A new framework for multi-parameter regularization

Silvia Gazzola, Lothar Reichel

January 10, 2015

Abstract

This paper proposes some new general strategies for the analysis and implementation of multi-parameter regularization methods. 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 dimensions. The latter methods regularize by choosing suitable 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.

1 Introduction

Consider linear least-squares problems of the form

$$\min_{x\in\mathbb{R}^N} \|Ax-b\|, \qquad A\in\mathbb{R}^{M\times N}, \quad b\in\mathbb{R}^M, \tag{1}$$

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 $e\in\mathbb{R}^M$, 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 [9]. Because of the ill-conditioning of A and the perturbation e in b, one has to employ some kind of regularization of (1) in order to be able to compute a meaningful approximation of the 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 positive 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 the properties of A, one usually chooses between direct or iterative regularization methods. The former methods are based on available factorizations of A and of the regularization 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 size 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

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

where $\lambda > 0$ is the regularization parameter and $L \in \mathbb{R}^{P\times N}$ the regularization matrix. Here and in the following $\|\cdot\|$ denotes the vector 2-norm and $(\cdot, \cdot)$ the standard inner product. When $L = I_N$ (i.e., the identity matrix of order N), the problem (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

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}. \tag{3}$$

Then the minimizer x_λ of (2) is unique.

It is clear from (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., [17, 22] for discussions.

Krylov subspace methods and generalizations thereof are popular iterative methods for the approximate solution of large-scale Tikhonov regularization problems of the form (2); see, e.g., [6, 11, 17, 20]. Regularization is achieved both by using the penalty term and by terminating the iterations early.

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) by a penalized minimization problem of the form

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \sum_{i=1}^{m} \lambda_i \|L_i x\|^2 \right\} ; \tag{4}$$

see, e.g., [1, 2, 4, 5, 15, 16] and references therein. Here the scalars $\lambda_i > 0$ 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 (4) has a unique solution, which we denote by x_Λ, where $\Lambda = (\lambda_1, \ldots, \lambda_m)$ is referred to as the regularization vector. When $m = 1$ the problem (4) simplifies to (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 the regularization vector. To the best of our knowledge, the first attempt to derive a systematic parameter choice strategy was proposed by Belge et al. [1], who introduce a generalization of the L-curve and describe 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] propose and analyze an approach based on the GCV method. Specifically, they solve m different one-parameter problems (one for each regularization term appearing in (4)) and then combine the m approximations of x^{ex} so obtained. More recently, Lu and Pereverzyev [15] and Lu et al. [16] proposed to employ the discrepancy principle to select a regularization vector for (4). Two approaches are investigated. Their first approach is based on approximating the Tikhonov functional and the discrepancy function by means of suitable model functions, and requiring the computed solution to satisfy the discrepancy principle; this method is further analyzed by Fornasier et al. [4]. The second approach [16] considers particular one-parameter problems, and applies a Newton zero-finder to the discrepancy function.

It is the purpose of the present paper to investigate a new approach to determine the regularization vector Λ of multi-parameter regularization methods (4). Similarly as Lu et al. [15, 16], we first determine a set of regularization vectors $\{\Lambda_j\}_{j=1,\ldots,h}$ that satisfy the discrepancy principle and then choose a vector from this set that solves the problem

$$\max_{j=1,\ldots,h} \Psi\left(x_{\Lambda_j}\right). \tag{5}$$

Common choices of the functional $\Psi(x_{\Lambda_j})$ are $\|x_{\Lambda_j}\|$, $\|Lx_{\Lambda_j}\|$, or $\|x_{\Lambda_j}\|^2 + \|Lx_{\Lambda_j}\|^2$, 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 [16].

In addition to multi-parameter Tikhonov regularization of the form (4), we also consider Krylov-Tikhonov regularization, which is intrinsically a multi-parameter method. Here one of the regularization parameters is the dimension of the solution subspace used. We propose to use the discrepancy principle to choose the Tikhonov regularization parameter at each iteration [11, 14]; the number of iterations, which equals the dimension of the solution subspace, is determined by solving an optimization problem similar to (5). We will illustrate this approach for a method that projects the regularization matrix in (2) into the Krylov subspace determined by the Arnoldi algorithm. However, the approach can be applied to other iterative Tikhonov regularization methods as well, for instance to the methods discussed in [3, 11, 20]. Recently, Gazzola and Novati [5] proposed an approach to determine Λ as well as the number of iterations with an Arnoldi-Tikhonov method for (4), which relies on sequentially updating an approximation of the discrepancy function. We will review this approach and discuss how the criterion (5) can be incorporated.

This paper is organized as follows: in Section 2 we discuss direct multi-parameter Tikhonov regularization (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 (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 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 (4) that satisfies the discrepancy principle, i.e., is such that

$$\|b - Ax_\Lambda\|^2 = \eta^2 \varepsilon^2 \,, \tag{6}$$

where $\varepsilon \geq \|e\|$ 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 $\|e\|$. Introduce the discrepancy function

$$\Phi(\Lambda) = \|b - Ax_\Lambda\|^2. \tag{7}$$

In the one-parameter case, i.e., when $\Lambda = (\lambda)$, one usually applies the discrepancy principle by solving the nonlinear equation (6) with respect to λ [18]. However, when $\Lambda \in \mathbb{R}^m$, $m \geq 2$, the problem (6) is underdetermined. Following [16], we define the discrepancy hypersurface

$$\mathcal{D} = \{\, \Lambda \in \mathbb{R}^m \,:\, \Lambda > 0,\ \|b - Ax_\Lambda\| = \eta\varepsilon \,\} \,,$$

where the condition $\Lambda > 0$ is to be interpreted component-wise. In this section we describe a strategy to impose additional constraints on the solution x_Λ of (4) in order to reduce the number of degrees of freedom in the choice of Λ.

A natural constraint is

$$\max_{\Lambda \in \mathcal{D}} \|x_\Lambda\| \,. \tag{8}$$

It can be justified in the following way. The discrepancy principle generally determines over-smoothed approximations of x^{ex}, cf. [9, §7.2] and references therein. By imposing both (6) and (8), we seek to determine the least over-smoothed approximation x_Λ of x^{ex} with $\Lambda \in \mathcal{D}$. Of course, generally the constraint (8) 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 [12] 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 result provides a sufficient condition for a vector x_Λ that satisfies both (6) and (8) 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 $\mathcal{I}$ denote the set indexing the regularization vectors belonging to $\mathcal{D}$.

Theorem 1. *Let $\{x_{\Lambda_i}\}_{i\in\mathcal{I}}$ be a set of solutions of (4) that satisfy the discrepancy principle (6) and define $\ell = \arg\max_{i\in\mathcal{I}} \|x_{\Lambda_i}\|$. Let $x_{\Lambda_i} = x_{\Lambda_\ell} + \delta x_i$. If*

$$(x^{ex}, \delta x_i) \leq 0 \tag{9}$$

and

$$(x_{\Lambda_\ell}, \delta x_i) \geq -\frac{\|\delta x_i\|^2}{2} + (x^{ex}, \delta x_i) \,, \tag{10}$$

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

$$\|x^{ex} - x_{\Lambda_\ell}\| \leq \|x^{ex} - x_{\Lambda_i}\| \quad \forall i \in \mathcal{I}. \tag{11}$$

Proof. The result follows directly by writing

$$\|x^{ex} - x_{\Lambda_i}\|^2 - \|x^{ex} - x_{\Lambda_\ell}\|^2 = 2\,(x_{\Lambda_\ell}, \delta x_i) + \|\delta x_i\|^2 - 2\,(x^{ex}, \delta x_i) \tag{12}$$

and by considering that, thanks to (10), the right-hand side of the above equality is nonnegative. Since the assumption $\|x_{\Lambda_\ell}\| \geq \|x_{\Lambda_i}\|$ implies

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

one can easily see that (11) holds under both the conditions (9) and (10). □

We may replace (8) by a constraint of the form

$$\max_{\Lambda \in \mathcal{D}} \sum_{i=1}^{m} \|L_i x_\Lambda\|^2\,, \tag{13}$$

where L_i, $i = 1, \ldots, m$, are the regularization matrices appearing in (4). Optimality properties analogous to Theorem 1 can be established also for the constraint (13). For instance, one can show the following result.

