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

A new framework for multi-parameter regularization

Silvia Gazzola · **Lothar Reichel**

Received: date / Accepted: date

Abstract This paper proposes a new approach for choosing the regularization parameters in multi-parameter regularization methods when applied to approximate the solution of linear discrete ill-posed problems. We consider both direct methods, such as Tikhonov regularization with two or more regularization terms, and iterative methods based on the projection of a Tikhonov-regularized problem onto Krylov subspaces of increasing dimension. The latter methods regularize by choosing appropriate regularization terms and the dimension of the Krylov subspace. Our investigation focuses on selecting a proper set of regularization parameters that satisfies the discrepancy principle and maximizes a suitable quantity, whose size reflects the quality of the computed approximate solution. Theoretical results are shown and illustrated by numerical experiments.

Keywords Ill-posed problems · multi-parameter Tikhonov method · Arnoldi–Tikhonov method · discrepancy principle

Mathematics Subject Classification (2000) 65F10 · 65F22 · 65N20 · 65N21

1 Introduction

Consider linear least-squares problems of the form

$$\min_{x\in\mathbb{R}^N} \|Ax-b\|, \qquad A\in\mathbb{R}^{M\times N}, \quad b\in\mathbb{R}^M, \tag{1.1}$$

where the singular values of the coefficient matrix quickly and smoothly decay to zero (in particular, A is severely ill-conditioned), and where the vector b is contaminated by unknown additive white noise

Research supported in part by the GNCS project "Giovani Ricercatori", by the "ex-60%" funds of the University of Padova, and by NSF grant DMS-1115385.

S. Gazzola
Dipartimento di Matematica, Università di Padova, Via Trieste 63, Padova, Italy
E-mail: gazzola@math.unipd.it

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

$e \in \mathbb{R}^M$, i.e., $b = b^{ex} + e$, where b^{ex} denotes the unknown exact (noise-free) vector associated with b. Least-squares problems of this kind are commonly referred to as linear discrete ill-posed problems, and arise in a variety of scientific and engineering applications linked to the solution of inverse problems [4, 10]. Because of the ill-conditioning of A and the perturbation e in b, one has to employ some kind of regularization of (1.1) in order to be able to compute a meaningful approximation of a desired solution $x^{ex} \in \mathbb{R}^N$ of the consistent noise-free linear system of equations $Ax^{ex} = b^{ex}$.

Regularization methods are determined by the choice of one or several nonnegative regularization parameters, which specify the amount of regularization, and by the associated regularization matrices. The latter impose some regularity properties on the computed approximation of x^{ex}. Depending on the size and properties of A, one usually chooses between direct or iterative regularization methods. The former methods compute factorizations of A and the regularization matrix or matrices (usually the singular value or generalized singular value decompositions), while the latter methods first reduce A and the regularization matrix or matrices to small sizes, and then solve the reduced problem by a direct method.

One-parameter Tikhonov regularization is possibly the best understood direct regularization method. It solves a penalized minimization problem of the form

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|Lx\|^2 \right\}, \tag{1.2}$$

where $\lambda \geq 0$ is the regularization parameter and $L \in \mathbb{R}^{P\times N}$ is the regularization matrix. Here and in the following $\|\cdot\|$ denotes the vector 2-norm and $(\cdot,\cdot)$ the standard inner product. When $L = I_N$ (the identity matrix of order N), the problem (1.2) is said to be in *standard form*; otherwise it is in *general form*. We assume that L is chosen so that the following relation between the null spaces of A and L holds:

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}. \tag{1.3}$$

Then the minimizer x_λ of (1.2) is unique for any $\lambda > 0$. It is clear from (1.2) that the component of the solution in $\mathcal{N}(L)$ is not affected by regularization. Therefore, an effective regularization matrix L is such that known important features of the desired solution x^{ex} belong to $\mathcal{N}(L)$; see, e.g., [22, 27] for discussions.

Recently, many authors have pointed out the need of going beyond the classical framework of one-parameter Tikhonov regularization and proposed the use of multi-parameter Tikhonov regularization methods. These methods replace the least-squares problem (1.1) by a penalized minimization problem of the form

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \sum_{i=1}^{m} \lambda_i \|L_i x\|^2 \right\}; \tag{1.4}$$

see, e.g., [1, 2, 5, 6, 15, 17, 19, 20, 26] and references therein. The scalars $\lambda_i \geq 0$ are regularization parameters and the $L_i \in \mathbb{R}^{P_i\times N}$ are regularization matrices for $i = 1,\ldots,m$. We assume in the following that

$$\mathcal{N}(A) \bigcap_{i=1,\ldots,m} \mathcal{N}(L_i) = \{0\}$$

to secure that (1.4) has a unique solution when at least one of the regularization parameters λ_i is positive. We denote the solution of (1.4) by x_Λ and refer to $\Lambda = (\lambda_1,\ldots,\lambda_m)$ as the *regularization vector*. When $m = 1$, problem (1.4) simplifies to (1.2). An advantage of multi-parameter Tikhonov regularization, when compared with one-parameter Tikhonov regularization, is that different features of the solution can be enhanced by using several regularization matrices with different null spaces. However, a drawback of multi-parameter Tikhonov regularization is that one has to define reliable strategies to determine an adequate regularization vector Λ. To the best of our knowledge, the first attempt to derive a systematic parameter choice strategy was proposed by Belge et al. [1], who introduced a generalization of the L-curve

and described an efficient algorithm to compute the regularization parameters corresponding to a point on the L-hypersurface where the curvature is approximately maximized. Brezinski et al. [2] proposed and analyzed an approach based on the GCV method. Specifically, they solved m different one-parameter problems (one for each regularization term appearing in (1.4)) and then combined the m approximations of x^{ex} so obtained. More recently, Lu and Pereverzev [19] and Lu et al. [20] investigated the application of the discrepancy principle to select a regularization vector for (1.4). They showed for square nonsingular regularization matrices L_i that any combination of nonnegative regularization parameters λ_i such that x_Λ satisfies the discrepancy principle constitutes a regularization method in a well-defined sense. Also a root-finder for computing a regularization vector Λ such that x_Λ satisfies the discrepancy principle is described. Further discussions on the discrepancy principle for multi-parameter Tikhonov regularization are provided by Fornasier et al. [5]. Multi-parameter Tikhonov regularization also is described by Ito et al. [15], who advocate the application of a balancing principle for determining the regularization vector. It has the advantage over the discrepancy principle that it delivers a unique regularization vector Λ. Moreover, while the discrepancy principle requires that an estimate of the norm of the error in b be available, see Section 2, the balancing principle does not. This can be an advantage in some applications. Parameter choice rules that do not use a bound for the error in b are referred to as "heuristic methods." Heuristic methods are known to fail to determine a suitable regularization vector for some problems; see, e.g., [4, 16] for discussions. This shortcoming of heuristic regularization methods is a reason for our interest in applying the discrepancy principle. Kunisch and Pock [17] describe a parameter learning approach to multi-parameter regularization for denoising problems. Thus, they restrict themselves to problems (1.4) with A the identity. Moreover, the proposed parameter choice method is heuristic and, hence, may fail. Mead [21] recently proposed an interesting statistically based multiparameter Tikhonov regularization method for linear discrete ill-posed problems (1.1) for which the desired solution x^{ex} is discontinuous.

It is the purpose of the present paper to investigate a new approach to determine the regularization vector Λ of multi-parameter regularization methods (1.4) based on the discrepancy principle. Similarly as Lu et al. [19,20], we first determine a set of regularization vectors $\{\Lambda_j\}_{j=1,\ldots,h}$ that satisfy the discrepancy principle. Then we choose a vector from this set that solves the problem

$$\max_{j=1,\ldots,h} \Psi\left(x_{\Lambda_j}\right) \tag{1.5}$$

for some functional Ψ. Functionals $\Psi(x_{\Lambda_j})$ of interest include $\|x_{\Lambda_j}\|$, $\|Lx_{\Lambda_j}\|$, and $\|x_{\Lambda_j}\|^2 + \|Lx_{\Lambda_j}\|^2$, where L is a regularization matrix. If L is a discretization of a derivative operator, then the latter choice corresponds to maximizing a discrete Sobolev norm of the regularized solution. In the available literature on multi-parameter Tikhonov regularization, the issue of choosing a particular regularization vector among the vectors that satisfy the discrepancy principle appears to be discussed just marginally in [20]. We are interested in studying the choice of a regularization vector Λ when one or several regularization matrices L_i in (1.4) has a nontrivial null space, which is the case when L_i is a projection matrix or a discretization of a derivative operator. These kinds of regularization matrices are commonly used and they are not covered by the analysis in [20]. In Section 4, we present computed examples that illustrate that the quality of the computed solution depends on the selection of Λ among the regularization vectors such that x_Λ satisfies the discrepancy principle. Therefore, the choice of a regularization vector among the set such that x_Λ satisfies the discrepancy principle can be important.

In addition to multi-parameter Tikhonov regularization (1.4), we also consider Krylov–Tikhonov methods for the solution of large-scale Tikhonov regularization problems of the form (1.2). These iterative solution methods first reduce the given large problem (1.2) to a small problem, and then determine an approximate solution in a solution subspace of small dimension. The quality of the computed solution depends on the dimension of the solution subspace used. This dimension, together with the parameter λ in (1.2), are regularization parameters. Therefore, these methods are intrinsically multi-parameter. They have received

considerable attention, see, e.g., [8,13,26,28]. We propose to use the discrepancy principle to choose the Tikhonov regularization parameter at each iteration [13,18]; the number of iterations, which equals the dimension of the solution subspace, is determined by solving an optimization problem similar to (1.5). Recently, Regińska [25] provided a theoretical investigation of this type of two-parameter regularization methods for Tikhonov regularization problems (1.2) with the matrix A determined by a partial singular value decomposition or a least-squares projection of a compact operator, and $L = I$. In our numerical investigation we reduce a large matrix A to a small one by applying a few steps of the Arnoldi process and allow a more general regularization matrix. A condition related to (1.5) is imposed. Our approach can be applied to other iterative Tikhonov regularization methods as well, such as to the methods discussed in [12,13,26,28]. Recently, Gazzola and Novati [6] proposed a scheme for determining Λ as well as the number of iterations with an Arnoldi–Tikhonov method for (1.4), which relies on sequentially updating an approximation of the discrepancy function. We will review this approach and discuss how the criterion (1.5) can be incorporated.

This paper is organized as follows: in Section 2 we discuss direct multi-parameter Tikhonov regularization (1.4). We analyze some properties of the corresponding one-parameter problem (i.e., we sequentially vary one regularization parameter at a time). Strategies for solving (1.5) are described, and we provide some insight into the choice of functional Ψ. Section 3 is concerned with multi-parameter Krylov–Tikhonov regularization. We report results from many numerical experiments in Section 4. They are concerned with the solution of discretized Fredholm integral equation of the first kind, and with image deblurring and denoising. Some concluding remarks can be found in Section 5.

2 Direct multi-parameter regularization

We would like to compute a solution x_Λ of (1.4) that satisfies the discrepancy principle, i.e., is such that

$$\|b - Ax_\Lambda\|^2 = \eta^2\varepsilon^2, \tag{2.1}$$

where $\varepsilon \geq \|e\|$ is an available upper bound for the error in b and $\eta > 1$ is a user-supplied safety factor independent of ε. In order for the discrepancy principle to yield an accurate approximation of x^{ex} it is usually necessary that η is close to unity and ε is close to $\|e\|$ and not too large. Introduce the discrepancy function

$$\Phi(\Lambda) := \|b - Ax_\Lambda\|^2. \tag{2.2}$$

In the one-parameter case, i.e., when $\Lambda = (\lambda)$, one usually applies the discrepancy principle by solving the nonlinear equation (2.1) with respect to λ [23]. However, when $\Lambda \in \mathbb{R}^m$, $m \geq 2$, the problem (2.1) is underdetermined. Following [20], we define the discrepancy hypersurface

$$\mathscr{D} = \{\Lambda \in \mathbb{R}^m : \Lambda \geq 0, \Lambda \not\equiv 0, \|b - Ax_\Lambda\| = \eta\varepsilon\},$$

where the condition $\Lambda \geq 0$ is to be interpreted component-wise. In this section we describe a strategy to impose additional constraints on the solution x_Λ of (1.4) in order to reduce the number of degrees of freedom in the choice of Λ.

A natural constraint is

$$x_\Lambda = \arg\max_{\bar{x}_\Lambda} \|\bar{x}_\Lambda\| \quad \text{subject to} \quad \Lambda \in \mathscr{D}. \tag{2.3}$$

It can be justified in the following way. The discrepancy principle generally determines over-smoothed approximations of x^{ex}, i.e., approximations of norm smaller than $\|x^{ex}\|$; cf. [10, §7.2] and references therein. By imposing (2.3) or, more in general, (1.5), we seek to determine the least over-smoothed approximation x_Λ of x^{ex} with $\Lambda \in \mathscr{D}$. Of course, generally the constraint (2.3) does not by itself determine

an accurate approximation of x^{ex}. Indeed, this constraint would by itself deliver an unregularized solution. We remark that a related approach has previously been considered in [14] for determining an improved approximation of x^{ex} given a set of regularized solutions computed by different methods such as TSVD or Tikhonov regularization. The following theoretical result provides a sufficient condition for a vector x_Λ that satisfies (2.3) to be an optimal approximation of x^{ex} in the sense that the error $\|x^{ex} - x_\Lambda\|$ is minimal. In the following theorem and below, we let $\mathscr{I}$ denote the set indexing the regularization vectors belonging to $\mathscr{D}$.

