

Embedded techniques for choosing the parameter in Tikhonov regularization

S. Gazzola, P. Novati^{*,†} and M. R. Russo

Departments of Mathematics University of Padua Padua Italy

SUMMARY

This paper introduces a new strategy for setting the regularization parameter when solving large-scale discrete ill-posed linear problems by means of the Arnoldi–Tikhonov method. This new rule is essentially based on the discrepancy principle, although no initial knowledge of the norm of the error that affects the right-hand side is assumed; an increasingly more accurate approximation of this quantity is recovered during the Arnoldi algorithm. Some theoretical estimates are derived in order to motivate our approach. Many numerical experiments performed on classical test problems as well as image deblurring problems are presented. Copyright © 2014 John Wiley & Sons, Ltd.

Received 10 June 2013; Revised 28 January 2014; Accepted 4 March 2014

KEY WORDS: linear discrete ill-posed problems; Tikhonov regularization; Arnoldi algorithm; Discrepancy principle

1. INTRODUCTION

Let us consider a linear discrete ill-posed problem of the form

$$Ax = b, \quad (1)$$

where $A \in \mathbb{R}^{N \times N}$ is severely ill-conditioned and may be of huge size. These sort of systems typically arise from the discretization of Fredholm integral equations of the first kind with compact kernel (for an exhaustive background on these class of problems, cf. [1, Chapter 1]). The right-hand side b is assumed to be affected by an unknown additive error e coming from the discretization process or measurements inaccuracies, that is,

$$b = b^{ex} + e, \quad (2)$$

where b^{ex} denotes the unknown exact right-hand side. We assume that the unperturbed system $Ax = b^{ex}$ is consistent and we denote by x^{ex} a desired solution (e.g., the solution of minimal Euclidean norm); the system (1) is not guaranteed to be consistent. Referring to the SVD of the matrix A ,

$$A = U \Sigma V^T, \quad (3)$$

we furthermore assume that the singular values σ_i quickly decay toward zero with no evident gap between two consecutive ones.

Because of the ill-conditioning of A and the presence of noise in b , in order to find a meaningful approximation of x^{ex} , we have to substitute the available system (1) with a nearby problem having better numerical properties: this process is called regularization. One of the most well-known and

*Correspondence to: P. Novati, Departments of Mathematics, University of Padua, Padua, Italy.

†E-mail: novati@math.unipd.it

well-established regularization technique is Tikhonov method that, in its most general form, can be written as

$$\min_{x \in \mathbb{R}^N} \{ \|Ax - b\|^2 + \lambda \|L(x - x_0)\|^2 \}, \quad (4)$$

where $L \in \mathbb{R}^{P \times N}$ is the regularization matrix, $\lambda > 0$ is the regularization parameter and $x_0 \in \mathbb{R}^N$ is an initial guess for the solution. We denote the solution of the problem (4) by x_λ . When $L = I_N$ (the identity matrix of order N) and $x_0 = 0$, the problem is said to be in standard form. In this paper, the norm $\|\cdot\|$ is always the Euclidean one. The use of a regularization matrix different from the identity may improve the quality of the reconstruction obtained by (4), especially when one wants to enhance some known features of the solution. In many situations, L is taken as a scaled finite differences approximation of a derivative operator (cf. Section 5).

A proper choice of the regularization parameter is crucial, because it specifies the amount of regularization to be imposed. Many techniques have been developed in order to set the regularization parameter in (4); we cite [2, 3] for a review of the classical ones along with some more recent ones. Here, we are concerned with the discrepancy principle, which suggests to set the parameter λ such that the nonlinear equation

$$\|b - Ax_\lambda\| = \eta \|e\|, \quad \eta \gtrsim 1,$$

is satisfied. Of course, this strategy can be applied only if a fairly accurate approximation of the quantity $\|e\|$ is known.

Denoting by $x_{m,\lambda}$, the approximation of x_λ computed at the m -th step of a certain iterative method applied to (4), and by $\phi_m(\lambda) = \|b - Ax_{m,\lambda}\|$ the corresponding discrepancy, each nonlinear solver for the equation

$$\phi_m(\lambda) = \eta \|e\|, \quad (5)$$

leads to a parameter choice rule associated with the iterative process. The basic idea of this paper, in which we assume $\|e\|$ to be unknown, is to consider (if possible) the approximation $\phi_k(0) \approx \|e\|$, where $k < m$, and then to solve

$$\phi_m(\lambda) = \eta \phi_k(0), \quad (6)$$

with respect to λ . The use of (6) as a parameter choice rule is motivated by the fact that many iterative solvers for $Ax = b$ produce approximations $x_m = x_{m,0}$ whose corresponding residual $\|b - Ax_m\|$ tends to stagnate around $\|e\|$. In other words, the information about the noise level can be recovered during the iterative process. Moreover, in many situations, the computational effort of the algorithm that delivers $x_{m,\lambda}$ can be exploited for forming $x_{m,0}$ (or vice versa). For this reason, we may refer to any iterative process, which simultaneously uses x_m to approximate $\|e\|$ and solves (6) to compute $x_{m,\lambda}$ as an embedded approach.

In this paper, we are mainly interested in solving (4) by means of the so-called Arnoldi–Tikhonov (AT) methods (originally introduced in [4] for the standard form regularization), which are based on the orthogonal projection of (4) onto the Krylov subspaces $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}$ of increasing dimensions. As well known, these methods typically show a fast superlinear convergence when applied to discrete ill-posed problems, and hence they are particularly attractive for large-scale problems. Dealing with this kind of methods, efficient algorithms based on the solution of (5) have been considered in [5] and [6]. More recently, in [7] a very simple strategy for solving (5), based on the linearization of $\phi_m(\lambda)$, has been presented. In this paper, we extend the latter approach by considering the approximation $\phi_{m-1}(0) \approx \|e\|$ where, in this setting, $\phi_{m-1}(0)$ is just the norm of the generalized minimal residual method (GMRES) residual computed at the previous iteration.

The paper is organized as follows. In Section 2, we survey the basic features of the AT methods. In Section 3, we review the linearization technique described in [7], and in Section 4, we explain the parameter choice rule based on an embedded approach, giving also a theoretical justification in the AT case. In the first part of Section 5, we write down the algorithm, in order to summarize the new method and to better describe some practical details; the remaining parts are devoted to display the results of some of the performed numerical tests.

2. THE ARNOLDI-TIKHONOV METHOD

The AT method was first proposed in [4] with the basic aims of reducing the problem (4) (in the particular case $L = I_N$ and $x_0 = 0$) to a problem of much smaller dimension and to avoid the use of A^T as in the Lanczos-type methods (see e.g., [8]). Then, in [7, 9, 10], the method has been extended to work with a general $L \in \mathbb{R}^{P \times N}$ and x_0 . Assuming $x_0 = 0$ (this assumption will hold throughout the paper), we consider the Krylov subspaces

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}, \quad m \geq 1. \quad (7)$$

In order to construct an orthonormal basis for this Krylov subspace, we can use the Arnoldi algorithm [11, Chapter 6], which leads to the associated decomposition

$$AW_m = W_m H_m + h_{m+1,m} w_{m+1} e_m^T \quad (8)$$

$$= W_{m+1} \bar{H}_m, \quad (9)$$

where $W_{m+1} = [w_1, \dots, w_{m+1}] \in \mathbb{R}^{N \times (m+1)}$ has orthonormal columns that span the Krylov subspace $\mathcal{K}_{m+1}(A, b)$, and $w_1 = b / \|b\|$. The matrices $H_m \in \mathbb{R}^{m \times m}$ and $\bar{H}_m \in \mathbb{R}^{(m+1) \times m}$ are upper Hessenberg.