Theorem 2. *Let* $\{x_{\Lambda_i}\}_{i \in \mathcal{I}}$ *be a set of solutions of (4) that satisfy the discrepancy principle (6). Define* $\ell = \arg\max_{i \in \mathcal{I}} \left(\|x_{\Lambda_i}\|^2 + \|L x_{\Lambda_i}\|^2 \right)$ *and let* $x_{\Lambda_i} = x_{\Lambda_\ell} + \delta x_i$. *If*

$$(x^{ex}, \delta x_i) \leq -\frac{\|L\delta x_i\|^2}{2} - (L x_{\Lambda_\ell}, L\delta x_i) \tag{14}$$

and

$$\left(x_{\Lambda_\ell}, (I + L^T L)\delta x_i\right) \geq -\frac{\|\delta x_i\|^2}{2} + \left(x^{ex} + L^T L x_{\Lambda_\ell}, \delta x_i\right)\,, \tag{15}$$

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

$$\|x^{ex} - x_{\Lambda_\ell}\| \leq \|x^{ex} - x_{\Lambda_i}\| \quad \forall i \in \mathcal{I}. \tag{16}$$

Proof. By writing the errors as in (12), we can conclude that (16) holds, since (15) implies (10). Since the assumption $\|x_{\Lambda_\ell}\| + \|L x_{\Lambda_\ell}\| \geq \|x_{\Lambda_i}\| + \|L x_{\Lambda_i}\|$ 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 (16) holds under both the conditions (14) and (15). □

Remark 3. 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 is equivalent to Theorem 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, (14) and (15) are equivalent to (9) and (10), respectively. Moreover, when the x_{Λ_i}'s are accurate approximations of x^{ex} and are close to $\mathcal{N}(L)$, then the constraint (13) with $L_i = L$ approximately reduces to constraint (8). 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

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda_1 \|L_1 x\|^2 + \lambda_2 \|L_2 x\|^2 \right\} . \tag{17}$$

The use of more than two regularization terms in (4) can be treated analogously. In the two-parameter case, $\mathcal{D}$ is a differentiable curve in $\mathbb{R}^2$; see [15]. A simple technique to impose the constraints (8) or (13) 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 determining λ_2 so that the discrepancy principle is satisfied (if possible, cf. Proposition 5). Thus, let λ_2 be the zero of the one-variable function

$$\lambda_2 \to \Phi(h(\hat{\lambda}_1), h(\lambda_2))^2 - \eta^2 \varepsilon^2 \,, \tag{18}$$

where Φ is given by (7) 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 (18). This change of variable for the function (18) is also considered by Lu et al. [16]. It is commonly used for one-parameter Tikhonov regularization problems (2). Lu et al. [16] consider 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 4. *If*

$$(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)}) < 0 \tag{19}$$

and

$$(L_1 \partial^2_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)}) = 2(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}) \,, \tag{20}$$

then $\Phi(h(\hat{\lambda}_1), h(\lambda_2))$ is a decreasing and convex function of (λ_1, 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

$$\Omega(\lambda_1, \lambda_2) = \|b - Ax_\Lambda\|^2 + \lambda_1 \|L_1 x_\Lambda\|^2 + \lambda_2 \|L_2 x_\Lambda\|^2, \qquad \Lambda = (\lambda_1, \lambda_2).$$

Following Lu and Pereverzyev [15], we consider the following relations

$$\Phi(\hat{\lambda}_1, \lambda_2) \;\; = \;\; \Omega(\hat{\lambda}_1, \lambda_2) - \hat{\lambda}_1 \partial_{\lambda_1} \Omega(\lambda_1, \lambda_2)|_{\lambda_1 = \hat{\lambda}_1} - \lambda_2 \partial_{\lambda_2} \Omega(\hat{\lambda}_1, \lambda_2) \,,$$

$$
\begin{aligned}
\partial_{\lambda_2}\Phi(\hat\lambda_1,\lambda_2) &= -\hat\lambda_1\partial_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2)|_{\lambda_1=\hat\lambda_1} - \lambda_2\partial^2_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2)\,,\\
\partial^2_{\lambda_2}\Phi(\hat\lambda_1,\lambda_2) &= -\hat\lambda_1\partial^2_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2)|_{\lambda_1=\hat\lambda_1} - \partial^2_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2) - \lambda_2\partial^3_{\lambda_2}\Omega(\hat\lambda_1,\lambda_2)\,. \quad (21)
\end{aligned}
$$

In the above expressions,

$$
\begin{aligned}
\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= \|L_1x_\Lambda\|^2\,,\\
\partial_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= \|L_2x_\Lambda\|^2\,,\\
\partial_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= 2\left(L_1\partial_{\lambda_2}x_\Lambda, L_1x_\Lambda\right)\,,\\
\partial^2_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= 2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2x_\Lambda\right)\,,\\
\partial^2_{\lambda_2}\partial_{\lambda_1}\Omega(\lambda_1,\lambda_2) &= 2\left(L_1\partial^2_{\lambda_2}x_\Lambda, L_1x_\Lambda\right) + 2\left(L_1\partial_{\lambda_2}x_\Lambda, L_1\partial_{\lambda_2}x_\Lambda\right)\,,\\
\partial^3_{\lambda_2}\Omega(\lambda_1,\lambda_2) &= 2\left(L_2\partial^2_{\lambda_2}x_\Lambda, L_2x_\Lambda\right) + 2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2\partial_{\lambda_2}x_\Lambda\right)\,.
\end{aligned} \quad (22)
$$

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

$$
\begin{aligned}
(A\partial_{\lambda_2}x_\Lambda, Az) + \lambda_1\left(L_1\partial_{\lambda_2}x_\Lambda, L_1z\right) + \lambda_2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2z\right) &= -\left(L_2x_\Lambda, L_2z\right)\,, \quad (23)\\
\left(A\partial^2_{\lambda_2}x_\Lambda, Az\right) + \lambda_1\left(L_1\partial^2_{\lambda_2}x_\Lambda, L_1z\right) + \lambda_2\left(L_2\partial^2_{\lambda_2}x_\Lambda, L_2z\right) &= -2\left(L_2\partial_{\lambda_2}x_\Lambda, L_2z\right) \quad (24)
\end{aligned}
$$

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

$$
\begin{aligned}
\partial_{\lambda_2}\Phi(h(\lambda_1),h(\lambda_2)) &= \frac{2}{\lambda_1\lambda_2^2}\left(L_1\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_1x_{h(\Lambda)}\right) \quad (25)\\
&\quad + \frac{2}{\lambda_2^3}\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_2x_{h(\Lambda)}\right)\,,
\end{aligned}
$$

for $\lambda_1=\hat\lambda_1$. By taking $z=\partial_{h(\lambda_2)}x_{h(\Lambda)}$ in (23), evaluated at $\left(h(\hat\lambda_1),h(\lambda_2)\right)$, we get

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

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 can conclude that

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

To prove the convexity of $\Phi(\hat\lambda_1,\lambda_2)$, the derivations are more cumbersome. By applying the chain rule to (21) and by exploiting the equalities (22), we get

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

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

for $\lambda_1 = \hat{\lambda}_1$. By evaluating (23) for $z = \partial^2_{h(\lambda_2)}x_{h(\Lambda)}$ and (24) for $z = \partial_{h(\lambda_2)}x_{h(\Lambda)}$, and by considering their difference, we get

$$
\left(L_2x_{h(\Lambda)}, L_2\partial^2_{h(\lambda_2)}x_{h(\Lambda)}\right) = 2\left(L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}, L_2\partial_{h(\lambda_2)}x_{h(\Lambda)}\right)\,. \tag{26}
$$

Substituting (20) and (26) 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 (19) and performing some algebraic manipulations on (23) 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 5. *Under the assumptions (19), (20), and*

$$
\left\|\left(I_N - A\left(A^TA + \frac{1}{\hat{\lambda}_1}L_1^TL_1\right)^{-1}A^T\right)b\right\|^2 < \eta^2\varepsilon^2 < \|b\|^2\,, \tag{27}
$$

equation (18) has a unique solution $\lambda_2 > 0$.

