Electronic Transactions on Numerical Analysis.
Volume 44, pp. 83–123, 2015.

ISSN 1068–9613.

ETNA
Kent State University
http://etna.math.kent.edu

ON KRYLOV PROJECTION METHODS AND TIKHONOV REGULARIZATION*

SILVIA GAZZOLA†, PAOLO NOVATI†, AND MARIA ROSARIA RUSSO†

Abstract. In the framework of large-scale linear discrete ill-posed problems, Krylov projection methods represent an essential tool since their development, which dates back to the early 1950's. In recent years, the use of these methods in a hybrid fashion or to solve Tikhonov regularized problems has received great attention especially for problems involving the restoration of digital images. In this paper we review the fundamental Krylov-Tikhonov techniques based on Lanczos bidiagonalization and the Arnoldi algorithms. Moreover, we study the use of the unsymmetric Lanczos process that, to the best of our knowledge, has just marginally been considered in this setting. Many numerical experiments and comparisons of different methods are presented.

Key words. discrete ill-posed problems, Krylov projection methods, Tikhonov regularization, Lanczos bidiagonalization, nonsymmetric Lanczos process, Arnoldi algorithm, discrepancy principle, generalized cross validation, L-curve criterion, Regińska criterion, image deblurring

AMS subject classifications. 65F10, 65F22, 65R32.

1. Introduction.

The solution of large-scale linear systems

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

obtained by suitably discretizing ill-posed operator equations that model many inverse problems arising in various scientific and engineering applications generally requires the use of iterative methods. In this setting, the coefficient matrix A is typically of ill-determined rank, i.e., the singular values of A quickly decay and cluster at zero with no evident gap between two consecutive ones to indicate numerical rank; in particular, A is ill-conditioned. Moreover, generally, the available right-hand side vector b is affected by error (noise), i.e., $b = b^{ex} + e$, where b^{ex} represents the unknown error-free right-hand side. In the following we just consider additive white noise. For an introduction to the solution of this kind of problems, we refer to [32, 36].

Historically, since the derivation of the Conjugate Gradient (CG) method [41], CG-like techniques such as the CGLS method and Craig's method (CGNE) [18], in which (1.1) is solved in terms of a least-squares problem, have been widely studied. After the famous paper [29], in which the authors define the so-called Lanczos bidiagonalization procedure by exploiting the Lanczos algorithm for the tridiagonalization of symmetric matrices [48], in [59] the LSQR method is introduced. This method is mathematically equivalent to CGLS but with better stability properties, and it is still widely used to solve least-squares problems. It is important to recall that these methods compare well with, and in some cases are preferable to, direct techniques for solving linear systems, especially for severely ill-conditioned problems. Indeed, as pointed out in [31], contrarily to the truncated singular value decomposition (TSVD), the projection attained with the CGLS (LSQR) method is tailored to the specific right-hand side b, providing more rapid convergence. All of the above mentioned CG-like techniques belong to the broad class of Krylov subspace methods, which in turn belong to the even broader class of projection methods: at each iteration of these methods a projection of problem (1.1)

*Received December 1, 2014. Accepted January 21, 2015. Published online on February 11, 2015. Recommended by L. Reichel. This work was partially supported by MIUR (project PRIN 2012 N. 2012MTE38N), and by the University of Padova (project CPDA124755 "Multivariate approximation with applications to image reconstruction").
†Department of Mathematics, University of Padova, Italy
({gazzola,novati,mrrusso}@math.unipd.it).

ETNA
Kent State University
http://etna.math.kent.edu

onto Krylov subspaces is performed, and the so obtained problem of reduced dimensions is solved (we refer to Section 2 for the details).

In the case of discrete ill-posed problems, a well-known basic property of Krylov iterative methods (which might be considered both an advantage or a disadvantage) is the so-called semi-convergence phenomenon, i.e., at the beginning of the iterative process the solution computed by a Krylov subspace method typically approaches the solution x^{ex} of the error-free problem $Ax = b^{ex}$, but after just a few iterations it is affected by the errors in b and rapidly deteriorates. This is due to the fact that the ill-conditioning of the problem is inherited by the projected problems after a certain number of steps. For this reason, Krylov subspace methods are regarded as iterative regularization methods, the number of iterations being the regularization parameter that should be properly set.

The first attempt to remedy the semiconvergence issue seems to be the one proposed in [58], where the TSVD of the projected problem obtained by Lanczos bidiagonalization is considered. The aim of this first hybrid technique was to regularize the projected problem, i.e., to stabilize the behavior of the error (defined, in this setting, as the norm of the difference between x^{ex} and the regularized solution computed at each iteration). The problem of choosing the correct number of iterations is then reformulated as a problem of singular value analysis. Similar approaches, coupled with parameter selection techniques such as the discrepancy principle, the generalized cross validation (GCV), and the L-curve were then studied in [2, 3, 31, 32, 47, 61].

Another well-established technique to stabilize the behavior of Krylov projection methods is to apply them in connection with Tikhonov regularization. Referring to the original problem (1.1), regularizing it by the Tikhonov method 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 called a regularization parameter and $L \in \mathbb{R}^{P\times N}$ is called a regularization matrix (see again [32, 36] for a background); the norms considered in this paper are always the vector 2-norm or the associated induced matrix norm. Assuming that $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$, we denote the solution of (1.2) by x_λ. It is known that a proper choice of both λ and L is crucial for a meaningful approximation of x^{ex}; in particular, the regularization matrix L can be suitably chosen if some information on the behavior of x^{ex} is available. The simplest regularization consists in taking $L = I_N$, where I_N is the identity matrix of order N: this method is typically referred to as standard form Tikhonov regularization.

The simultaneous use of Krylov methods and Tikhonov regularization for approximating the exact solution of (1.1) can be formulated in two ways. The first one (hybrid methods) consists in regularizing the projected problem. From now on in this paper, the word hybrid will always refer to the process of applying Tikhonov regularization on a projected problem. The second one (Krylov-Tikhonov methods) consists in projecting the regularized problem, i.e., in solving (1.2) by projections. To the best of our knowledge, the use of hybrid methods has been first suggested in [58, 59] with the aim of regularizing the Lanczos-bidiagonalization-based methods with the identity matrix. This technique has then been used in a number of papers (e.g., [2, 4, 10, 11, 12, 13, 17, 47, 62, 64]) in connection with many techniques for the definition of the sequence of the regularization parameters (one for each iteration of the underlying iterative method). Hybrid methods based on the Arnoldi process (i.e., regularization of the GMRES method [70]) have a more recent history: they were first introduced in [10], and then studied (also with the Range-Restricted implementation) in [42, 49]. We remark that a hybrid Arnoldi method has even been implicitly used in [51], where the symmetric Lanczos process is used for A being symmetric positive semidefinite in connection with the

ETNA
Kent State University
http://etna.math.kent.edu

Lavrentiev (Franklin) regularization

$$(A + \lambda I_N)x = b, \quad \lambda > 0.$$

Again, in [15] the same algorithm is applied for A being symmetric with the standard Tikhonov regularization.