The AT method searches for approximations $x_{m,\lambda}$ of the solution of problem (4) belonging to $\mathcal{K}_m(A, b)$. Therefore, replacing $x = W_m y$, $y \in \mathbb{R}^m$, into (4), yields the reduced minimization problem

$$y_{m,\lambda} = \arg \min_{y \in \mathbb{R}^m} \left\{ \|\bar{H}_m y - c\|^2 + \lambda \|L W_m y\|^2 \right\}, \quad (10)$$

where $c = \|b\|e_1$, being e_1 the first vector of the canonical basis of \mathbb{R}^{m+1} . The aforementioned problem is equivalent to

$$y_{m,\lambda} = \arg \min_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} L W_m \end{pmatrix} y - \begin{pmatrix} c \\ 0 \end{pmatrix} \right\|^2. \quad (11)$$

Obviously, $y_{m,\lambda}$ is also the solution of the normal equation

$$(\bar{H}_m^T \bar{H}_m + \lambda W_m^T L^T L W_m) y_{m,\lambda} = \bar{H}_m^T c. \quad (12)$$

We remark that, when dealing with standard form problems ($L = I_N$ and $x_0 = 0$), the Arnoldi-Tikhonov formulation considerably simplifies thanks again to the orthogonality of the columns of W_m and, instead of (11), we can consider

$$y_{m,\lambda} = \arg \min_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} I_m \end{pmatrix} y - \begin{pmatrix} c \\ 0 \end{pmatrix} \right\|^2. \quad (13)$$

In (13), the dimension of the problem is fully reduced because at each iteration we deal with a $(2m+1) \times m$ matrix. On the other side, considering (11), there is still track of the original dimensions of the problem. Anyway, because the AT method can typically recover a meaningful approximation of the exact solution after just a few iterations of the Arnoldi algorithm have been performed, the computational cost is still low. Assuming that $P \leq N$ in (4) and defining a new matrix L obtained by appending $N - P$ zero rows to the original one, we can also consider the following new formulation:

$$y_m = \arg \min_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} L_m \end{pmatrix} y - \begin{pmatrix} c \\ 0 \end{pmatrix} \right\|^2, \quad \text{where } L_m = W_m^T L W_m. \quad (14)$$

The aforementioned problem is not equivalent to (11) anymore but can be justified by the fact that L_m is the orthogonal projection of L onto $\mathcal{K}_m(A, b)$, and hence, in some sense, L_m inherits the properties of L (see [12] for a discussion). Alternatively, one can obtain a full dimensional reduction by considering the QR factorization of LW_m (see [9]). Some preliminary numerical experiments have revealed that this two approaches yield very similar results.

3. THE PARAMETER CHOICE STRATEGY

As said in the Introduction, the discrepancy principle is a well-known and quite successful parameter selection strategy that, when applied to Tikhonov regularization method (4), prescribes to choose the regularization parameter $\lambda > 0$ such that $\|Ax_\lambda - b\| = \eta\|e\|$, where the parameter η is greater than 1, although very close to it.

An algorithm exploiting the discrepancy principle has been first considered for the Arnoldi–Tikhonov method in [5], where the authors suggest to solve, at each iteration m , the nonlinear equation

$$\phi_m(\lambda) := \|\tilde{H}_m y_{m,\lambda} - c\| = \eta\|e\|, \tag{15}$$

employing a special zero-finder described in [6]. In order to decide when to stop the iterations, a preliminary condition should be satisfied and then some adjustments should be made.

Considering the normal equations associated to (14), we write

$$\phi_m(\lambda) = \left\| c - \tilde{H}_m (\tilde{H}_m^T \tilde{H}_m + \lambda L_m^T L_m)^{-1} \tilde{H}_m^T c \right\|. \tag{16}$$

Denoting by $r_m = b - Ax_m$ the GMRES residual, we have that $\phi_m(0) = \|r_m\|$. In this setting, in [7], the authors solve (15) after considering the linear approximation

$$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m, \tag{17}$$

where, at each iteration, the scalar β_m is defined by the ratio

$$\beta_m = \frac{\phi_m(\lambda_{m-1}) - \phi_m(0)}{\lambda_{m-1}}. \tag{18}$$

In (18), $\phi_m(\lambda_{m-1})$ is obtained by solving the m -dimensional problem (14) using the parameter $\lambda = \lambda_{m-1}$, which is computed at the previous step.

Therefore, to select $\lambda = \lambda_m$ for the next step of the AT algorithm, we can approximate $\phi_m(\lambda_m)$ by (17) and impose

$$\phi_m(\lambda_m) = \eta\|e\|. \tag{19}$$

Substituting in the linear approximation of $\phi_m(\lambda_m)$, the expression derived in (18), and using the condition (19), we obtain

$$\lambda_m = \frac{\eta\|e\| - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \lambda_{m-1}. \tag{20}$$

When $\phi_m(0) > \eta\|e\|$, formula (20) produces a negative value for λ_m . Thus, in order to keep $\lambda_m > 0$, we consider the relation

$$\lambda_m = \left| \frac{\eta\|e\| - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \right| \lambda_{m-1}. \tag{21}$$

In this procedure, λ_0 must be set to an initial value by the user, but the numerical experiments show that this strategy is very robust with respect to this choice (typically one may set $\lambda_0 = 1$).

Remark 1

We remark that the use of the absolute value in (21) can be avoided by forcing initially $\lambda = 0$, that is, working with the GMRES, and then switching to the AT method equipped with (20) as soon as $\phi_m(0) < \eta \|e\|$.

In [7], this scheme has been called secant update method, because at each iteration of the Arnoldi algorithm, it basically performs just one step of a secant-like zero-finder applied to the equation $\phi_m(\lambda) = \eta \|e\|$. Numerically, formula (21) is very stable, in the sense that after the discrepancy principle is satisfied, λ_m is almost constant for growing values of m .

4. EXPLOITING THE GMRES RESIDUAL

We now try to generalize the secant update approach, dropping the hypothesis that the quantity $\|e\|$ is available. In this situation, one typically employs other well-known techniques, such as the L-curve criterion or the generalized cross validation (GCV); both have already been used in connection with the AT or Lanczos-hybrid methods [4, 12–14]. The strategy we are going to describe is to be considered different because we still want to apply the discrepancy principle, starting with no information on $\|e\|$ and trying to recover an estimate of it during the iterative process.

Our basic assumption is that, after just a few iterations of the Arnoldi algorithm, the norm of the residual associated to the GMRES method lies around the threshold $\|e\|$ and, despite being slightly decreasing, stabilizes during the following iterations (cf. Figure 5). This motivates the use of the following strategy to choose the regularization parameter at the m -th iteration

$$\lambda_m = \frac{\eta \phi_{m-1}(0) - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \lambda_{m-1}, \quad \eta > 1, \quad (22)$$

where we have replaced the quantity $\|e\|$ in (21) by $\phi_{m-1}(0) = \|r_{m-1}\|$. We remark that, from a theoretical point of view, formula (22) cannot produce negative values because $\phi_m(0) = \|r_m\| \leq \|r_{m-1}\| = \phi_{m-1}(0)$ and $\phi_m(\lambda)$ is an increasing function with respect to λ . In what follows, we provide a theoretical justification for this approach, giving also some numerical experiments using test problems taken from [15]; in the first subsection, we focus on the case $b = b^{ex}$, while in the second subsection, we treat the case $b = b^{ex} + e$.