Proof. Immediate thanks to Lemma 4. □

Remark 6. The assumptions of Lemma 4 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, (20) can be regarded as a condition on the higher order derivatives of x_Λ. Also the inequalities (27) of Proposition 5 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 (2) with $\lambda = 1/\hat{\lambda}_1$ and $L = L_1$.

It follows from Lemma 4 that Newton's method applied to determining the smallest zero of (18) is guaranteed to give quadratic and monotonic convergence if the initial approximation of the zero is smaller than the desired zero; see also [21]. To implement Newton's method, one needs to compute $\partial_{h(\lambda_2)}\Phi(\hat{\lambda}_1, \lambda_2)$. An expression for $\partial_{h(\lambda_2)}\Phi(\hat{\lambda}_1, \lambda_2)$ is given in (25) and, according to (23), one can compute $\partial_{h(\lambda_2)} x_{h(\Lambda)}$ in the following way

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

or, equivalently, by taking

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

We remark that the matrix in the above formulation is the same as the one associated with the problem (17).

The following result sheds some light on the maximization problem (8).

Proposition 7. *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 (23) that

$$\partial_{\lambda_2} \|x_\Lambda\|^2 = 2\left(\partial_{\lambda_2} x_\Lambda, x_\Lambda\right) = -2\left(\left(A^T A + \hat{\lambda}_1 L_1^T L_1 + \lambda_2 L_2^T L_2\right)^{-1} (L_2^T L_2) x_\Lambda, x_\Lambda\right).$$

Since $L_2^T L_2$ 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 (2) with $L = L_1$, and $\lambda = \hat{\lambda}_1$ chosen so that the discrepancy principle is satisfied.

If we assume that the discrepancy curve can be explicitly expressed as

$$\lambda_1 = g(\lambda_2) , \tag{28}$$

where λ_2, $g(\lambda_2) > 0$, and g is differentiable, then we can extend the derivations in [15] to obtain (23). More precisely, one can see that $\partial_{\lambda_2} x_\Lambda$ solves the following problem

$$\begin{aligned} \left(A^T A \partial_{\lambda_2} x_\Lambda, z\right) &+ g(\lambda_2)\left(L_1^T L_1 \partial_{\lambda_2} x_\Lambda, z\right) + \lambda_2\left(L_2^T L_2 \partial_{\lambda_2} x_\Lambda, z\right) \\ &= -\left(\left(L_2^T L_2 + \partial_{\lambda_2} g(\lambda_2) L_1^T L_1\right) x_\Lambda, z\right) \end{aligned} \tag{29}$$

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 gets

$$\partial_{\lambda_2} \|x_\Lambda\|^2 = -2 \left(\left(A^\sharp \right)^{-1} (L_2^T L_2 + \partial_{\lambda_2} g(\lambda_2) L_1^T L_1) x_\Lambda, x_\Lambda \right) . \tag{30}$$

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 7 hold when replacing the maximization (8) by (13). Of course, λ_1 and λ_2 can be interchanged in the above discussion.

3 Krylov-Tikhonov multi-parameter regularization

Krylov-Tikhonov methods are obtained by projecting a Tikhonov-regularized problem (2) or (4) onto Krylov subspaces of increasing dimensions; different Krylov-Tikhonov methods are obtained by varying the original problem (2) or (4), or the Krylov subspaces. For instance, the authors of [11] project the general form Tikhonov problem ($L \neq I_N$) onto the subspaces $\mathcal{K}_k(A^T A, A^T b)$, $k \geq 1$, generated by the Golub-Kahan bidiagonalization algorithm; in [3] the problem (2) with $L = I_N$ is projected onto subspaces $\mathcal{K}_k(A, b)$, $k \geq 1$, generated by the Arnoldi algorithm. Range restricted Arnoldi methods that generate solution subspaces $\mathcal{K}_k(A, Ab)$, $k \geq 1$, are discussed in [14] and [13]; the latter reference discusses a Petrov-Galerkin implementation. Krylov-Tikhonov methods are intrinsically multi-parameter methods, since (at least) one regularization parameter λ and the dimension k of the Krylov subspace has to be chosen. The dimension k may also be considered a regularization parameter. Generally, λ depends on the dimension.

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 [23, §6.3]. Application of k steps of this algorithm to A with initial vector $b/\|b\|$ yields the decomposition

$$AW_k = W_{k+1} \bar{H}_k \,, \quad \text{where} \quad W_{k+1} \in \mathbb{R}^{N \times (k+1)}, \; W_k = W_{k+1} \begin{bmatrix} I_k \\ 0 \end{bmatrix} \in \mathbb{R}^{N \times k}, \tag{31}$$

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. Modifications of the decomposition (31) that can be applied in the range restricted Arnoldi method case will be commented on below. We describe a method that is analogous to the scheme in [11], but with $\mathcal{K}_k(A, b)$ as solution subspaces. We refer to this method as the Arnoldi-Tikhonov method. The special case when $L = I_N$ is described in [3]. Substituting the Arnoldi decomposition (31) into (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

$$y_{k,\lambda} = \arg \min_{y \in \mathbb{R}^k} \left\| \begin{bmatrix} \bar{H}_k \\ \lambda^{1/2} R_k \end{bmatrix} - \begin{bmatrix} e_1 \|b\| \\ 0 \end{bmatrix} \right\|^2 . \tag{32}$$

We note that due to (3), the coefficient matrix in (32) has full column rank. An approximate solution of (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 $\|Ax_{k,\lambda} - b\| = \eta^2\varepsilon^2$. In order to secure that Newton's method converges monotonically, one carries out the change of variable $\mu = \lambda^{-1}$, see [11, 14], and introduces the function

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

The solution μ_k of $\Phi^{(k)}(\mu)$ yields the desired value $\lambda_k = 1/\mu_k$ of the regularization parameter. It can be shown that $\Phi^{(k)}(0)$ is a decreasing function of k, and the existence of the solution μ_k requires k to be large enough. Computed examples in [14] illustrate that carrying out more than the minimal number of iterations k may result in improved approximations of x^{ex}. We propose that the number of additional iterations be determined by maximizing (8) or (13). Computed examples of Section 4 illustrate this kind of stopping criterion.

An alternative stopping rule is to terminate the iterations when the regularization parameter λ_k does not change much with k, i.e., when

$$\frac{|\lambda_{k+1} - \lambda_k|}{\lambda_k} < \tau \tag{33}$$

for some user-chosen tolerance τ. This stopping criterion is meaningful because we would like x_{k,λ_k} to be an accurate approximation of the solution of (2). It is therefore natural to monitor the stabilization of λ_k.

In [5], the authors derive a multi-parameter Arnoldi-Tikhonov (mP-AT) method by projecting problem (4) onto the Krylov subspaces $\mathcal{K}_k(A, b)$, $k \geq 1$. We briefly describe this method and focus on the case with two regularization matrices ($m = 2$). Analogously to (32), we determine an approximate solution of (4) of the form $x_{k,\Lambda} = W_k y_{k,\Lambda}$ and obtain the minimization problem

$$y_{k,\Lambda} = \arg\min_{y\in\mathbb{R}^k} \left\| \begin{bmatrix} \bar{H}_k \\ (\lambda_1^{(k)})^{1/2} R_k^{(1)} \\ (\lambda_2^{(k)})^{1/2} R_k^{(2)} \end{bmatrix} y - \begin{bmatrix} e_1\|b\| \\ 0 \\ 0 \end{bmatrix} \right\|^2, \tag{34}$$

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

$$\bar{\Phi}^{(k)}(\Lambda) = \left\| \|b\|e_1 - \bar{H}_k y_{k,\Lambda} \right\|, \tag{35}$$

where $\Lambda = (\lambda_1, \lambda_2)$. Let k^* be the smallest integer such that

$$\bar{\Phi}^{(k^*)}(0, 0) < \eta\varepsilon\,. \tag{36}$$

Since $\bar{\Phi}^{(k)}(0, 0)$ is the residual norm of the kth iterate determined by GMRES applied to (1) with initial iterate $x_0 = 0$ (and $M = N$), the integer k^* typically exists and is quite small; see [7]. Following [5], we approximate for $k > k^*$ the function (35) by the linear function

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

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

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

The parameters $(\lambda_1^{(k)}, \lambda_2^{(k)})$ in the above definitions are the ones employed in (34) 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"

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

This yields

$$\lambda_1 = \frac{\eta\varepsilon - \alpha_0^{(k)}}{\alpha_1^{(k)}} - \frac{\alpha_2^{(k)}}{\alpha_1^{(k)}}\lambda_2 =: \gamma^{(k)} - \delta^{(k)}\lambda_2. \tag{37}$$

It follows from (36) 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

$$0 \le \lambda_2 \le \frac{\gamma^{(k)}}{\delta^{(k)}}\,. \tag{38}$$

In the terminology of Lu and Pereverzyev [15], $\bar{\Phi}_{\rm lin}^{(k)}(\Lambda)$ 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 just imposing three interpolation conditions

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

The discrepancy curve associated with $\bar{\Phi}_{\rm lin}^{(k)}(\Lambda)$ 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 (28) is determined by (37) for each $k > k^*$.