Theorem 2.1 *Let $\{x_{\Lambda_i}\}_{i\in\mathscr{I}}$ be a set of solutions of (1.4) that satisfy the discrepancy principle (2.1) and define $\ell = \arg\max_{i\in\mathscr{I}} \left\|x_{\Lambda_i}\right\|$. Let $x_{\Lambda_i} = x_{\Lambda_\ell} + \delta x_i$. If*

$$(x^{ex}, \delta x_i) \leq 0 \tag{2.4}$$

and

$$\left(x_{\Lambda_\ell}, \delta x_i\right) \geq -\frac{\|\delta x_i\|^2}{2} + (x^{ex}, \delta x_i)\,, \tag{2.5}$$

for all $i \in \mathscr{I}$, $i \neq \ell$, then

$$\left\|x^{ex} - x_{\Lambda_\ell}\right\| \leq \left\|x^{ex} - x_{\Lambda_i}\right\| \quad \forall i \in \mathscr{I}. \tag{2.6}$$

Proof The result follows directly by writing

$$\left\|x^{ex} - x_{\Lambda_i}\right\|^2 - \left\|x^{ex} - x_{\Lambda_\ell}\right\|^2 = 2\left(x_{\Lambda_\ell}, \delta x_i\right) + \|\delta x_i\|^2 - 2\left(x^{ex}, \delta x_i\right) \tag{2.7}$$

and by considering that, thanks to (2.5), the right-hand side of the above equality is nonnegative. Since the assumption $\|x_{\Lambda_\ell}\| \geq \|x_{\Lambda_i}\|$ implies

$$\left(x_{\Lambda_\ell}, \delta x_i\right) \leq -\frac{\|\delta x_i\|^2}{2},$$

one can easily see that (2.6) holds under both the conditions (2.4) and (2.5). □

We may replace (2.3) by a constraint of the form

$$\max_{\Lambda\in\mathscr{D}} \sum_{i=1}^{m} \|L_i x_\Lambda\|^2\,, \tag{2.8}$$

where L_i, $i = 1, \ldots, m$, are the regularization matrices in (1.4). Theoretical optimality properties analogous to Theorem 2.1 can be established also for the constraint (2.8). For instance, one can show the following result.

Theorem 2.2 *Let $\{x_{\Lambda_i}\}_{i\in\mathscr{I}}$ be a set of solutions of (1.4) that satisfy the discrepancy principle (2.1). Define $\ell = \arg\max_{i\in\mathscr{I}} \left(\left\|x_{\Lambda_i}\right\|^2 + \left\|Lx_{\Lambda_i}\right\|^2\right)$ and let $x_{\Lambda_i} = x_{\Lambda_\ell} + \delta x_i$. If*

$$(x^{ex}, \delta x_i) \leq -\frac{\|L\delta x_i\|^2}{2} - (Lx_{\Lambda_\ell}, L\delta x_i) \tag{2.9}$$

and

$$\left(x_{\Lambda_\ell}, (I + L^T L)\delta x_i\right) \geq -\frac{\|\delta x_i\|^2}{2} + \left(x^{ex} + L^T L x_{\Lambda_\ell}, \delta x_i\right)\,, \tag{2.10}$$

for all $i \in \mathscr{I}$, $i \neq \ell$, then

$$\left\|x^{ex} - x_{\Lambda_\ell}\right\| \leq \left\|x^{ex} - x_{\Lambda_i}\right\| \quad \forall i \in \mathscr{I}. \tag{2.11}$$

Proof By writing the errors as in (2.7), we conclude that (2.11) holds, because (2.10) implies (2.5). Since the assumption $\|x_{\Lambda_\ell}\| + \|Lx_{\Lambda_\ell}\| \geq \|x_{\Lambda_i}\| + \|Lx_{\Lambda_i}\|$ implies

$$\left(x_{\Lambda_\ell}, (I + L^T L)\delta x_i\right) \leq -\frac{\|L\delta x_i\|^2}{2} - \frac{\|\delta x_i\|^2}{2},$$

one can easily see that (2.11) holds under both the conditions (2.9) and (2.10). □

Remark 2.1 In the special case where $\delta x_i \in \mathcal{N}(L)$ for all $i \in \mathcal{I}$, we can immediately see that Theorem 2.2 is equivalent to Theorem 2.1. Specifically,

$$\ell = \arg\max_{i\in\mathcal{I}} \left(\|x_{\Lambda_i}\|^2 + \|Lx_{\Lambda_i}\|^2\right) = \arg\max_{i\in\mathcal{I}} \left(\|x_{\Lambda_i}\|^2 + \|Lx_{\Lambda_\ell}\|^2\right) = \arg\max_{i\in\mathcal{I}} \|x_{\Lambda_i}\|^2.$$

Further, the inequalities (2.9) and (2.10) are equivalent to (2.4) and (2.5), respectively. Moreover, when the x_{Λ_i}'s are accurate approximations of x^{ex} and are close to $\mathcal{N}(L)$, the constraint (2.8) with $L_i = L$ approximately reduces to the constraint (2.3). The situation when the x_{Λ_i}'s are close to $\mathcal{N}(L)$ is of interest in applications; see the discussion of Section 1.

To keep our discussion simple, we focus on the two-parameter case

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda_1 \|L_1 x\|^2 + \lambda_2 \|L_2 x\|^2 \right\}. \tag{2.12}$$

The use of more than two regularization terms in (1.4) can be treated analogously. In the two-parameter case, $\mathcal{D}$ is a differentiable curve in $\mathbb{R}^2$; see [19]. A simple technique to impose the constraints (2.3) or (2.8) is to sample the quantities $\|L_i x_\Lambda\|$ for logarithmically equispaced values of λ_1 or λ_2. For instance, we may define the sampling space by first keeping $\lambda_1 = \hat{\lambda}_1$ fixed and determine (if possible) λ_2 so that the discrepancy principle is satisfied; cf. Proposition 2.1 below. Thus, let λ_2 be the zero of the one-variable function

$$\lambda_2 \to \Phi(h(\hat{\lambda}_1), h(\lambda_2))^2 - \eta^2 \varepsilon^2, \tag{2.13}$$

where Φ is given by (2.2) and $h(\lambda_i) = \lambda_i^{-1}$, $i = 1, 2$. The purpose of using $h(\lambda_i)$ instead of λ_i is to secure convexity of the function (2.13). This change of variable for the function (2.13) is also considered by Lu et al. [20]. It is commonly used for one-parameter Tikhonov regularization problems (1.2). Lu et al. [20] considered fixed linear combinations of regularization matrices. Then the computation of the multi-parameter Tikhonov regularization problem simplifies to one-parameter Tikhonov regularization. We consider both regularization matrices independent and this leads to a somewhat different zero-finder. We will need the following result.

Lemma 2.1 *If*

$$(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)}) < 0 \tag{2.14}$$

and

$$(L_1 \partial^2_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)}) = 2(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}), \tag{2.15}$$

then $\Phi(h(\hat{\lambda}_1), h(\lambda_2))$ is a decreasing and convex function of λ_2. Here $\partial_{h(\lambda_2)} x_{h(\Lambda)}$ denotes the partial derivative with respect to the variable $h(\lambda_2)$.

Proof Introduce the function

$$\Omega(\lambda_1,\lambda_2) = \|b - Ax_\Lambda\|^2 + \lambda_1\|L_1x_\Lambda\|^2 + \lambda_2\|L_2x_\Lambda\|^2, \qquad \Lambda = (\lambda_1,\lambda_2).$$

Following Lu and Pereverzev [19], we consider the following relations

$$\begin{aligned}
\Phi(\hat\lambda_1,\lambda_2) &= \Omega(\hat\lambda_1,\lambda_2) - \hat\lambda_1\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2)|_{\lambda_1=\hat\lambda_1} - \lambda_2\partial_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2)\,,\\
\partial_{\lambda_2}\Phi(\hat\lambda_1,\lambda_2) &= -\hat\lambda_1\partial_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2)|_{\lambda_1=\hat\lambda_1} - \lambda_2\partial^2_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2)\,,\\
\partial^2_{\lambda_2}\Phi(\hat\lambda_1,\lambda_2) &= -\hat\lambda_1\partial^2_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2)|_{\lambda_1=\hat\lambda_1} - \partial^2_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2) - \lambda_2\partial^3_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2)\,.
\end{aligned} \tag{2.16}$$

In the above expressions,

$$\begin{aligned}
\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= \|L_1x_\Lambda\|^2\,,\\
\partial_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= \|L_2x_\Lambda\|^2\,,\\
\partial_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= 2\left(L_1\partial_{\lambda_2}x_\Lambda, L_1x_\Lambda\right)\,,\\
\partial^2_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= 2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2x_\Lambda\right)\,,\\
\partial^2_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= 2\left(L_1\partial^2_{\lambda_2}x_\Lambda, L_1x_\Lambda\right) + 2\left(L_1\partial_{\lambda_2}x_\Lambda, L_1\partial_{\lambda_2}x_\Lambda\right)\,,\\
\partial^3_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= 2\left(L_2\partial^2_{\lambda_2}x_\Lambda, L_2x_\Lambda\right) + 2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2\partial_{\lambda_2}x_\Lambda\right)\,.
\end{aligned} \tag{2.17}$$

We also remark that $\partial_{\lambda_2}x_\Lambda$ and $\partial^2_{\lambda_2}x_\Lambda$ solve the following problems

$$\left(A\partial_{\lambda_2}x_\Lambda, Az\right) + \lambda_1\left(L_1\partial_{\lambda_2}x_\Lambda, L_1z\right) + \lambda_2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2z\right) = -\left(L_2x_\Lambda, L_2z\right)\,, \tag{2.18}$$

$$\left(A\partial^2_{\lambda_2}x_\Lambda, Az\right) + \lambda_1\left(L_1\partial^2_{\lambda_2}x_\Lambda, L_1z\right) + \lambda_2\left(L_2\partial^2_{\lambda_2}x_\Lambda, L_2z\right) = -2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2z\right)\,, \tag{2.19}$$

for all $z\in\mathbb{R}^N$, respectively. Now, to evaluate the derivatives of $\Phi(h(\lambda_1),h(\lambda_2))$, we employ the chain rule. Concerning the first derivative we get

$$\begin{aligned}
\partial_{\lambda_2}\Phi(h(\lambda_1),h(\lambda_2)) = &\frac{2}{\lambda_1\lambda_2^2}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_1x_{h(\Lambda)}\right)\\
&+\frac{2}{\lambda_2^3}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_2x_{h(\Lambda)}\right)\,,
\end{aligned} \tag{2.20}$$

for $\lambda_1=\hat\lambda_1$. By taking $z=\partial_{h(\lambda_2)}x_{h(\Lambda)}$ in (2.18), evaluated at $\left(h(\hat\lambda_1),h(\lambda_2)\right)$, we get

$$\begin{aligned}
-\left(L_2x_{h(\Lambda)}, L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right) = &\left(A\partial_{h(\lambda_2)}x_{h(\Lambda)}, A\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\
&+h(\hat\lambda_1)\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\
&+h(\lambda_2)\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right) > 0.
\end{aligned}$$

This implies that

$$\frac{2}{\lambda_2^3}\left(L_2x_{h(\Lambda)}, L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right) < 0\,.$$

Since by assumption $\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_1x_{h(\Lambda)}\right) < 0$, we conclude that

$$\partial_{\lambda_2}\Phi(h(\hat\lambda_1),h(\lambda_2)) < 0\,.$$

It is more cumbersome to show the convexity of $\Phi(\hat{\lambda}_1,\lambda_2)$. By applying the chain rule to (2.16) and by exploiting the equalities (2.17), we obtain

$$\begin{aligned}\partial^2_{\lambda_2}\Phi(h(\lambda_1),h(\lambda_2)) = &-\frac{2}{\lambda_1\lambda_2^4}\left(L_1\partial^2_{h(\lambda_2)}x_{h(\Lambda)},L_1x_{h(\Lambda)}\right)\\ &-\frac{2}{\lambda_1\lambda_2^4}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)},L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\ &-\frac{4}{\lambda_1\lambda_2^3}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)},L_1x_{h(\Lambda)}\right)\\ &-\frac{2}{\lambda_2^5}\left(L_2\partial^2_{h(\lambda_2)}x_{h(\Lambda)},L_2x_{h(\Lambda)}\right)\\ &-\frac{2}{\lambda_2^5}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)},L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\ &-\frac{6}{\lambda_2^4}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)},L_2x_{h(\Lambda)}\right)\end{aligned}$$

for $\lambda_1=\hat{\lambda}_1$. By evaluating (2.18) for $z=\partial^2_{h(\lambda_2)}x_{h(\Lambda)}$ and (2.19) for $z=\partial_{h(\lambda_2)}x_{h(\Lambda)}$, and by considering their difference, we get

$$\left(L_2x_{h(\Lambda)},L_2\partial^2_{h(\lambda_2)}x_{h(\Lambda)}\right)=2\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)},L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right). \tag{2.21}$$

Substituting (2.15) and (2.21) into the expression for $\partial^2_{\lambda_2}\Phi(h(\lambda_1),h(\lambda_2))$, we obtain

$$\begin{aligned}\partial^2_{\lambda_2}\Phi(h(\lambda_1),h(\lambda_2)) = &-\frac{6}{\lambda_1\lambda_2^4}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)},L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\ &-\frac{4}{\lambda_1\lambda_2^3}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)},L_1x_{h(\Lambda)}\right)\\ &-\frac{6}{\lambda_2^5}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)},L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\\ &-\frac{6}{\lambda_2^4}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)},L_2x_{h(\Lambda)}\right).\end{aligned}$$

Exploiting (2.14) and performing some algebraic manipulations on (2.18) evaluated for $z=\partial_{h(\lambda_2)}x_{h(\Lambda)}$ (namely, after moving to the left-hand side all terms except for $(A\partial_{h(\lambda_2)}x_{h(\Lambda)},A\partial_{h(\lambda_2)}x_{h(\Lambda)})$ and multiplying both sides by $6/\lambda_2^4$), we get

$$\partial^2_{\lambda_2}\Phi(h(\hat{\lambda}_1),h(\lambda_2))>0,$$

which concludes the proof. □

Proposition 2.1 *Under the assumptions (2.14), (2.15), and*

$$\left\|\left(I_N-A\left(A^TA+\frac{1}{\hat{\lambda}_1}L_1^TL_1\right)^{-1}A^T\right)b\right\|^2<\eta^2\varepsilon^2<\|b\|^2, \tag{2.22}$$

equation (2.13) has a unique solution $\lambda_2>0$.

