1
phillips	3.1e-2	4.8e-2	1.6e-0	6.7e-2	7.3e-2	1.1e-0	1.1e-1	1.0e-1	9.7e-1
shaw	1.5e-1	1.3e-1	8.9e-1	1.8e-1	1.7e-1	9.6e-1	2.0e-1	1.9e-1	9.4e-1

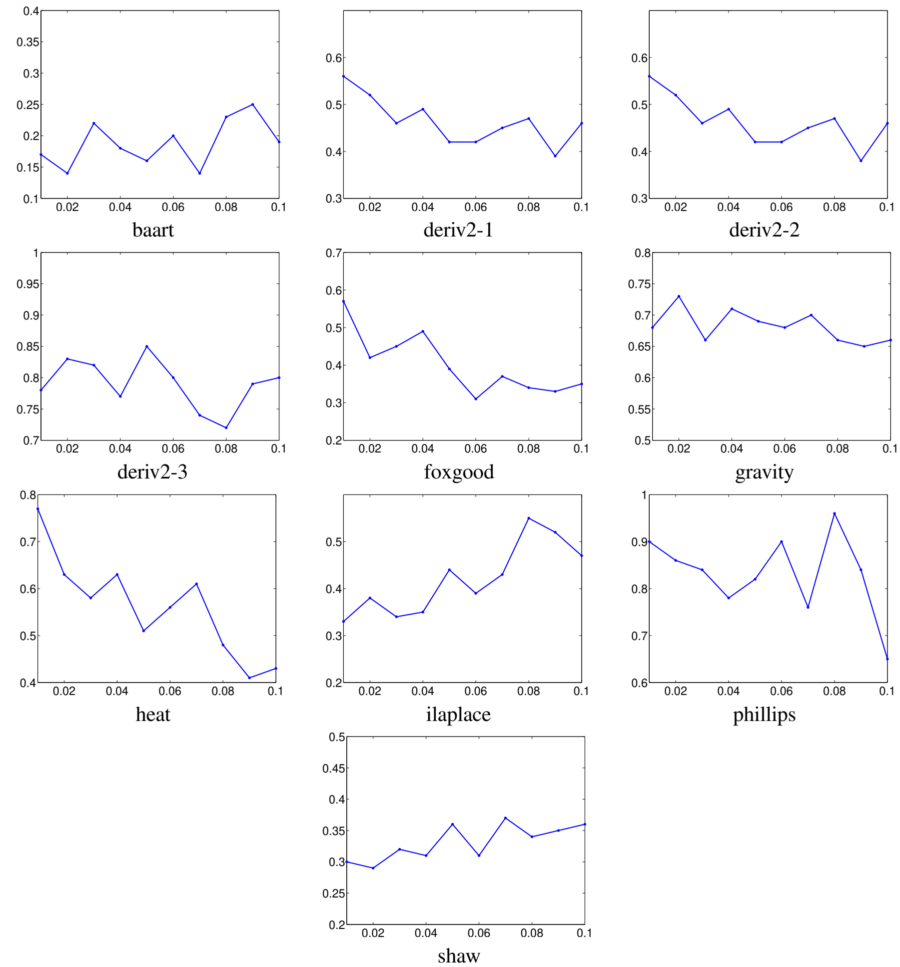

Fig. 6.1 Optimal α (vertical axes) for error-levels 1%, 2%, ..., 10% (horizontal axes) for the test problems of Tables 6.1–6.3 (average taken over 100 random error vectors).

havior, in terms of the power of σ, at the origin. Table 6.4 shows fractional Tikhonov to usually render solutions of higher quality.

Experiment 6.2 We illustrate the performance of the method of Section 4 for large problems. The same problems as in the previous tables are considered, but now for $m = n = 5000$. We project these large problems onto the Krylov space (4.2) of dimension $\ell = 20$. This value of ℓ is quite arbitrary; other values typically give similar results. Computed results are reported in Table 6.5. Table 6.5 shows that the fractional Tikhonov approach with $\alpha = 0.6$ is better than the standard Tikhonov method in all cases reported.

Table 6.5 Qualities of Tikhonov, fractional Tikhonov, and their ratios for various $n = 5000$ examples projected onto 20-dimensional Lanczos bidiagonalization spaces for different error-levels (1%, 5%, 10%) and $\alpha = 0.6$.