4.1. The unperturbed problem

Thanks to a number of results in literature (see, e.g., [16]), we know that the GMRES exhibits superlinear convergence when solving problems in which the singular values rapidly decay to 0. Indeed, in this situation, the Krylov subspaces tend to become A -invariant after few iterations. In general, the fast convergence of a Krylov subspace method applied to an ill-posed system (1) can be explained by monitoring the behavior of the sequence $\{h_{m+1,m}\}_m$. The following theorem (proved in [12, Proposition 3.3]) gives us an estimate for the quantities $\{h_{m+1,m}\}_m$ whenever we work with the exact right-hand side b^{ex} , and A is assumed to be severely ill-conditioned, that is, with singular values that decay exponentially (cf. [17]).

Theorem 2

Assume that A has full rank with singular values of the type $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$, and that b^{ex} satisfies the discrete Picard condition [18], that is, $|u_j^T b^{ex}| \sim \sigma_j$, where u_j is the j -column of the matrix U in (3). Then, if b^{ex} is the starting vector of the Arnoldi process, we have

$$h_{m+1,m} = O\left(m^{3/2} \sigma_m\right). \quad (23)$$

In order to assess the estimate (23), in Figure 1, we report a couple of numerical experiments.

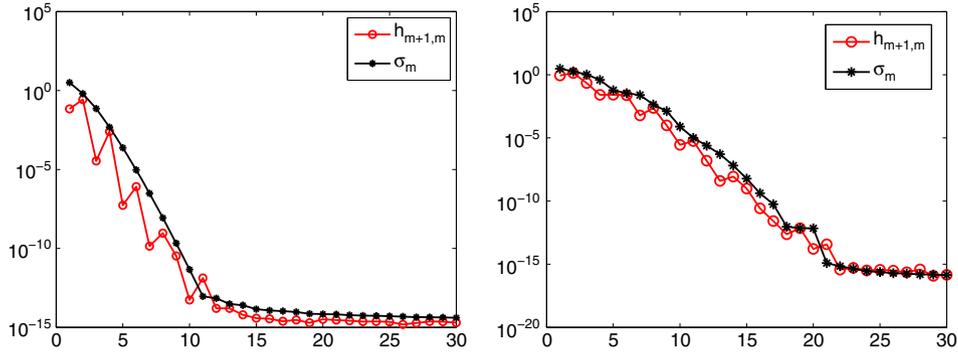


Figure 1. Behavior of the sequences $\{h_{m+1,m}\}_m$ and $\{\sigma_m\}_m$ for the test problems baart (left) and shaw (right) from [15].

Remark 3

The hypothesis $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$, apparently limits the aforementioned results to severely ill-conditioned problems. Actually, in [12, Remark 3.1], it is explained that this assumption is just used to have

$$\sum_{j \geq m+1} \sigma_j = O(e^{-\alpha m}) = O(\sigma_m).$$

In this sense, the results can be extended to mildly ill-conditioned problems, in which $\sigma_j = O(j^{-\alpha})$, $\alpha > 1$. In this situation, we would have

$$\sum_{j \geq m+1} \sigma_j = O(m^{1-\alpha}),$$

so that, for α sufficiently large, the results of Theorem 2 and Corollary 5, can be extended to mildly ill-conditioned problems by replacing σ_m with $O(m^{1-\alpha})$.

Relation (23) can be exploited to estimate the decay of the GMRES residual when dealing with the unperturbed problem. Indeed, it is well known that the GMRES residual is related to the full orthogonalization method (FOM) residual ρ_m as follows [11, Chapter 6]:

$$\|r_m\| \leq \|\rho_m\| = h_{m+1,m} |e_m^T H_m^{-1} c|, \tag{24}$$

where H_m is as in (8) and $c = \|b^{ex}\|e_1 \in \mathbb{R}^m$. Assuming that the FOM solutions $y_m = H_m^{-1}c$, $m = 1, \dots, N$ do not ‘explode’, that is,

$$\exists M > 0 \quad \text{such that} \quad |e_m^T H_m^{-1} c| \leq M, \quad \forall 1 \leq m \leq N, \tag{25}$$

and recalling the relation

$$\|r_m\|^2 = \frac{1}{\frac{1}{\|\rho_m\|^2} + \frac{1}{\|r_{m-1}\|^2}},$$

which expresses the well-known peak–plateau phenomenon (see [19]), we can conclude that the GMRES residuals decay as the quantities $h_{m+1,m}$. In Figure 2, we report the FOM residual history for some test problems in order to show the behavior described by (24).

Remark 4

Employing the SVD of the matrix H_m , that is,

$$H_m = U_m^{(m)} \Sigma_m^{(m)} \left(V_m^{(m)} \right)^T,$$

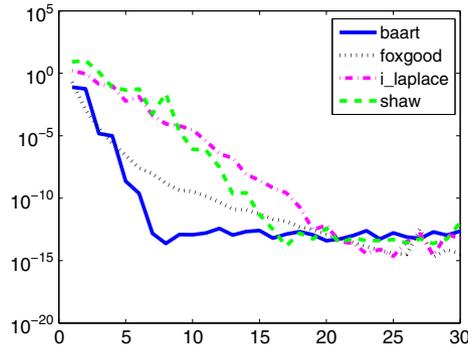


Figure 2. FOM residual history for some common test problems taken from [15].

where $\Sigma_m^{(m)} = \text{diag}(\sigma_1^{(m)}, \dots, \sigma_m^{(m)})$, we have

$$H_m^{-1} c = V_m^{(m)} (\Sigma_m^{(m)})^{-1} U_m^{(m)T} c,$$

so that assuming condition (25) is equivalent to requiring the discrete Picard condition to hold for the projected problem. It is known that, if $\bar{\sigma}_j^{(m)}$, $j = 1, \dots, m$, are the singular values approximations arising from the SVD of \tilde{H}_m , then $\bar{\sigma}_m^{(m)} \geq \bar{\sigma}_{m+1}^{(m+1)} \geq \sigma_N > 0$ (cf. [4]). Because $h_{m+1,m}$ goes rapidly to 0, we also have that after a few iterations $\sigma_j^{(m)} \approx \bar{\sigma}_j^{(m)}$, $j = 1, \dots, m$ so that we can expect that $\sigma_m^{(m)} \geq \sigma_N$. In general, however, we do not have guarantees that the constant M in (25) is small, so that (26) may be quantitatively not much useful. Everything is closely related to the SVD approximation that we can achieve with the Arnoldi algorithm (see [12] for some theoretical results). It is known that, if the matrix A is highly nonsymmetric, then the SVD approximation may be poor so that the discrete Picard condition may be badly inherited by the projected problem. In Figure 3, we show the values of the quantities $|e_m^T H_m^{-1} c|$ for increasing values of m and some Picard plots (cf. [20, Chapter 3]) relative to the projected quantities H_m and c for a couple of test problems taken from [15].

