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

A GCV based Arnoldi-Tikhonov regularization method

Paolo Novati · **Maria Rosaria Russo**

Received: date / Accepted: date

Abstract For the solution of linear discrete ill-posed problems, in this paper we consider the Arnoldi-Tikhonov method coupled with the Generalized Cross Validation for the computation of the regularization parameter at each iteration. We study the convergence behavior of the Arnoldi method and its properties for the approximation of the (generalized) singular values, under the hypothesis that Picard condition is satisfied. Numerical experiments on classical test problems and on image restoration are presented.

Mathematics Subject Classification (2000) 65F10 · 65F22 · 65R32

Keywords Linear discrete ill-posed problem · Tikhonov regularization · Arnoldi algorithm · Generalized Cross Validation

1 Introduction

In this paper we consider discrete ill-posed problems,

$$Ax = b, \quad A \in \mathbb{R}^{N\times N},\ b \in \mathbb{R}^{N}, \tag{1.1}$$

in which the right-hand side b is assumed to be affected by noise, caused by measurement or discretization errors. These systems typically arise from the discretization of linear ill-posed problem, such as Fredholm integral equations of the first kind with compact kernel (see e.g. [15, Chapter 1] for a background). A common property of

P. Novati, M.R. Russo
Dipartimento di Matematica, Università di Padova
Via Trieste 63, Padova, Italy
Tel.: +39-049-8271424
E-mail: novati@math.unipd.it

M.R. Russo
E-mail: mrrusso@math.unipd.it

these kind of problems, is that the singular values of the kernel rapidly decay and cluster near zero. In this situation, provided that the discretization which leads to (1.1) is consistent with the continuous problem, this property is inherited by the matrix A.

Because of the ill conditioning of A and the presence of noise in b, some sort of regularization is generally employed for solving this kind of problems. In this framework, a popular and well established regularization technique is the Tikhonov method, which consists in solving the minimization problem

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

where $\lambda > 0$ is the regularization parameter and $L \in \mathbb{R}^{P\times N}$ is the regularization matrix (see e.g. [14] and [15] for a background). We denote the solution of (1.2) by x_λ. For a discussion about the choice of L we may quote here the recent work [5] and the references therein. As well known, the choice of the parameter λ is crucial in this setting, since it defines the amount of regularization one wants to impose. Many techniques have been developed to determine a suitable value for the regularizing parameter and we can refer to the recent papers [31,2,11,22] for the state of the art, comparison and discussions. We remark that in (1.2) and throughout the paper, the norm used is always the Euclidean norm.

Assuming that $b = \overline{b} + e$, where $\overline{b}$ represents the unknown error-free right-hand side, in this paper we assume that no information is available on the error e. In such a situation, the most popular and established techniques for the definition of λ in (1.2), as for instance the L-curve criterion and the Generalized Cross Validation (GCV), typically requires the computation of the Generalized Singular Value Decomposition (GSVD) of the matrix pair (A,L). Of course this decomposition may represent a serious computational drawback for large-scale problems, such as the image deblurring. In order to overcome this problem, Krylov projection methods such as the ones based on the Lanczos bidiagonalization [1,12,20,21] and the Arnoldi algorithm [3,24] are generally used. Pure iterative methods such as the GMRES or the LSQR, eventually implemented in a hybrid fashion ([15, § 6.6]) can also be considered in this framework.

In this paper we analyze the Arnoldi method for the solution of (1.2) (the so called Arnoldi-Tikhonov method, introduced in [3]), coupled with the GCV as parameter choice rule. Similarly to what made in [4] for the Lanczos bidiagonalization process, we show that the resulting algorithm can be fruitfully used for large-scale regularization. Being based on the orthogonal projection of the matrix A onto the Krylov subspaces $\mathscr{K}_m(A,b) = \text{span}\{b, Ab, \ldots, A^{m-1}b\}$, we shall observe that for discrete ill-posed problems, the Arnoldi algorithm is particularly efficient for the approximation of the GCV curve, after a very few number of iterations.

Indeed, under the hypothesis that Picard condition is satisfied [13], we provide some theoretical results about the convergence of the Arnoldi-Tikhonov methods and its properties for the approximation of the singular values of A. These properties allow us to consider approximation of the GCV curve which can be obtained working in small dimension (similarly to what made in [3] where a "projected" L-curve criterion is used). The GCV curve approximation leads to the definition of a sequence of regularization parameters (one for each step of the algorithm), which are fairly

good approximation of the regularization parameter arising from the exact SVD (or GSVD).

The paper is organized as follows. In Section 2 we present a brief outline about the Arnoldi-Tikhonov method for the iterative solution of (1.2). In Section 3 and 4 we provide some theoretical results concerning the convergence of the Arnoldi algorithm and the SVD (GSVD) approximation. In Section 5 we explain the use of the AT method with the GCV criterion. Some numerical experiments are presented in Section 6 and 7.

2 The Arnoldi-Tikhonov method

Denoting by $\mathscr{K}_m(A,b) = \text{span}\{b, Ab, \ldots, A^{m-1}b\}$ the Krylov subspaces generated by A and the vector b, the Arnoldi algorithm (see e.g. [32] for a background) computes an orthonormal basis $\{w_1, ..., w_m\}$ of $\mathscr{K}_m(A,b)$. Setting $W_m = [w_1, ..., w_m] \in \mathbb{R}^{N\times m}$, the algorithm can be written in matrix form as

$$AW_m = W_m H_m + h_{m+1,m} w_{m+1} e_m^T, \tag{2.1}$$

where $H_m = (h_{i,j}) \in \mathbb{R}^{m\times m}$ is an upper Hessenberg matrix which represents the orthogonal projection of A onto $\mathscr{K}_m(A,b)$, and $e_m = (0, ..., 0, 1)^T \in \mathbb{R}^m$. Equivalently, the relation (2.1) can be written as

$$AW_m = W_{m+1}\overline{H}_m, \tag{2.2}$$

where

$$\overline{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}. \tag{2.3}$$

In exact arithmetics the Arnoldi process terminates whenever $h_{m+1,m} = 0$, which means that $\mathscr{K}_{m+1}(A,b) = \mathscr{K}_m(A,b)$.

If we consider the constrained minimization

$$\min_{x\in\mathscr{K}_m(A,b)} \left\{ \|Ax-b\|^2 + \lambda^2 \|Lx\|^2 \right\}, \tag{2.4}$$

writing $x = W_m y_m$, $y_m \in \mathbb{R}^m$, and using (2.2), we obtain

$$\min_{y_m\in\mathbb{R}^m} \left\{ \left\| \overline{H}_m y_m - \|b\| e_1 \right\|^2 + \lambda^2 \|LW_m y_m\|^2 \right\}, \tag{2.5}$$

which is known as the Arnoldi-Tikhonov (AT) method. Dealing with Krylov type solvers, one generally hopes that a good approximation of the exact solution can be achieved for $m \ll N$, which, in other words, means that the spectral properties of the matrix A are rapidly simulated by the ones of $\overline{H}_m$. This method has been introduced in [3] in the case of $L = I_N$ (where I_N is the identity matrix of order N, so that $\|LW_m y_m\| = \|y_m\|$) and then used in [7, 28] with $L \neq I_N$, with the basic aim of reducing the dimension of the original problem and to avoid the matrix-vector multiplication with A^T used by Lanczos type schemes (see [1, 12] and the references therein).

It is worth noting that (2.5) can also be interpreted as an hybrid method. Indeed, the minimization (2.5) with $L = I_N$ is equivalent to the inner regularization of the

GMRES [21]. We remark however, that for $L \neq I_N$, the philosophy is completely different, since (2.5) represents the projection of a regularization, while the hybrid approach aims to regularize the projected problem. As we shall see, this difference can be appreciated more clearly whenever a parameter choice rule for λ is adopted.