Proof Immediate thanks to Lemma 2.1. □

Remark 2.2 The assumptions of Lemma 2.1 are satisfied for many linear discrete ill-posed problems, and for many regularization matrices and regularization parameters of different sizes. We provide an illustration in Section 4. Indeed, when L_1 is a discretized derivative operator, (2.15) can be regarded as a condition on the higher order derivatives of x_Λ. Also the inequalities (2.22) of Proposition 2.1 are typically satisfied: the upper bound limits the amount of noise in b, and the lower bound is a condition on the one-parameter problem (1.2) with $\lambda = 1/\hat{\lambda}_1$ and $L = L_1$.

It follows from Lemma 2.1 that Newton's method applied to determining the zero of (2.13) is guaranteed to give quadratic and monotonic convergence if the initial approximation of the zero is smaller than the desired zero. Application of Newton's method requires the computation of $\partial_{h(\lambda_2)}\Phi(\hat{\lambda}_1,\lambda_2)$. An expression for $\partial_{h(\lambda_2)}\Phi(\hat{\lambda}_1,\lambda_2)$ is given in (2.20) and, according to (2.18), one can compute $\partial_{h(\lambda_2)}x_{h(\Lambda)}$ in the following way

$$\partial_{h(\lambda_2)}x_{h(\Lambda)} = -\left(A^TA + \frac{1}{\lambda_1}L_1^TL_1 + \frac{1}{\lambda_2}L_2^TL_2\right)^{-1}\left(L_2^TL_2x_{h(\Lambda)}\right),$$

or, equivalently, by solving the least-squares problem,

$$\partial_{h(\lambda_2)}x_{h(\Lambda)} = \arg\min_{x\in\mathbb{R}^N}\left\|\begin{bmatrix} A \\ \lambda_1^{-1/2}L_1 \\ \lambda_2^{-1/2}L_2 \end{bmatrix}x - \begin{bmatrix} 0 \\ 0 \\ -\lambda_2^{1/2}L_2x_{h(\Lambda)} \end{bmatrix}\right\|^2.$$

The solution of this least-squares problem typically is the numerically preferable approach to compute $\partial_{h(\lambda_2)}x_{h(\Lambda)}$. We remark that the matrix in the above formulation is the same as the one associated with the problem (2.12).

The following result sheds some light on the maximization problem (2.3).

Proposition 2.2 *Assume that $\lambda_1 = \hat{\lambda}_1$ is fixed, that L_2 has full column rank, and that $\|x_\Lambda\| > 0$. Then $\|x_\Lambda\|^2$ is a decreasing function of λ_2.*

Proof It follows from (2.18) that

$$\partial_{\lambda_2}\|x_\Lambda\|^2 = 2\left(\partial_{\lambda_2}x_\Lambda, x_\Lambda\right) = -2\left(\left(A^TA + \hat{\lambda}_1L_1^TL_1 + \lambda_2L_2^TL_2\right)^{-1}(L_2^TL_2)x_\Lambda, x_\Lambda\right).$$

Since $L_2^TL_2$ is nonsingular, one can easily see that the matrix appearing in the above scalar product is positive definite, and therefore $\partial_{\lambda_2}\|x_\Lambda\|^2 < 0$ when $\|x_\Lambda\| > 0$. □

The assumptions of the above proposition hold when, for instance, $L_2 = I_N$. In this case, to maximize $\|x_\Lambda\|$, one should let $\lambda_2 = 0$, i.e., one should consider the one-parameter Tikhonov regularization problem (1.2) with $L = L_1$, and $\lambda = \hat{\lambda}_1$ chosen so that the discrepancy principle is satisfied.

If the discrepancy curve can be explicitly expressed as

$$\lambda_1 = g(\lambda_2), \tag{2.23}$$

where λ_2, $g(\lambda_2) > 0$, and g is differentiable, then we can extend the derivations in [19] to obtain (2.18). More precisely, one can see that $\partial_{\lambda_2}x_\Lambda$ solves the following problem

$$\begin{aligned}\left(A^TA\partial_{\lambda_2}x_\Lambda, z\right) &+ g(\lambda_2)\left(L_1^TL_1\partial_{\lambda_2}x_\Lambda, z\right) + \lambda_2\left(L_2^TL_2\partial_{\lambda_2}x_\Lambda, z\right) \\ &= -\left(\left(L_2^TL_2 + \partial_{\lambda_2}g(\lambda_2)L_1^TL_1\right)x_\Lambda, z\right)\end{aligned} \tag{2.24}$$

for all $z \in \mathbb{R}^N$. Therefore, after defining

$$A^\sharp = A^T A + g(\lambda_2) L_1^T L_1 + \lambda_2 L_2^T L_2 \,,$$

one obtains

$$\partial_{\lambda_2} \|x_\Lambda\|^2 = -2 \left(\left(A^\sharp \right)^{-1} (L_2^T L_2 + \partial_{\lambda_2} g(\lambda_2) L_1^T L_1) x_\Lambda \,, x_\Lambda \right) . \tag{2.25}$$

We can explicitly determine the sign of $\partial_{\lambda_2} \|x_\Lambda\|^2$ for certain functions $g(\lambda_2)$, for instance when $L_2^T L_2 + \partial_{\lambda_2} g(\lambda_2) L_1^T L_1$ is positive definite or negative definite. We analyze the behavior of this matrix in the following section, assuming $g(\lambda_2)$ to be a linear function.

Results analogous to Proposition 2.2 hold when replacing the maximization (2.3) by (2.8). Of course, λ_1 and λ_2 can be interchanged in the above discussion.

The theoretical results of this section have been formulated for two-parameter Tikhonov regularization so as to keep the notation and formulas simple. It is straightforward to extend the discussion to the situation when there are $m > 2$ regularization parameters. The set $\mathscr{D}$ in (2.3) and (2.8) is a hypersurface of dimension $m-1$, and the analogue of equation (2.23) becomes

$$\lambda_1 = g(\lambda_2, \ldots, \lambda_m).$$

We omit the details.

3 Krylov–Tikhonov multi-parameter regularization

Krylov–Tikhonov methods are obtained by projecting a Tikhonov-regularized problem (1.2) or (1.4) onto a Krylov subspace. Different Krylov–Tikhonov methods are defined by varying the original problem (1.2) or (1.4), or the Krylov subspace. For instance, the authors of [13] project the general form Tikhonov problem ($L \neq I_N$) onto the subspaces $\mathscr{K}_k(A^T A, A^T b)$, $k \geq 1$, generated by the Golub–Kahan bidiagonalization algorithm. In [3] problem (1.2) with $L = I_N$ is projected onto subspaces $\mathscr{K}_k(A, b)$, $k \geq 1$, generated by the Arnoldi algorithm, while in [6] problem (1.4) is projected onto the same subspaces. Range restricted Arnoldi methods that generate solution subspaces $\mathscr{K}_k(A, Ab)$, $k \geq 1$, are discussed in [12,18]. In these ways, one can efficiently approximate the regularized solution of large-scale regularized problems (1.2) and (1.4). Krylov–Tikhonov methods are intrinsically multi-parameter methods since, in addition to the dimension k of the Krylov subspace, a regularization parameter λ or a regularization vector Λ have to be chosen for the problems (1.2) or (1.4), respectively. Generally, λ or Λ depend on the dimension k. Recently, Regińska [25] analyzed this kind of regularization methods in Hilbert spaces for two particular projections of A, and for L the identity.

In the following we restrict ourselves to the case when $A \in \mathbb{R}^{N \times N}$ and consider methods based on the Arnoldi algorithm [29, §6.3]. The Arnoldi algorithm typically requires fewer matrix-vector product evaluations to deliver approximate solutions of roughly the same quality than methods based on Golub–Kahan bidiagonalization; see, e.g., [18] for illustrations. Application of k steps of the Arnoldi algorithm to A with initial vector $b/\|b\|$ yields the decomposition

$$AW_k = W_{k+1} \bar{H}_k \,, \quad \text{where} \quad W_{k+1} \in \mathbb{R}^{N \times (k+1)}, \, W_k = W_{k+1} \begin{bmatrix} I_k \\ 0 \end{bmatrix} \in \mathbb{R}^{N \times k}, \tag{3.1}$$

have orthonormal columns, $W_k e_1 = b/\|b\|$, and $\bar{H}_k \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg. We assume here that k is small enough so that no breakdown takes place. Therefore, the columns of W_k form an orthonormal basis for the space $\mathscr{K}_k(A, b)$. We describe a method that is analogous to the scheme in [13], but with

$\mathscr{K}_k(A,b)$ as solution subspace. We refer to this method as the Arnoldi–Tikhonov (AT) method. The special case when $L = I_N$ is described in [3]. Modifications of the decomposition (3.1) required when using the range restricted Arnoldi–Tikhonov (RRAT) method are commented on below. Substituting the Arnoldi decomposition (3.1) into (1.2) and letting $R_k \in \mathbb{R}^{k\times k}$ be the upper triangular matrix in the QR factorization of LW_k yields the minimization problem

$$y_{k,\lambda} = \arg\min_{y\in\mathbb{R}^k} \left\| \begin{bmatrix} \bar{H}_k \\ \lambda^{1/2} R_k \end{bmatrix} y - \begin{bmatrix} \|b\|e_1 \\ 0 \end{bmatrix} \right\|. \tag{3.2}$$

We note that, due to (1.3), the coefficient matrix in (3.2) has full column rank. An approximate solution of (1.2) is given by $x_{k,\lambda} = W_k y_{k,\lambda}$. The regularization parameter λ is determined so that $x_{k,\lambda}$ satisfies the discrepancy principle. Hence, we would like

$$\|Ax_{k,\lambda} - b\| = \left\|\bar{H}_k y_{k,\lambda} - \|b\|e_1\right\| = \eta\varepsilon.$$

To secure that Newton's method converges monotonically, one carries out the change of variable $\mu = \lambda^{-1}$, see [13, 18], and introduces the function

$$\Phi^{(k)}(\mu) := \left\|\bar{H}_k y_{k,1/\mu} - \|b\|e_1\right\|^2 - \eta^2\varepsilon^2. \tag{3.3}$$

The zero $\mu^{(k)}$ of $\Phi^{(k)}(\mu)$ yields the desired value $\lambda^{(k)} = 1/\mu^{(k)}$ of the regularization parameter in (3.2). It can be shown that $\Phi^{(k)}(0)$ is a decreasing function of k, and the existence of the zero $\mu^{(k)}$ requires k to be large enough. Computed examples in [18] illustrate that carrying out more than the minimal number of iterations k may result in improved approximations of x^{ex}. We propose the number of additional iterations to be determined by maximizing (2.3) or (2.8) with respect to the number of iterations. Computed examples of Section 4 illustrate this approach to choosing the discrete regularization parameter k in (3.2).

An alternative stopping rule is to terminate the iterations when the regularization parameter $\lambda^{(k)}$ does not change much with k, i.e., when

$$\frac{|\lambda^{(k+1)} - \lambda^{(k)}|}{\lambda^{(k)}} < \tau \tag{3.4}$$

for some user-chosen tolerance τ. This stopping criterion is meaningful because the goal of Krylov–Tikhonov methods is to compute an accurate approximation $x_{k,\lambda^{(k)}}$ of the solution of (1.2), where λ is properly chosen. Therefore, after projecting (1.2) onto the spaces $\mathscr{K}_k(A,b)$ with increasing k, one should stop when some stabilization in the approximations $x_{k,\lambda^{(k)}}$ occurs. Typically, $x_{k,\lambda^{(k)}}$ stabilizes when $\lambda^{(k)}$ does; see [8] for comments. This stopping criterion is in Section 4 shown to give results comparable to the optimal choice of k.