Thanks to the aforementioned derivations and to Theorem 2, we can immediately state the following:

Corollary 5

Under the hypothesis of Theorem 2 and assuming that (25) is satisfied, the GMRES residuals are of the type

$$\|r_m\| = O\left(m^{3/2} \sigma_m\right). \tag{26}$$

4.2. The perturbed problem

When the right-hand side of (1) is affected by noise, we can give the following preliminary estimate for the norm of the GMRES residual.

Proposition 6

Let $b = b^{ex} + e$ and let $r_m^{ex} = p_m^{ex}(A)b^{ex}$ be the residual of the GMRES applied to the system $Ax = b^{ex}$. Assume that for $m \geq m^*$, $\|p_m^{ex}(A)\| \leq \eta^*$ (cf. [21]). Then the m -th residual of the GMRES applied to $Ax = b$ satisfies

$$\|r_m\| \leq \eta \|e\|,$$

where

$$\eta = \frac{\|r_{m^*}^{ex}\|}{\|e\|} + \eta^*.$$

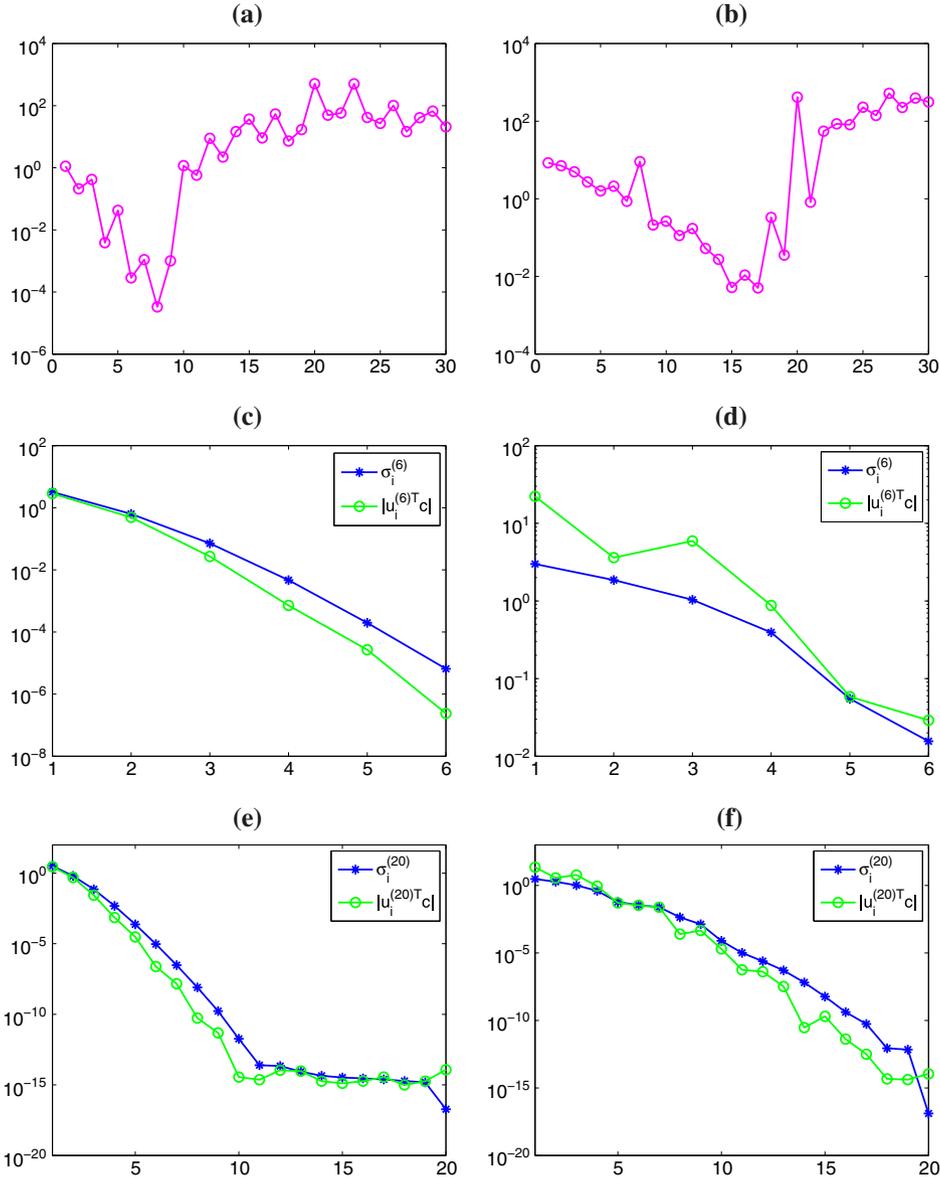


Figure 3. Plot of the quantities $|e_m^T H_m^{-1} c|$ versus the number of iterations m for the problems baart (a) and shaw (b). Plot of $\sigma_i^{(m)}$ and the Fourier coefficients $\left| \left(u_i^{(m)} \right)^T c \right|$ versus the index i for the unperturbed projected systems of size $m = 6$ (problem baart (c) and problem shaw (d)) and of size $m = 20$ (problem baart (e) and problem shaw (f)).

Proof

Thanks to the optimality property of the GMRES residual,

$$\|r_m\| = \min_{p_m(0)=1} \|p_m(A)b\| \leq \|p_m^{ex}(A)b\|.$$

Because $b = b^{ex} + e$,

$$\|r_m\| \leq \|p_m^{ex}(A)b^{ex}\| + \|p_m^{ex}(A)e\| \leq \|r_m^{ex}\| + \eta^* \|e\|.$$

The result follows from $\|r_m^{ex}\| \leq \|r_{m^*}^{ex}\|$, which holds for $m \geq m^*$. □

In the remaining part of this section, we try to give some additional information about the value of the constant η of Proposition 6. Let

$$\tilde{V}_m = \left[\frac{b}{\|b\|}, \frac{Ab}{\|Ab\|}, \dots, \frac{A^{m-1}b}{\|A^{m-1}b\|} \right], \quad \tilde{V}_m^{ex} = \left[\frac{b^{ex}}{\|b^{ex}\|}, \frac{Ab^{ex}}{\|Ab^{ex}\|}, \dots, \frac{A^{m-1}b^{ex}}{\|A^{m-1}b^{ex}\|} \right].$$

With these notations, we can write

$$\|r_m\| = \min_{s \in \mathbb{R}^{m+1}, s_1=0} \|b - \tilde{V}_{m+1}s\|,$$

where s_1 is the first component of vector s .

Proposition 7

For the GMRES residual, we have

$$\|r_m\| \leq \eta(m)\|e\|,$$

where

$$\eta(m) = 1 + \frac{\|r_m^{ex}\| + \|(\tilde{V}_{m+1} - \tilde{V}_{m+1}^{ex})s^{ex}\|}{\|e\|}$$

and s^{ex} ($s_1^{ex} = 0$) is such that $\|r_m^{ex}\| = \|b - \tilde{V}_{m+1}^{ex}s^{ex}\|$.