Differently from the approach adopted in [5], we choose the pair $(\lambda_1^{(k+1)}, \lambda_2^{(k+1)})$ so that (37) 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

$$\bar{H}_k^\sharp = \bar{H}_k^T\bar{H}_k + \lambda_1 (R_k^{(1)})^T R_k^{(1)} + \lambda_2 (R_k^{(2)})^T R_k^{(2)}\,, \tag{39}$$

and performing some standard manipulations, we get

$$\partial_{\lambda_2}\|y_{k,\Lambda}\|^2 = 2\left(\partial_{\lambda_2} y_{k,\Lambda}, y_{k,\Lambda}\right)\,, \tag{40}$$

where

$$\partial_{\lambda_2} y_{k,\Lambda} = -\left(\bar{H}_k^\sharp\right)^{-1}\left((R_k^{(2)})^T R_k^{(2)} - \delta^{(k)}(R_k^{(1)})^T R_k^{(1)}\right) y_{k,\Lambda}\,. \tag{41}$$

Note that (40) and (41) are particular cases of (30) and (29), respectively, obtained by taking $g(\lambda_2) = \gamma^{(k)} - \delta^{(k)}\lambda_2$.

The behavior of (40) is quite easy to analyze. Let $\zeta_j^{(k)}$, $j = 1, \ldots, k$, denote the generalized singular values of the matrix pair $(R_k^{(1)}, R_k^{(2)})$. Then 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 (38), $\lambda_2 = 0$;
- 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 (38), $\lambda_2 = \gamma^{(k)}\left(\delta^{(k)}\right)^{-1}$;
- if none of the previous conditions is 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 (38).

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

$$\bar{\Phi}^{(k)}(\lambda_1^{(k)}, \lambda_2^{(k)}) < \eta\varepsilon. \tag{42}$$

This approach overcomes the main shortcoming of the strategy developed in [5], namely the dependence of the computed approximate solution on the order of the regularization matrices in (34). For instance, by exchanging the order of the regularization matrices in (34), i.e., by expressing λ_2 as a function of λ_1, one gets

$$\lambda_2 = \frac{\gamma^{(k)}}{\delta^{(k)}} - \frac{1}{\delta^{(k)}}\lambda_1 \tag{43}$$

with the bounds $0 \leq \lambda_1 \leq \gamma^{(k)}$, instead of (37) and (38), respectively. Computing $\partial_{\lambda_1}\|y_{k,\Lambda}\|^2$ as done in (40), one obtains

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

where $\bar{H}_k^{\sharp}$ is defined as in (39). 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 (43), $\lambda_2 = 0$. 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.

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 [8], 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 (17). The second set of experiments illustrates the performance of the multi-parameter Tikhonov, AT, RRAT, and mP-AT methods when applied to many popular problems still taken from [8]. The third set of experiments considers image deblurring and denoising problems, and we apply the direct multi-parameter Tikhonov regularization and the mP-AT method. Depending on the problem, we take as regularization matrices I_N or the following

$$D_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(N-1)\times N}, \tag{44}$$

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2)\times N}, \tag{45}$$

which represent scaled finite difference approximations of the first and the second derivative operators, respectively. All the computations have been executed using Matlab 8.1 with 16 significant digits.

Example 1. The coefficient matrices considered for the direct regularization method (17) 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 white noise is added in such a way that the noise level $\widetilde{\varepsilon} = {\|e\|}/{\|b\|}$ is 10^{-2}. We use the safety factor $\eta = 1.01$ in the discrepancy principle. As regularization matrices, we take $L_i = I_N$, D_1, $i = 1, 2$, so that $x^{ex} \in \mathcal{N}(D_1)$. Table 1 reports the averages of the results obtained running 50 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 first regularization parameter λ_1 is varied: the values of λ_1 are logarithmically equispaced between 10^{-8} and 10^2. For each λ_1, a value of λ_2 such that (6) is satisfied is determined by applying the Newton zero finder described in Section 2. For each test problem we display the relative error obtained with the one-parameter Tikhonov method (2) with $L = D_1$ (third column), the minimum relative error obtained considering different combinations of the vectors (λ_1, λ_2) (fourth column), the relative error obtained when selecting the regularization parameters according to (8) (fifth column), together with the corresponding (λ_1, λ_2) (sixth and seventh columns), and the relative error obtained when selecting the regularization parameters according to (13) (eighth column).

Figure 1 displays the behavior of various quantities associated to the problem `phillips` with $x^{ex} = [1, 1, \ldots, 1]^T$ and $\widetilde{\varepsilon} = 10^{-2}$. In particular, we plot the relative errors for all the admissible pairs (λ_1, λ_2), sequentially varying both λ_1 or λ_2 and determining the corresponding parameter λ_2 or λ_1 such that (6) 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

	Reg.M.	**error (1P)**	**error (opt)**	**error (8)**	λ_1	λ_2	**error (13)**
`baart`	(I_N, D_1)	6.7210e-04	3.4362e-04	6.7210e-04	1.0000e-08	1.0000e+08	6.7210e-04
`baart`	(D_1, I_N)	6.7210e-04	8.8168e-03	8.8168e-03	9.8815e+01	6.5757e-03	8.8168e-03
`deriv2`	(I_N, D_1)	7.5030e-04	3.4723e-04	7.5011e-04	1.0000e-08	1.0000e+08	7.5011e-04
`deriv2`	(D_1, I_N)	7.5030e-04	1.4748e-03	1.4748e-03	1.0000e+02	1.3657e-05	1.4748e-03
`i_laplace`	(I_N, D_1)	9.6223e-04	6.5678e-04	9.6223e-04	1.0000e-08	1.0000e+08	9.6223e-04
`i_laplace`	(D_1, I_N)	9.6223e-04	3.7590e-02	3.7590e-02	9.7657e+01	1.8152e-03	3.7590e-02
`phillips`	(I_N, D_1)	6.7026e-04	3.7505e-04	6.7026e-04	1.0000e-08	1.0000e+08	6.7026e-04
`phillips`	(D_1, I_N)	6.7026e-04	5.3248e-03	5.3248e-03	9.7657e+01	7.8922e-02	5.3248e-03
`shaw`	(I_N, D_1)	9.1441e-04	4.7361e-04	9.1441e-04	1.0000e-08	1.0000e+08	9.1441e-04
`shaw`	(D_1, I_N)	9.1441e-04	9.5181e-03	9.5181e-03	9.6526e+01	8.6195e-03	9.5181e-03

Table 1: Results obtained considering problems (2) and (17) with $x^{ex} = [1, 1, \ldots, 1]^T$ and $\widetilde{\varepsilon} = 10^{-2}$. The regularization matrix pairs are reported in the second column.

error (black circle), and the quantities satisfying the criteria (8) (black square) and (13) (black hexagram).

When applying the 2P-AT method (34), 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 (37) and (8). 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 2 reports the averages of the minimum relative errors obtained running 50 tests (third column): more precisely, at each iteration of the Arnoldi algorithm, criteria (37) and (8) 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 (42) is satisfied (sixth column), as well as the components of the vector (λ_1, λ_2) obtained by (37), (8) and (42) (seventh and eighth columns), are also reported. The average number of iterations is displayed in parentheses.