As well known, in many applications the use of a suitable regularization operator $L \neq I_N$, may substantially improve the quality of the approximate solution with respect to the choice of $L = I_N$. Anyway, we need to observe that with a general $L \in \mathbb{R}^{P\times N}$, the minimization (2.5) is equivalent to

$$\min_{y_m\in\mathbb{R}^m} \left\| \begin{pmatrix} \overline{H}_m \\ \lambda L W_m \end{pmatrix} y_m - \begin{pmatrix} \|b\| e_1 \\ 0 \end{pmatrix} \right\|^2, \tag{2.6}$$

so that, for $P \approx N$, the dimension of (2.6) inherits the dimension of the original problem. Computationally, the situation can be efficiently faced by means of the "skinny" QR factorization. Anyway, assuming that $P \leq N$, in order to work with reduced dimension problems, we add $N-P$ zero rows to L (which does not alter (2.4)) and consider the orthogonal projection of L onto $\mathscr{K}_m(A,b)$, that is,

$$L_m := W_m^T L W_m \in \mathbb{R}^{m\times m}. \tag{2.7}$$

This modification leads to the reduced minimization

$$\begin{aligned} &\min_{y_m\in\mathbb{R}^m} \left\{ \left\| \overline{H}_m y_m - \|b\| e_1 \right\|^2 + \lambda^2 \|L_m y_m\|^2 \right\} \\ &= \min_{x\in\mathscr{K}_m(A,b)} \left\{ \|Ax-b\|^2 + \lambda^2 \|W_m^T L x\|^2 \right\}, \end{aligned} \tag{2.8}$$

which is not equivalent to (2.4) anymore. Anyway, the use of L_m appears natural in this framework, and it is also justified by the fact that

$$\left\| W_m^T L x \right\| \leq \|Lx\|,$$

since $\left\| W_m^T L x \right\| = \left\| W_m W_m^T L x \right\|$ and $\left\| W_m W_m^T \right\| = 1$, being $W_m W_m^T$ an orthogonal projection. We observe moreover that L_m would be the regularization operator of the projection of a Franklin type regularization [6]

$$(A + \lambda L) x = b.$$

In order to reduce completely the dimension of (2.6), instead of considering the projection (2.7) one may even consider the QR factorization $LW_m = Q_m R_m$ as in [17]. In terms of convergence rate and accuracy, to our experience these approaches perform about the same.

3 Convergence analysis for discrete ill-posed problems

In what follows we denote by $A = U\Sigma V^T \in \mathbb{R}^{N\times N}$ the SVD of A where $\Sigma = diag(\sigma_1, ..., \sigma_N)$, and by $A_m := U_m \Sigma_m V_m^T$ the truncated SVD. We remember that the matrix $\Delta_m := A - A_m$ is such that $\|\Delta_m\| = \sigma_{m+1}$.

An important property of the methods based on orthogonal projections such as the Arnoldi algorithm, is the fast theoretical convergence ($h_{m+1,m} \to 0$) if the matrix A comes from the discretization of operators whose spectrum is clustered around zero. Denote by λ_j, $j \geq 1$ the eigenvalues of A and assume that $|\lambda_j| \geq |\lambda_{j+1}|$ for $j \geq 1$. We have the following result (cf. [27, Theorem 5.8.10]), in which we assume N arbitrarily large.

Theorem 3.1 *Assume that* $1 \notin \sigma(A)$ *and*

$$\sum_{j \geq 1} \sigma_j^p < \infty \text{ for a certain } 0 < p \leq 1. \tag{3.1}$$

Let $p_m(z) = \prod_{i=1}^m (z - \lambda_i)$. *Then*

$$\|p_m(A)\| \leq \left(\frac{\eta e}{m}\right)^{m/p}, \tag{3.2}$$

where

$$\eta(p) \leq (1+p) \sum_{j \geq 1} \sigma_j^p. \tag{3.3}$$

Since

$$\prod_{i=1}^m h_{i+1,i} \leq \|p_m(A)b\|, \tag{3.4}$$

for each monic polynomial p_m of exact degree m (see [36, p. 269]), Theorem 3.1 reveals that the rate of decay of $\prod_{i=1}^m h_{i+1,i}$ is superlinear and depends on the p-summability of the singular values of A. We remark that the superlinear convergence of certain Krylov subspace methods when applied to linear equations involving compact operators is known in literature (see e.g. [26] and the references therein). The rate of convergence depends on the degree of compactness of the operator, which can be measured in terms of the decay of the singular values.

Here, dealing with severely ill-posed problems, the typical situation is $\sigma_j = O(e^{-\alpha j})$, where $\alpha > 0$ handles the degree of ill-conditioning [19, Definition 2.42]. In this situation, the following result expresses more clearly the fast decay of $h_{i+1,i}$ with respect to the value of α.

Proposition 3.1 *Let* $\sigma_j = O(e^{-\alpha j})$. *Then, for* $m \to \infty$,

$$\left(\prod_{i=1}^m h_{i+1,i}\right)^{1/m} \leq k e^{-\frac{m\alpha}{e^2} + \frac{\alpha+2}{2} + O\left(\frac{1}{m}\right)}, \tag{3.5}$$

where k *is a constant independent of* m.

Proof Let k be a constant such that $\sigma_j \leq k e^{-\alpha j}$ for each j. Then for $p > 0$

$$\eta(p) \leq (1+p) \sum_{j \geq 1} \sigma_j^p \leq k^p \frac{(1+p)}{1 - e^{-\alpha p}}, \tag{3.6}$$

(cf. (3.3)). Now consider the approximation

$$k^p \frac{(1+p)}{1 - e^{-\alpha p}} \approx \frac{1}{\alpha p} =: \widetilde{\eta}(p),$$

which is fairly accurate for $p \approx 0$. Using this approximation in (3.2), we find that the minimum of

$$\left(\frac{\widetilde{\eta}(p)e}{m}\right)^{m/p},$$

is attained for $p^* = \frac{e^2}{m\alpha}$. Using this value, the bound (3.6), and defining $t := \frac{e^2}{m}$, we obtain

$$\begin{aligned}
\left(\frac{\eta(p^*)e}{m}\right)^{m/p^*} &\leq k^m \left(\frac{(1+p^*)}{1-e^{-\alpha p^*}}\frac{e}{m}\right)^{m/p^*} \\
&= k^m \exp\left(\frac{m\alpha}{t}\ln\left(\frac{1+\frac{t}{\alpha}}{1-e^{-t}}\frac{t}{e}\right)\right) \\
&= k^m \exp\left(\frac{m\alpha}{t}\left(-1+t\left(\frac{1}{\alpha}+\frac{1}{2}\right)+O(t^2)\right)\right) \quad \text{for } t \to 0 \\
&= k^m \exp\left(-\frac{m^2\alpha}{e^2}+m\left(\frac{\alpha+2}{2}\right)+O(1)\right) \quad \text{for } m \to \infty.
\end{aligned}$$

The result immediately follows from (3.4) and (3.2).

In Figure 3.1 (a)-(b) we experimentally test the bound (3.5) working with test problems SHAW and WING, taken from Hansen's Regularization Toolbox [16]. For these two problems it is known that $\alpha = 2$ and $\alpha = 4.5$ respectively.

In the following results we assume to work with problems in which the discrete Picard condition (see [13]) is satisfied , that is, $u_m^T b = O(\sigma_m)$, where u_m denotes the m-th column of U, and b is assumed to be the exact right-hand side.

Proposition 3.2 *Assume that the singular values of A are of the type $\sigma_j = O(e^{-\alpha j})$. Assume moreover that the discrete Picard condition is satisfied. Let $\widetilde{V}_m := [\widetilde{v}_0, ..., \widetilde{v}_{m-1}] \in \mathbb{R}^{N\times m}$ where $\widetilde{v}_k := A^k b / \left\|A^k b\right\|$. If $\widetilde{V}_m$ has full column rank, then there exists $C_m \in \mathbb{R}^{m\times m}$ nonsingular, $E_m, F_m \in \mathbb{R}^{N\times m}$, such that*

$$\widetilde{V}_m = U_m C_m + E_m, \quad \|E_m\| = O(m^{1/2}\sigma_m), \tag{3.7}$$

$$U_m = \widetilde{V}_m C_m^{-1} + F_m, \quad \|F_m \Sigma_m\| = O(m^{3/2}\sigma_m). \tag{3.8}$$