We briefly describe the multi-parameter Arnoldi–Tikhonov (mP-AT) method proposed in [6] to project problem (1.4) onto the Krylov subspaces $\mathscr{K}_k(A,b)$, $k \geq 1$. We only focus on the case $m = 2$, i.e., when we have two regularization parameters λ_1 and λ_2 and regularization matrices L_1 and L_2. Analogously to (3.2), we determine an approximate solution of (1.4) of the form $x_{k,\Lambda} = W_k y_{k,\Lambda}$ and obtain the minimization problem

$$y_{k,\Lambda} = \arg\min_{y\in\mathbb{R}^k} \left\| \begin{bmatrix} \bar{H}_k \\ (\lambda_1^{(k)})^{1/2} R_k^{(1)} \\ (\lambda_2^{(k)})^{1/2} R_k^{(2)} \end{bmatrix} y - \begin{bmatrix} \|b\|e_1 \\ 0 \\ 0 \end{bmatrix} \right\|, \tag{3.5}$$

where the $R_k^{(i)} \in \mathbb{R}^{k\times k}$ are the upper triangular matrices in the QR factorizations of $L_i W_k$, $i = 1, 2$. We define the associated discrepancy function

$$\bar{\Phi}^{(k)}(\Lambda) := \left\| \|b\|e_1 - \bar{H}_k y_{k,\Lambda} \right\|, \tag{3.6}$$

where $\Lambda = (\lambda_1, \lambda_2)$. Note that the method [6] does not require the computation of the zero of (3.3). Therefore, the squaring of the discrepancy norm and the substitution $\lambda = 1/\mu$ are not necessary. Let k^* be the smallest integer such that

$$\bar{\Phi}^{(k^*)}(0,0) < \eta\varepsilon\,. \tag{3.7}$$

Since $\bar{\Phi}^{(k)}(0,0)$ is the residual norm of the kth iterate determined by GMRES applied to (1.1) with initial iterate $x_0 = 0$ (and $M = N$), the integer k^* typically exists and is quite small; see [7]. Following [6], we approximate for $k > k^*$ the function (3.6) by the linear function

$$\bar{\Phi}^{(k)}_{\text{lin}}(\lambda_1, \lambda_2) = \alpha_0^{(k)} + \alpha_1^{(k)}\lambda_1 + \alpha_2^{(k)}\lambda_2\,,$$

where $\alpha_0^{(k)} = \bar{\Phi}^{(k)}(0,0)$,

$$\alpha_1^{(k)} = \frac{\bar{\Phi}^{(k)}(\lambda_1^{(k)}, 0) - \bar{\Phi}^{(k)}(0,0)}{\lambda_1^{(k)}}\,,$$
$$\alpha_2^{(k)} = \frac{\bar{\Phi}^{(k)}(0, \lambda_2^{(k)}) - \bar{\Phi}^{(k)}(0,0)}{\lambda_2^{(k)}}\,.$$

The parameters $(\lambda_1^{(k)}, \lambda_2^{(k)})$ in the above definitions are the ones employed in (3.5) in the kth iteration. The pair $(\lambda_1^{(k+1)}, \lambda_2^{(k+1)})$ used in the next iteration step is determined by imposing the "approximate discrepancy principle"

$$\bar{\Phi}^{(k)}_{\text{lin}}(\lambda_1, \lambda_2) = \eta\varepsilon\,.$$

This yields

$$\lambda_1 = \frac{\eta\varepsilon - \alpha_0^{(k)}}{\alpha_1^{(k)}} - \frac{\alpha_2^{(k)}}{\alpha_1^{(k)}}\lambda_2 =: \gamma^{(k)} - \delta^{(k)}\lambda_2. \tag{3.8}$$

It follows from (3.7) and the fact that the one-variable functions $\lambda_1 \to \bar{\Phi}^{(k)}(\lambda_1, 0)$ and $\lambda_2 \to \bar{\Phi}^{(k)}(0, \lambda_2)$ are increasing, that the coefficients $\gamma^{(k)}$ and $\delta^{(k)}$ are positive for $k > k^*$. In order for λ_1 and λ_2 to be nonnegative, we require that

$$0 \le \lambda_2 \le \frac{\gamma^{(k)}}{\delta^{(k)}}\,. \tag{3.9}$$

In the terminology of Lu and Pereverzev [19], $\bar{\Phi}^{(k)}_{\text{lin}}(\Lambda)$ is a model function approximation of $\bar{\Phi}^{(k)}(\Lambda)$. At each step of the Arnoldi algorithm, we consider the linear model function obtained by imposing three interpolation conditions

$$\bar{\Phi}^{(k)}_{\text{lin}}(0,0) = \bar{\Phi}^{(k)}(0,0)\,, \quad \bar{\Phi}^{(k)}_{\text{lin}}(\lambda_1^{(k)}, 0) = \bar{\Phi}^{(k)}(\lambda_1^{(k)}, 0)\,, \quad \bar{\Phi}^{(k)}_{\text{lin}}(0, \lambda_2^{(k)}) = \bar{\Phi}^{(k)}(0, \lambda_2^{(k)})\,.$$

The discrepancy curve associated with $\bar{\Phi}^{(k)}_{\text{lin}}(\Lambda)$ is the line connecting the points $(\lambda_1^{(k)}, 0)$ and $(0, \lambda_2^{(k)})$ in the (λ_1, λ_2)-plane. Therefore, the generic function $g(\lambda_2)$ defined in (2.23) is determined by (3.8) for each $k > k^*$.

Differently from the approach adopted in [6], we choose the pair $(\lambda_1^{(k+1)}, \lambda_2^{(k+1)})$ so that (3.8) holds and $\|x_{k,\Lambda}\| = \|y_{k,\Lambda}\|$ is maximized. We can derive an explicit expression for $\partial_{\lambda_2}\|y_{k,\Lambda}\|^2$. Letting

$$\bar{H}_k^{\sharp} = \bar{H}_k^T\bar{H}_k + \lambda_1(R_k^{(1)})^T R_k^{(1)} + \lambda_2(R_k^{(2)})^T R_k^{(2)}\,, \tag{3.10}$$

and performing some standard manipulations, we get

$$\partial_{\lambda_2} \left\| y_{k,\Lambda} \right\|^2 = 2 \left(\partial_{\lambda_2} y_{k,\Lambda}, y_{k,\Lambda} \right) , \tag{3.11}$$

where

$$\partial_{\lambda_2} y_{k,\Lambda} = - \left(\bar{H}_k^{\sharp} \right)^{-1} \left((R_k^{(2)})^T R_k^{(2)} - \delta^{(k)} (R_k^{(1)})^T R_k^{(1)} \right) y_{k,\Lambda} . \tag{3.12}$$

Note that (3.11) and (3.12) are particular cases of (2.25) and (2.24), respectively, obtained by taking $g(\lambda_2) = \gamma^{(k)} - \delta^{(k)} \lambda_2$.

The behavior of (3.11) is quite easy to analyze. Let $\zeta_j^{(k)}$, $j = 1, \ldots, k$, denote the generalized singular values of the matrix pair $(R_k^{(1)}, R_k^{(2)})$. We distinguish the following cases:

- if $(\zeta_k^{(k)})^2 \leq \left(\delta^{(k)} \right)^{-1}$, then $\partial_{\lambda_2} \|y_{k,\Lambda}\|^2 \leq 0$, and we maximize $\|x_{k,\Lambda}\|$ by letting λ_2 be as small as possible, i.e., according to (3.9), $\lambda_2 = 0$;
- if $(\zeta_1^{(k)})^2 \geq \left(\delta^{(k)} \right)^{-1}$, then $\partial_{\lambda_2} \|y_{k,\Lambda}\|^2 \geq 0$, and we maximize $\|x_{k,\Lambda}\|$ by letting λ_2 be as large as possible, i.e., according to (3.9), $\lambda_2 = \gamma^{(k)} \left(\delta^{(k)} \right)^{-1}$;
- if none of the above conditions are satisfied, then no conclusion about the sign of $\partial_{\lambda_2} \|y_{k,\Lambda}\|^2$ can be drawn, and we evaluate $\|y_{k,\Lambda}\|^2$ at logarithmically equispaced values of λ_2 in the interval (3.9).

The behavior of the quantities $\|L_i x_{k,\Lambda}\|$ and $\|L_i x_{k,\Lambda}\|^2 + \|L_j x_{k,\Lambda}\|^2$, $i, j = 1, 2$, can be analyzed in an analogous way.

We terminate the computations as soon as

$$\bar{\Phi}^{(k)}(\lambda_1^{(k)}, \lambda_2^{(k)}) < \eta \varepsilon. \tag{3.13}$$

This approach overcomes the main shortcoming of the strategy developed in [6], namely the dependence of the computed approximate solution on the order of the regularization matrices in (3.5). For instance, by exchanging the order of the regularization matrices in (3.5), i.e., by expressing λ_2 as a function of λ_1, one gets

$$\lambda_2 = \frac{\gamma^{(k)}}{\delta^{(k)}} - \frac{1}{\delta^{(k)}} \lambda_1 \tag{3.14}$$

with the bounds $0 \leq \lambda_1 \leq \gamma^{(k)}$, instead of (3.8) and (3.9), respectively. Computing $\partial_{\lambda_1} \|y_{k,\Lambda}\|^2$ as done in (3.11), one obtains

$$\partial_{\lambda_1} \left\| y_{k,\Lambda} \right\|^2 = -2 \left(\left(\bar{H}_k^{\sharp} \right)^{-1} \left((R_k^{(1)})^T R_k^{(1)} - (\delta^{(k)})^{-1} (R_k^{(2)})^T R_k^{(2)} \right) y_{k,\Lambda}, y_{k,\Lambda} \right) ,$$

where $\bar{H}_k^{\sharp}$ is defined as in (3.10). One can easily see that $\|y_{k,\Lambda}\|$ is increasing if $(\zeta_k^{(k)})^2 \leq \left(\delta^{(k)} \right)^{-1}$. Therefore, in this case, in order to maximize $\|x_{k,\Lambda}\|$ one should take the largest possible λ_1, i.e., $\lambda_1 = \gamma^{(k)}$. Correspondingly, thanks to (3.14), $\lambda_2 = 0$. These values agree with the ones chosen when λ_1 is expressed as a function of λ_2. Indeed, if $(\zeta_k^{(k)})^2 \leq \left(\delta^{(k)} \right)^{-1}$, then $\|x_{k,\Lambda}\|$ is still maximized by taking $\lambda_2 = 0$ and, correspondingly, $\lambda_1 = \gamma^{(k)}$. We can conclude that, at least in this situation, the approximate solution is invariant with respect to the order of the regularization matrices. Computed examples of Section 4 illustrate this.

4 Numerical experiments

We report the results of some numerical tests with the parameter choice methods discussed. The first set of experiments is concerned with test problems from [9], and we customize the problems so that the exact solution lies in the null space of one of the regularization matrices L_1 or L_2 in (2.12). Although this is a very particular situation, similar tests were already devised in [2,6] to validate the proposed parameter choice strategies. The second set of experiments illustrates the performance of the multi-parameter Tikhonov, AT, RRAT, and *m*P-AT methods when applied to many popular problems still taken from [9]. Whenever appropriate, we compare the results of our discrepancy-based strategy and the GCV-based approach described in [2]. The third set of experiments considers image deblurring and denoising problems, and we apply the direct multi-parameter Tikhonov regularization and the *m*P-AT methods. Depending on the problem, we take as regularization matrices the identity $I_N \in \mathbb{R}^{N\times N}$, the bidiagonal matrix

$$D_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(N-1)\times N}, \tag{4.1}$$

or the tridiagonal matrix

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2)\times N}. \tag{4.2}$$

The matrices D_1 and D_2 represent scaled finite difference approximations of the first and the second derivative operators in one space-dimension, respectively. All computations have been executed using MATLAB 8.1 with about 15 significant decimal digits.