	Reg.M.	**error (8,37,opt)**	λ_1	λ_2	**error (8,37,42)**	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
`i_laplace`	(D_2, I_N)	2.9994e-02 (13.9)	3.36e+07	2.70e-06	5.5257e-01 (5.1)	2.99e+00	1.00e-10
`i_laplace`	(I_N, D_2)	1.8750e-01 (9.3)	3.82e-10	6.47e+05	4.8774e-01 (5.4)	1.68e-06	1.76e+00
`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

Table 2: Results obtained considering problem (34) 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.

Figure 2 considers one single test performed with the coefficient matrix `baart`,

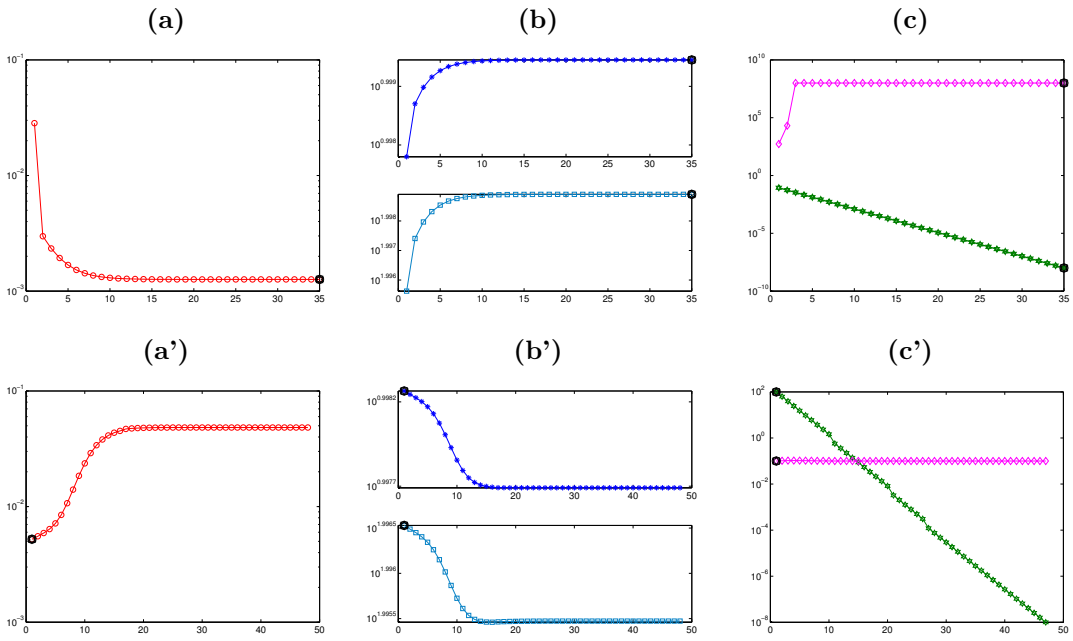

Figure 1: Test problem `phillips` with $x^{ex} = [1, 1, \ldots, 1]^T$. The first and second rows display quantities related to problem (17) with (I_N, D_1) and (D_1, I_N), respectively. **(a)**, **(a')**: relative errors for the combinations of λ_i's displayed in the third frames; **(b)**, **(b')**: 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).

$x^{ex} = [1, 2, \ldots, N]^T$, and $\widetilde{\varepsilon} = 10^{-1}$. We plot the relative errors, $(\lambda_1^{(k)}, \lambda_2^{(k)})$, and the discrepancy function versus the number of Arnoldi steps k. The iteration delivering the best relative error is highlighted by a black circle, while the iteration satisfying the stopping criterion is highlighted by a black square.

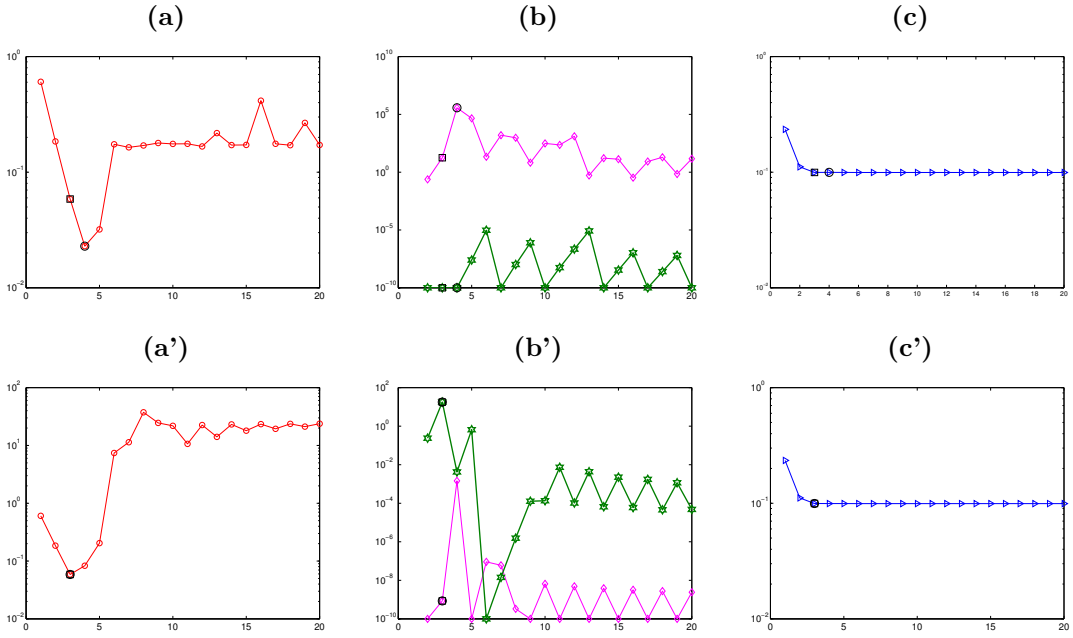

Figure 2: Test problem `baart` with $x^{ex} = [1, 2, \ldots, N]^T$. The first and second rows refer to the regularization matrix pairs (I_N, D_2) and (D_2, I_N), respectively. **(a)**, **(a')**: relative errors versus number of steps; **(b)**, **(b')**: pairs of values λ_1 (hexagram) and λ_2 (diamond) 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 L_i such that $x^{ex} \in \mathcal{N}(L_i)$), and to weigh it with a large regularization parameter. Moreover, the results are quite invariant with respect to the order of the regularization matrices. Two-parameter Tikhonov regularization gives computed solutions of higher quality than one-parameter Tikhonov regularization for some problems. For many problems the 2P-AT method determines approximate solutions whose quality is comparable to the optimal attainable one.

Example 2. Some popular test problems with the solutions given in [8] (or slightly modified versions of them) are used. Namely, two versions of the test problem `deriv2` are used. 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 $f(t) = \exp(t)$. The test problem `i_laplace` is modified: a discretization of the function $f(t) = 1 + \exp(-t/2)$ is taken as the exact solution. A reason for this modification is that the criterion (8) fails when the exact solution is sparse, i.e., with many components close to zero (such as the discretized version of

$\exp(-t/2)$); indeed, by maximizing the norm of a sparse solution, one would recover a quite corrupted and noisy approximation of it, even if the discrepancy principle holds.