Proof Let $U_m^{\perp} := [u_{m+1}, ..., u_N] \in \mathbb{R}^{N\times(N-m)}$. Defining $C_m := U_m^T \widetilde{V}_m \in \mathbb{R}^{m\times m}$ and $E_m := U_m^{\perp}\left(U_m^{\perp}\right)^T \widetilde{V}_m \in \mathbb{R}^{N\times m}$ we have $\widetilde{V}_m = U_m C_m + E_m$. Now we observe that for $0 \leq k \leq m-1$

$$\left|u_j^T \widetilde{v}_k\right| \sim \sigma_j. \tag{3.9}$$

For $k = 0$ the above relation is ensured by the Picard Condition, whereas for $k \geq 1$ it holds since

$$\widetilde{v}_k = \frac{\left\|A^{k-1}b\right\|}{\|A^k b\|} A\widetilde{v}_{k-1}.$$

Therefore, using $\sigma_j = O(e^{-\alpha j})$, we immediately obtain

$$\|E_m\| = \left\|\left(U_m^{\perp}\right)^T \widetilde{V}_m\right\| = O(m^{1/2}\sigma_m), \tag{3.10}$$

We observe that the matrix C_m can be written as

$$C_m = U_m^T W_m S_m,$$

where S_m is upper triangular and nonsingular if $\widetilde{V}_m$ has full rank. Now, from the relation [10, §2.6.3]

$$\sigma_{\min}(U_m^T W_m)^2 = 1 - \left\| \left(U_m^{\perp}\right)^T W_m \right\|^2,$$

the quantity $\left\| \left(U_m^{\perp}\right)^T W_m \right\|$, which express the distance between $R(U_m)$ and $R(W_m)$, is strictly less than one if the Picard Condition is satisfied. Thus, by (3.7), we can write

$$U_m = \widetilde{V}_m C_m^{-1} - E_m C_m^{-1}, \tag{3.11}$$

and since $E_m = U_m^{\perp}\left(U_m^{\perp}\right)^T \widetilde{V}_m$ we have that

$$E_m C_m^{-1} = U_m^{\perp}\left(U_m^{\perp}\right)^T \widetilde{V}_m \left(U_m^T \widetilde{V}_m\right)^{-1}. \tag{3.12}$$

By (3.9), using the Cramer rule to compute $\left(U_m^T \widetilde{V}_m\right)^{-1} \Sigma_m \in \mathbb{R}^{m\times m}$ we can see that each element of this matrix is of the type $O(1)$, so that

$$\left| \left(U_m^{\perp}\right)^T \widetilde{V}_m \left(U_m^T \widetilde{V}_m\right)^{-1} \Sigma_m \right| \sim m \begin{pmatrix} \sigma_{m+1} & \cdots & \sigma_{m+1} \\ \vdots & & \vdots \\ \sigma_N & \cdots & \sigma_N \end{pmatrix} \in \mathbb{R}^{(N-m)\times m},$$

and hence

$$\left\| \left(U_m^{\perp}\right)^T \widetilde{V}_m \left(U_m^T \widetilde{V}_m\right)^{-1} \Sigma_m \right\| = O(m^{3/2}\sigma_m), \tag{3.13}$$

using again $\sigma_j = O(e^{-\alpha j})$. Defining $F_m = -E_m C_m^{-1}$ we obtain (3.8) by (3.11), (3.12) and (3.13).

Remark 3.1 The hypothesis $\sigma_j = O(e^{-\alpha j})$ of Proposition 3.2 is just used to have $\left\|\varepsilon^{(0)}\right\| = O(\sigma_m)$ by

$$\left(\sum_{j\geq m+1} e^{-2\alpha j} \right)^{1/2} \leq \frac{1}{\sqrt{2\alpha}} e^{-\alpha m}. \tag{3.14}$$

The result of the proposition can be extended to work with moderately ill-posed problems, in which $\sigma_j = O(j^{-\alpha})$, provided that α is large enough. As consequence in this situation we would have a slower decay of $\|E_m\|$ and $\|F_m \Sigma_m\|$.

The following result improves the one of Proposition 3.1 (which holds without hypothesis on b).

Proposition 3.3 *Under the hypothesis of Proposition 3.2*

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

Proof By (2.1)

$$\begin{aligned} h_{m+1,m} &= w_{m+1}^T A w_m \\ &= w_{m+1}^T \Delta_m w_m + w_{m+1}^T A_m w_m \\ &= O(\sigma_{m+1}) + w_{m+1}^T U_m \Sigma_m V_m^T w_m, \end{aligned}$$

since $\|\Delta_m\| = \sigma_{m+1}$. Therefore, using (3.8) we obtain

$$h_{m+1,m} = O(\sigma_{m+1}) + w_{m+1}^T (\widetilde{V}_m C_m^{-1} + F_m) \Sigma_m V_m^T w_m.$$

which concludes the proof, since $w_{m+1}^T \widetilde{V}_m = 0$ and $\|F_m \Sigma_m\| = O(m^{3/2}\sigma_m)$.

In Figure 3.1 (c)-(d) we compare the decay of the sequence $\{h_{m+1,m}\}_{m\geq 1}$ with that of the singular values, working again with the test problems SHAW and WING.

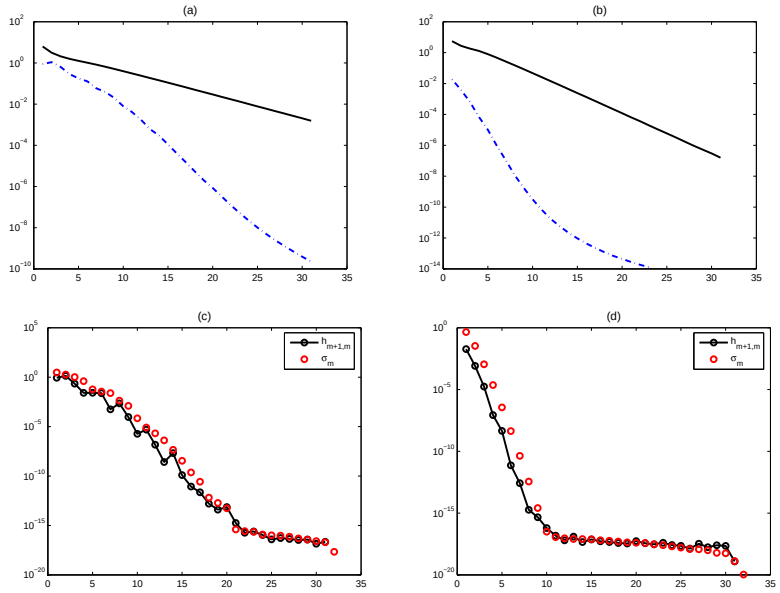

Fig. 3.1 *(a)-(b) decay rate of* $(\prod_{i=1}^{m} h_{i+1,i})^{1/m}$ *(dash-dot line) and bound (3.5) (solid line), (c)-(d) decay of* $h_{m+1,m}$ *and* σ_m*. On the left the results for SHAW and on the right the results for WING. In each experiment* $N = 32$.

We need to remark that the results of Figure 3.1 are obtained working with the Householder implementation of the Arnoldi algorithm and hence simulating what happens in exact arithmetics.

4 The approximation of the SVD

The use of the Arnoldi algorithm as a method to approximate the marginal values of the spectrum of a matrix is widely known in literature. We may refer to [33, Chapter

6] for an exhaustive background. Using similar arguments, in this section we analyze the convergence of the singular values of the matrices $\overline{H}_m$ to the largest singular values of A. For the Lanczos bidiagonalization method [1,30], the analysis can be done by exploiting the connection between this method and the symmetric Lanczos process (see e.g. [8]). The use of the Lanczos bidiagonalization to construct iteratively the GSVD of (A,L) has been studied in [20].

Let us consider the SVD factorization of $\overline{H}_m$, that is, $\overline{H}_m = U^{(m)}\Sigma^{(m)}V^{(m)T}$, $U^{(m)} \in \mathbb{R}^{(m+1)\times(m+1)}$, $V^{(m)} \in \mathbb{R}^{m\times m}$ and

$$\Sigma^{(m)} = \begin{pmatrix} \sigma_1^{(m)} & & \\ & \ddots & \\ & & \sigma_m^{(m)} \\ 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{(m+1)\times m}.$$

We can state the following results.