Beyond the hybrid approaches, the use of Krylov projection methods for solving (1.2) (i.e., Krylov-Tikhonov methods) with $L \neq I_N$ is even more recent. Of course, this approach is potentially more effective. Indeed, since no information on the main features of the true solution are in principle inherited by the solutions of the projected problems, for hybrid methods one is somehow forced to use the identity matrix to regularize them. Lanczos bidiagonalization for solving (1.2) has been used in [46], where an algorithm for the simultaneous bidiagonalization of A and L is introduced, and in [43], where the skinny QR factorization of the penalty term is used to fully project the problem. The Arnoldi algorithm for (1.2) has been used in [27, 28, 56, 57]; in [26] it is used in the multiparameter setting. A nice method based on the generalized Arnoldi algorithm applied to the matrix pair (A, L) is presented in [63]. We remark that, starting from [2], many authors have suggested to bridge the gap between hybrid methods and the solution of (1.2) by Krylov projection: indeed, after transforming (1.2) into standard form [21], the smoothing effect of L is incorporated into the hybrid process; see also [37, 42, 49]. However, unless L is invertible or has a special structure, it is not easy to use this transformation; probably, for this reason, this transformation is often not implemented and tested in the papers where it is mentioned. Some computationally-friendly approaches to define L as a smoothing preconditioner for the system (1.1) have been proposed in [14]. Other efficient ways to define square regularization matrices are described, for instance, in [19, 65].

The aim of the present paper is to review the basic properties and the computational issues of the methods based on the Lanczos bidiagonalization and the Arnoldi algorithm for solving both (1.1) and (1.2) with particular attention to the parameter choice rules (for both λ and the number of iterations). We also consider the use of the unsymmetric Lanczos process, which underlies some well-known linear system solvers such as BiCG, CGS, QMR, and BiCGstab (see [69, Chapter 7] and the references therein) but has never been used in the framework of Tikhonov regularization: indeed, in [6], these methods have been briefly addressed as iterative regularization methods, but they have never been employed to project a Tikhonov-regularized problem. While Krylov methods are mainly interesting for large-scale problems, we shall compare the three approaches primarily on moderate-size test problems taken from [35].

This paper is organized as follows: in Section 2 we address the Krylov projection methods considered in this paper and, more precisely, we outline some methods based on the Lanczos bidiagonalization algorithm (Section 2.1), the Arnoldi algorithm (Section 2.2), and the nonsymmetric Lanczos algorithm (Section 2.3); we also prove some theoretical properties. In the first part of Section 3 (Section 3.1), we introduce a common framework that embraces all the methods considered in Section 2. In order to assess the regularizing performances of the described Krylov subspace methods, in the second part of Section 3 (Section 3.2) we include the results of numerous numerical experiments. Then, in Section 4, we describe in a general framework the hybrid methods and the Krylov-Tikhonov methods employing the discrepancy principle as parameter choice strategy. Theoretical considerations as well as meaningful results of numerical experiments are proposed. In Section 5, we review (and we comment on) the use of various parameter choice methods in the Krylov-Tikhonov setting. Most of them are commonly employed when performing Tikhonov or iterative regularization and, except for the Regińska criterion, all of them have already been considered in connection with the Krylov-Tikhonov methods. Finally, in Section 6, we analyze the performance of the different Krylov-Tikhonov methods when applied to image deblurring and denoising problems:

ETNA
Kent State University
http://etna.math.kent.edu

we consider a medical and an astronomical test image, and all the parameter choice strategies described in Section 5 are taken into account. In this paper, we will use the following

Notations: we denote the SVD of the full-dimensional matrix A by

$$A = U^{\mathcal{S}} \Sigma^{\mathcal{S}} \left(V^{\mathcal{S}}\right)^T, \tag{1.3}$$

where $U^{\mathcal{S}}$, $V^{\mathcal{S}} \in \mathbb{R}^{N\times N}$ are orthogonal, and $\Sigma^{\mathcal{S}} = \text{diag}(\sigma_1, \ldots, \sigma_N) \in \mathbb{R}^{N\times N}$ is diagonal. We denote the TSVD of A by

$$A_m^{\mathcal{S}} = U_m^{\mathcal{S}} \Sigma_m^{\mathcal{S}} (V_m^{\mathcal{S}})^T, \tag{1.4}$$

where $U_m^{\mathcal{S}}$, $V_m^{\mathcal{S}} \in \mathbb{R}^{N\times m}$ are obtained by extracting the first m columns of the matrices $U^{\mathcal{S}}$, $V^{\mathcal{S}}$ in (1.3), respectively, and $\Sigma_m^{\mathcal{S}}$ is the leading $m \times m$ submatrix of $\Sigma^{\mathcal{S}}$ in (1.3). We also denote by $x_m^{\mathcal{S}}$ the TSVD solution of (1.1), i.e.,

$$x_m^{\mathcal{S}} = V_m^{\mathcal{S}} \left(\Sigma_m^{\mathcal{S}}\right)^{-1} (U_m^{\mathcal{S}})^T b. \tag{1.5}$$

The Generalized Singular Values Decomposition (GSVD) of the matrix pair (A, L) is defined by the factorizations

$$AX^{\mathcal{G}} = U^{\mathcal{G}} S^{\mathcal{G}}, \quad LX^{\mathcal{G}} = V^{\mathcal{G}} C^{\mathcal{G}}, \tag{1.6}$$

where $S^{\mathcal{G}} = \text{diag}(s_1, \ldots, s_N)$ and $C^{\mathcal{G}} = \text{diag}(c_1, \ldots, c_N)$, $X^{\mathcal{G}} \in \mathbb{R}^{N\times N}$ is nonsingular, and $U^{\mathcal{G}}, V^{\mathcal{G}} \in \mathbb{R}^{N\times N}$ are orthogonal. The generalized singular values γ_i of (A, L) are defined by the ratios

$$\gamma_i = \frac{s_i}{c_i}, \qquad i = 1, \ldots, N. \tag{1.7}$$

To keep the notation simpler, in (1.6) and (1.7) we have assumed $L \in \mathbb{R}^{N\times N}$. We underline that the superscripts $\mathcal{S}$ and $\mathcal{G}$ have been introduced to better distinguish the SVD of A and the GSVD of (A, L), respectively, from the SVD and GSVD of the matrices associated to the projected problems that we will consider in the following sections.

Test problems: in order to numerically describe the properties of the methods considered in the paper, we make use of the test problems available from Hansen's toolbox *Regularization Tools* [35]. Some test problems, such as `i_laplace`, are implemented with more than one choice for the right-hand side, so that the corresponding solution may have different regularity. Coherently with the switches used in the toolbox, we denote by "`problem` - s" the s-th test of the Matlab code.

2. Krylov projection methods. As mentioned in the introduction, in this paper we review some Krylov methods as a tool for the regularization of ill-conditioned linear systems. Given a matrix $C \in \mathbb{R}^{N\times N}$ and a vector $d \in \mathbb{R}^N$, the Krylov subspace $\mathcal{K}_m(C, d)$ is defined by

$$\mathcal{K}_m(C, d) = \text{span}\{d, Cd, \ldots, C^{m-1}d\}.$$

Typically, in this paper, $C = A$, A^T, A^TA, AA^T and $d = b$, A^Tb, Ab. Given two Krylov subspaces $\mathcal{K}'_m$ and $\mathcal{K}''_m$ both of dimension m, Krylov projection methods are iterative methods in which the m-th approximation x_m is uniquely determined by the conditions

$$x_m \in x_0 + \mathcal{K}'_m, \tag{2.1}$$

$$b - Ax_m \perp \mathcal{K}''_m, \tag{2.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

where x_0 is the initial guess. In order to simplify the exposition, from now on we assume $x_0 = 0$. Denoting by $W_m \in \mathbb{R}^{N\times m}$ the matrix whose columns span $\mathcal{K}'_m$, i.e., $\mathcal{R}(W_m) = \mathcal{K}'_m$, we are interested in methods where $x_m = W_m y_m$ (approximately) solves

$$\min_{x\in\mathcal{K}'_m} \|b - Ax\| = \min_{y\in\mathbb{R}^m} \|b - AW_m y\| = \|b - AW_m y_m\|. \tag{2.3}$$

Before introducing the methods considered in this paper we recall the following definition.