To test the two-parameter Tikhonov method (17), coefficient matrices of size 100×100 and $\widetilde{\varepsilon} = 10^{-2}$ are considered. As regularization matrices, the pairs (D_1, D_2) and (D_2, D_1) are employed. In Figure 3, the assumptions of Lemma 4 are checked for the test problem `phillips`. More precisely, λ_1 and λ_2 are varied (on the horizontal and vertical axis, respectively), and for each pair (λ_1, λ_2) the value

$$\left| \left(L_1 \partial^2_{h(\lambda_2)} x_{h(\Lambda)}, L_1 x_{h(\Lambda)} \right) - 2 \left(L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)}, L_1 \partial_{h(\lambda_2)} x_{h(\Lambda)} \right) \right| \tag{46}$$

is displayed. 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 3. Similarly to Example 1, Table 3 displays

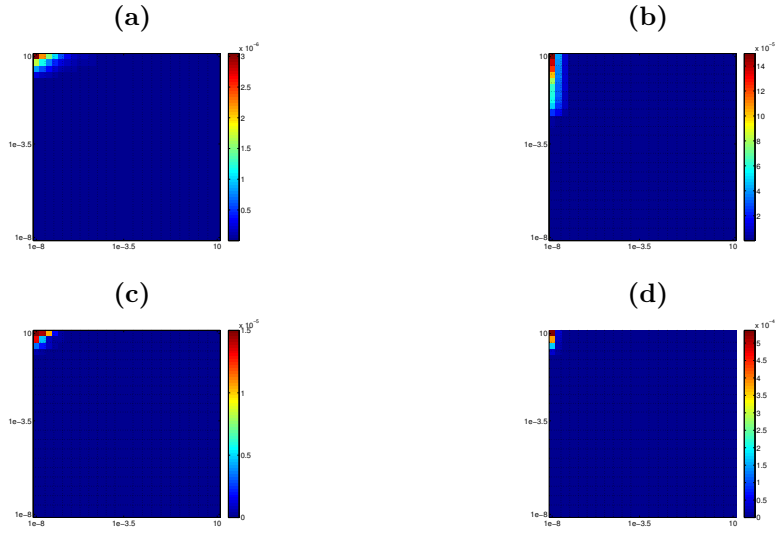

Figure 3: Values of the differences (46), test problem `phillips`. The considered regularization matrices are: **(a)** (I_N, D_2), **(b)** (D_2, I_N), **(c)** (D_1, D_2), **(d)** (D_2, D_1).

the average results obtained running each test problem 50 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. Figure 4 focuses on the test problem `shaw`. Using the layout of Figure 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 (8) and (13) with $(L_1, L_2) = (D_1, D_2)$ (black square and hexagram, respectively). The top left frame of Figure 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-parameters methods are improved with respect to the one-parameter case. Often criteria (8) and (13) deliver solutions of similar quality, whose accuracy is close to the optimal one. Moreover, the behavior

	Reg.M.	**error (1P)**	**error (opt)**	**error (8)**	λ_1	λ_2	**error (13)**
`baart`	(D_2, D_1)	8.0635e-02	1.3453e-01	1.3453e-01	1.6472e-01	1.0864e-03	1.3453e-01
`baart`	(D_1, D_2)	8.0635e-02	7.8716e-02	7.8716e-02	1.0959e-08	1.7387e+02	7.8716e-02
`deriv2,1`	(D_2, D_1)	3.1010e-03	6.3631e-02	6.3631e-02	2.5941e-03	2.2372e-06	6.3631e-02
`deriv2,1`	(D_1, D_2)	3.1010e-03	1.9937e-03	1.0704e-02	1.3493e-08	5.2958e-04	2.7687e-02
`deriv2,2`	(D_2, D_1)	2.5946e-02	5.1406e-02	5.1406e-02	7.1663e-03	3.1757e-06	5.1406e-02
`deriv2,2`	(D_1, D_2)	2.5946e-02	2.5946e-02	2.5946e-02	1.0000e-08	2.5395e+01	2.5946e-02
`i_laplace`	(D_2, D_1)	4.0132e-01	1.4544e-02	1.3883e-01	6.8407e-03	1.1369e+01	1.3883e-01
`i_laplace`	(D_1, D_2)	4.0132e-01	2.5835e-02	4.0099e-01	1.0638e-08	1.3742e+01	4.0132e-01
`phillips`	(D_2, D_1)	3.0705e-02	2.8188e-02	2.8189e-02	2.9757e+00	4.0418e-02	2.8188e-02
`phillips`	(D_1, D_2)	3.0705e-02	3.0668e-02	3.0705e-02	1.0000e-08	9.2435e+01	3.0705e-02
`shaw`	(D_2, D_1)	2.2179e-01	1.7871e-01	2.3344e-01	3.0639e-02	1.2488e-08	1.7972e-01
`shaw`	(D_1, D_2)	2.2179e-01	1.9440e-01	2.1487e-01	2.3807e-08	2.0072e+00	1.9449e-01

Table 3: Results obtained by Tikhonov regularization problems (2) and (17) with the solution given in [8] (except for `i_laplace`) and $\widetilde{\varepsilon} = 10^{-2}$. The regularization matrix pairs are reported in the second column, and (13) is implemented with $(L_1, L_2) = (D_1, D_2)$.

of the solution is usually almost invariant with respect to the order of regularization matrix pair.

Table 4 reports the average relative errors obtained running the AT (with $L = D_2$) and RRAT methods 50 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 λ_k. The relative errors are displayed when the final dimension k of the Krylov subspaces $\mathcal{K}_k(A, b)$ or $\mathcal{K}_k(A, Ab)$ is selected according to an optimal criterion (i.e., the relative error is minimized) (third column), according to (8) (fourth column), according to (13) (just for general form problems) (fifth column), according to the stabilization of the values of λ_k (33) (with $\tau = 1$) (sixth column), and according to the stopping rule proposed in [14, 11] (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. (32)) drops below the threshold $\eta\varepsilon$. The relative errors delivered by the stopping strategies (8) or (13) are often close to the optimal ones, and one commonly gets an improved reconstruction with respect to the criterion proposed in [14, 11]. Figure 5 displays the history of the relative errors, of the quantities $\|x_{k,\lambda}\|$, and of the regularization parameters λ_k for the test problems `phillips` and `baart`, when the 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 7, respectively.

Finally, we test the performance of the 2P-AT method. Analogously to Example 1, we compare the results obtained when the approximation (37) together with the rule (8) is equipped with the discrepancy-based stopping criterion (42), and when the stopping iteration is selected in order to minimize the relative error (i.e., optimal stopping criterion). Similarly to Table 2, Table 5 reports the averages of the minimum relative errors obtained running 50 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

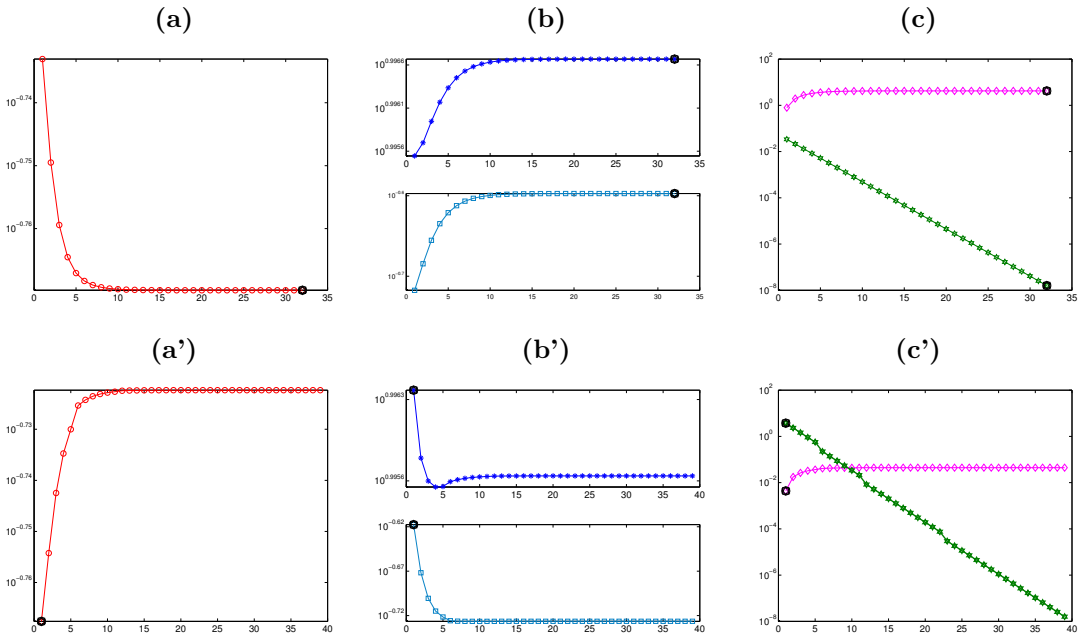

Figure 4: Test problem `shaw`. The first and second rows display quantities relative to problem (17) with (D_1, D_2) and (D_2, D_1), respectively. **(a)**, **(a')**: relative errors for the combinations of λ_i's displayed in the third frames; **(b)**, **(b')**: values of $\|x_\Lambda\|$ (upper frame) 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).