Proposition 4.1 *Let $\overline{U}_{m+1} = W_{m+1}U^{(m)} \in \mathbb{R}^{N\times(m+1)}$ and $\overline{V}_m = W_m V^{(m)} \in \mathbb{R}^{N\times m}$. Then*

$$\left\| A - \overline{U}_{m+1}\Sigma^{(m)}\overline{V}_m^T \right\| = \left\| A(I - W_m W_m^T) \right\|.$$

Proof Using (2.2), we have

$$\begin{aligned} A - \overline{U}_{m+1}\Sigma^{(m)}\overline{V}_m^T &= A - W_{m+1}U^{(m)}\Sigma^{(m)}V^{(m)T}W_m^T \\ &= A - W_{m+1}\overline{H}_m W_m^T \\ &= A - AW_m W_m^T. \end{aligned}$$

Observe that since $\overline{U}_{m+1}\Sigma^{(m)} = W_{m+1}\widetilde{U}^{(m)}\widetilde{\Sigma}^{(m)}$, where $\widetilde{\Sigma}^{(m)} \in \mathbb{R}^{m\times m}$ is just $\Sigma^{(m)}$ without the last row, and $\widetilde{U}^{(m)} \in \mathbb{R}^{(m+1)\times m}$ is $U^{(m)}$ without the last column, the above result states that the triplet $\left(W_{m+1}\widetilde{U}^{(m)}, \widetilde{\Sigma}^{(m)}, W_m V^{(m)}\right)$ defines an approximation of the truncated SVD of A, which cannot be too bad since $\left\|A(I - W_m W_m^T)\right\| \le \|A\|$. Moreover, it states that if the Arnoldi algorithm does not terminate before N iterations, then it produces the complete SVD. The following result gives some additional information.

Proposition 4.2 *Let $u_k^{(m)} \in \mathbb{R}^{m+1}$ and $v_k^{(m)} \in \mathbb{R}^m$ be respectively the right and left singular vectors relative to the singular value $\sigma_k^{(m)}$ of $\overline{H}_m$, that is, $\overline{H}_m v_k^{(m)} = \sigma_k^{(m)} u_k^{(m)}$ and $\overline{H}_m^T u_k^{(m)} = \sigma_k^{(m)} v_k^{(m)}$, with $1 \le k \le m$. Then defining $\overline{u}_k = W_{m+1}u_k^{(m)}$ and $\overline{v}_k = W_m v_k^{(m)}$ we have that*

$$A\overline{v}_k - \sigma_k^{(m)}\overline{u}_k = 0, \tag{4.1}$$

$$W_m^T(A^T\overline{u}_k - \sigma_k^{(m)}\overline{v}_k) = 0. \tag{4.2}$$

Proof (4.1) follows directly by (2.2). Moreover, since

$$\overline{H}_m^T u_k^{(m)} - \sigma_k^{(m)} v_k^{(m)} = 0,$$

using $\overline{H}_m^T = W_m^T A^T W_{m+1}$, and the definition of $\overline{u}_k$ and $\overline{v}_k$, we easily obtain (4.2).

Remark 4.1 Using the square matrix H_m to approximate the singular values of A, that is, computing the SVD $H_m = U^{(m)}\Sigma^{(m)}V^{(m)T}$, where now $U^{(m)}, \Sigma^{(m)}, V^{(m)} \in \mathbb{R}^{m\times m}$, if $H_m v_k^{(m)} = \sigma_k^{(m)} u_k^{(m)}$ then

$$\left\| A\overline{v}_k - \sigma_k^{(m)}\overline{u}_k \right\| \leq h_{m+1,m} \quad \text{with} \quad \overline{u}_k = W_m u_k^{(m)}, \overline{v}_k = W_m v_k^{(m)}. \tag{4.3}$$

The above relation is very similar to the one which arises when using the eigenvalues of H_m (the Ritz values) to approximate the eigenvalues of A [33, §6.2]. Note moreover that whenever $h_{m+1,m} \approx 0$, and hence very quickly for linear ill-posed problems (see Section 3), the use of $\overline{H}_m$ or H_m is almost equivalent to approximate the largest singular values of A.

The Galerkin condition (4.2) is consequence of the fact that the Arnoldi algorithm does not work with the transpose. Obviously, if $A = A^T$, the algorithm reduces to the symmetric Lanczos process and, under the hypothesis of Proposition 4.2, we easily obtain $A^T\overline{u}_k - \sigma_k^{(m)}\overline{v}_k = 0$. In the general case of $A \neq A^T$, Proposition 4.2 ensures that since $\overline{v}_k = W_m v_k^{(m)} \in \mathscr{K}_m(A,b)$, by (4.2) the vector $\sigma_k^{(m)}\overline{v}_k$ is just the orthogonal projection of $A^T\overline{u}_k$ onto $\mathscr{K}_m(A,b)$, that is, $\sigma_k^{(m)}\overline{v}_k = W_m W_m^T A^T \overline{u}_k$, which implies

$$\left\| A^T\overline{u}_k - \sigma_k^{(m)}\overline{v}_k \right\| \leq \left\| (I - W_m W_m^T) A^T W_m W_m^T \right\|. \tag{4.4}$$

This means that the approximation is good if $A^T\overline{u}_k$ is close to $\mathscr{K}_m(A,b)$. It is interesting to observe that (4.4) is just the "transpose version" of (4.3) since

$$h_{m+1,m} = \left\| (I - W_m W_m^T) A W_m W_m^T \right\|,$$

which can be easily proved using again (2.1) (cf. [33, Chapter 4]).

Experimentally, one observes that the Arnoldi algorithm seems to be very efficient for approximating the largest singular values for discrete ill-posed problems. In order to have a-posteriori strategy to monitor step-by-step the quality of approximation, we can state the following.

Proposition 4.3 *Assume that the matrix A has full rank. Then*

$$\left\| A^T\overline{u}_k - \sigma_k^{(m)}\overline{v}_k \right\| \leq \left\| W_{m+1}^T A W_m^{\perp} \right\|, \tag{4.5}$$

where $\overline{u}_k$, $\overline{v}_k$, $\sigma_k^{(m)}$ *are defined as in Proposition 4.2, and* $W_m^{\perp} = [w_{m+1}, \ldots, w_N]$.

Proof Since $\overline{v}_k \in \mathscr{K}_m(A,b)$, and $\overline{u}_k = W_{m+1} u_k^{(m)}$, by (4.2)

$$\left\| A^T\overline{u}_k - \sigma_k^{(m)}\overline{v}_k \right\| \leq \left\| \left(W_m^{\perp}\right)^T A^T W_{m+1} \right\|. \tag{4.6}$$

Formula (4.5) is rather interesting because since $h_{ij} = w_i^T A w_j$ from the Arnoldi algorithm,

$$W_{m+1}^T A W_m^\perp = \begin{bmatrix} h_{1,m+1} & \cdots & h_{1,N} \\ \vdots & & \vdots \\ h_{m+1,m+1} & \cdots & h_{m+1,N} \end{bmatrix}.$$

Since in many cases the elements of the projected matrix H_m tend to annihilate departing from the diagonal (this is the basic assumption of the methods based on the incomplete orthogonalization, see e.g. [34]), one may obtain useful estimates for the bound (4.5) working with few columns of $W_{m+1}^T A W_m^\perp$, that is, with few columns of $W_m^\perp$, and hence obtaining a-posteriori estimates for the quality of the SVD approximation. In order to have an experimental confirmation of this statement, in Figure 4.1 we show the behavior of $\left\| A - \overline{U}_{m+1} \Sigma^{(m)} \overline{V}_m^T \right\|$ and $\left\| W_{m+1}^T A w_{m+1} \right\|$, for some test problems. Note that $\left\| W_{m+1}^T A w_{m+1} \right\|$ comes from the bound (4.5) with $W_m^\perp$ replaced by w_{m+1}.

We remark that Proposition 3.3 and 4.3 can be used to stop the procedure whenever the noise level ε is known, since it is generally useless to continue with the SVD approximation if we find $\sigma_k^{(m)} << \varepsilon$, for a certain k and m. Indeed, in this situation the Picard condition is no longer satisfied since typically $U_m^T b \approx \varepsilon$ for m large enough.

For what concerns the generalized SVD of the matrix pair (A,L), let $AX = US$ and $LX = VC$, where $S = diag(s_1,...,s_N)$ and $C = diag(c_1,...,c_N)$, $X \in \mathbb{R}^{N\times N}$ is nonsingular and $U,V \in \mathbb{R}^{N\times N}$ are orthogonal. Moreover let $\overline{H}_m X^{(m)} = U^{(m)} S^{(m)}$ and $L_m X^{(m)} = V^{(m)} C^{(m)}$, where $S^{(m)} = diag(s_1^{(m)},...,s_m^{(m)})$ and $C^{(m)} = diag(c_1^{(m)},...,c_m^{(m)})$, be the generalized SVD of the matrix pair $(\overline{H}_m, L_m)$. In this situation, for the convergence of the approximated generalized singular values and vectors, we can state the following result.