Proof

We have

$$\begin{aligned} \|r_m\| &= \min_{s \in \mathbb{R}^m} \|b - \tilde{V}_{m+1}s\| \leq \|b - \tilde{V}_{m+1}s^{ex}\| \\ &= \|b^{ex} + e - \tilde{V}_{m+1}s^{ex} + \tilde{V}_{m+1}^{ex}s^{ex} - \tilde{V}_{m+1}s^{ex}\| \\ &\leq \|r_m^{ex}\| + \|e\| + \|(\tilde{V}_{m+1} - \tilde{V}_{m+1}^{ex})s^{ex}\|. \end{aligned}$$

□

The fast decay of the singular values of A ensures that, for $k \geq 1$ (note that $s_1^{ex} = 0$)

$$\frac{1}{\|e\|} \left\| \frac{A^k b}{\|A^k b\|} - \frac{A^k b^{ex}}{\|A^k b^{ex}\|} \right\| \ll 1, \tag{27}$$

so that, whenever $\|r_m^{ex}\| \approx 0$, we have $\eta(m) \approx 1$. Condition (27) is also at the basis of the so-called range-restricted approach for Krylov type methods (see [5]). Moreover, we remark that the relation (27) can be interpreted as the discrete analogous of the Riemann–Lebesgue Lemma (see e.g.

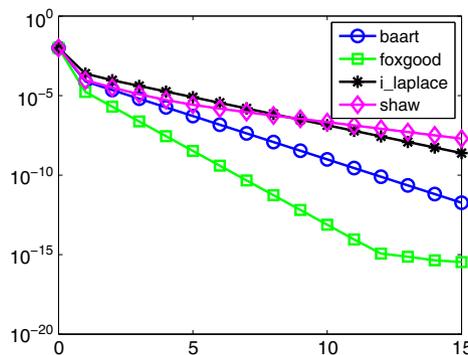


Figure 4. Decay of the quantities $\frac{1}{\|e\|} \left\| \frac{A^k b}{\|A^k b\|} - \frac{A^k b^{ex}}{\|A^k b^{ex}\|} \right\|$ versus the value of $k \geq 0$. The right-hand side is affected by 1% Gaussian noise.

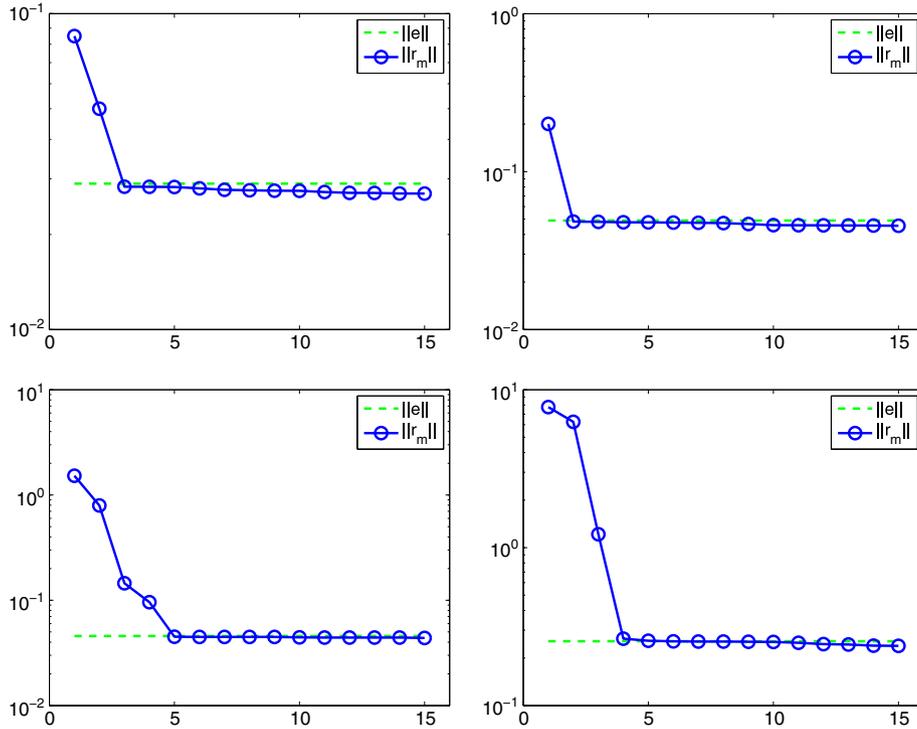


Figure 5. GMRES residual history when the right-hand side is affected by 1% noise. In clockwise order, the problem considered are baart, foxgood, shaw, and i_laplace.

[1, p.6)), whenever we assume that the noise e does not involve low frequencies. We give some examples of this behavior in Figure 4.

Finally, in Figure 5, we prove experimentally our main assumption, that is, $\|r_m\| \approx \|e\|$ for m sufficiently large, which justifies the use of formula (22).

5. ALGORITHM AND NUMERICAL EXPERIMENTS

Comparing the parameter selection strategies (21) and (22), we can state that (22) generalizes the approach described in Section 3, because no knowledge of $\|e\|$ is assumed. However, on the downside, scheme (21) can simultaneously determine the value of the regularization parameter at each iteration and the number of iterations to be performed, while this is no more possible considering the rule (22). In order to determine when to stop the iterations of the Arnoldi algorithm, we have to consider a separate stopping criterion. Because both $\phi_m(\lambda_{m-1})$ and $\|r_m\|$ exhibit a stable behavior going on with the iterations, a way to set m is to monitor when such stability occurs, that is, to evaluate the relative difference between the norm of the residuals and the relative difference between the discrepancy functions. Therefore, once two thresholds τ_{res} and τ_{discr} have been set, we decide to stop the iterations as soon as

$$\frac{\|r_m\| - \|r_{m-1}\|}{\|r_{m-1}\|} < \tau_{\text{res}}, \tag{28}$$

and

$$\frac{\phi_m(\lambda_{m-1}) - \phi_{m-1}(\lambda_{m-2})}{\phi_{m-1}(\lambda_{m-2})} < \tau_{\text{discr}}. \tag{29}$$

This approach is very similar to the one adopted in [13] for the GCV method in a hybrid setting. Also, in [4], the authors decide to terminate the Arnoldi process when the corners of two consecutive

projected L-curves are pretty close. Moreover, we can expect the value of λ_m obtained at the end of the iterations to be suitable for the original problem (4).

The method so far described can be summarized in Algorithm 1.

Algorithm 1 AT method equipped with the parameter choice rule (22)

Inputs: $A, b, L, x_0, \lambda_0, \eta, \tau_{\text{res}}, \tau_{\text{discr}}$

For $m = 1, 2, \dots$, until (28) and (29) are both fulfilled

1. Update W_m and \bar{H}_m by the Arnoldi algorithm (9).
2. Compute the reduced-dimension GMRES solution $y_{m,0}$ (cf. (10)) and the corresponding residual r_m .
3. Compute the solution $y_{m,\lambda}$ of (14), taking

$$\begin{cases} \lambda = \lambda_0 & \text{if } m = 1, 2, \\ \lambda = \lambda_{m-1} & \text{otherwise.} \end{cases}$$

4. Compute the discrepancy $\phi_m(\lambda_{m-1}) = \|\bar{H}_m y_{m,\lambda_{m-1}} - c\|$.
5. if $m \geq 2$ update λ_m by formula (22).

end

Compute $x_{m,\lambda_{m-1}} = W_m y_{m,\lambda_{m-1}}$.

To illustrate the behavior of this algorithm, we treat three different kinds of test problems. All the experiments have been carried out using Matlab 7.10 with 16 significant digits on a single processor computer (Intel Core i7). The algorithm is implemented with $\lambda_0 = 1$, $\eta = 1.02$, and $\tau_{\text{res}} = \tau_{\text{discr}} = 5 \cdot 10^{-2}$.