DEFINITION 2.1. *Assume that $b = b^{ex}$ in (1.1) is the exact right-hand side. Let u_m be the m-th column of $U^{\mathcal{S}}$. Then the Discrete Picard Condition (DPC, cf. [33]) is satisfied if $\{|u_m^T b|\}_{1\le m\le N}$, on the average, decays faster than $\{\sigma_m\}_{1\le m\le N}$.*

More generally, for continuous problems, the Picard Condition ensures that a square integrable solution exists; see [36, p. 9]. For discrete problems, the DPC ensures that the TSVD solutions of (1.1) are uniformly bounded. If we assume to work with severely ill-conditioned problems, i.e., $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$ (cf. [44]), and that the coefficients $u_m^T b$, $1 \le m \le N$, satisfy the model

$$|u_m^T b| = \sigma_m^{1+\beta}, \quad \beta > 0,$$

(cf. [36, p. 81]), then the TSVD solutions (1.5) are bounded as

$$\left\|x_m^{\mathcal{S}}\right\|^2 \le \sum_{j=1}^{m} \sigma_j^{2\beta} \le C \sum_{j=1}^{m} e^{-2\beta\alpha j} \le C\frac{1}{1-e^{-2\beta\alpha}}.$$

Similar bounds can be straightforwardly obtained when dealing with mildly ill-conditioned problems, in which $\sigma_j = O(j^{-\alpha})$ provided that α is large enough. Of course, whenever the solution of a given problem is bounded, then the DPC is automatically verified.

2.1. Methods based on the Lanczos bidiagonalization algorithm. The Lanczos bidiagonalization algorithm [29] computes two orthonormal bases $\{w_1, \ldots, w_m\}$ and $\{z_1, \ldots, z_m\}$ for the Krylov subspaces $\mathcal{K}_m(A^TA, A^Tb)$ and $\mathcal{K}_m(AA^T, b)$, respectively. In Algorithm 1 we summarize the main computations involved in the Lanczos bidiagonalization procedure.

Algorithm 1 Lanczos bidiagonalization algorithm.

Input: A, b.
Initialize: $\nu_1 = \|b\|$, $z_1 = b/\nu_1$.
Initialize: $w = A^T z_1$, $\mu_1 = \|w\|$, $w_1 = w/\mu_1$.
For $j = 2, \ldots, m+1$
1. Compute $z = Aw_{j-1} - \mu_{j-1} z_{j-1}$.
2. Set $\nu_j = \|z\|$.
3. Take $z_j = z/\nu_j$.
4. Compute $w = A^T z_j - \nu_j w_{j-1}$.
5. Set $\mu_j = \|w\|$.
6. Take $w_j = w/\mu_j$.

Setting $W_m = [w_1, \ldots, w_m] \in \mathbb{R}^{N\times m}$ and $Z_m = [z_1, \ldots, z_m] \in \mathbb{R}^{N\times m}$, the Lanczos bidiagonalization algorithm can be expressed in matrix form by the following relations

$$AW_m = Z_{m+1}\bar{B}_m, \tag{2.4}$$

$$A^T Z_{m+1} = W_m \bar{B}_m^T + \mu_{m+1} w_{m+1} e_{m+1}^T, \tag{2.5}$$

ETNA
Kent State University
http://etna.math.kent.edu

where

$$\bar{B}_m = \begin{bmatrix} \mu_1 & & & & \\ \nu_2 & \mu_2 & & & \\ & \ddots & \ddots & & \\ & & \nu_m & \mu_m & \\ & & & \nu_{m+1} & \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}$$

and e_{m+1} denotes the $(m+1)$-st canonical basis vector of $\mathbb{R}^{m+1}$.

The most popular Krylov subspace method based on Lanczos bidiagonalization is the LSQR method, which is mathematically equivalent to the CGLS but with better numerical properties. Referring to (2.1) and (2.2), the LSQR method has $\mathcal{K}'_m = \mathcal{K}_m(A^TA, A^Tb)$ and $\mathcal{K}''_m = A\mathcal{K}_m(A^TA, A^Tb)$. This method consists in computing, at the m-th iteration of the Lanczos bidiagonalization algorithm,

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| \|b\| e_1 - \bar{B}_m y \right\| \tag{2.6}$$

and in taking $x_m = W_m y_m$ as approximate solution of (1.1). Indeed, for this method,

$$\begin{aligned} \min_{x\in\mathcal{K}_m} \|b - Ax\| &= \min_{y\in\mathbb{R}^m} \|b - AW_m y\| \\ &= \min_{y\in\mathbb{R}^m} \left\| \|b\| Z_{m+1}e_1 - Z_{m+1}\bar{B}_m y \right\| \\ &= \min_{y\in\mathbb{R}^m} \left\| \|b\| e_1 - \bar{B}_m y \right\|. \end{aligned}$$

As already addressed in the introduction, Lanczos-bidiagonalization-based regularization methods have historically been the first Krylov subspace methods to be employed with regularization purposes in a purely iterative fashion (cf. [71]), as hybrid methods (cf. [58]), and to approximate the solution of (1.2) (cf. [4]). In the remaining part of this section we prove some propositions that are useful to better understand the regularizing properties of the LSQR method. The following proposition deals with the rate of convergence of the method.

PROPOSITION 2.2. *Assume that* (1.1) *is severely ill-conditioned, i.e.,* $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$. *Assume moreover that* b *satisfies the DPC. Then, for* $m = 1, \dots, N-2$,

$$\mu_{m+1}\nu_{m+1} = O(m\sigma_m^2), \tag{2.7}$$

$$\mu_{m+1}\nu_{m+2} = O(m\sigma_{m+1}^2), \tag{2.8}$$

Proof. Concerning estimate (2.7), by (2.4) and (2.5) we have that

$$(A^TA)W_m = W_m(\bar{B}_m^T\bar{B}_m) + \mu_{m+1}\nu_{m+1}w_{m+1}e_m^T,$$

so that W_m is the matrix generated by the symmetric Lanczos process applied to the system $A^TAx = A^Tb$ and its columns span $\mathcal{K}_m(A^TA, A^Tb)$. After recalling that the singular values of A^TA are the scalars σ_i^2, $i = 1, \dots, N$, and that the super/sub-diagonal elements of the symmetric tridiagonal matrix $\bar{B}_m^T\bar{B}_m \in \mathbb{R}^{m\times m}$ are of the form $\mu_{m+1}\nu_{m+1}$, $m = 1, \dots, N-1$, the estimate (2.7) directly follows by applying Proposition 2.6 reported in Section 2.2 (for a proof, see [57, Proposition 3.3]) and the refinement given in [24, Theorem 6]).