Proposition 4.4 *Let $u_k^{(m)}$, $v_k^{(m)}$ and $x_k^{(m)}$ be the k-th column of the matrices $U^{(m)} \in \mathbb{R}^{(m+1)\times m}$, $V^{(m)} \in \mathbb{R}^{m\times m}$ and $X^{(m)} \in \mathbb{R}^{m\times m}$ respectively. Then defining $\overline{u}_k = W_{m+1} u_k^{(m)}$, $\overline{v}_k = W_m v_k^{(m)}$ and $\overline{x}_k = W_m x_k^{(m)}$, we have*

$$A\overline{x}_k - s_k^{(m)} \overline{u}_k = 0, \tag{4.7}$$

$$W_m^T (L\overline{x}_k - c_k^{(m)} \overline{v}_k) = 0. \tag{4.8}$$

Proof Similarly to Proposition 4.2, (4.7) and (4.8) follows immediately from the basic relation (2.2).

As before the proposition ensures that if the matrix A has full rank, than the Arnoldi algorithm allows to construct the GSVD of (A,L). Step by step, the quality of the approximation depends on the distance between $span\{Lw_1,...,Lw_m\}$ and $\mathscr{K}_m(A,b)$. Similarly to (4.4) and (4.6), since $\overline{v}_k = W_m v_k^{(m)} \in \mathscr{K}_m(A,b)$, we have

$$\left\| L\overline{x}_k - c_k^{(m)} \overline{v}_k \right\| \leq \left\| (I - W_m W_m^T) L W_m W_m^T \right\|.$$

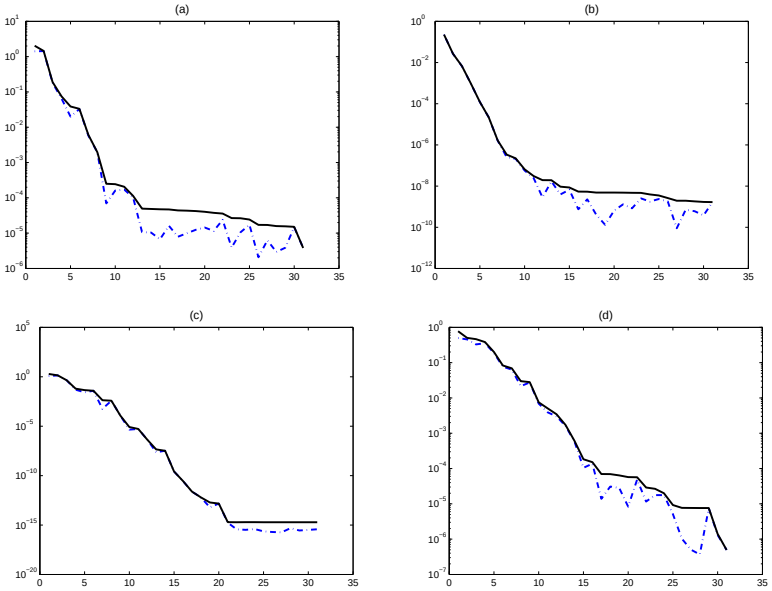

Fig. 4.1 *Decay behavior of* $\left\|A-\overline{U}_{m+1}\Sigma^{(m)}\overline{V}_m^T\right\|$ *(solid line) and lower bound* $\left\|W_{m+1}^T A w_{m+1}\right\|$ *arising from Proposition 4.3 (dash-dot line) for BAART (a), WING (b), SHAW (c) and I_LAPLACE (d). The dimension of each problem is* $N = 32$.

and

$$\left\|L\overline{x}_k - \sigma_k^{(m)}\overline{v}_k\right\| \leq \left\|\left(W_m^{\perp}\right)^T LW_m\right\|.$$

In Figure 4.2 we show the convergence of the singular values of $\overline{H}_m$, and the generalized singular values of the matrix pair $\left(\overline{H}_m, L_m\right)$, with

$$L = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \\ 0 & \cdots & \cdots & 0 \end{pmatrix}, \tag{4.9}$$

working with the test problems SHAW and BAART. The results show that the approximations are quite accurate. It is interesting to observe that, in both cases, after 8-9 iterations the algorithm starts to generate spurious approximations. This is due to the loss of orthogonality of the Krylov vectors, since in these experiments (and in what follows) we have used the Gram-Schmidt implementation. Working with the Householder version of the algorithm the problem is fixed. Anyway in the framework of the regularization, a more accurate approximation of the smallest singular values is useless because of the error in b.

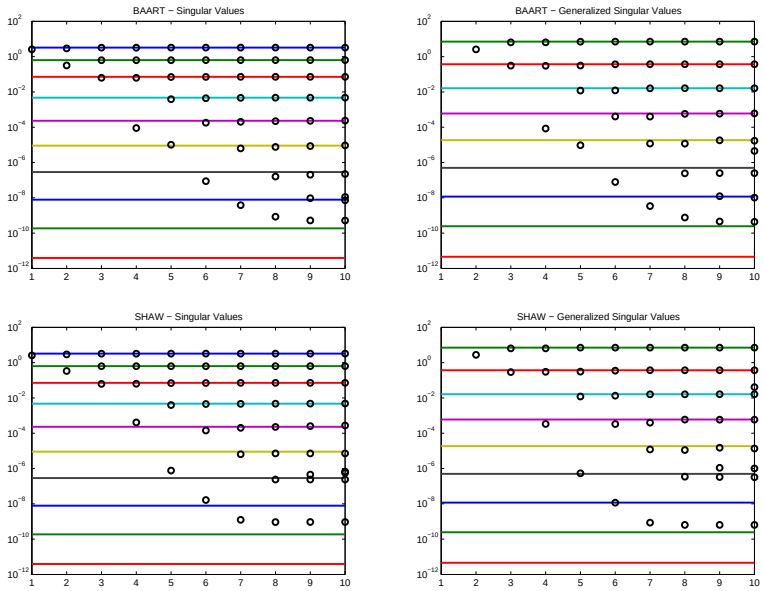

Fig. 4.2 *Plots of the singular values (circle) of the matrix $\overline{H}_m$ (left) and the generalized singular values of the matrix pair $(\overline{H}_m, L_m)$ (right) versus the iteration number k, for the problem BAART and SHAW with $N = 32$. The solid lines represent the singular values of the matrix A (left) and the generalized singular values of the matrix pair (A,L) (right).*

5 Generalized Cross-Validation

A popular method for choosing the regularization parameter, which does not require the knowledge of the noise properties nor its norm $\|e\|$, is the Generalized Cross-Validation (GCV) [9,37]. The major idea of the GCV is that a good choice of λ should predict missing values, so that the model is not sensitive to the elimination of one data point. This means that the regularized solution should predict a datum fairly well, even if that datum is not used in the model. This viewpoint leads to minimization with respect to λ of the GCV function

$$G(\lambda) = \frac{\|b - Ax_\lambda\|^2}{[\text{trace}(I - AA_\lambda)]^2},$$

where $A_\lambda = (A^TA + \lambda^2 L^TL)^{-1}A^T$ is the matrix that gives the regularized solutions of (1.2) from the normal equations

$$(A^TA + \lambda^2 L^TL)x_\lambda = A^Tb.$$

Using the GSVD of the matrix pair (A,L), with a general $A \in \mathbb{R}^{M\times N}, L \in \mathbb{R}^{P\times N}$, let $A = USX^{-1}$ and $L = VCX^{-1}$, where $S = diag(s_1,\ldots,s_P)$ and $C = diag(c_1,\ldots,c_P)$, the generalized singular values γ_i of (A,L) are defined by the ratios

$$\gamma_i = \frac{s_i}{c_i}, \qquad i = 1,\ldots,P.$$

Therefore, one can show that the expression of the GCV function is given by

$$G(\lambda) = \frac{\sum_{i=1}^{N} \left(\frac{\lambda^2}{\gamma_i^2+\lambda^2} u_i^T b \right)^2}{\left(M-(N-P) - \sum_{i=1}^{P} \frac{\gamma_i^2}{\gamma_i^2+\lambda^2} \right)^2}. \tag{5.1}$$

For the square case $M = N$, and $P \leq N$, rearranging the sum at the denominator we obtain

$$G(\lambda) = \frac{\sum_{i=1}^{N} \left(\frac{\lambda^2}{\gamma_i^2+\lambda^2} u_i^T b \right)^2}{\left(\sum_{i=1}^{P} \frac{\lambda^2}{\gamma_i^2+\lambda^2} \right)^2}. \tag{5.2}$$

The GCV criterion is then based on the choice of λ which minimizes $G(\lambda)$. It is well known that this minimization problem is generally ill-conditioned, since the function $G(\lambda)$ is typically flat in a relatively wide region around the minimum. As a consequence, this criterion may even lead to a poor regularization [23,25,35].