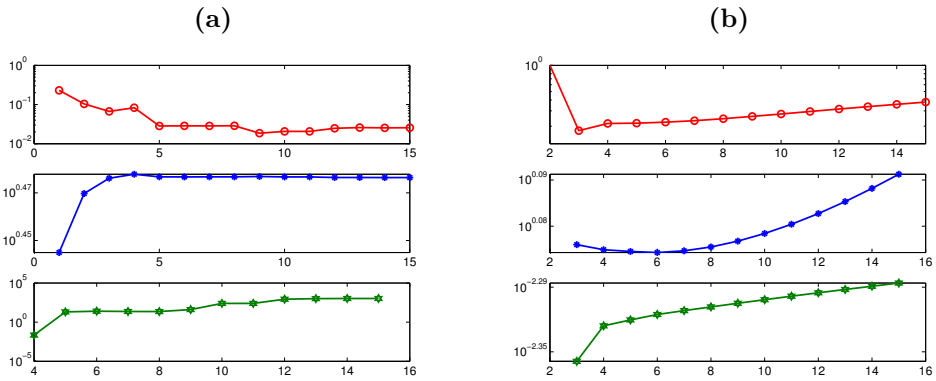

Figure 5: AT method (column **(a)**), and RRAT method (column **(b)**) applied to the test problem `phillips`. The first box displays the relative errors versus the number of iterations, the second box displays the quantities $\|x_{k,\lambda}\|$ versus the number of iterations k, and the third box displays the values λ_k versus the number of iterations k.

	Meth.	(opt)	(8)	(13)	(33)	[14, 11]
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)
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)
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)
deriv2,2	RRAT	2.64e-01 (8.2)	2.64e-01 (6.0)		2.70e-01 (5.0)	2.87e-01 (4.0)
i_laplace	AT	8.72e-02 (6.6)	5.55e-01 (9.2)	3.81e-01 (6.3)	1.92e-01 (7.3)	1.38e-01 (5.4)
i_laplace	RRAT	1.07e-01 (4.7)	1.20e-01 (4.8)		1.74e-01 (5.1)	1.38e-01 (4.1)
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)
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)
shaw	RRAT	1.46e-01 (10.1)	1.46e-01 (9.1)		1.48e-01 (5.1)	1.69e-01 (4.0)

Table 4: 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.

reported.

	Reg.M.	error (8,37,opt)	error (8,37,42)	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
deriv2,1	(D_1, D_2)	2.4581e-01 (6.4)	2.7423e-01 (5.2)	4.2869e-04	6.2413e-07
i_laplace	(D_2, D_1)	6.6196e-02 (6.9)	1.2223e-01 (6.3)	7.3493e+00	9.3721e-01
i_laplace	(D_1, D_2)	6.5610e-02 (7.4)	1.2765e-01 (6.4)	1.0593e+00	7.5247e+00
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

Table 5: Results obtained considering problem (34) with the solution given in [8] (except for i_laplace) 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.

Similarly to Figures 2 and 4, Figure 6 displays one single test performed with the coefficient matrix i_laplace. 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 7. We can still conclude that the quality of the reconstruction is stable with respect to the regularization matrix pairs (see in particular Figure 6), and that the relative errors obtained by applying simultaneously the strategies (37), (8) and (42) are close to the optimal ones.

Example 3. The test images employed for this set of experiments are available in the Matlab package [19], and are of size 256×256 pixels. To restore the blurred and

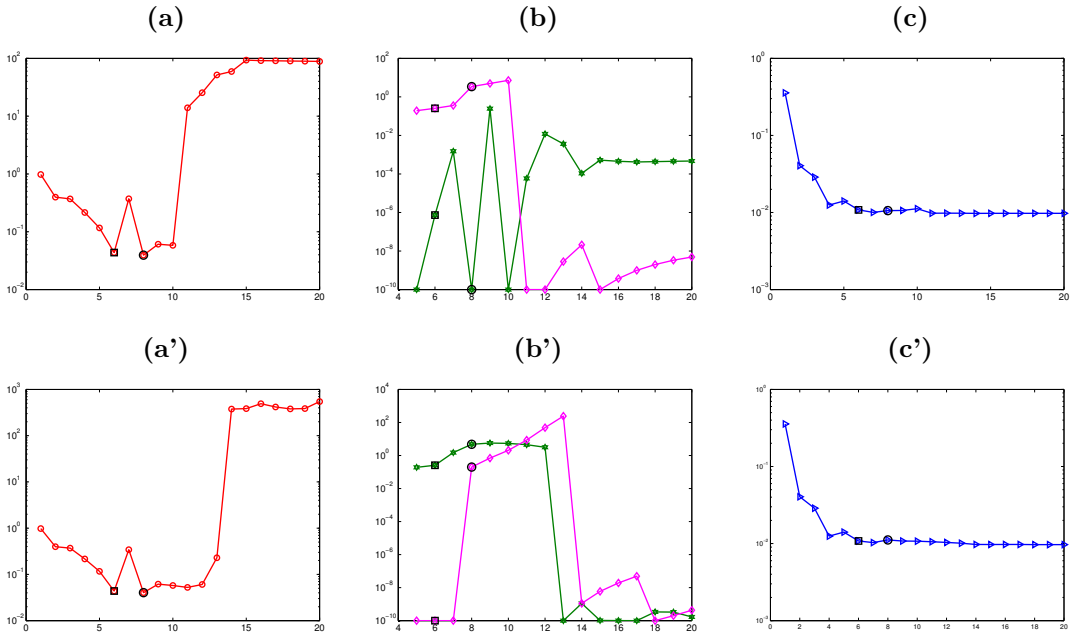

Figure 6: Test problem `i_laplace`. The first and second rows display quantities relative to problem (34) with (D_2, D_1) and (D_1, D_2), respectively. **(a)**, **(a')**: history of the relative errors; **(b)**, **(b')**: history of the regularization parameters $\lambda_1^{(k)}$ (hexagram) and $\lambda_2^{(k)}$ (diamond); **(c)**, **(c')**: history of the values of the discrepancy function.