Concerning estimate (2.8), using again (2.5) and (2.4), we have that

$$(AA^T)Z_{m+1} = Z_{m+1}(\bar{B}_m\bar{B}_m^T) + \mu_{m+1}Aw_{m+1}e_{m+1}^T.$$

ETNA
Kent State University
http://etna.math.kent.edu

By step 1 of Algorithm 1,

$$\begin{aligned}(AA^T)Z_{m+1} &= Z_{m+1}(\bar{B}_m\bar{B}_m^T) + \mu_{m+1}[\mu_{m+1}z_{m+1} + \nu_{m+2}z_{m+2}]e_{m+1}^T \\ &= Z_{m+1}(\bar{B}_m\bar{B}_m^T + \mu_{m+1}^2 e_{m+1}e_{m+1}^T) + \mu_{m+1}\nu_{m+2}z_{m+2}e_{m+1}^T,\end{aligned}$$

so that Z_{m+1} is the matrix generated by the symmetric Lanczos process applied to the system $AA^Tx = b$ and its columns span $\mathcal{K}_{m+1}(AA^T, b)$. Since the singular values of AA^T are the scalars σ_i^2, $i = 1, \ldots, N$, and the super/sub-diagonal elements of the symmetric tridiagonal matrix $(\bar{B}_m\bar{B}_m^T + \mu_{m+1}^2 e_{m+1}e_{m+1}^T) \in \mathbb{R}^{(m+1)\times(m+1)}$ are of the form $\mu_{m+1}\nu_{m+2}$, $m = 0, \ldots, N-2$, the estimate (2.7) follows again by applying [24, Theorem 6]. ∎

One of the main reasons behind the success of Lanczos bidiagonalization as a tool for regularization [1, 45] is basically due to the ability of the projected matrices $\bar{B}_m$ to approximate the largest singular values of A. Indeed we have the following result.

PROPOSITION 2.3. *Let $\bar{B}_m = \bar{U}_m\bar{\Sigma}_m\bar{V}_m^T$ be the SVD of $\bar{B}_m$, and let $U_m = Z_{m+1}\bar{U}_m$, $V_m = W_m\bar{V}_m$. Then*

$$AV_m - U_m\bar{\Sigma}_m = 0\,, \tag{2.9}$$

$$\left\|A^TU_m - V_m\bar{\Sigma}_m^T\right\| \le \mu_{m+1}\,. \tag{2.10}$$

Proof. Relation (2.9) immediately follows from (2.4) and $\bar{B}_m\bar{V}_m = \bar{U}_m\bar{\Sigma}_m$. Moreover, by employing (2.5),

$$\begin{aligned}A^TU_m &= A^TZ_{m+1}\bar{U}_m = W_m\bar{B}_m^T\bar{U}_m + \mu_{m+1}w_{m+1}e_{m+1}^T\bar{U}_m \\ &= W_m\bar{V}_m\bar{\Sigma}_m^T + \mu_{m+1}w_{m+1}e_{m+1}^T\bar{U}_m = V_m\bar{\Sigma}_m^T + \mu_{m+1}w_{m+1}e_{m+1}^T\bar{U}_m,\end{aligned}$$

which, since $\|w_{m+1}\| = \|e_{m+1}\| = \|\bar{U}_m\| = 1$, leads to (2.10). ∎

Provided that $\mu_m \to 0$, relations (2.9) and (2.10) ensure that the triplet $\left(U_m, \bar{\Sigma}_m, V_m\right)$ represents an increasingly better approximation of the TSVD of A: for this reason, Lanczos-bidiagonalization-based methods have always proved very successful when employed with regularization purposes; cf. [1, 31, 32, 43] and [36, Chapter 6]. Indeed, looking at Algorithm 1, we have that $\mu_j = \|w\|$, where $w \in \mathcal{K}_j(A^TA, A^Tb)$ and $w \perp \mathcal{K}_{j-1}(A^TA, A^Tb)$. If A represents a compact operator, we know that quite rapidly $\mathcal{K}_j(A^TA, A^Tb)$ becomes almost A^TA-invariant, i.e., $\mathcal{K}_j(A^TA, A^Tb) \approx \mathcal{K}_{j-1}(A^TA, A^Tb)$; see, e.g., [50] and the references therein.

PROPOSITION 2.4. *Under the same hypotheses of Proposition 2.2, we have that ν_m, $\mu_m \to 0$ and*

$$\begin{aligned}\left\|AA^Tz_m - \mu_m^2 z_m\right\| &= O(\nu_m)\,, \\ \left\|A^TAw_m - \nu_m^2 w_m\right\| &= O(\mu_{m-1})\,.\end{aligned}$$

Proof. We start by assuming that $\nu_m \nrightarrow 0$. Then, by Proposition 2.2, we immediately have that $\mu_m = O(m\sigma_m^2)$. Thus, by step 1 of Algorithm 1,

$$z = Aw_{m-1} + d_{m-1}, \quad \|d_{m-1}\| = O(m\sigma_{m-1}^2), \tag{2.11}$$

for m large enough. Then, by step 4 and by (2.11),

$$w = A^Tz_m - \nu_m w_{m-1} = A^T\left(\frac{Aw_{m-1} + d_{m-1}}{\nu_m}\right) - \nu_m w_{m-1},$$

ETNA
Kent State University
http://etna.math.kent.edu

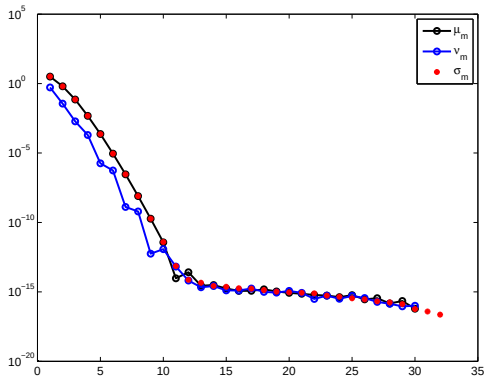

FIG. 2.1. *Problem* `baart`*: decay behavior of the sequences* $\{\nu_m\}_m$ *and* $\{\mu_m\}_m$ *with respect to the singular values of* A*.*

which implies

$$\mu_m \nu_m w_m = A^T A w_{m-1} - \nu_m^2 w_{m-1} + A^T d_{m-1},$$

and hence

$$\left\| A^T A w_{m-1} - \nu_m^2 w_{m-1} \right\| = O(m\sigma_{m-1}^2). \tag{2.12}$$

The above relation means that, asymptotically, ν_m behaves like a singular value of A, so that $\nu_m \to 0$. Still by step 4 of Algorithm 1, we have that

$$w = A^T z_m + d'_m, \quad \|d'_m\| = \nu_m.$$

Then at the next step 1,

$$z = A w_m - \mu_m z_m = A \left(\frac{A^T z_m + d'_m}{\mu_m} \right) - \mu_m z_m,$$

so that

$$\nu_{m+1} \mu_m z_{m+1} = A A^T z_m - \mu_m^2 z_m + A d'_m,$$

and hence

$$\left\| A A^T z_m - \mu_m^2 z_m \right\| = O(\nu_m).$$

The above relation means that μ_m asymptotically behaves like a singular value of A, so that $\mu_m \to 0$. At this point of the proof, we have demonstrated that $\nu_m \to 0$and consequently that $\mu_m \to 0$. Finally, rewriting the right-hand side of equality (2.12) by replacing $\|d_{m-1}\| = O(m\sigma_{m-1}^2) = \mu_{m-1}$, we obtain the result. □

Proposition 2.4 states that, for severely ill-conditioned problems, we can expect that the sequences $\{\nu_m\}_m$ and $\{\mu_m\}_m$ behave similarly, and that their rate of decay is close to the one of the singular values of A. An example of this behavior is reported in Figure 2.1. Thanks to this proposition, we can state that the approximation of the singular values of A attainable with the singular values of $\bar{B}_m$ is expected to be very accurate; see Proposition 2.3.