5.1. Test problems from regularization tools

We consider again some classical test problems taken from Hansen's Regularization Tools [15]. In particular, in Figure 6, we report the results for the problems `baart`, `shaw`, `foxgood`, `i_laplace`; the right-hand side b is affected by additive 0.1% Gaussian noise e , such that the noise level $\varepsilon = \|e\|/\|b^{\text{ex}}\|$ is equal to 10^{-3} . The dimension of each problem is $N = 120$. The regularization operator used is the discrete first derivative L_1 for `shaw` and `i_laplace`, and the discrete second derivative L_2 for `baart` and `foxgood`, augmented with one or two zero rows, respectively, in order to make it square, that is,

$$L_1 := \begin{pmatrix} 1 & -1 & & & \\ & \ddots & \ddots & & \\ & & 1 & -1 & \\ 0 & \dots & \dots & 0 & \end{pmatrix}, \quad L_2 := \begin{pmatrix} 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 \\ 0 & \dots & \dots & \dots & 0 & \\ 0 & \dots & \dots & \dots & 0 & \end{pmatrix}. \quad (30)$$

For each experiment, we show the following: (i) the approximate solution; (ii) the relative residual and error history; and (iii) the value of the regularization parameter computed at each iteration by the secant update method (λ_{sec}) given by formula (21), the embedded method (λ_{emb}) computed by (22), the ones arising from the L-curve criterion [4] ($\lambda_{\text{L-curve}}$), and the optimal one (λ_{opt}) for the original, full dimensional regularized problem (4) obtained by the minimization of the distance between the regularized and the exact solution [22]

$$\min_{\lambda} \|x_{\lambda} - x^{\text{ex}}\|^2 = \min_{\lambda} \left\| \sum_{i=1}^P \frac{\lambda^2}{(\gamma_i^2 + \lambda^2)} \frac{\bar{u}_i^T b}{\sigma_i} x_i + \sum_{i=P+1}^N (u_i^T b) x_i - \sum_{i=1}^N \frac{u_i^T b^{\text{ex}}}{\sigma_i} v_i \right\|,$$

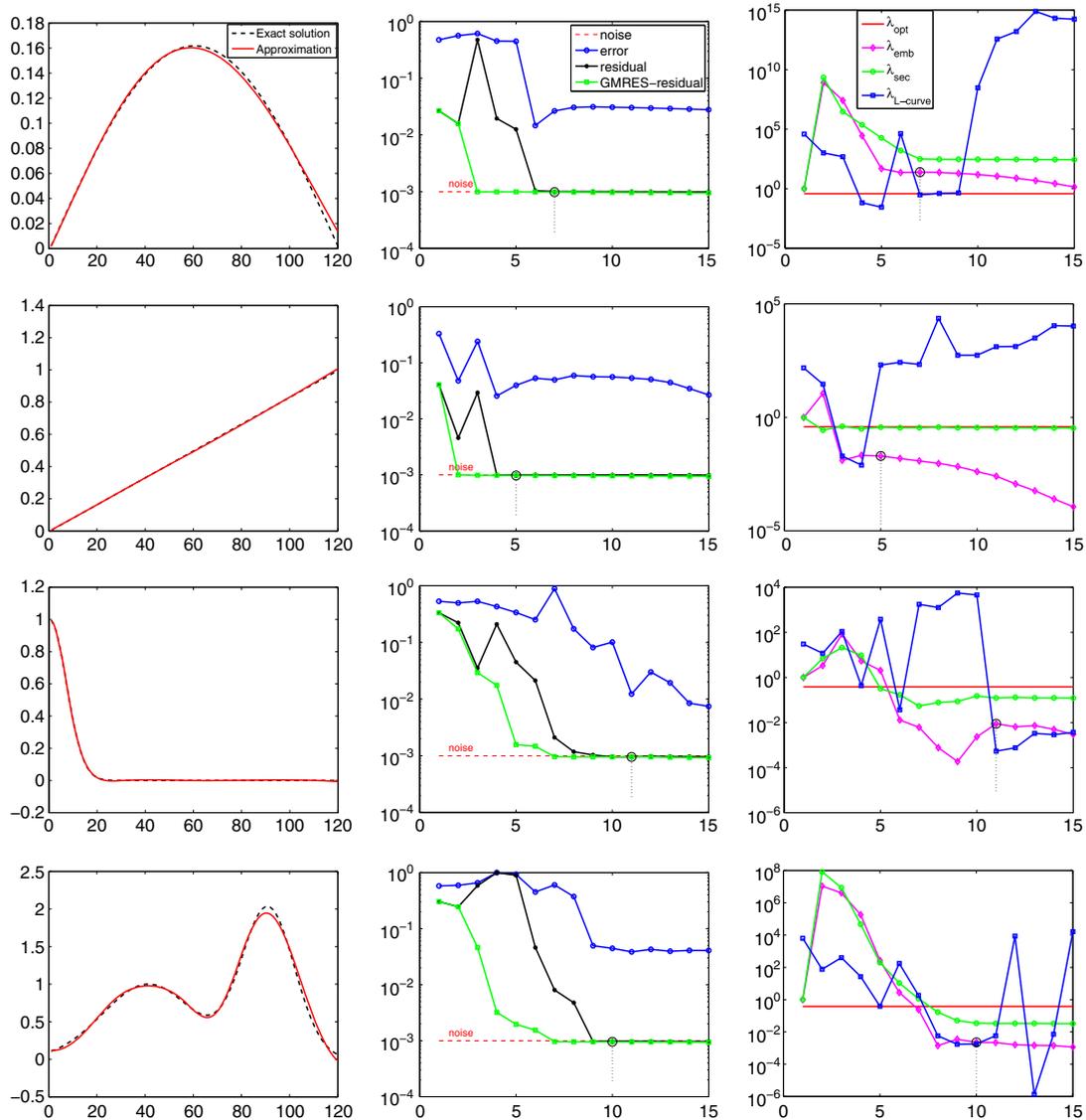


Figure 6. From top to bottom: results for *baart*, *foxgood*, *i_laplace*, and *shaw*. On the left column, we display the computed approximate solution. In the middle column, we show the convergence behavior of the new method (error, discrepancy, and GMRES residual) with the noise level highlighted by a dashed line. On the right, we compare different parameter choice strategies. The thick circle displayed in all the frame of the middle and the rightmost columns marks the iteration at which we would stop, according to the rule (28) and (29). The approximate solutions refer to this iteration.

where $\gamma_i, \bar{u}_i, i = 1, \dots, P$ are, respectively, the generalized singular values and left generalized singular vectors of (A, L) , and $x_i, i = 1, \dots, N$ are the right generalized singular vectors of (A, L) . Finally, in Figure 7, we more carefully compare the behavior of the AT method coupled with the embedded, the GCV [12], and the L-curve [4] parameter choice rules. The graphs are obtained running 50 times each test problem (the noise level is $\varepsilon = 10^{-2}$, and for each test, a different noise realization is considered); the values of the relative errors and the number of iterations resulting from the stopping rules (28) and (29) are displayed. We underline the mean values using a thick square for each rule. The results show our method to produce meaningful approximate solutions whose quality is comparable with the other existing approaches employed when no information about the norm of the noise is available.