Example 1. The coefficient matrices considered for the direct regularization method (2.12) are of size 100×100. We take the particular solution $x^{ex} = [1,1,\ldots,1]^T$. The unperturbed right-hand side is obtained by computing $b^{ex} = Ax^{ex}$, and Gaussian white noise is added in such a way that the noise level

$$\widetilde{\varepsilon} = \frac{\|e\|}{\|b^{ex}\|}$$

is 10^{-2}. We use the safety factor $\eta = 1.01$ in the discrepancy principle. As regularization matrices, we take $L_i \in \{I_N, D_1\}$, $i = 1, 2$. Note that $x^{ex} \in \mathcal{N}(D_1)$. Table 4.1 reports the averages of the results obtained running 100 tests. Each test uses a different realization of the noise e. In detail, we consider the two regularization matrix pairs (I_N, D_1) and (D_1, I_N). For both pairs the second regularization parameter λ_2 is varied: we take 50 values of λ_2 logarithmically equispaced between 10^{-8} and 10^2. For each λ_2, a value of λ_1 such that (2.1) is satisfied is determined by applying the Newton zero finder described in Section 2. The initial guess for the zero finder is $\lambda_1 = 10^6$, which is also the largest possible value for λ_1. For each test problem we display the minimum relative error obtained for different combinations of the vectors (λ_1, λ_2) (second and fifth columns), the relative error obtained when selecting the regularization parameters according to (2.3) (third and sixth columns) and (2.8) (fourth and seventh column), and the relative error obtained with the one-parameter Tikhonov method (1.2) with $L = D_1$ (eighth column). A comparison of columns 2-4 of Table 4.1 shows that both parameter selection approaches (2.3) and (2.8) determine approximations of the desired solution x^{ex} of about the same quality as the "optimal combination" of regularization parameters. This also can be seen when comparing the columns 5-7. The difference between the results in columns 2-4 and 5-7 stems from differences in the discretizations of the regularization parameters for I_N and D_1. A finer discretization of the the λ_2 parameter would reduce the difference. The last column of Table 4.1 indicates that for these problems one-parameter Tikhonov regularization with

the regularization matrix D_1 gives approximate solutions of the same quality as two-parameter Tikhonov regularization with the regularization matrix pair (D_1, I_N). When solving an ill-posed problem (1.1), it is generally not known whether it is advantageous to use more than one regularization matrix. It is therefore desirable that the inclusion of a second regularization term does not reduce the accuracy significantly when this term is superfluous. Column 8 of Table 4.1 shows that the quality of the approximate solution does not deteriorate by using the regularization matrix pair (D_1, I_N), when compared with the use of the regularization matrix D_1 only. We remark that the methods described in [2, 6] may give lower accuracy when including more regularization matrices. Of course, it is desirable that the quality of the computed approximate solution improves by including more than one regularization matrix. Illustrations of this are provided below.

In Table 4.2 we compare the relative errors obtained when selecting the regularization parameters (λ_1, λ_2) according to the strategy (2.3) and according to the GCV approach described in [2]. The values of the regularization parameters and the average running times for each test are also displayed.

Table 4.1 Results obtained considering problems (1.2) and (2.12) with $x^{ex} = [1, 1, \ldots, 1]^T$ and $\widetilde{\varepsilon} = 10^{-2}$.

Reg.M.	(I_N, D_1)			(D_1, I_N)			D_1
	error (opt)	**error (2.3)**	**error (2.8)**	**error (opt)**	**error (2.3)**	**error (2.8)**	**error (1P)**
`baart`	7.5722e-03	7.5722e-03	7.5722e-03	5.1508e-04	8.0616e-04	8.0616e-04	8.0643e-04
`deriv2`	1.6085e-03	1.6085e-03	1.6085e-03	3.5001e-04	6.5138e-04	6.5138e-04	6.5138e-04
`phillips`	5.2806e-03	5.2806e-03	5.2806e-03	3.5391e-04	6.2711e-04	6.2711e-04	6.2525e-04
`shaw`	8.7007e-03	8.7007e-03	8.7007e-03	4.8346e-04	7.5421e-04	7.5421e-04	7.5419e-04

Table 4.2 Results obtained considering problem (2.12) with $x^{ex} = [1, 1, \ldots, 1]^T$, $\widetilde{\varepsilon} = 10^{-2}$, and (D_1, I_N). The discrepancy-based strategy (2.3) and the GCV-based approach [2] are compared. The timings are expressed in seconds.

	error (2.3)	λ_1 **(2.3)**	λ_2 **(2.3)**	**time**	**error, GCV**	λ_1**, GCV**	λ_2**, GCV**	**time**
`baart`	8.0616e-04	1.0000e+06	1.0000e-08	6.3e-01	3.7111e-02	1.9558e+02	2.9729e-04	4.8e-02
`deriv2`	6.5138e-04	1.0000e+06	1.0000e-08	9.5e-01	4.1505e-02	1.4808e-01	5.9095e-07	5.6e-02
`phillips`	6.2711e-04	1.0000e+06	1.0000e-08	6.4e-01	3.7959e-03	9.7416e+03	8.3653e-03	5.8e-02
`shaw`	7.5421e-04	1.0000e+06	1.0000e-08	6.4e-01	7.5518e-03	7.9495e+02	7.4205e-05	4.6e-02

Figure 4.1 displays the behavior of various quantities associated to the problem `phillips` with $x^{ex} = [1, 1, \ldots, 1]^T$ and $\widetilde{\varepsilon} = 10^{-2}$. With respect to the results displayed in the previous tables, here we only consider a single test (i.e., only one realization of the noise e). Both the regularization matrix pairs (D_1, I_N) and (I_N, D_1) are tested. In particular, we plot the relative errors for all the admissible pairs (λ_1, λ_2), sequentially varying both λ_2 and determining the corresponding parameter λ_1 such that (2.1) is satisfied. In a similar way, we display the quantities $\|x_\Lambda\|$ and $\|x_\Lambda\|^2 + \|D_1 x_\Lambda\|^2$, and the values of λ_1 and λ_2. We use special markers to highlight the quantities delivering the minimum attainable relative error (black circle), and the quantities satisfying the criteria (2.3) (black square) and (2.8) (black hexagram). When considering the regularization matrices (D_1, I_N), only 35 values (out of 50) of λ_2 between 10^{-8} and 10^2 have a matching λ_1 that satisfies (2.1). In particular, it is not possible to compute λ_1 for the largest λ_2.

When applying the 2P-AT method (3.5), the size of the coefficient matrices is $N = 200$ and $x^{ex} = [1, 2, \ldots, N]^T$ is a vector of increasing linearly equispaced values; the noise level is $\widetilde{\varepsilon} = 10^{-1}$ and $\eta = 1.1$. The regularization matrices employed are (I_N, D_2) and (D_2, I_N). The regularization parameters λ_i, $i = 1, 2$, are selected according to (3.8) and (2.3). Analogously to the previous experiments, the first regularization parameter λ_1 assumes logarithmically equispaced values between 10^{-10} and 10^2. The maximum number of Arnoldi steps allowed is 20. Table 4.3 reports the averages of the minimum relative errors obtained over 100 tests (third column). More precisely, at each iteration of the Arnoldi algorithm,

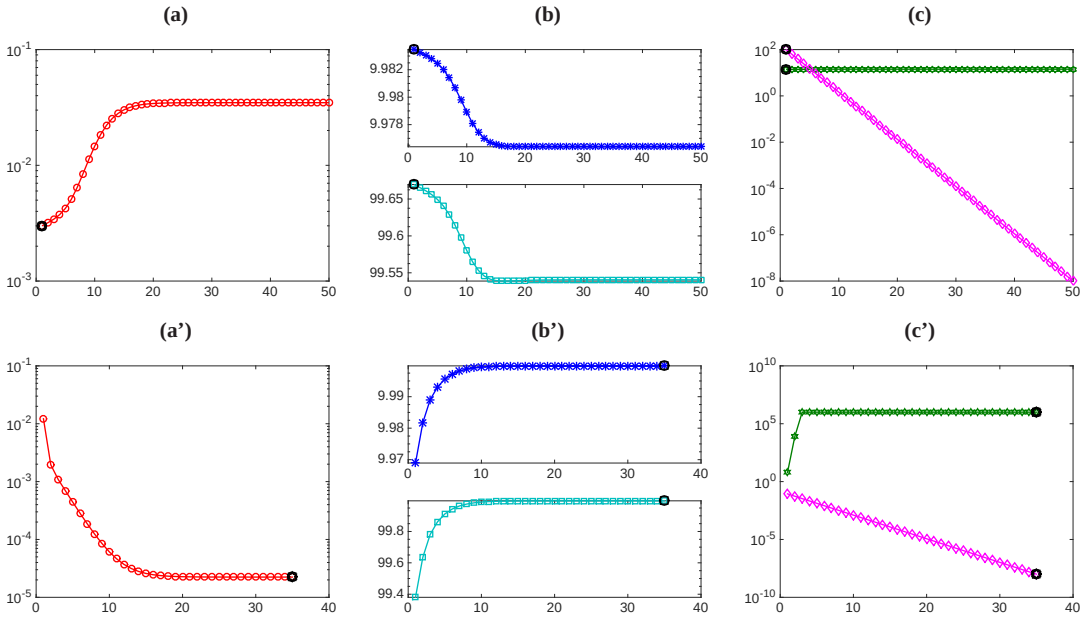

Fig. 4.1 Test problem `phillips` with $x^{ex} = [1,1,\ldots,1]^T$. The first and second rows display quantities related to problem (2.12) with (I_N, D_1) and (D_1, I_N), respectively. **(a)**, **(a')**: relative errors for the combinations of λ_i's displayed in the third frames; **(b)**, **(b')**: values of $\|x_\Lambda\|$ (upper frame) and $\|x_\Lambda\|^2 + \|D_1 x_\Lambda\|^2$ (lower frame) for the combinations of λ_i's displayed in the third frames; **(c)**, **(c')**: pairs of values λ_1 (hexagram) and λ_2 (diamond).

the criteria (3.8) and (2.3) are applied, and the error is minimized with respect to the number of Arnoldi steps; the associated regularization parameter pairs are displayed in the fourth and fifth columns. The relative errors obtained when the discrepancy-based stopping criterion (3.13) is satisfied (sixth column), as well as the components of the vector (λ_1, λ_2) obtained by (3.8), (2.3) and (3.13) (seventh and eighth columns), are also reported. The average number of iterations is displayed in parentheses. The direct two-parameter Tikhonov regularization methods (2.12) equipped with the discrepancy principle (2.3) and the GCV criterion [2] are compared in Table 4.4 for problems of size $N = 100$. The approximate solutions determined by the GCV approach [2] can be seen to deliver less accurate approximations of the desired solution x^{ex} than the discrepancy-based strategy (2.3).

Table 4.3 Results obtained considering problem (3.5) with $x^{ex} = [1,2,\ldots,N]^T$ and $\widetilde{\varepsilon} = 10^{-1}$. The regularization matrix pairs are reported in the second column. The criteria used are recalled in the third and sixth column headings.

	Reg.M.	**error (2.3,3.8,opt)**	λ_1	λ_2	**error (2.3,3.8,3.13)**	λ_1	λ_2
`baart`	(D_2, I_N)	4.0561e-02 (4.4)	1.07e+10	2.30e-05	6.5831e-02 (3.0)	1.41e+03	1.00e-10
`baart`	(I_N, D_2)	4.8238e-02 (3.5)	3.63e-05	1.93e+06	6.4423e-02 (3.0)	1.36e-08	1.01e+04
`deriv2`	(D_2, I_N)	3.7113e-01 (4.7)	9.20e-01	5.85e-06	4.5094e-01 (3.0)	3.38e-01	1.00e-10
`deriv2`	(I_N, D_2)	3.6711e-01 (4.8)	8.89e-06	7.10e-01	4.5115e-01 (3.0)	1.83e-10	3.42e-01
`phillips`	(D_2, I_N)	8.2781e-02 (13.8)	3.01e+05	1.76e-01	2.0457e-01 (2.1)	8.35e-01	1.00e-10
`phillips`	(I_N, D_2)	1.7842e-02 (6.6)	6.72e-02	4.45e+04	2.0056e-01 (3.1)	9.56e-08	7.33e-01
`shaw`	(D_2, I_N)	2.6984e-01 (7.7)	3.33e+04	9.42e-04	4.8510e-01 (4.1)	2.87e+01	2.05e-03
`shaw`	(I_N, D_2)	2.6447e-01 (7.8)	8.72e-04	5.09e+04	4.9258e-01 (4.1)	2.12e-03	1.53e+01

Figure 4.2 considers one single test (i.e., only one realization of e). The coefficient matrix `baart`, the exact solution $x^{ex} = [1,2,\ldots,N]^T$, the noise level $\widetilde{\varepsilon} = 10^{-1}$, and the 2P-AT method (3.5) are considered.

Table 4.4 Results obtained considering problem (2.12) with $x^{ex} = [1,2,\dots,N]^T$, $\widetilde{\varepsilon} = 10^{-1}$, and (I_N, D_2). The discrepancy-based strategy (2.3) and the GCV-based approach [2] are compared. The timings are expressed in seconds.

	error (2.3)	λ_1 **(2.3)**	λ_2 **(2.3)**	**time**	**error, GCV**	λ_1**, GCV**	λ_2**, GCV**	**time**
`baart`	3.4665e-03	1.0000e-08	1.0000e+06	6.6e-01	1.6998e-01	1.4086e-02	4.4994e+02	4.9e-02
`deriv2`	3.8149e-03	1.0000e-08	1.0000e+06	1.0e+00	7.9608e-02	9.4812e-06	7.0416e+00	4.9e-02
`phillips`	1.8433e-03	1.0000e-08	1.0000e+06	5.7e-01	1.7401e-02	6.7284e-02	1.3884e+06	5.0e-02
`shaw`	1.8740e-03	1.0000e-08	1.0000e+06	6.5e-01	4.7313e-02	7.2550e-02	1.1133e+06	5.2e-02

We plot the relative errors, $(\lambda_1^{(k)}, \lambda_2^{(k)})$, and the discrepancy function versus the number of Arnoldi steps k. The criteria (3.8) and (2.3) are applied. The iteration delivering the best relative error is highlighted by a black circle, while the iteration satisfying the stopping criterion (3.13) is highlighted by a black square.