PROPOSITION 2.5. *If the full-dimensional system* (1.1) *satisfies the DPC, then the DPC is inherited by the projected problems* (2.6)*, for* $1 \le m \le N$.

ETNA
Kent State University
http://etna.math.kent.edu

Proof. Recalling that LSQR is mathematically equivalent to CG applied to the normal equations $A^TAx = A^Tb$ and thanks to the relations derived in [41, Theorem 6.1] and elaborated in [36, Chapter 6], we can state that

$$\|x_m\| \leq \|x_{m+1}\|, \quad m = 1, \ldots, N-1.$$

Since the DPC holds for the problem (1.1), $\|y_N\| = \|x_N\| = \|x^{ex}\| = c < \infty$. Moreover, since

$$\|x_m\| = \|W_m y_m\| = \|y_m\|, \quad m = 1, \ldots, N,$$

we can state that

$$\|y_m\| \leq c, \quad m = 1, \ldots, N,$$

which proves the result. □

2.2. Methods based on the Arnoldi algorithm. The Arnoldi algorithm computes an orthonormal basis $\{w_1, \ldots, w_m\}$ for the Krylov subspace $\mathcal{K}_m(A, b)$. In Algorithm 2 we summarize the main computations involved in the Arnoldi orthogonalization scheme.

Algorithm 2 Arnoldi algorithm.

Input: A, b.
Initialize: $w_1 = b/\|b\|$.
For $j = 1, 2, \ldots, m$
1. For $i = 1, \ldots, j$: compute $h_{i,j} = (Aw_j, w_i)$.
2. Compute $w = Aw_j - \sum_{i=1}^{j} h_{i,j} w_i$.
3. Define $h_{j+1,j} = \|w\|$.
4. If $h_{j+1,j} = 0$ stop; else take $w_{j+1} = w/h_{j+1,j}$.

Setting $W_m = [w_1, \ldots, w_m] \in \mathbb{R}^{N \times m}$, the Arnoldi algorithm can be written in matrix form as

$$AW_m = W_m H_m + h_{m+1,m} w_{m+1} e_m^T, \tag{2.13}$$

where $H_m = [h_{i,j}]_{i,j=1,\ldots,m} \in \mathbb{R}^{m \times m}$ is an upper Hessenberg matrix that represents the orthogonal projection of A onto $\mathcal{K}_m(A, b)$, i.e., $W_m^T A W_m = H_m$, and e_m is the m-th canonical basis vector of $\mathbb{R}^m$. Equivalently, relation (2.13) can be written as

$$AW_m = W_{m+1} \bar{H}_m,$$

where

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

The basic steps outlined in Algorithm 2 are only indicative. There are some important variants of the algorithm as, for instance, the modified Gram-Schmidt or the Householder implementation [69, §6.3], which may considerably improve its accuracy measured in terms of the quantity $\left\|W_m^T W_m - I_m\right\|$. It is known that, when using the standard Gram-Schmidt process, the theoretical orthogonality of the basis vectors is almost immediately lost. On the other side, when using the Householder orthogonalization, the orthogonality is guaranteed at the machine precision level. Throughout this section we are mainly interested in the theoretical properties of the methods based on the Arnoldi algorithm, so that we assume to work in exact arithmetic.

ETNA
Kent State University
http://etna.math.kent.edu

2.2.1. The GMRES method. The most popular Krylov subspace method based on the Arnoldi algorithm is the GMRES method [70]. Referring to (2.1) and (2.2), the GMRES method works with $\mathcal{K}'_m = \mathcal{K}_m(A,b)$ and $\mathcal{K}''_m = A\mathcal{K}_m(A,b)$. Similarly to LSQR, we have

$$\begin{aligned}\min_{x\in\mathcal{K}_m} \|b-Ax\| &= \min_{y\in\mathbb{R}^m} \|b-AW_m y\| \\ &= \min_{y\in\mathbb{R}^m} \left\| \|b\|\, W_{m+1}e_1 - W_{m+1}\bar{H}_m y\right\| \\ &= \min_{y\in\mathbb{R}^m} \left\| \|b\|\, e_1 - \bar{H}_m y\right\|,\end{aligned}$$

so that at the m-th iteration of the Arnoldi algorithm, the GMRES method prescribes to compute

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| \|b\|\, e_1 - \bar{H}_m y\right\|$$

and to take $x_m = W_m y_m$ as an approximate solution of (1.1).

The theoretical analysis of the regularizing properties of the GMRES method applied to the solution of ill-conditioned linear systems has been fully performed in [9], where the authors show that the approximate solutions tend to the exact solution whenever the norm of the error of the right hand side of the system goes to 0 and a stopping criterion based on the residual is employed.

It is well known that the rate of convergence of the method is closely related to the behavior of the sequence $\{h_{m+1,m}\}_m$ since $h_{m+1,m} = \|w\|$ (cf. step 3 of Algorithm 2) is a measure of the extendibility of the Krylov subspaces. Moreover, it is also known that the residual of GMRES can be bounded using the Full Orthogonalization Method (FOM, see, e.g., [69, §6.4]) residual as follows

$$\|r_m\| \le h_{m+1,m} \left| e_m^T H_m^{-1} e_1 \right| \|b\|.$$

In the case of severely ill-conditioned problems, the following result has been proved in [57] (cf. Figure 2.2).

PROPOSITION 2.6. *Assume that A has full rank with singular values of the form $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$, and that b satisfies the DPC. Then, if b is the starting vector of the Arnoldi process, we obtain*

$$h_{m+1,m} = O\left(m\,\sigma_m\right).$$

The authors of [57] show that the Arnoldi algorithm can be regarded as a tool for approximating the TSVD of the matrix A similarly to what is done when one employs the Lanczos bidiagonalization algorithm; cf. Section 2.1 and [1, 31]. Moreover, the authors of [8] show that, in some situations, GMRES equipped with a suitable stopping rule can deliver more accurate approximations than TSVD. In [57] the following result was proved.

PROPOSITION 2.7. *Let $\bar{H}_m = \bar{U}_m \bar{\Sigma}_m \bar{V}_m^T$ be the SVD of $\bar{H}_m$, and let $U_m = W_{m+1}\bar{U}_m$ and $V_m = W_m \bar{V}_m$. Then*

$$\begin{aligned} AV_m - U_m\bar{\Sigma}_m &= 0, \\ W_m^T(A^T U_m - V_m \bar{\Sigma}_m^T) &= 0. \end{aligned} \tag{2.14}$$

Since the Arnoldi algorithm does not involve A^T, unless the matrix is symmetric, we cannot expect that the approximation of the largest singular values of A is as good as the one attainable with the Lanczos bidiagonalization algorithm. The different approximation

ETNA
Kent State University
http://etna.math.kent.edu

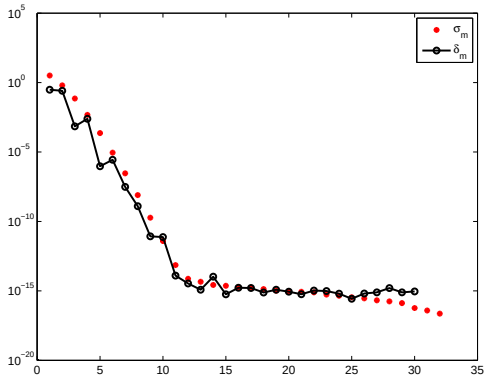

FIG. 2.2. *Problem* `baart`*: decay behavior of the sequence* $\{h_{m+1,m}\}_m$ *with respect to the singular values of* A.

capabilities of the two algorithms can also be understood by comparing (2.10) and (2.14): the latter represents a Galerkin condition that only guarantees that, if A is nonsingular, at the end of the process the Arnoldi algorithm provides the complete SVD of A.