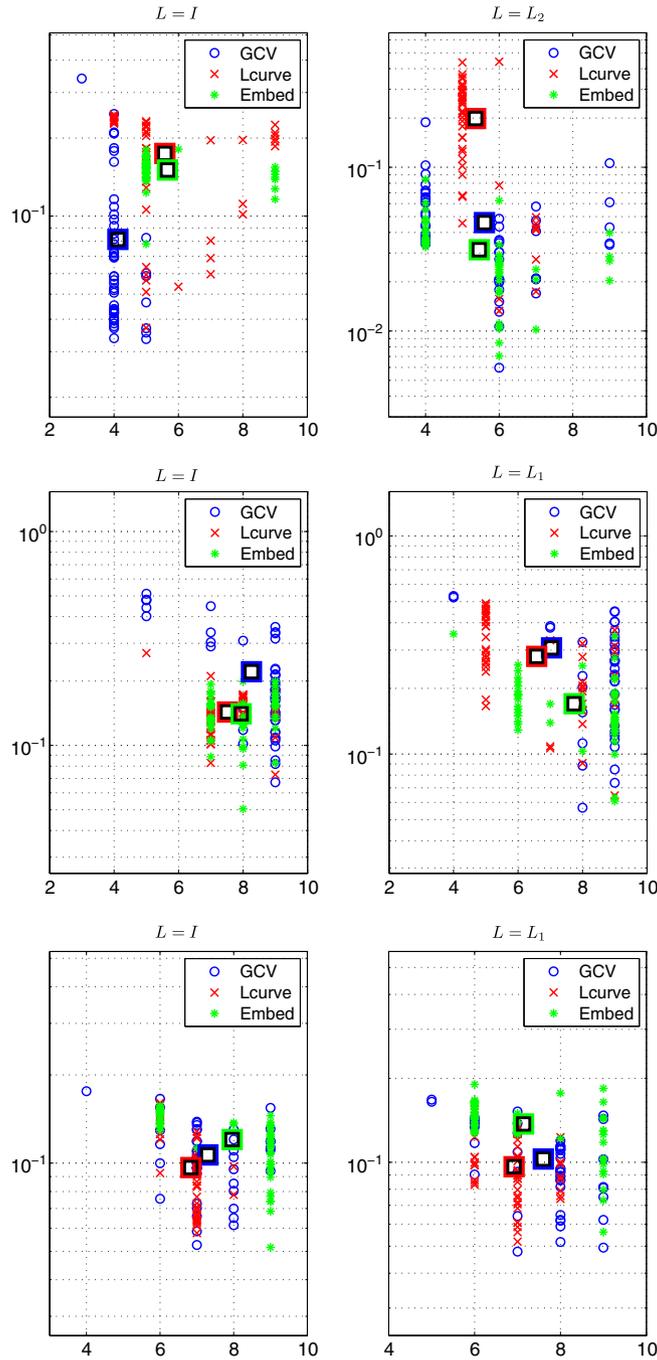


Figure 7. Arnoldi–Tikhonov method coupled with different parameter choice rules for the test problems *baart*, *i_laplace* and *shaw* (from top to bottom), and different regularization matrices. Relative errors versus the number of iterations obtained running 50 times each test (one marker is displayed for each of them). In each test, the number of iterations is the one resulting from the stopping criterion. The thick squares display the mean values.

5.2. Results for image restoration

To test the performance of our algorithm in the image restoration context, a number of experiments were carried out, some of which are presented here.

Let X be a $n \times n$ two dimensional image. The vector x^{ex} of dimension $N = n^2$ obtained by stacking the columns of the image X and the associated blurred and noise-free image b^{ex} is generated by multiplying x^{ex} by a blurring matrix $A \in \mathbb{R}^{N \times N}$. The matrix A is block Toeplitz with Toeplitz blocks and is implemented in the function `blur` from [15], which has two parameters, `band` and `sigma`; the former specifies the half-bandwidth of the Toeplitz blocks, and the latter the variance of the Gaussian point spread function. We generate a blurred and noisy image $b \in \mathbb{R}^N$ by adding a noise-vector $e \in \mathbb{R}^N$, so that $b = Ax^{ex} + e$. We assume the blurring operator A and the corrupted image b to be available while no information is given on the error e .

In the example, the original image is the `cameraman.tif` test image from Matlab, a 256×256 , 8-bit gray-scale image, commonly used in image deblurring experiments. The image is blurred with parameters `band = 7` and `sigma = 2`. We further corrupt the blurred images with 0.1% additive Gaussian noise. The blurred and noisy image is shown in the center column of Figure 8, the regularization operator is defined as

$$L = I_n \otimes L_1 + L_1 \otimes I_n \in \mathbb{R}^{N \times N}, \quad (31)$$

(cf. [23, §5]). The restored image is shown in the right column of Figure 8. The result has been obtained in $m = 8$ iterations of the Arnoldi algorithm; the CPU-time required for this experiment is around 1.2 s. Many other experiments on image restoration have shown similar performances.



Figure 8. Restoration of `cameraman.tif`. From left to right: original image; blurred and noisy image with blur parameters `band = 7`, `sigma = 2`, and noise level $\varepsilon = 10^{-3}$; and restored image. From top to bottom: original size image and two zooms.