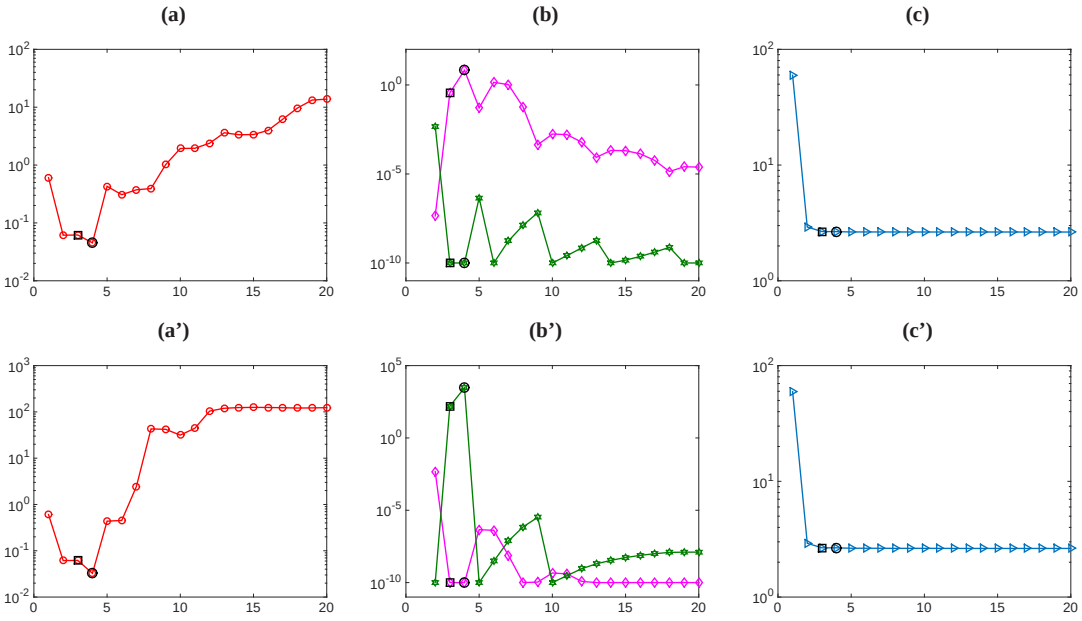

Fig. 4.2 Test problem `baart` with $x^{ex} = [1,2,\dots,N]^T$. The first and second rows refer to the regularization matrix pairs (D_2, I_N) and (I_N, D_2), respectively. **(a)**, **(a')**: relative errors versus number of steps; **(b)**, **(b')**: pairs of values λ_1 (hexagram) and λ_2 (diamond) versus the number of steps; **(c)**, **(c')**: values of the discrepancy function versus the number of steps.

Looking at the results of the first set of experiments, we can conclude that the proposed strategies are able to detect the most meaningful regularization term (i.e., the one involving the matrix D_i such that $x^{ex} \in \mathcal{N}(D_i)$, $i = 1, 2$), and to weigh it with a large regularization parameter. In particular, looking at Tables 4.2 and 4.4, we can see that the selected regularization parameter is always the largest one (i.e., 10^6) if $x^{ex} \in \mathcal{N}(D_i)$, $i = 1, 2$, and the smallest one (i.e., 10^{-8}) otherwise. In this peculiar situation, problem (2.12) essentially reduces to (1.2). The results are also quite invariant with respect to the order of the regularization matrices. Moreover, the quality of the regularized solution and the values of the regularization parameters seem to be almost equal for all the test problems, independently of the additional constraint (2.3) or (2.8) imposed on the discrepancy curve. Indeed, since the computed solutions are quite accurate approximations of x^{ex}, and $x^{ex} \in \mathcal{N}(D_i)$, $i = 1, 2$, the criterion (2.8) essentially coincides with (2.3). The results delivered by the GCV-based two-parameter Tikhonov method [2] are usually slightly worse than the discrepancy-based ones. A reason for this may be that GCV does not require an estimate

of $\|e\|$. Moreover, the regularization parameters are chosen sequentially by GCV, while the proposed discrepancy-based strategy simultaneously evaluates many different regularization vectors. The latter is a reason for the new strategy being slower than GCV. Finally, looking at Table 4.3, we see that for many problems the discrepancy-based 2P-AT method determines approximate solutions whose quality is comparable to the optimal attainable one. We again underline that this set of experiments is quite artificial, and its purpose is to validate the new discrepancy-based parameter choice strategy.

Example 2. Some popular test problems with the solutions given in [9] (or slightly modified versions of them) are used, among them two versions of the test problem `deriv2`. The exact solution of the test denoted by `deriv2,1` is a discretization of the function $f(t) = t$, while the exact solution of the test denoted by `deriv2,2` is a discretization of the function $f(t) = \exp(t)$.

To test the two-parameter Tikhonov method (2.12), coefficient matrices of size 100×100 and $\widetilde{\varepsilon} = 10^{-2}$ are considered. As regularization matrices, the pairs (I_N, D_1), (D_1, I_N), (D_1, D_2), and (D_2, D_1) are employed. In Figure 4.3, the assumptions of Lemma 2.1 are checked for the test problem `phillips`. More precisely, λ_1 and λ_2 are varied (on the horizontal plane), and for each pair (λ_1, λ_2) the value

$$\left| \left(L_1 \partial^2_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)} \right) - 2 \left(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)} \right) \right| \tag{4.3}$$

is displayed along the vertical axis. We can see that the above difference often is negligible; the largest values are attained when λ_2 is much larger than the corresponding λ_1. Analogous results hold for all the test problems listed in Table 4.5.

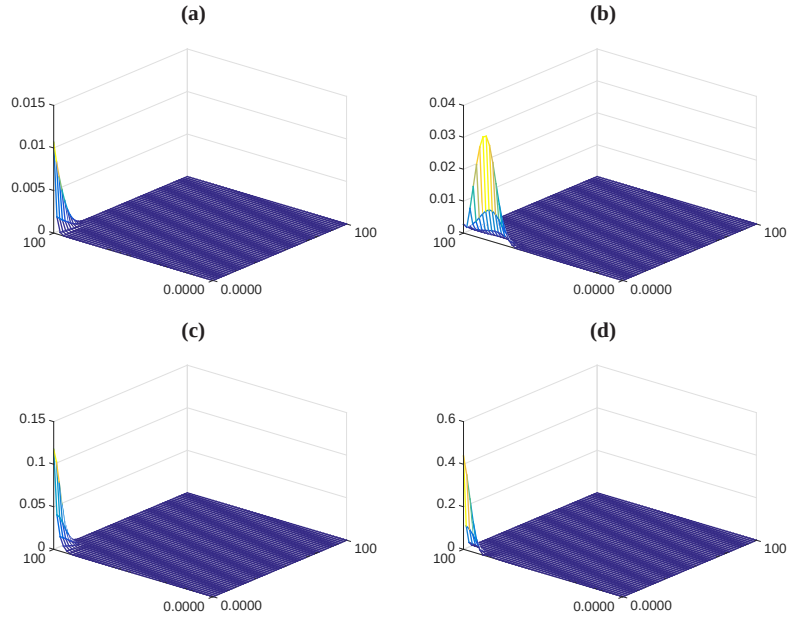

Fig. 4.3 Differences (4.3) for test problem `phillips`. The considered regularization matrices are: **(a)** (I_N, D_1), **(b)** (D_1, I_N), **(c)** (D_1, D_2), **(d)** (D_2, D_1). Note that the figures have different vertical scales.

Similarly to Example 1, Table 4.5 displays the average results obtained running each test problem 100 times with different noise realizations. The strategy employed to compute the regularized solutions, and the layout of the table, have already been described for the first set of experiments. Table 4.6 compares

the present strategy with the GCV-based strategy derived in [2]. We can again see that the discrepancy principle delivers more accurate solutions than the GCV criterion.

Table 4.5 Results obtained by Tikhonov regularization problems (1.2) and (2.12) with the solution given in [9] and $\widetilde{\varepsilon} = 10^{-2}$. The regularization matrix pairs are reported in the table header, and (2.8) is implemented with $(L_1, L_2) = (D_1, I_N)$.

Reg.M.	(D_1, I_N)			(I_N, D_1)			I_N
	error (opt)	**error (2.3)**	**error (2.8)**	**error (opt)**	**error (2.3)**	**error (2.8)**	**error (1P)**
`baart`	1.5081e-01	1.5081e-01	1.5081e-01	1.5676e-01	1.5676e-01	1.5676e-01	1.8541e-01
`deriv2,1`	5.9526e-02	5.9526e-02	5.9526e-02	6.1646e-02	6.1646e-02	6.1646e-02	2.6407e-01
`deriv2,2`	5.4627e-02	5.4627e-02	5.4627e-02	5.6260e-02	5.6260e-02	5.6260e-02	2.5011e-01
`phillips`	2.6177e-02	2.8088e-02	2.8088e-02	2.5375e-02	2.7546e-02	2.7546e-02	2.7783e-02
`shaw`	1.2194e-01	1.6726e-01	1.6726e-01	1.1288e-01	1.3535e-01	1.3535e-01	1.6726e-01

Table 4.6 Results obtained by Tikhonov regularization problems (1.2) and (2.12) with the solution given in [9], $\widetilde{\varepsilon} = 10^{-2}$, and (D_2, D_1) as regularization matrices. Comparisons with the multi-parameter GCV method proposed in [2]. The timings are expressed in seconds.

	error (2.3)	λ_1 **(2.3)**	λ_2 **(2.3)**	**time**	**error, GCV**	λ_1**, GCV**	λ_2**, GCV**	**time**
`baart`	8.3648e-02	1.8944e+02	1.0000e-08	1.8e+00	1.0793e-01	5.5998e+00	1.1107e-02	5.4e-02
`deriv2,1`	3.2009e-03	1.0000e+06	1.0000e-08	1.7e+00	2.1343e-02	5.8910e+00	3.2295e-04	5.3e-02
`deriv2,2`	2.7612e-02	2.6670e+05	4.4887e-04	1.6e+00	4.4470e-02	1.6180e+00	6.5289e-04	5.5e-02
`phillips`	3.0854e-02	1.0104e+02	1.0000e-08	2.8e+00	3.1964e-01	2.1120e+00	6.8051e-02	5.3e-02
`shaw`	2.0864e-01	3.3745e+00	2.7494e-02	2.1e+00	2.1099e-01	3.3002e-01	3.1482e-03	5.3e-02

Figure 4.4 focuses on the test problem `deriv2,2`. Using the layout of Figure 4.1, the relative errors for all the admissible pairs (λ_1, λ_2) are plotted. In a similar way, the quantities $\|x_\Lambda\|$ and $\|D_1 x_\Lambda\|^2 + \|D_2 x_\Lambda\|^2$, and the values of λ_1 and λ_2 are displayed. Special markers are employed to highlight the quantities minimizing the relative error (black circle), and satisfying (2.3) and (2.8) with $(L_1, L_2) = (D_1, D_2)$ (black square and hexagram, respectively). We can see that, in the (D_2, D_1) case, the largest values of λ_2 do not allow a matching λ_1 satisfying the discrepancy principle (2.1). The top left frame of Figure 4.7 displays the reconstruction obtained by the (D_2, D_1) regularization. Looking at these results, we can see that for many problems the reconstructions computed by the two-parameter methods are improved with respect to the one-parameter case. Often the criteria (2.3) and (2.8) deliver solutions of similar quality with accuracy close to the optimal one. Moreover, the behavior of the solution is usually almost invariant with respect to the order of regularization matrix pair.