As for the Discrete Picard Condition, up to our knowledge the question whether this condition is inherited by the projected problem (cf. Proposition 2.5) is still open. Computationally it is quite evident that it is in fact inherited, but the theoretical proof is still unavailable. The same holds also for the other methods considered below.

2.2.2. The Range-Restricted GMRES method. The Range-Restricted GMRES (RRGMRES) method was first introduced in [5] and then used in [7] with the aim of reducing the presence of the error in the starting vector of the Arnoldi algorithm. Indeed, this method prescribes to look for approximate solutions belonging to the Krylov subspaces $\mathcal{K}_m(A, Ab)$ and therefore to run the Arnoldi algorithm with starting vector $w_1 = Ab/\|Ab\|$. Thanks to the smoothing properties of A, many high-frequency noise components are removed in w_1, and therefore the propagation of the noise in the RRGMRES basis vectors is less severe than in the GMRES ones. However, on the downside, the vector b might be important for the reconstruction especially if the exact solution is intrinsically not very smooth: not including b in the solution subspace can lead to a loss of information; cf. the discussion in [7]. More recently, in [20] RRGMRES has been generalized to work with starting vector $A^s b$, $s \geq 1$.

Let $W_m = [w_1, \ldots, w_m] \in \mathbb{R}^{N\times m}$ be the orthogonal basis of $\mathcal{K}_m(A, Ab)$ computed by the Arnoldi algorithm. Then relation (2.13) still holds, i.e.,

$$AW_m = W_{m+1}\bar{H}_m, \tag{2.15}$$

where $\bar{H}_m$ is an upper Hessenberg matrix. Writing

$$b = W_{m+1}W_{m+1}^T b + \left(I - W_{m+1}W_{m+1}^T\right) b = W_{m+1}W_{m+1}^T b + W_{m+1}^{\perp}\left(W_{m+1}^{\perp}\right)^T b\,,$$

we have

$$\begin{aligned}
\min_{x\in\mathcal{K}_m(A,Ab)} \|b - Ax\|^2 &= \min_{y\in\mathbb{R}^m} \|b - AW_m y\|^2 \\
&= \min_{y\in\mathbb{R}^m} \left\|W_{m+1}W_{m+1}^T b - W_{m+1}\bar{H}_m y\right\|^2 + \left\|W_{m+1}^{\perp}\left(W_{m+1}^{\perp}\right)^T b\right\|^2 \\
&= \min_{y\in\mathbb{R}^m} \left\|W_{m+1}^T b - \bar{H}_m y\right\|^2 + \left\|\left(W_{m+1}^{\perp}\right)^T b\right\|^2,
\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

so that, at the m-th iteration of the Arnoldi algorithm, the RRGMRES method prescribes to compute

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| W_{m+1}^T b - \bar{H}_m y \right\| .$$

Proposition 2.7 is still valid since it only involves the Arnoldi decomposition (2.15): this assures that RRGMRES can still be interpreted as a method able to approximate the singular values of A.

We remark that the above derivations are only meaningful from a theoretical point of view since improved implementations of RRGMRES (and other methods related to it) were proposed in [54, 55]. In particular, the most recent implementations do not rely on the explicit computation of the quantities $W_{m+1}^T b$ and $(W_{m+1}^\perp)^T b$, and therefore they are more stable with respect to the loss of orthogonality in the columns of W_{m+1}.

2.3. Methods based on the Nonsymmetric Lanczos algorithm. The Nonsymmetric Lanczos algorithm (also referred to as two-sided Lanczos process, or Lanczos biorthogonalization procedure) is employed to compute two bases $\{w_1, \dots, w_m\}$ and $\{k_1, \dots, k_m\}$ for the Krylov subspaces $\mathcal{K}_m(A, b)$ and $\mathcal{K}_m(A^T, b)$, respectively, satisfying the biorthogonality condition $w_i^T k_j = \delta_{ij}$, $i, j = 1, \dots, m$, where δ_{ij} is the Kronecker delta. In Algorithm 3 we summarize the main computations involved in the Lanczos biorthogonalization procedure.

Algorithm 3 Lanczos biorthogonalization algorithm.

Input: A, b.
Initialize: $w_1 = b/\|b\|$, $k_1 = w_1$ so that $(w_1, k_1) = 1$.
Initialize: $\beta_1 = \delta_1 = 0$, $w_0 = k_0 = 0$.
For $j = 1, \dots, m$
1. $\alpha_j = (Aw_j, k_j)$.
2. Compute $w = Aw_j - \alpha_j w_j - \beta_j w_{j-1}$.
3. Compute $k = A^T k_j - \alpha_j k_j - \delta_j k_{j-1}$.
4. Set $\delta_{j+1} = |(w, k)|^{1/2}$. If $\delta_{j+1} = 0$ stop.
5. Set $\beta_{j+1} = (w, k)/\delta_{j+1}$.
6. Take $k_{j+1} = k/\beta_{j+1}$.
7. Take $w_{j+1} = w/\delta_{j+1}$.

Setting $W_m = [w_1, \dots, w_m]$ and $K_m = [k_1, \dots, k_m]$, the Lanczos biorthogonalization algorithm can be expressed in matrix form by the following relations

$$\begin{aligned} AW_m &= W_m T_m + \delta_{m+1} w_{m+1} e_m^T, \\ A^T K_m &= K_m T_m^T + \beta_{m+1} k_{m+1} e_m^T, \end{aligned} \tag{2.16}$$

where $T_m \in \mathbb{R}^{m\times m}$ is the tridiagonal matrix

$$T_m = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \end{bmatrix}.$$

Because of the biorthogonality property, relation (2.16) yields

$$K_m^T A W_m = T_m \quad \text{and} \quad W_m^T A^T K_m = T_m^T.$$

ETNA
Kent State University
http://etna.math.kent.edu

It is well known that, if the matrix A is symmetric, then this method reduces to the symmetric Lanczos process: indeed, in this case, $W_m = K_m$ have orthogonal columns, and T_m is symmetric.

The matrix T_m can be regarded as the projection of A obtained from an oblique projection process onto $\mathcal{K}_m(A, b)$ and orthogonal to $\mathcal{K}_m(A^T, b)$. Relation (2.16) can be written as

$$AW_m = W_{m+1}\bar{T}_m, \tag{2.17}$$

where

$$\bar{T}_m = \begin{bmatrix} T_m \\ \delta_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}.$$

We remark that the definition of $\delta_{j+1} = \left|k^T w\right|^{1/2}$ at step 4 of Algorithm 3 only represents a common choice since it leads to $\delta_{j+1} = \pm\beta_{j+1}$; cf. step 5 of the same algorithm. More generally, to build the two bases, it is only necessary that $\delta_{j+1}\beta_{j+1} = k^T w$.

The most popular Krylov subspace methods based on Lanczos biorthogonalization are the BiCG and QMR methods; cf. [69, Chapter 7] and the references therein. In the following we focus just on the QMR method, and we always assume that the Lanczos nonsymmetric algorithm does not breakdown at or before the m-th step.