As already said in the Introduction, for large-scale problems the GCV approach for (1.2) is too expensive since it requires the SVD (GSVD). In this setting, our idea is to fully exploit the approximation properties of the Arnoldi algorithm investigated in Section 3 and 4. In particular, our aim is to define a sequence of regularization parameters $\{\lambda_m\}$, i.e., one for each iteration of the Arnoldi algorithm, obtained by the minimization of the following GCV function approximations

$$G_m(\lambda) = \frac{\left\| \overline{H}_m y_{m,\lambda} - \|b\| e_1 \right\|^2}{\left(N-m+\sum_{i=1}^{m} \frac{\lambda^2}{\gamma_i^{(m)2}+\lambda^2} \right)^2}, \tag{5.3}$$

where $y_{m,\lambda}$ solves the reduced minimization (2.8), and $\gamma_i^{(m)}$, $i = 1,\ldots,m$, are the approximations of the generalized singular values, obtained with the Arnoldi process. Note that

$$\left\| \overline{H}_m y_{m,\lambda} - \|b\| e_1 \right\|^2 = \sum_{i=1}^{m} \left(\frac{\lambda^2}{\gamma_i^{(m)2}+\lambda^2} u_i^{(m)^T} c \right)^2 + \left(u_{m+1}^{(m)^T} c \right)^2,$$

where $u_i^{(m)}$ is defined as in Proposition 4.4 and $c = \|b\| e_1$, so that the construction of $G_m(\lambda)$ can be obtained working in reduced dimension. The basic idea which leads to the approximation $G_m(\lambda) \approx G(\lambda)$, is to set equal to 0 the generalized singular values that are not approximated by the Arnoldi algorithm, and that are expected to be close to 0 after few iterations. This is justified by the analysis and the experiments of Section 3 and 4.

We remark that in a hybrid approach [21], one aims to regularize the projected problem

$$\min_{y \in \mathbb{R}^m} \left\{ \left\| \overline{H}_m y - \|b\| e_1 \right\| \right\}. \tag{5.4}$$

Since no geometrical information on the solution of (5.4) can be inherited from the solution of the original problem, the choice of $L_m = I_m$ as regularization operator is

somehow forced (this is a standard strategy for hybrid methods [15, §6.7]). In this framework, if the GCV criterion is used to regularize (5.4), the basic difference with respect to (5.3) is at the denominator, where $N-m$ is replaced by m. We observe moreover that (5.3) is similar to the GCV approximation commonly used for iterative methods, in which the denominator is simply $N-m$ [15, §7.4].

In the following we show the algorithm that has been used for the tests of the next sections.

AT - GCV Algorithm

given $A \in \mathbb{R}^{N\times N}$, $b \in \mathbb{R}^N$, δ
while $\big|\, \|r_m\| - \|r_{m-1}\| \,\big| / \|r_m\| \geq \delta$
update $\overline{H}_m$ and L_m from (2.3) and (2.7)
compute GSVD($\overline{H}_m, L_m$)
compute $\lambda_m = \arg\min_\lambda G_m(\lambda)$
solve $\min_{y_m\in\mathbb{R}^m} \left\| \begin{pmatrix} \overline{H}_m \\ \lambda_m L_m \end{pmatrix} y_m - \begin{pmatrix} \|b\| e_1 \\ 0 \end{pmatrix} \right\|^2$
compute the corresponding residual r_m
end
compute $x_m = W_m y_m$

The stopping rule used in the algorithm is just based on the residual. As an alternative, one may even employ the strategy adopted in [4], based on the observation of the GCV approximations.

6 Numerical results

In order to test the performance of the proposed method, we consider again some classical test problems taken from the Regularization Tools [16]. In particular in Figures 6.1-6.2, we consider the problems BAART, SHAW, FOXGOOD, I_LAPLACE, with right-hand side affected by 0.1% or 1% Gaussian noise. The regularization operator is always the discretized first derivative (4.9), augmented with a zero row at the bottom to make it square (cf. (2.7)). For each experiment we show: (a) the approximation of $G(\lambda)$ obtained with the functions $G_m(\lambda)$ for some values of m, with a graphical comparison of the local minima; (b) the approximate solution; (c) the relative residual and error history; (d) the sequence of selected parameters $\{\lambda_m\}$, with respect to the one obtained with the minimization of $G(\lambda)$ (denoted by λ_A in the pictures) and the optimal one (λ_{opt}) obtained by the minimization of the distance between the regularized and the true solution [29]

$$\min_\lambda \left\| x_{reg} - x_{true} \right\|^2 \equiv \min_\lambda f(\lambda),$$

where

$$f(\lambda) = \left\{ \sum_{i=1}^{p} \left(\frac{\lambda^2}{(\gamma_i^2+\lambda^2)} \frac{u_i^T b}{\sigma_i} x_i - \sum_{i=p+1}^{N} (u_i^T b) x_i \right) - \sum_{i=1}^{N} \frac{u_i^T b}{\sigma_i} v_i \right\}^2 .$$