Error	1%			5%			10%		
Problem	Tikh	Frac	Ratio	Tikh	Frac	Ratio	Tikh	Frac	Ratio
baart	2.1e-1	1.9e-1	9.2e-1	3.3e-1	3.0e-1	9.1e-1	3.7e-1	3.5e-1	9.4e-1
deriv2-1	2.8e-1	2.6e-1	9.2e-1	3.8e-1	3.5e-1	9.3e-1	4.4e-1	4.1e-1	9.3e-1
deriv2-2	2.7e-1	2.5e-1	9.2e-1	3.7e-1	3.4e-1	9.2e-1	4.2e-1	3.9e-1	9.1e-1
deriv2-3	4.4e-2	3.1e-2	7.0e-1	8.8e-2	6.7e-2	7.7e-1	1.1e-1	9.5e-2	8.6e-1
foxgood	4.7e-2	4.1e-2	8.8e-1	1.5e-1	1.2e-1	7.6e-1	2.2e-1	1.8e-1	7.9e-1
gravity	4.2e-2	3.1e-2	7.4e-1	7.7e-2	6.3e-2	8.2e-1	1.1e-1	9.1e-2	8.5e-1
heat	1.1e-1	9.3e-2	8.1e-1	2.6e-1	2.3e-1	9.0e-1	3.6e-1	3.3e-1	9.1e-1
ilaplace	2.5e-1	2.4e-1	9.6e-1	2.8e-1	2.7e-1	9.6e-1	2.9e-1	2.8e-1	9.7e-1
phillips	2.5e-2	1.8e-2	7.0e-1	6.1e-2	4.8e-2	7.8e-1	9.5e-2	8.0e-2	8.4e-1
shaw	1.6e-1	1.3e-1	8.5e-1	1.9e-1	1.8e-1	9.4e-1	2.3e-1	2.1e-1	9.3e-1

7 Conclusions

We have studied a family of fractional Tikhonov regularization methods which depend on a parameter $\alpha > 0$. Standard Tikhonov regularization is obtained for $\alpha = 1$. We have shown how the solution depends on α and, in particular, investigated how the choice of α affects solutions that satisfy the discrepancy principle. The norm of these solution has a local minimum for $\alpha = 1$. Analogously, if the computed solution is required to be of specified norm, the norm of the residual error has a local minimum for $\alpha = 1$. This indicates that the choice $\alpha = 1$ is quite natural. However, it is known that standard Tikhonov gives over-smoothed solutions and we propose to remedy this by choosing $\alpha < 1$. Extensive numerical experiments suggest that letting α be smaller than, but close to unity, such as $\alpha = 0.8$, gives better results than $\alpha = 1$ for almost all examples. Smaller values of α, such as $\alpha = 0.5$, may work even better for some examples, at the cost of rendering worse results for others. A general rule-of-thumb is that the larger the error-level, the more advantageous it is to let $\alpha < 1$. Our computed examples illustrate the performance of the method in conjunction with the discrepancy principle. We remark that fractional Tikhonov can be applied with other selection rules for the regularization parameter as well. Finally, the techniques also can be used for large-scale problems by first projecting them onto low-dimensional Krylov spaces.

Acknowledgements. LR would like to thank MH for an enjoyable visit to TU/e during which work on this paper was carried out. The authors thank the referee for useful suggestions.

References

1. Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
2. D. Calvetti and L. Reichel, Lanczos-based exponential filtering for discrete ill-posed problems, Numer. Algorithms, 29 (2002), pp. 45–65.

3. D. Calvetti and L. Reichel, Tikhonov regularization with a solution constraint, SIAM J. Sci. Comput., 26 (2004), pp. 224–239.
4. G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
5. P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
6. P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007), pp. 189–194
7. E. Klann and R. Ramlau, Regularization by fractional filter methods and data smoothing, Inverse Problems, 24 (2008), 025018.
8. J. Lampe, M. Rojas, D. Sorensen, and H. Voss, Accelerating the LSTRS algorithm, Bericht 138, Institute of Numerical Simulation, Hamburg University of Technology, Hamburg, Germany, July 2009.
9. V. A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer-Verlag, New York, 1984.
10. M. Rojas and D. C. Sorensen, A trust-region approach to regularization of large-scale discrete forms of ill-posed problems, SIAM J. Sci. Comput., 23 (2002), pp. 1842–1860.