At the m-th iteration of the nonsymmetric Lanczos algorithm, the QMR [23] method prescribes to compute

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| \|b\| e_1 - \bar{T}_m y \right\| \tag{2.18}$$

and to take $x_m = W_m y_m$ as approximate solution of (1.1). Since the matrix W_{m+1} is not orthogonal, it is known that $\left\| \|b\| e_1 - \bar{T}_m y_m \right\|$ is just a pseudo-residual since

$$\|b - Ax_m\| = \left\| W_{m+1} \left(\|b\| e_1 - \bar{T}_m y_m \right) \right\|.$$

Exploiting the QR factorization of W_{m+1} and hence the relation between QMR and GMRES, it can be proved that (cf. [23])

$$\left\| r_m^{\text{QMR}} \right\| \le \kappa(W_{m+1}) \left\| r_m^{\text{GMRES}} \right\|,$$

where r_m^{QMR} and r_m^{GMRES} are the residuals of QMR and GMRES, respectively. Of course, if A is symmetric, then QMR and GMRES are mathematically equivalent.

In the remaining part of this section we make some considerations that are helpful to gain some insight into the use of the QMR method for regularization purposes. Since the matrix W_{m+1} is not orthogonal, it is difficult to theoretically demonstrate that QMR can be efficiently used as a tool for regularization. Indeed, it is not easy to provide relations which show that the matrix $\bar{T}_m$ reproduces the singular value properties of A. We only know (see [68, Chapter 6]) that for m large enough, the matrix $\bar{T}_m$ contains some of the spectral information of A since it can be used to approximate the left and right eigenvalues. For this reason, we may expect that, if A is not much far from symmetric, then $\bar{T}_m$ can also be used to approximate its singular values. To study the convergence of the nonsymmetric Lanczos process, we recall the following proposition originally proved in [57].

PROPOSITION 2.8. *Let us assume that the singular values A are of the form $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$. Let us moreover assume that the discrete Picard condition is satisfied. Let*

$$\widetilde{V}_m = [\widetilde{v}_0, \ldots, \widetilde{v}_{m-1}] \in \mathbb{R}^{N\times m}, \quad \textit{where} \quad \widetilde{v}_k = A^k b / \left\| A^k b \right\|.$$

ETNA
Kent State University
http://etna.math.kent.edu

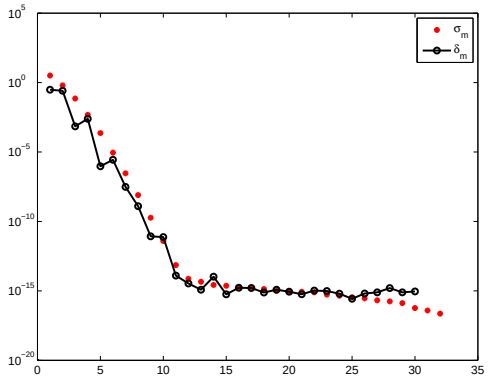

FIG. 2.3. *Problem* `baart`*: decay behavior of the sequence* $\{\delta_m\}_m$ *with respect to the singular values of* A.

If $\widetilde{V}_m$ *has full column rank, then there exist* $C_m \in \mathbb{R}^{m\times m}$ *nonsingular and* $E_m, F_m \in \mathbb{R}^{N\times m}$*, such that*

$$\begin{aligned} \widetilde{V}_m &= U_m^{\mathcal{S}} C_m + E_m, \quad \|E_m\| = O(\sigma_m), \\ U_m^{\mathcal{S}} &= \widetilde{V}_m C_m^{-1} + F_m, \quad \left\|F_m \Sigma_m^{\mathcal{S}}\right\| = O(m\sigma_m). \end{aligned} \tag{2.19}$$

At this point, we can prove the following result (cf. Figure 2.3).

PROPOSITION 2.9. *Under the same hypothesis of Proposition 2.8, for* $m = 1, \ldots, N-1$

$$\delta_{m+1} = O(m\sigma_m). \tag{2.20}$$

Proof. Directly from relation (2.17), we have that $K_{m+1}^T A W_m = \bar{T}_m$ and that $\delta_{m+1} = k_{m+1}^T A w_m$. Thanks to [30, §2.5.5], we can write $A = A_m^{\mathcal{S}} + \Delta_m$, where $A_m^{\mathcal{S}}$ is defined in (1.4) and $\|\Delta_m\| = \sigma_{m+1}$. Therefore

$$\begin{aligned} \delta_{m+1} &= k_{m+1}^T A w_m = k_{m+1}^T A_m^{\mathcal{S}} w_m + k_{m+1}^T \Delta_m w_m \\ &= k_{m+1}^T U_m^{\mathcal{S}} \Sigma_m^{\mathcal{S}} (V_m^{\mathcal{S}})^T w_m + k_{m+1}^T \Delta_m w_m \\ &= k_{m+1}^T (\widetilde{V}_m C_m^{-1} + F_m) \Sigma_m^{\mathcal{S}} (V_m^{\mathcal{S}})^T w_m + k_{m+1}^T \Delta_m w_m, \end{aligned}$$

where we have used (2.19). Since $\mathcal{R}(\widetilde{V}_m) = \mathcal{R}(W_m) = \mathcal{K}_m(A, b)$, we can immediately conclude that $k_{m+1}^T \widetilde{V}_m = 0$. Therefore

$$\begin{aligned} \delta_{m+1} &= k_{m+1}^T (F_m \Sigma_m^{\mathcal{S}}) (V_m^{\mathcal{S}})^T w_m + k_{m+1}^T \Delta_m w_m \\ &\le (O(m\sigma_m) + \sigma_{m+1}) \|k_{m+1}\| \|w_{m+1}\|. \end{aligned}$$

Since $\|k_{m+1}\|\,\|w_m\|$ does not depend on the rate of the decay of σ_m, we obtain (2.20). ∎

As it is well known, a disadvantage of the methods based on the nonsymmetric Lanczos process is that they can break down for several reasons even in exact arithmetic. More precisely, the procedure outlined in Algorithm 3 may break down as soon as a vector k is found to be orthogonal to the corresponding w, so that δ_{j+1} as defined in line 4 of Algorithm 3 vanishes. If this occurs when both k and w are different from zero, then we are dealing with a so-called serious breakdown. Although such exact breakdowns are very rare in practice, near breakdowns (i.e., $k^T w \approx 0$) can cause severe numerical stability problems in subsequent iterations. The possibility of breakdowns has brought the nonsymmetric Lanczos process into discredit. The term "look-ahead" Lanczos is commonly used to denote extensions of the

ETNA
Kent State University
http://etna.math.kent.edu

standard Lanczos method that skip over breakdowns and near-breakdowns. In our setting, since the convergence is generally very fast, the situation $k^T w \approx 0$ is somehow less expectable, and hence, as we will see, the QMR method actually represents a valid alternative to LSQR and GMRES. More precisely, the search subspaces for QMR and GMRES are the same while the constraints imposed on the approximate solutions differ. Furthermore, the Lanczos biorthogonalization process is based on two three-term recurrences (cf. lines 2 and 3 of Algorithm 3) involving the columns of W_m and K_m, respectively, and therefore the storage requirements are potentially less demanding with respect to GMRES. However, using the basic implementation of Algorithm 3, two matrix-vector products (one with A and one with A^T) are required at each iteration.

In some of the following numerical experiments (Section 6) we also consider a range-restricted version of the nonsymmetric Lanczos algorithm, where $x_m \in \mathcal{K}_m(A, Ab)$. The reasons for considering such a method for the regularization of (1.1) are analogous to the ones explained in Section 2.2.2 for RRGMRES.