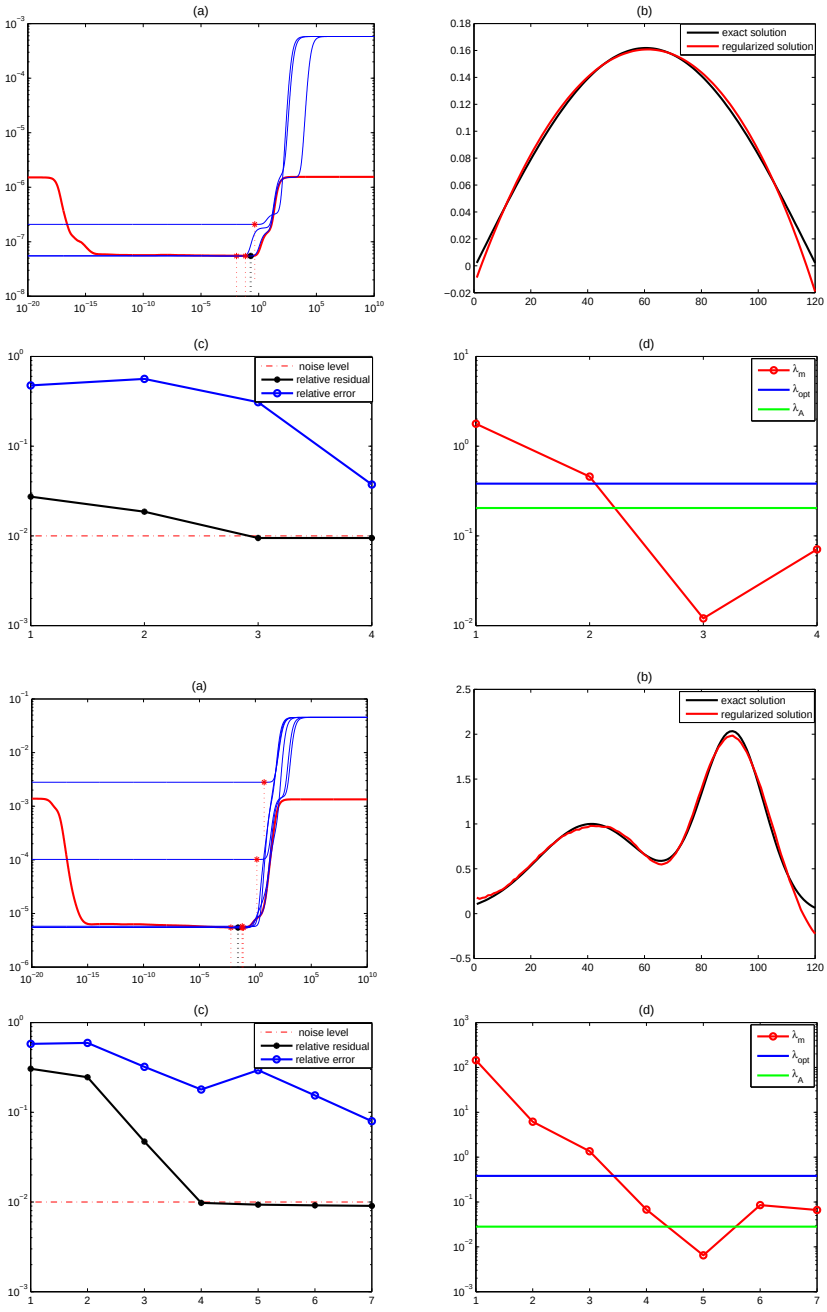

Fig. 6.1 *Results for BAART (top) and SHAW (bottom). The dimension of each problem is* $N = 120$*. Noise level* $\varepsilon = 10^{-2}$*. In subfigures (a) the tick red line indicates the function* $G(\lambda)$*. In both cases the regularization operator is (4.9).*

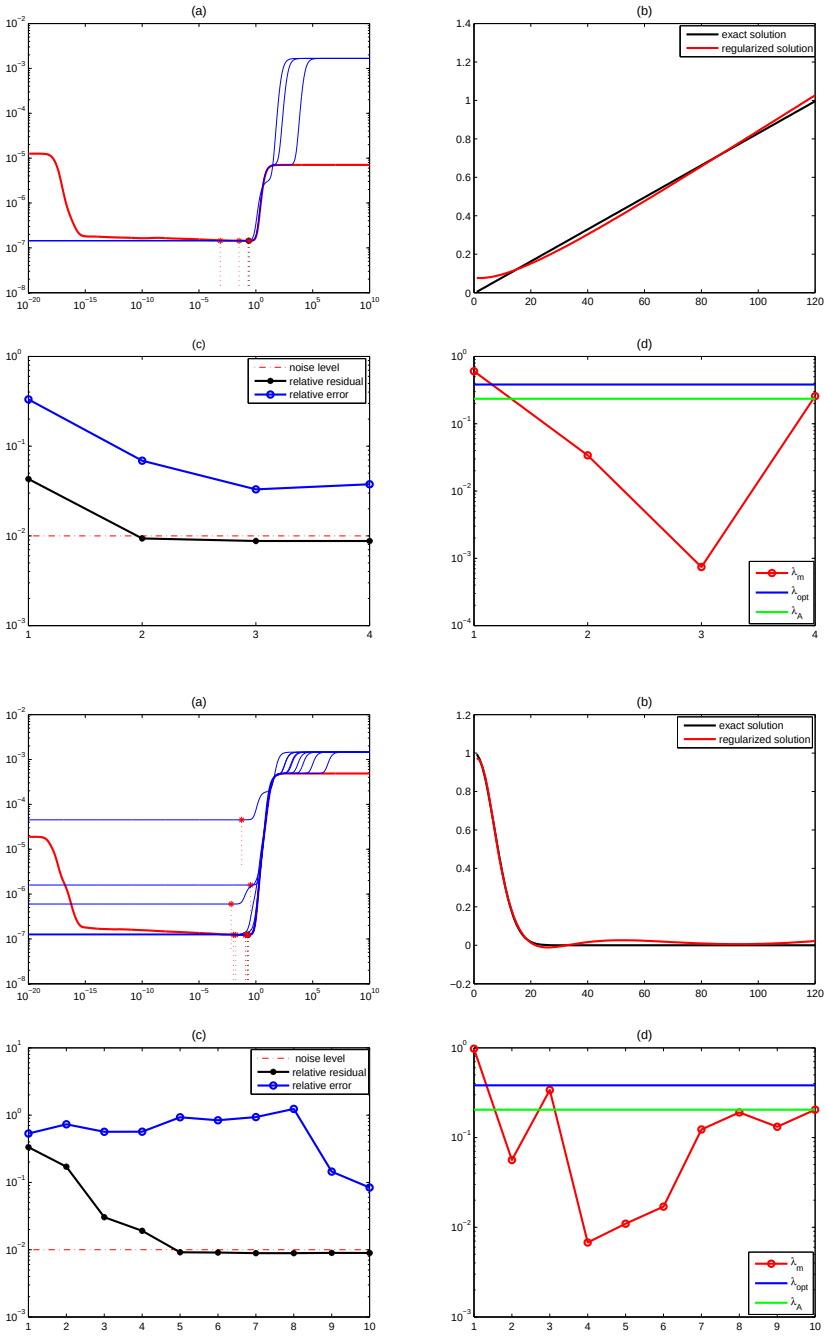

Fig. 6.2 *Results for FOXGOOD (top) and I_LAPLACE (bottom). The dimension of each problem is* $N = 120$. *Noise level* $\varepsilon = 10^{-3}$. *In subfigures (a) the tick red line indicates the function* $G(\lambda)$. *In both cases the regularization operator is (4.9).*

7 An example of image restoration

We conclude with an illustration of the performance of the GCV-Arnoldi approach on a 2D image deblurring problem which consists in recovering the original $n \times n$ image from a blurred and noisy observed image.

Let X be a $n \times n$ two dimensional image. The vector $\overline{x}$ of dimension $N = n^2$ obtained by stacking the columns of the image X represents a blur-free and noise-free image. We generate an associated blurred and noise-free image $\overline{b}$ by multiplying $\overline{x}$ by a block Toeplitz matrix $A \in \mathbb{R}^{N\times N}$ with Toeplitz blocks, implemented in the function `blur.m` from the Regularization Tools [16]. This Matlab function 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. The blur and noise contaminated image $b \in \mathbb{R}^N$ is obtained by adding a noise-vector $e \in \mathbb{R}^N$, so that $b = A\overline{x} + e$. We assume the blurring operator A and the corrupted image b to be available while no information is given on the error e, we would like to determine a restoration which accurately approximates the blur-free and noise-free image $\overline{x}$.

We consider the restoration of a corrupted version of the 256×256 test image `mri.png`. Contamination is by 1% white Gaussian noise and space-invariant Gaussian blur. The latter is generated as described above with blur parameters `band`=7, `sigma`=2, so that the condition number of A is around 10^{13}. Figure 7.1 displays the performance of the AT-GCV. On the left the blur-free and noise-free image, on the middle the corrupted image, on the right the restored image. From top to bottom the image in original size and two different zooms. The regularization operator is defined as (cf. [7])

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

where $L_1 \in \mathbb{R}^{n\times n}$ is the discretized first derivative with a zero row at the bottom as in (4.9) (cf. also [20, §5]). The experiment has been carried out using Matlab 7.10 on a single processor computer (Intel Core i7). The result has been obtained in 5 iterations of the Arnoldi algorithm, in around 0.5 seconds.

8 Conclusion

The fast convergence of the Arnoldi algorithm when applied to compact operators makes the AT method particularly attractive for the regularization of discrete ill-posed problems. The projected problems rapidly inherit the basic features of the original one, so that the rate of convergence is closely related to the decay rate of the singular values of A.