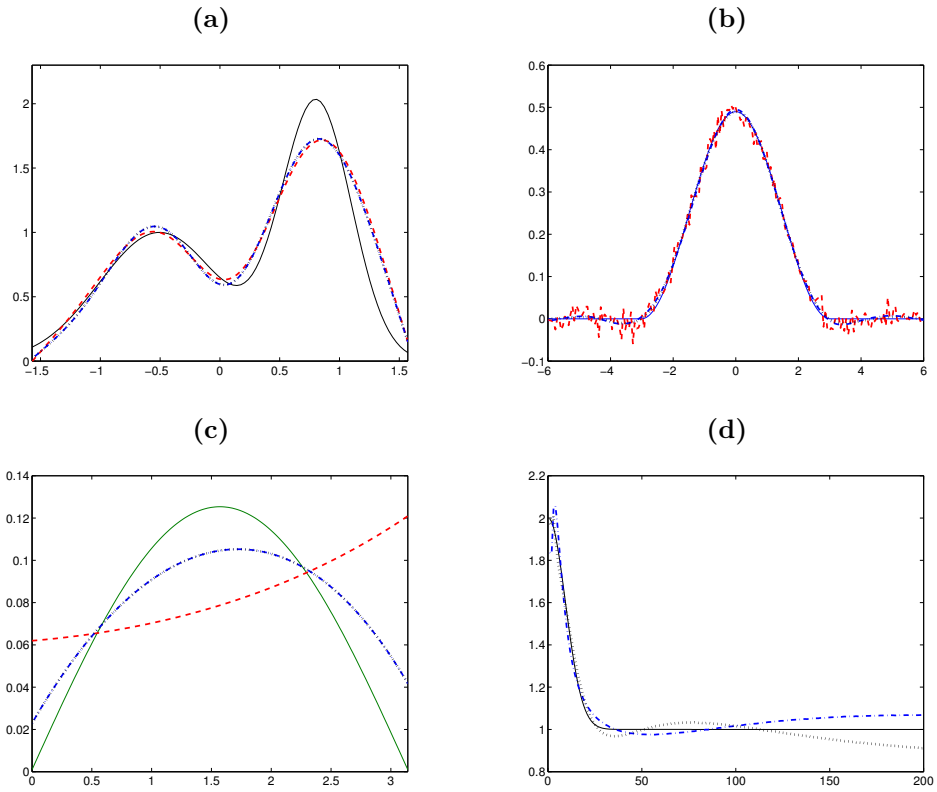

Figure 7: Reconstructions obtained by the two-parameter Tikhonov method (frame **(a)**), by the AT method (frame **(b)**), by the RRAT method (frame **(c)**), and by the 2P-AT method (frame **(d)**). In each frame, the exact solution is plotted by a thin solid line. In frame **(a)**, the reconstruction obtained by (2) is plotted by a dashed line, the reconstruction obtained by (17) and the criterion (8) is plotted by a dashed-dotted line, and the optimal reconstruction obtained by (17) is plotted by a dotted line. In frames **(b)** and **(c)**, the reconstruction obtained by the stopping criterion employed in [14, 11] is plotted by a dashed line, the reconstruction obtained applying the stopping criterion (8) is plotted by a dashed-dotted line, and the optimal reconstruction is plotted by a dotted line. In frame **(d)**, the reconstruction obtained by (37) and (8) is plotted by a dashed-dotted line, and the optimal reconstruction obtained by (37) is plotted by a dotted line.

noisy image shown in Figure 8, frame **(b)**, the two-parameter Tikhonov method (17) is applied, equipped with the criterion (8). The blur is defined by a symmetric Gaussian point spread function (PSF), given analytically by the function

$$k(s,t) = \frac{1}{2\pi\alpha^2} \exp\left(-\frac{1}{2\alpha^2}(s^2+t^2)\right), \tag{47}$$

where $\alpha = 2.5$. Periodic boundary conditions are considered, so that the matrix A in (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 [10, Chapter 4] for the details). The noise level is 10^{-2}. The regularization 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 $n = 256$, and D_1 and D_2 are defined by (44) and (45) with $N = n = 256$. The reconstruction obtained is displayed in Figure 8, frame **(c)**. The relative restoration error is $2.5209 \cdot 10^{-1}$, and the corresponding regularization vector is $(\lambda_1, \lambda_2) = (2.5594 \cdot 10^{-3}, 7.4964 \cdot 10^{-2})$.

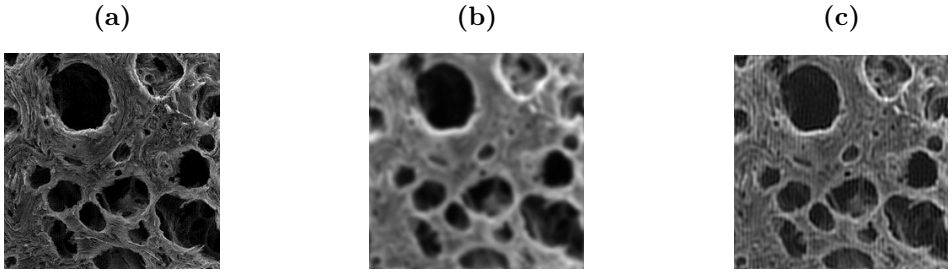

Figure 8: **(a)** exact image; **(b)** blurred and noisy ($\widetilde{\varepsilon} = 10^{-2}$) image; **(c)** reconstruction obtained by (17) with (8).

The 2P-AT method (34), equipped with the rules (37) and (8), is applied to restore the blurred and noisy image shown in Figure 9, frame **(b)**. A symmetric Gaussian PSF, with $\alpha = 2$ in (47), 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 9, frame **(c)**. The relative restoration error is $1.8660 \cdot 10^{-1}$, and the corresponding regularization vector is $(\lambda_1, \lambda_2) = (5.5782 \cdot 10^{-9}, 1.0071 \cdot 10^{-4})$.

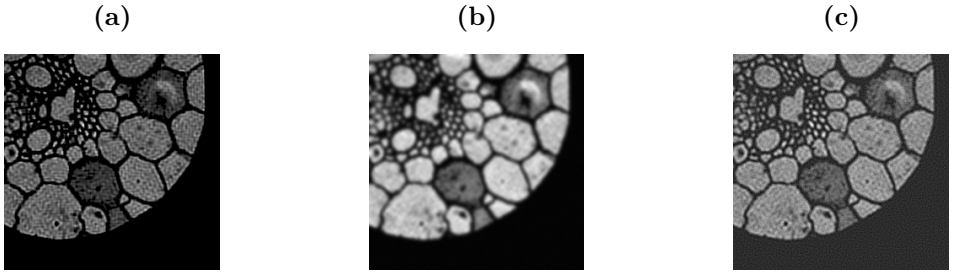

Figure 9: **(a)** Exact image; **(b)** blurred and noisy ($\widetilde{\varepsilon} = 10^{-2}$) image; **(c)** reconstruction obtained by (34) and (8).

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 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.

References

[1] M. Belge, M. E. Kilmer, and E. L. Miller. Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18:1161–1183, 2002.

[2] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numer. Math.*, 94:203–228, 2003.

[3] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *J. Comput. Appl. Math.*, 123:423–446, 2000.

[4] M. Fornasier, V. Naumova, and S. V. Pereverzyev. Parameter choice strategies for multipenalty regularization. *SIAM J. Numer. Anal.* To appear.

[5] S. Gazzola and P. Novati. Multi-parameter Arnoldi-Tikhonov methods. *Electron. Trans. Numer. Anal.*, 40:452–475, 2013.

[6] S. Gazzola, P. Novati, and M. R. Russo. On Krylov projection methods and Tikhonov regularization. *Electron. Trans. Numer. Anal.* To appear.

[7] S. Gazzola, P. Novati, and M. R. Russo. Embedded techniques for choosing the parameter in Tikhonov regularization. *Numer. Linear Algebra Appl.*, 21(6):796–812, 2014.

[8] P. C. Hansen. Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms*, 6:1–35, 1994.

[9] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM, Philadelphia, PA, 1998.

[10] P. C. Hansen, J. G. Nagy, and D. P. O'Leary. *Deblurring Images: Matrices, Spectra and Filtering.* SIAM, Philadelphia, PA, 2006.

[11] M. Hochstenbach and L. Reichel. An iterative method for Tikhonov regularization with a general linear regularization operator. *J. Integral Equations Appl.*, 22:463–480, 2010.

[12] M. Hochstenbach and L. Reichel. Combining approximate solutions for linear discrete ill-posed problems. *J. Comput. Appl. Math.*, 236:2179–2185, 2012.

[13] M. E. Hochstenbach, N. McNinch, and L. Reichel. Discrete ill-posed least-squares problems with a solution norm constraint. *Linear Algebra Appl.*, 436:3801–3818, 2012.

[14] B. Lewis and L. Reichel. Arnoldi-Tikhonov regularization methods. *J. Comput. Appl. Math.*, 226:92–102, 2009.

[15] S. Lu and S. V. Pereverzyev. Multi-parameter regularization and its numerical realization. *Numer. Math.*, 118:1–31, 2011.

[16] S. Lu, S. V. Pereverzyev, Y. Shao, and U. Tautenhahn. Discrepancy curves for multi-parameter regularization. *J. Inv. Ill-Posed Problems*, 18:655–676, 2010.

[17] S. Morigi, L. Reichel, and F. Sgallari. Orthogonal projection regularization operators. *Numer. Algorithms*, 44:99–114, 2007.

[18] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[19] J. G. Nagy, K. M. Palmer, and L. Perrone. 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.

[20] L. Reichel, F. Sgallari, and Q. Ye. Tikhonov regularization based on generalized Krylov subspace methods. *Appl. Numer. Math.*, 62:1215–1228, 2012.

[21] L. Reichel and A. Shyshkov. A new zero-finder for Tikhonov regularization. *BIT*, 48:627–643, 2008.

[22] L. Reichel and Q. Ye. Simple square smoothing regularization operators. *Electron. Trans. Numer. Anal.*, 33:63–83, 2009.

[23] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd Ed.* SIAM, Philadelphia, PA, 2003.