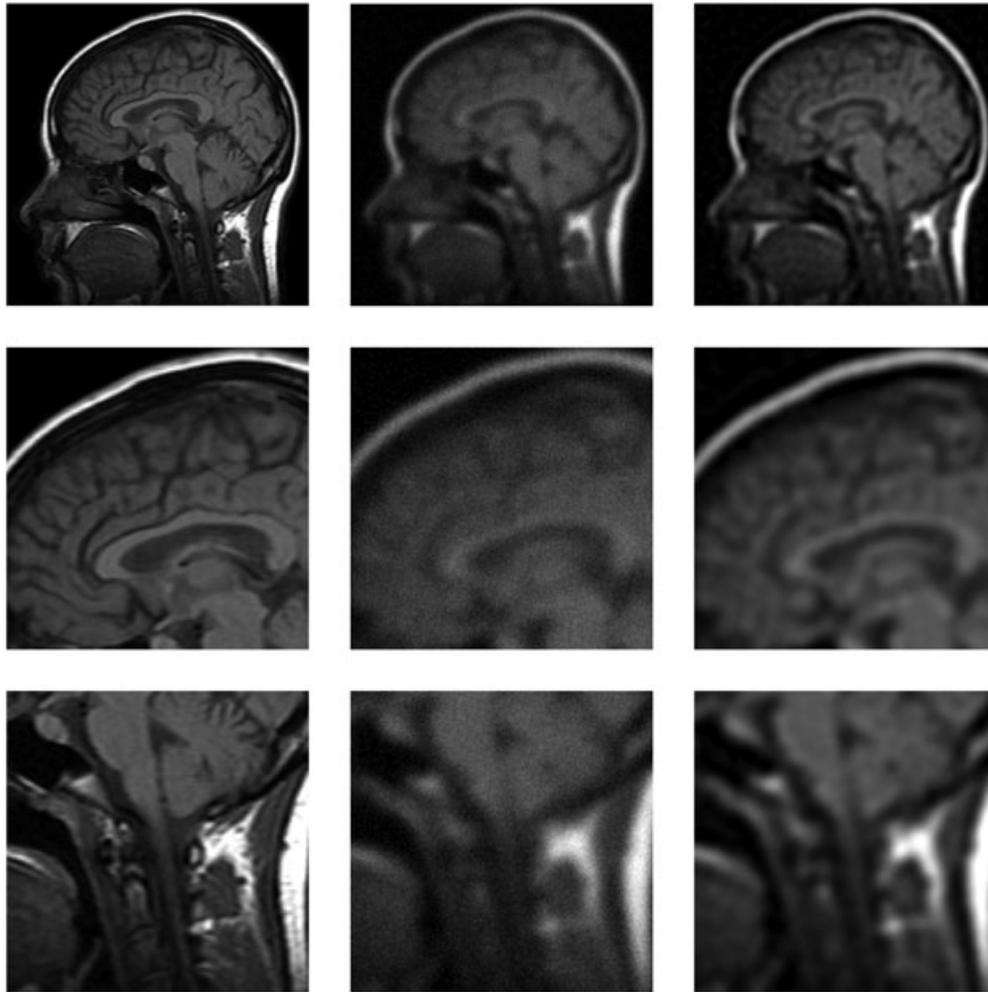


Figure 9. Restoration of the test image `mri.png`. From left to right: original image; blurred and noisy image with noise level $\varepsilon = 10^{-1}$ and blur parameters `band = 9`, `sigma = 2.5`; restored image. From top to bottom: original size image and two zooms.

5.3. Results for MRI reconstruction

The treatment of different kinds of medical images such as MRI, computed tomography, positron emission tomography often requires the usage of image processing techniques to remove various types of degradations such as noise, blur, and contrast imperfections. Our experiments focus on MRI medical images affected by Gaussian blur and noise. Typically, when blur and noise affect the MRI images, the visibility of small components in the image decreases, and therefore, image deblurring techniques are extensively employed to grant the image a sharper appearance.

In our test, we blur a synthetic MRI 256×256 image, with Gaussian blur (`band = 9`, `sigma = 2.5`), and we add 10% Gaussian white noise, because the noise level of a real problem may be expected to be quite high.

Figure 9 displays the performance of the algorithm. On the left column, we show the blur-free and noise-free image; on the middle column, we show the corrupted image; on the right column, we show the restored image. The regularization operator employed is again (31). The result has been obtained in $m = 5$ iterations of the algorithm, in around 0.7 s.

6. CONCLUSIONS

In this paper, we have proposed a very simple method to define the sequence of regularization parameters for the AT method, in absence of information on the percentage of error that affects the

right-hand side vector. The numerical results have shown that this technique is rather stable and the results are comparable with other existing approaches (e.g., GCV and L-curve). To describe this procedure, we have used the term ‘embedded’ because the construction of the Krylov subspaces is used, at the same time, both as an error estimator by means of the GMRES residual and for solving (4) by the AT method. We remark that, in principle, the idea can be extended to any basic iterative method that is able to simultaneously approximate $\|e\|$ and the Tikhonov regularized solution. Therefore, the proposed approach could be potentially employed in connection with Lanczos bidiagonalization and the range-restricted Arnoldi algorithm [5, 24, 25]: these extensions deserve a further investigation. In particular, adopting methods based on the range-restricted Arnoldi algorithm could improve the performance of the method here presented, especially for problems whose unknown solution is expected to be smooth. Of course, in the range-restricted framework, the analysis of Section 4 is no longer valid and should be modified accordingly. However, some preliminary experiments have shown that the embedded approach can be fruitfully coupled with the range-restricted approach, and this topic is currently under study.

ACKNOWLEDGEMENTS

The authors are grateful to the reviewers for the valuable comments. This work was partially supported by the University of Padova, Project 2010 No. CPDA104492.

REFERENCES

1. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion*. SIAM: Philadelphia (PA), 1998.
2. Bauer F, Lukas M. Comparing parameter choice methods for regularization of ill-posed problems. *Mathematics and Computers in Simulation* 2011; **81**:1795–1841.
3. Reichel L, Rodriguez G. Old and new parameter choice rules for discrete ill-posed problems. *Numerical Algorithms* 2013; **63**:65–87.
4. Calvetti D, Morigi S, Reichel L, Sgallari F. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics* 2000; **123**:423–446.
5. Lewis B, Reichel L. Arnoldi-Tikhonov regularization methods. *Journal of Computational and Applied Mathematics* 2009; **226**:92–102.
6. Reichel L, Shyshkov A. A new zero-finder for Tikhonov regularization. *BIT Numerical Mathematics* 2008; **48**:627–643.
7. Gazzola S, Novati P. Automatic parameter setting for Arnoldi-Tikhonov methods. *Journal of Computational and Applied Mathematics* 2014; **256**:180–195.
8. O’Leary DP, Simmons JA. A bidiagonalization-regularization procedure for large-scale discretizations of ill-posed problems. *SIAM Journal on Scientific and Statistical Computing* 1981; **2**:474–489.
9. Hochstenbach M, Reichel L. An iterative method for Tikhonov regularization with a general linear regularization operator. *Journal of Integral Equations and Applications* 2010; **22**:463–480.
10. Novati P, Russo MR. Adaptive Arnoldi-Tikhonov regularization for image restoration. *Numerical Algorithms*. DOI:10.1007/s11075-013-9712-0, in press.
11. Saad Y. *Iterative Methods for Sparse Linear Systems* ((2nd edn)). SIAM: Philadelphia (PA), 2003.
12. Novati P, Russo MR. A GCV based Arnoldi-Tikhonov regularization method. *BIT Numerical Mathematics*. DOI:10.1007/s10543-013-0447-z, in press.
13. Chung J, Nagy JG, O’Leary DP. A weighted-GCV method for Lanczos-hybrid regularization. *Electronic Transaction on Numerical Analysis* 2008; **28**:149–167.
14. Kilmer ME, O’Leary DP. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on Matrix Analysis and Applications* 2001; **22**:1204–1221.
15. Hansen PC. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms* 1994; **6**:1–35.
16. Moret I. A note on the superlinear convergence of GMRES. *SIAM Journal on Numerical Analysis* 1997; **34**:513–516.
17. Hofmann B. *Regularization for Applied Inverse and Ill-Posed Problems*. Teubner: Stuttgart (Germany), 1986.
18. Hansen PC. The discrete Picard condition for discrete ill-posed problems. *BIT Numerical Mathematics* 1990; **30**:658–672.
19. Meurant G. On the residual norm in FOM and GMRES. *SIAM Journal on Matrix Analysis and Applications* 2011; **32**:394–411.
20. Hansen PC. *Discrete Inverse Problems: Insight and Algorithms*. SIAM: Philadelphia (PA), 2010.
21. Calvetti D, Lewis B, Reichel L. On the regularizing properties of the GMRES method. *Numerische Mathematik* 2002; **91**:605–625.

22. O'Leary DP. Near-optimal parameters for Tikhonov and other regularization methods. *SIAM Journal on Scientific Computing* 2001; **23**:1161–1171.
23. Kilmer ME, Hansen PC, Español MI. A projection-based approach to general-form Tikhonov regularization. *SIAM Journal on Scientific Computing* 2007; **29**:315–330.
24. Hansen PC, Jensen TK. Noise propagation in regularizing iterations for image deblurring. *Electronic Transaction on Numerical Analysis* 2008; **31**:204–220.
25. Hochstenbach M, McNinch N, Reichel L. Discrete ill-posed least-squares problems with a solution norm constraint. *Linear Algebra and its Applications* 2012; **436**:3801–3818.