In this paper, in absence of information on the noise which affects the right-hand side of the system, we have employed the GCV criterion. Contrary to the hybrid techniques, the sequence of regularization parameters $\{\lambda_m\}_{m\geq 1}$ is defined in order to regularize the original problem instead of the projected one, leading to GCV approximations which are similar to the ones used for pure iterative methods ([15, §7.4]). Notwithstanding the intrinsic difficulties concerning the GCV criterion, the arising algorithm has shown to be quite robust. Of course there are cases in which the method

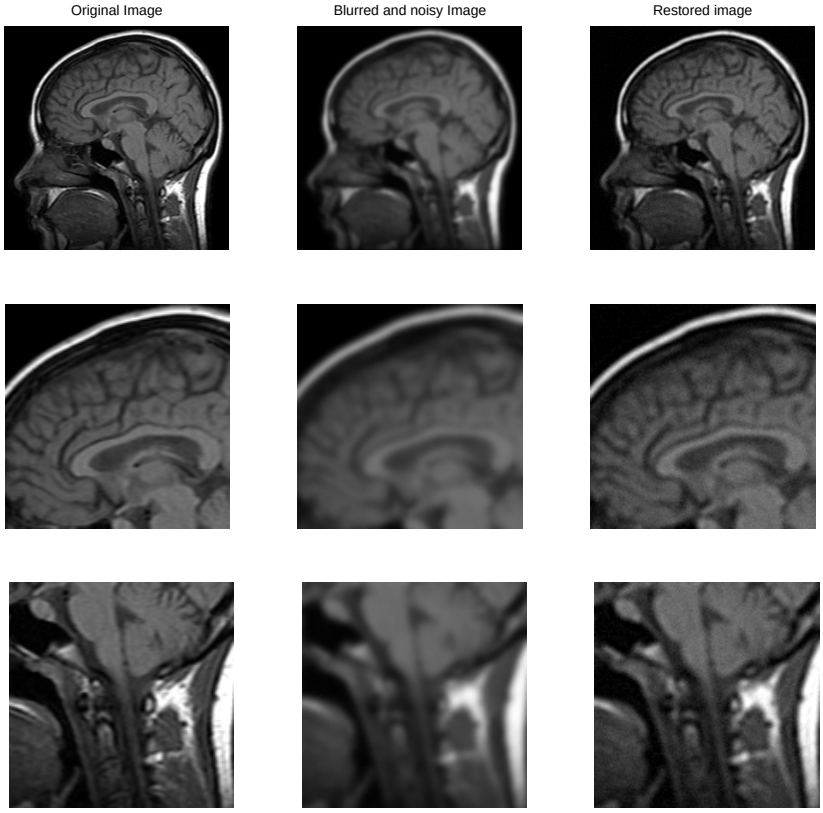

Fig. 7.1 *Restoration of* `mri.png`*. Original image, blurred and noisy image with noise level* $\varepsilon = 10^{-2}$ *and blur parameters* `band=7, sigma=2`*, restored image. From top to bottom original size and two zoom.*

fails, but the numerical experiments presented are rather representative of what happens in general.

While not considered in the paper, the Range Restricted Arnoldi method [24, 18] represent a potential improvement of the method here presented, especially for problems in which the noise level is rather high and if the regularization matrix is little effective as noise removal (as for instance the identity matrix). Of course the analysis of Section 3 and 4 should be modified accordingly. In particular the Picard condition should be no longer necessary to prove Proposition 3.2 (suitably modified), since the starting vector of the Arnoldi process would be Ab. The arising approximation of the dominating singular values and consequently, the approximation of the GCV function, should be analized.

References

1. A. Bjorck, A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations, BIT **28**, 659–670 (1988).

2. F. Bauer, M. Lukas, Comparing parameter choice methods for regularization of ill-posed problems, Math. Comput. Simulation **81**(9), 1795–1841 (2011).
3. D. Calvetti, S. Morigi, L. Reichel, F.Sgallari, Tikhonov regularization and the L-curve for large discrete ill-posed problems, J. Comput. Appl. Math. **123**, 423–446 (2000).
4. J.Chung, J.G. Nagy, D.P. O'Leary, A weighted-GCV method for Lanczos-hybrid regularization, Electron. Trans. Numer. Anal. **28**, 149–167 (2007/08).
5. M. Donatelli, A. Neuman, and L. Reichel, Square regularization matrices for large linear discrete ill-posed problems, Numer. Linear Algebra Appl. **19**, 896–913 (2012).
6. J. N. Franklin, Minimum Principles for Ill-Posed Problems, SIAM J. Math. Anal. **9**(4), 638–650 (1978).
7. S. Gazzola, P. Novati, Automatic parameter setting for Arnoldi-Tikhonov methods, J. Comput. Appl. Math. **256**, 180–195 (2014).
8. G.H. Golub, F. T. Luk, M. L. Overton, A Block Lánczos Method for Computing the Singular Values of Corresponding Singular Vectors of a Matrix, ACM Trans. Math. Software, **7**(2), 149–169 (1981).
9. G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics **21**(2), 215–223 (1979).
10. G.H. Golub, C.F. Van Loan, Matrix Computations. Johns Hopkins University Press, Baltimore (MD), 3rd edition, 1996.
11. U. Hämarik, R. Palm, T. Raus, A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level, J. Comput. Appl. Math. **236**(8), 2146–2157 (2012).
12. M. Hanke, On Lanczos based methods for the regularization of discrete ill-posed problems, BIT **41**, 1008–1018 (2001).
13. P.C. Hansen, The discrete Picard condition for discrete ill-posed problems, BIT **30**, 658–672 (1990).
14. M. Hanke, P.C. Hansen, Regularization methods for large-scale problems, Surv. Math. Ind. **3**, 253–315 (1993).
15. P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia (1998).
16. P.C. Hansen, Regularization Tools Version 4.0 for Matlab 7.3, Numer. Algorithms **46**, 189–194 (2007).
17. M. Hochstenbach, L. Reichel, An iterative method for Tikhonov regularization with a general linear regularization operator, J. Integral Equations Appl. **22**, 463–480 (2010).
18. M. Hochstenbach, N. McNinch, L. Reichel, Discrete ill-posed least-squares problems with a solution norm constraint, Linear Algebra Appl. **436**, 3801–3818 (2012).
19. B. Hofmann, Regularization for Applied Inverse and Ill-Posed Problems. Teubner, Stuttgart, Germany (1986).
20. M.E. Kilmer, P.C. Hansen, M.I. Español, A projection-based approach to general-form Tikhonov regularization, SIAM J. Sci. Comput. **29**(1), 315–330 (2007).
21. M.E. Kilmer, D.P. O'Leary, Choosing regularization parameters in iterative methods for ill-posed problems, SIAM J. Matrix Anal. Appl. **22**, 1204–1221 (2001).
22. S. Kindermann, Convergence analysis of minimization-based noise levelfree parameter choice rules for linear ill-posed problems, Electron. Trans. Numer. Anal. **38**, 233–257 (2011).
23. R. Kohn, C.F. Ansley, D. Tharm, The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters, J. Am. Stat. Assoc. **86** , 1042–1050 (1991).
24. B. Lewis, L. Reichel, Arnoldi-Tikhonov regularization methods, J. Comput. Appl. Math. **226**, 92–102 (2009).
25. M.A. Lukas, Robust generalized cross-validation for choosing the regularization parameter, Inverse Probl. **22**, 1883–1902 (2006).
26. I, Moret. A note on the superlinear convergence of GMRES, SIAM J. Numer. Anal. **34**(2), 513–516 (1997).
27. O. Nevanlinna, Convergence of Iterations for Linear Equations. Birkhäuser, Basel (1993).
28. P. Novati, M.R. Russo, Adaptive Arnoldi-Tikhonov regularization for image restoration, Numerical Algorithms, in press (2013). DOI: 10.1007/s11075-013-9712-0.
29. D.P. O'Leary, Near-optimal parameters for Tikhonov and other regularization methods, SIAM J. Sci. Comput. **23**(4), 1161–1171 (2011).
30. D.P. O'Leary, J.A. Simmons, A bidiagonalization-regularization procedure for large-scale discretizations of ill-posed problems, SIAM J. Sci. Statist. Comput. **2**, 474–489 (1981).
31. L. Reichel, G. Rodriguez, Old and new parameter choice rules for discrete ill-posed problems, Numer. Algorithms **63**, 65–87 (2013) .

32. Y. Saad, Iterative methods for Sparse Linear Systems, 2nd edition. SIAM, Philadelphia (PA), 2003.
33. Y. Saad, Numerical methods for large eigenvalue problems. Algorithms and Architectures for Advanced Scientific Computing. Manchester University Press, Manchester, Halsted Press, New York (1992).
34. Y. Saad, K. Wu, DQGMRES: a direct quasi-minimal residual algorithm based on incomplete orthogonalization, Numer. Linear Algebra Appl. **3**, 329–343 (1996).
35. A.M. Thompson, J.W. Kay, D.M. Titterington, A cautionary note about crossvalidatory choice, J. Stat. Comput. Simul. **33**, 199–216 (1989).
36. L.N. Trefethen, D. Bau, Numerical Linear Algebra. SIAM, Philadelphia (1997).
37. G. Wahba, A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, Ann. Stat. **13**, 1378–1402 (1985).