Table 4.7 reports the average relative errors obtained when applying the AT (with $L = D_2$) and RRAT methods 100 times with different noise realizations. The size of the coefficient matrices is 200×200, $\widetilde{\varepsilon} = 10^{-2}$, and $\eta = 1.1$. Recalling the explanations given at the beginning of Section 3, at each Arnoldi iteration a Newton zero-finder is employed to compute $\lambda^{(k)}$. The relative errors are displayed when the final dimension k of the Krylov subspaces $\mathscr{K}_k(A, b)$ or $\mathscr{K}_k(A, Ab)$ is selected according to an optimal criterion (i.e., the relative error is minimized) (third column), according to (2.3) (fourth column), according to (2.8) (just for general form problems) (fifth column), according to the stabilization of the values of $\lambda^{(k)}$ (3.4) (with $\tau = 1$) (sixth column), and according to the stopping rule proposed in [18, 13] (seventh column). The latter criterion consists in stopping the iterations as soon as the quantity $\|\bar{H}_k y_{k,0} - e_1 \|b\|\|$ (cf. (3.2)) drops below the threshold $\eta\varepsilon$. The relative errors delivered by the stopping strategies (2.3) or (2.8) are often close to the optimal ones, and one commonly gets an improved reconstruction with respect to the criterion proposed in [18, 13]. Figure 4.5 displays the history of the relative errors, of the quantities $\|x_{k,\lambda^{(k)}}\|$, and of the regularization parameters $\lambda^{(k)}$ for the test problems `phillips` and `baart`, when the

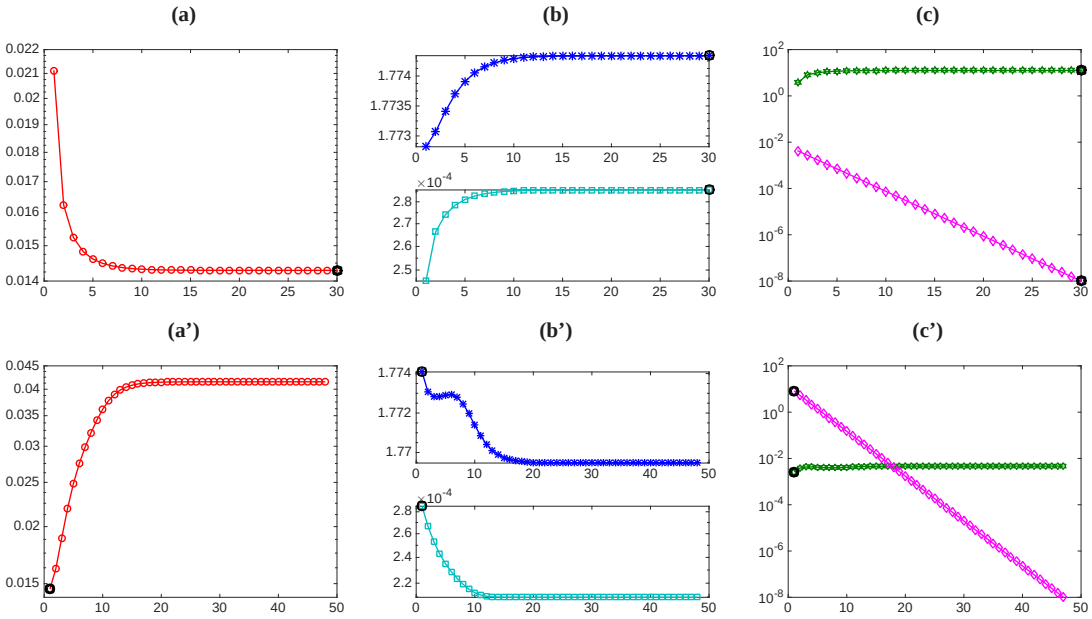

Fig. 4.4 Test problem `deriv2,2`. The first and second rows display quantities relative to problem (2.12) with (D_2, D_1) and (D_1, D_2), respectively. **(a)**, **(a')**: relative errors for the combinations of λ_i's displayed in the third frames; **(b)**, **(b')**: values of $\|x_\lambda\|$ (upper frame) and $\|D_1 x_\lambda\|^2 + \|D_2 x_\lambda\|^2$ (lower frame); **(c)**, **(c')**: pairs of values λ_1 (hexagram) and λ_2 (diamond).

AT and RRAT methods are performed, respectively. The reconstructions obtained by applying different stopping criteria are shown in the second and third frames of Figure 4.7, respectively.

Table 4.7 Averages of the relative errors obtained with the AT and RRAT methods (specified in the second column); different stopping criteria (listed in the column headings) are applied. The average number of iterations is reported in parentheses.

	Meth.	**(opt)**	**(2.3)**	**(2.8)**	**(3.4)**	[18, 13]	**1P** (D_1)
`baart`	AT	1.15e-02 (20.8)	1.15e-02 (14.8)	2.76e-01 (3.0)	4.60e-02 (5.4)	2.76e-01 (3.0)	2.05e-01 (4.0)
`baart`	RRAT	1.78e-01 (3.0)	2.77e-01 (2.7)	-	2.08e-01 (3.9)	5.27e-01 (2.4)	
`deriv2,1`	AT	7.86e-02 (7.8)	1.42e-01 (1.12)	1.65e-01 (5.3)	1.23e-01 (7.8)	1.74e-01 (5.1)	2.46e-01 (7.2)
`deriv2,1`	RRAT	2.80e-01 (9.9)	2.82e-01 (7.7)	-	2.86e-01 (5.1)	3.11e-01 (4.0)	
`deriv2,2`	AT	2.54e-01 (6.9)	2.84e-01 (17.2)	3.41e-01 (4.1)	3.16e-01 (5.1)	3.41e-01 (4.0)	2.37e-01 (19.8)
`deriv2,2`	RRAT	2.64e-01 (8.2)	2.64e-01 (6.0)	-	2.70e-01 (5.0)	2.87e-01 (4.0)	
`phillips`	AT	2.38e-02 (13.4)	2.94e-02 (8.4)	8.67e-02 (4.0)	2.85e-02 (6.1)	8.67e-02 (4.0)	2.38e-01 (20.0)
`phillips`	RRAT	2.79e-02 (6.8)	2.86e-02 (9.7)	-	2.89e-02 (5.0)	2.92e-02 (4.0)	
`shaw`	AT	7.94e-02 (7.8)	1.44e-01 (11.2)	1.64e-01 (5.3)	1.14e-01 (7.7)	1.71e-01 (5.1)	1.27e-01 (6.2)
`shaw`	RRAT	1.46e-01 (10.1)	1.46e-01 (9.1)	-	1.48e-01 (5.1)	1.69e-01 (4.0)	

Finally, we test the performance of the 2P-AT method. Analogously to Example 1, we compare the results obtained when the approximation (3.8) together with the rule (2.3) is equipped with the discrepancy-based stopping criterion (3.13), and when the stopping iteration is selected in order to minimize the relative error (i.e., optimal stopping criterion). Similarly to Table 4.3, Table 4.8 reports the averages of the minimum relative errors obtained running 100 times each test problem, and varying the noise realization. The coefficient matrices are of size 200×200, and $\widetilde{\varepsilon} = 10^{-2}$, $\eta = 1.1$. The average number of iterations is displayed in parentheses; the components of the vector (λ_1, λ_2) also are reported.

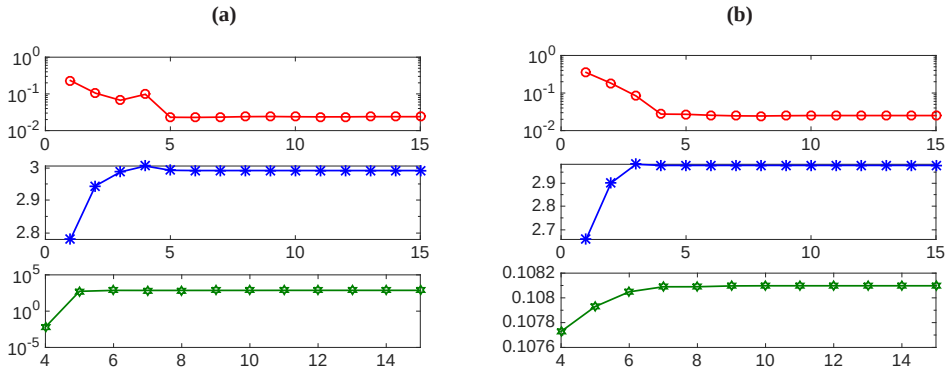

Fig. 4.5 The AT method (column **(a)**) and the RRAT method (column **(b)**) applied to the test problem `phillips`. The first box displays the relative errors versus the number of iterations k, the second box displays the quantities $\|x_{k,\lambda^{(k)}}\|$ versus the number of iterations k, and the third box displays the values $\lambda^{(k)}$ versus the number of iterations k.

Table 4.8 Results obtained considering problem (3.5) with the solution given in [9] and $\widetilde{\varepsilon} = 10^{-2}$. The regularization matrix pairs are reported in the second column. The criteria used are recalled in the third and fourth column headings.

	Reg.M.	**error (2.3,3.8,opt)**	**error (2.3,3.8,3.13)**	λ_1	λ_2
`baart`	(D_2,D_1)	3.9490e-02 (4.6)	4.3014e-02 (4.0)	4.0058e+00	2.5516e-02
`baart`	(D_1,D_2)	5.2746e-02 (4.1)	5.5152e-02 (4.0)	2.5382e-02	3.9275e+00
`deriv2,1`	(D_2,D_1)	2.4281e-01 (7.0)	2.7480e-01 (5.9)	2.3625e-07	1.0438e-04
`deriv2,1`	(D_1,D_2)	2.4822e-01 (6.9)	2.6928e-01 (5.9)	1.0439e-04	9.0103e-07
`deriv2,2`	(D_2,D_1)	2.3914e-01 (6.9)	2.8082e-01 (5.2)	5.2474e-08	4.2868e-04
`deriv2,2`	(D_1,D_2)	2.4581e-01 (6.4)	2.7423e-01 (5.2)	4.2869e-04	6.2413e-07
`phillips`	(D_2,D_1)	1.0888e-02 (11.3)	2.6077e-02 (8.1)	2.3893e+02	3.6592e+00
`phillips`	(D_1,D_2)	1.0927e-02 (11.3)	2.2764e-02 (8.1)	4.9763e+00	2.3893e+02
`shaw`	(D_2,D_1)	1.1856e-01 (6.0)	1.7966e-01 (4.0)	1.0000e-10	1.1298e-01
`shaw`	(D_1,D_2)	1.1939e-01 (6.1)	1.7966e-01 (4.0)	1.1298e-01	1.0000e-10

Similarly to Figures 4.2 and 4.4, Figure 4.6 displays one single test performed with the coefficient matrix `deriv2,1`. Both the regularization matrix pairs (D_1,D_2) and (D_2,D_1) are considered: for each combination of the regularization matrices, the relative errors, $(\lambda_1^{(k)},\lambda_2^{(k)})$, and the discrepancy function versus the number of Arnoldi steps k are plotted. The iteration delivering the best relative error is highlighted by a black circle, while the iteration satisfying the discrepancy-based stopping criterion is highlighted by a black square. The reconstructions associated to the matrices (D_2,D_1) are displayed in the fourth frame of Figure 4.7. We can conclude that the quality of the reconstruction is stable with respect to the regularization matrix pairs (see in particular Figure 4.6), and that the relative errors obtained by applying simultaneously the strategies (3.8), (2.3), and (3.13) are close to the optimal ones.

Example 3. The test images employed for this set of experiments are available in the MATLAB package [24], and are of size 256×256 pixels. To restore the blurred and noisy image shown in Figure 4.8, frame **(b)**, the two-parameter Tikhonov method (2.12) is applied, equipped with the criterion (2.3). The blur is defined by a symmetric Gaussian point spread function (PSF) given by

$$k(s,t) = \frac{1}{2\pi\alpha^2} \exp\left(-\frac{1}{2\alpha^2}(s^2+t^2)\right), \tag{4.4}$$

where $\alpha = 2.5$. Periodic boundary conditions are considered, so that the matrix A in (1.1), which represents the blur, is block circulant with circulant blocks. Matrix multiplications and inversions are efficiently performed with the aid of the FFT (see [11, Chapter 4] for the details). The noise level is 10^{-2}. The reg-

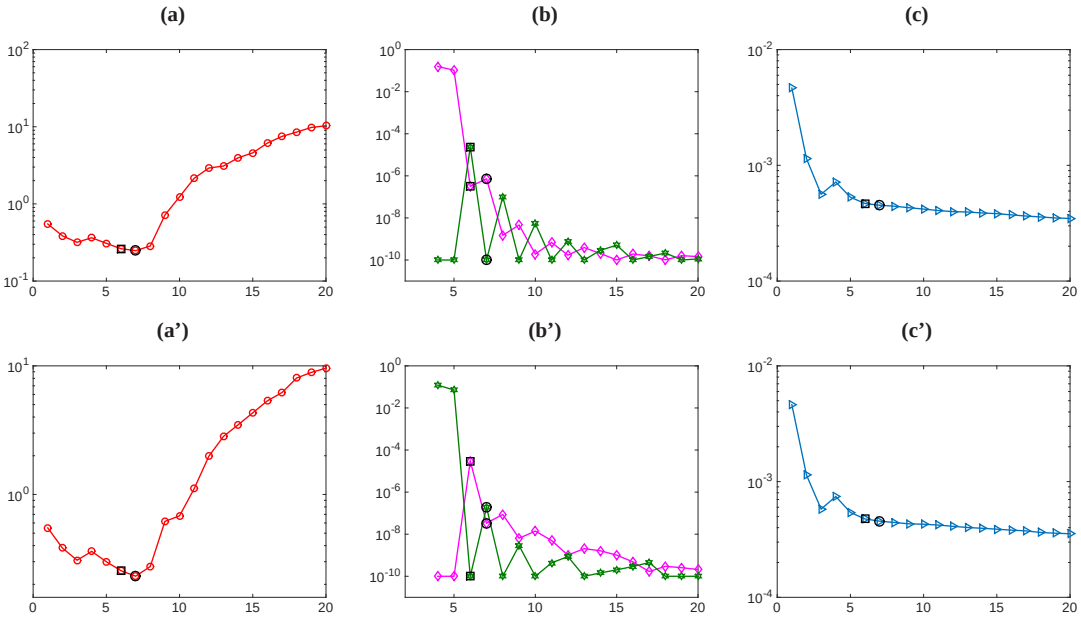

Fig. 4.6 Test problem `deriv2,1`. The first and second rows display quantities relative to problem (3.5) with (D_2, D_1) and (D_1, D_2), respectively. **(a)**, **(a')**: history of the relative errors; **(b)**, **(b')**: history of the regularization parameters $\lambda_1^{(k)}$ (hexagram) and $\lambda_2^{(k)}$ (diamond); **(c)**, **(c')**: history of the values of the discrepancy function.

ularization matrices employed are discretizations of the second and first derivative operators, which act along the vertical and horizontal directions of the image, respectively. More precisely, the regularization matrices are defined by

$$L_1 = I_n \otimes D_2 \quad \text{and} \quad L_2 = D_1 \otimes I_n \,,$$

where D_1 and D_2 are defined by (4.1) and (4.2) with $N = n = 256$. The reconstruction obtained is displayed in Figure 4.8, frame **(c)**. The relative restoration error is $2.5209 \cdot 10^{-1}$, and the corresponding regularization vector is $(\lambda_1, \lambda_2) = (2.5594 \cdot 10^{-3}, 7.4964 \cdot 10^{-2})$.

The 2P-AT method (3.5), equipped with the rules (3.8) and (2.3), is applied to restore the blurred and noisy image shown in Figure 4.9, frame **(b)**. A symmetric Gaussian PSF, with $\alpha = 2$ in (4.4), is considered. In this case, reflexive boundary conditions are employed. The noise level is 10^{-2}. The regularization matrices are

$$L_1 = \begin{bmatrix} I_n \otimes D_1 \\ D_1 \otimes I_n \end{bmatrix} \in \mathbb{R}^{2n(n-1)\times N} \quad \text{and} \quad L_2 = I_N \in \mathbb{R}^{N\times N} \,,$$

where $n = 256$, $N = n^2$; in particular, L_1 is defined by stacking the discretized vertical and horizontal first derivative operators. The reconstruction obtained at the 6th iteration of the Arnoldi algorithm is displayed in Figure 4.9, frame **(c)**. The relative restoration error is $1.8660 \cdot 10^{-1}$, and the corresponding regularization vector is $(\lambda_1, \lambda_2) = (5.5782 \cdot 10^{-9}, 1.0071 \cdot 10^{-4})$.

5 Final remarks

In this paper we proposed and analyzed a new strategy to perform multi-parameter regularization. The basic idea is to define regularized solutions that simultaneously satisfy the discrepancy principle and maximize some norm or seminorm. The numerical experiments with the two-parameter Tikhonov method, the

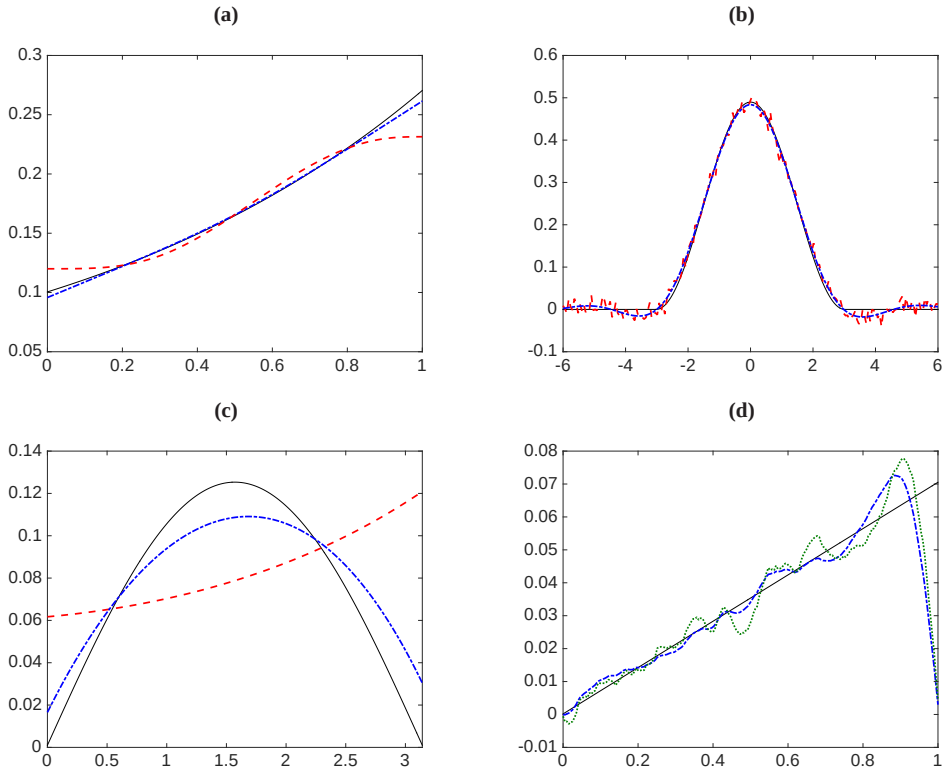

Fig. 4.7 Reconstructions obtained by the two-parameter Tikhonov method (frame **(a)**), by the AT method (frame **(b)**), by the RRAT method (frame **(c)**), and by the 2P-AT method (frame **(d)**). In each frame, the exact solution is plotted by a thin solid line. In frame **(a)**, the reconstruction obtained by (1.2) is plotted by a dashed line, and the reconstruction obtained by (2.12) and the criterion (2.3) is plotted by a dashed-dotted line. In frames **(b)** and **(c)**, the reconstruction obtained by the stopping criterion employed in [18, 13] is plotted by a dashed line, and the reconstruction obtained applying the stopping criterion (2.3) is plotted by a dashed-dotted line. In frame **(d)**, the reconstruction obtained by (3.8) and (2.3) is plotted by a dashed-dotted line, and the optimal reconstruction obtained by (3.8) is plotted by a dotted line.

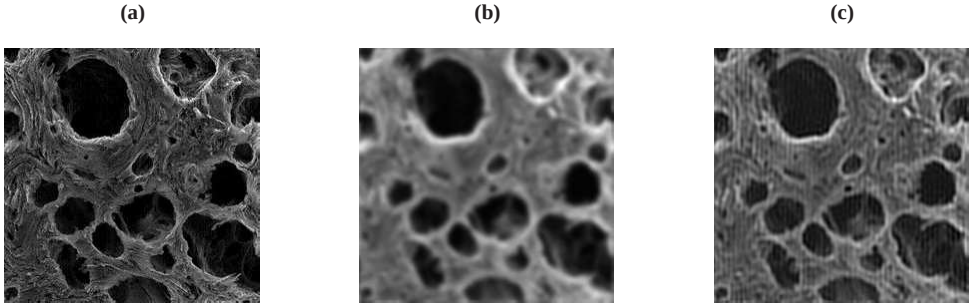

Fig. 4.8 **(a)** exact image; **(b)** blurred and noisy ($\widetilde{\varepsilon} = 10^{-2}$) image; **(c)** reconstruction obtained by (2.12) with (2.3).

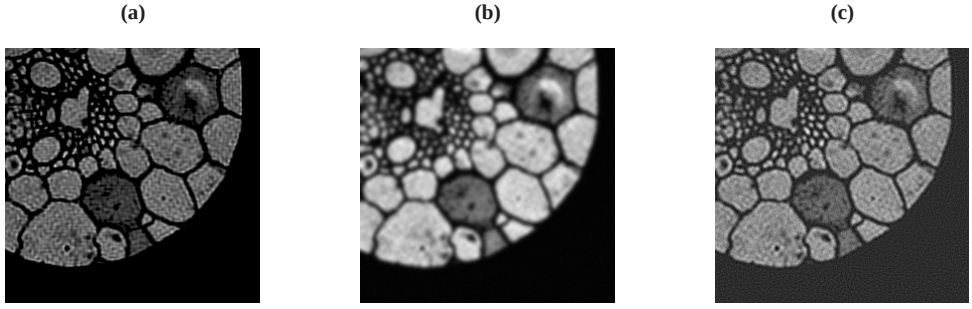

Fig. 4.9 **(a)** Exact image; **(b)** blurred and noisy ($\widetilde{\varepsilon} = 10^{-2}$) image; **(c)** reconstruction obtained by (3.5) and (2.3).

AT and RRAT methods, and the 2P-AT method show that the new strategy can deliver coherent and improved (with respect to the one-parameter methods) results, and therefore can be regarded as a valid alternative to other popular schemes employed so far to implement multi-parameter regularization.

Acknowledgments

We are grateful to Giuseppe Rodriguez for providing us with the GCV-based multi-parameter regularization software, as well as to the editor and referees for remarks that lead to improvements of the presentation.

References

1. Belge, M., Kilmer, M.E., Miller, E.L.: Efficient determination of multiple regularization parameters in a generalized L-curve framework. Inverse Problems **18**, 1161–1183 (2002)
2. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Multi-parameter regularization techniques for ill-conditioned linear systems. Numer. Math. **94**, 203–228 (2003)
3. Calvetti, D., Morigi, S., Reichel, L., Sgallari, F.: Tikhonov regularization and the L-curve for large discrete ill-posed problems. J. Comput. Appl. Math. **123**, 423–446 (2000)
4. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht (1996)
5. Fornasier, M., Naumova, V., Pereverzyev, S.V.: Parameter choice strategies for multipenalty regularization. SIAM J. Numer. Anal. **52**, 1770–1794 (2014)
6. Gazzola, S., Novati, P.: Multi-parameter Arnoldi-Tikhonov methods. Electron. Trans. Numer. Anal. **40**, 452–475 (2013)
7. Gazzola, S., Novati, P., Russo, M.R.: Embedded techniques for choosing the parameter in Tikhonov regularization. Numer. Linear Algebra Appl. **21**, 796–812 (2014)
8. Gazzola, S., Novati, P., Russo, M.R.: On Krylov projection methods and Tikhonov regularization. Electron. Trans. Numer. Anal. **44**, 83–123 (2015)
9. Hansen, P.C.: Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems. Numer. Algorithms **6**, 1–35 (1994)
10. Hansen, P.C.: Rank-Deficient and Discrete Ill-Posed Problems. SIAM, Philadelphia (1998)
11. Hansen, P.C., Nagy, J.G., O'Leary, D.P.: Deblurring Images: Matrices, Spectra and Filtering. SIAM, Philadelphia (2006)
12. Hochstenbach, M.E., McNinch, N., Reichel, L.: Discrete ill-posed least-squares problems with a solution norm constraint. Linear Algebra Appl. **436**, 3801–3818 (2012)
13. Hochstenbach, M.E., Reichel, L.: An iterative method for Tikhonov regularization with a general linear regularization operator. J. Integral Equations Appl. **22**, 463–480 (2010)
14. Hochstenbach, M.E., Reichel, L.: Combining approximate solutions for linear discrete ill-posed problems. J. Comput. Appl. Math. **236**, 2179–2185 (2012)
15. Ito, K., Jin, B., Takeuchi, T.: Multi-parameter Tikhonov regularization. Meth. Appl. Anal. **18**, 31–46 (2011)
16. Kindermann, S.: Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. Electron. Trans. Numer. Anal. **38**, 233–257 (2011)

17. Kunisch, K., Pock, T.: A bilevel optimization approach for parameter learning in variational methods. SIAM J. Imaging Sci. **6**, 938–983 (2013)
18. Lewis, B., Reichel, L.: Arnoldi-Tikhonov regularization methods. J. Comput. Appl. Math. **226**, 92–102 (2009)
19. Lu, S., Pereverzev, S.V.: Multi-parameter regularization and its numerical realization. Numer. Math. **118**, 1–31 (2011)
20. Lu, S., Pereverzev, S.V., Shao, Y., Tautenhahn, U.: Discrepancy curves for multi-parameter regularization. J. Inv. Ill-Posed Problems **18**, 655–676 (2010)
21. Mead, J.L.: Discontinuous parameter estimates with least squares estimators. Appl. Math. Comput. **219**, 5210–5223 (2013)
22. Morigi, S., Reichel, L., Sgallari, F.: Orthogonal projection regularization operators. Numer. Algorithms **44**, 99–114 (2007)
23. Morozov, V.A.: On the solution of functional equations by the method of regularization. Soviet Math. Dokl. **7**, 414–417 (1966)
24. Nagy, J.G., Palmer, K.M., Perrone, L.: Iterative methods for image deblurring: a Matlab object oriented approach. Numer. Algorithms **36**, 73–93 (2004).
See also: http://www.mathcs.emory.edu/~nagy/RestoreTools
25. Regińska, T.: Two-parameter discrepancy principle for combined projection and Tikhonov regularization of ill-posed problems. J. Inverse Ill-Posed Probl. **21**, 561–577 (2013)
26. Reichel, L., Sgallari, F., Ye, Q.: Tikhonov regularization based on generalized Krylov subspace methods. Appl. Numer. Math. **62**, 1215–1228 (2012)
27. Reichel, L., Ye, Q.: Simple square smoothing regularization operators. Electron. Trans. Numer. Anal. **33**, 63–83 (2009)
28. Reichel, L., Yu, X.: Matrix decompositions for Tikhonov regularization. Electron. Trans. Numer. Anal. **43**, 223–243 (2015)
29. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd Ed. SIAM, Philadelphia (2003)