3. General formulation. In this section we provide a general formulation that embraces the Krylov methods considered in this work.

3.1. Theoretical framework. The methods considered in the previous section are all based on algorithms that are able to construct three sequences of matrices $W_m, Z_m, K_m \in \mathbb{R}^{N\times m}$, $m \geq 1$, such that

$$AW_m = Z_{m+1}\bar{D}_m, \quad K_m^T W_m = I_m, \tag{3.1}$$

where $\bar{D}_m \in \mathbb{R}^{(m+1)\times m}$ has a simple structure. In this way, the solution x of (1.1) is approximated by $W_m y_m$, where y_m solves the projected least squares problem

$$\min_{y\in\mathbb{R}^m} \left\|d - \bar{D}_m y\right\| \approx \min_{y\in\mathbb{R}^m} \left\|b - AW_m y\right\|, \tag{3.2}$$

and where $d \in \mathbb{R}^{m+1}$ depends on the method. Considering the "skinny" QR factorization of the matrix Z_{m+1}, i.e.,

$$Z_{m+1} = Q_{m+1}R_{m+1}, \quad Q_{m+1} \in \mathbb{R}^{N\times(m+1)}, \ R_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}, \tag{3.3}$$

we can state the following general result.

PROPOSITION 3.1. *Given a Krylov subspace method based on the decomposition* (3.1)*, for each $y \in \mathbb{R}^m$ we have*

$$\left\|b - AW_m y\right\|^2 = \left\|Q_{m+1}^T b - R_{m+1}\bar{D}_m y\right\|^2 + \left\|(Q_{m+1}^{\perp})^T b\right\|^2. \tag{3.4}$$

Proof. Considering the factorizations (3.1) and (3.3) and writing

$$b = Q_{m+1}Q_{m+1}^T b + \left(I - Q_{m+1}Q_{m+1}^T\right) b = Q_{m+1}Q_{m+1}^T b + Q_{m+1}^{\perp}(Q_{m+1}^{\perp})^T b,$$

we have

$$\begin{aligned}\left\|b - AW_m y\right\| = \left\|b - Z_{m+1}\bar{D}_m y\right\|^2 &= \left\|Q_{m+1}\left(Q_{m+1}^T b - R_{m+1}\bar{D}_m y\right)\right\|^2 \\ &\quad + \left\|Q_{m+1}^{\perp}(Q_{m+1}^{\perp})^T b\right\|^2.\end{aligned}$$

Thanks to the orthonormality of the columns of Q_{m+1} and $Q_{m+1}^{\perp}$, we immediately obtain (3.4). ∎

Depending on the properties of the considered Krylov method, expression (3.4) can assume simpler forms. In particular:

ETNA
Kent State University
http://etna.math.kent.edu

- For LSQR we have $\bar{D}_m = \bar{B}_m$. Moreover, $W_m = K_m$ and Z_m have orthonormal columns: therefore, $Q_{m+1} = Z_{m+1}$, $R_{m+1} = I_{m+1}$. Since $\mathcal{R}(Z_m) = \mathcal{K}_m(AA^T, b)$, we also have $Q_{m+1}^T b = \|b\| e_1$ and $(Q_{m+1}^\perp)^T b = 0$; referring to (3.2), we have $d = \|b\| e_1$.
- For GMRES we have $\bar{D}_m = \bar{H}_m$. Moreover, $Q_m = Z_m = W_m = K_m$ have orthonormal columns, and $\mathcal{R}(W_m) = \mathcal{K}_m(A, b)$. Therefore, $Q_{m+1}^T b = \|b\| e_1$ and $(Q_{m+1}^\perp)^T b = 0$; referring to (3.2), we have $d = \|b\| e_1$.
- For RRGMRES we have $\bar{D}_m = \bar{H}_m$. Moreover, $Q_m = Z_m = W_m = K_m$, and $\mathcal{R}(W_m) = \mathcal{K}_m(A, Ab)$. Anyway, in general, $(Q_{m+1}^\perp)^T b \neq 0$; referring to (3.2), we have $d = Q_{m+1}^T b$.
- For QMR we have $\bar{D}_m = \bar{T}_m$ and $Z_m = W_m$. Unless A is symmetric, the QR factorization (3.3) is such that $R_{m+1} \neq I_{m+1}$. Since $b \in \mathcal{R}(Z_{m+1}) = \mathcal{R}(Q_{m+1})$ and more precisely $b = \|b\| Z_{m+1} e_1 = \|b\| Q_{m+1} e_1$, we have that $Q_{m+1}^T b = \|b\| e_1$ and $(Q_{m+1}^\perp)^T b = 0$; referring to (3.2), we have $d = \|b\| e_1$. Moreover, the matrix Q_m is just the orthogonal matrix W_m generated by the Arnoldi algorithm. By comparing (3.2) and (2.18) with (3.4), it is clear that for QMR the matrix $R_{m+1} \neq I_{m+1}$ is discarded.

All the Krylov methods studied in this paper are based on the solution of (3.2) with $d = Q_{m+1}^T b$. Observe, however, that none of them makes use of the QR decomposition (3.3) because, except for RRGMRES, we have $Q_{m+1}^T b = \|b\| e_1$, and, for RRGMRES, $Q_{m+1} = W_{m+1}$. Using the above general formulation, we have that the corresponding residual norm $\|b - Ax_m\|$ is in general approximated by a pseudo-residual

$$\|b - Ax_m\| \approx \left\| Q_{m+1}^T b - \bar{D}_m y_m \right\|. \tag{3.5}$$

The following proposition expresses the residual and the pseudo-residual in terms of the SVD decomposition of the projected matrix $\bar{D}_m$; its proof is straightforward. It will be used in Section 4.

PROPOSITION 3.2. *Let y_m be the solution of (3.2), and let $x_m = W_m y_m$ be the corresponding approximate solution of (2.3). Let moreover $\bar{D}_m = \bar{U}_m \bar{\Sigma}_m \bar{V}_m^T$ be the SVD decomposition of $\bar{D}_m$. Then*

$$\left\| Q_{m+1}^T b - \bar{D}_m y_m \right\| = \left| e_{m+1}^T \bar{U}_m^T Q_{m+1}^T b \right|.$$

3.2. Some numerical experiments.

3.2.1. The SVD approximation. As already addressed, the regularization properties of the considered methods are closely related to the ability of the projected matrices $\bar{D}_m$ to simulate the SVD properties of the matrix A. Indeed, the SVD of A is commonly considered the most useful tool for the analysis of discrete ill-posed problem (see, e.g., [36, Chapter 2]), and the TSVD is a commonly used tool for regularization (see again [36, Chapter 5]). Denoting by $A_m^{\mathcal{S}}$ the truncated singular value decomposition of A (1.4), the TSVD regularized solution of $Ax = b$ is given by the solution of the least squares problem

$$\min_{x \in \mathbb{R}^N} \left\| b - A_m^{\mathcal{S}} x \right\|.$$

When working with Krylov methods that satisfy (3.1), we have that the least-square solution of (1.1) is approximated by the solution of

$$\begin{aligned} \min_{x \in \mathcal{K}_m} \|b - Ax\| &= \min_{y \in \mathbb{R}^m} \|b - AW_m y\| = \min_{x \in \mathbb{R}^N} \left\| b - AW_m K_m^T x \right\| \\ &= \min_{x \in \mathbb{R}^N} \left\| b - Z_{m+1} \bar{D}_m K_m^T x \right\|, \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

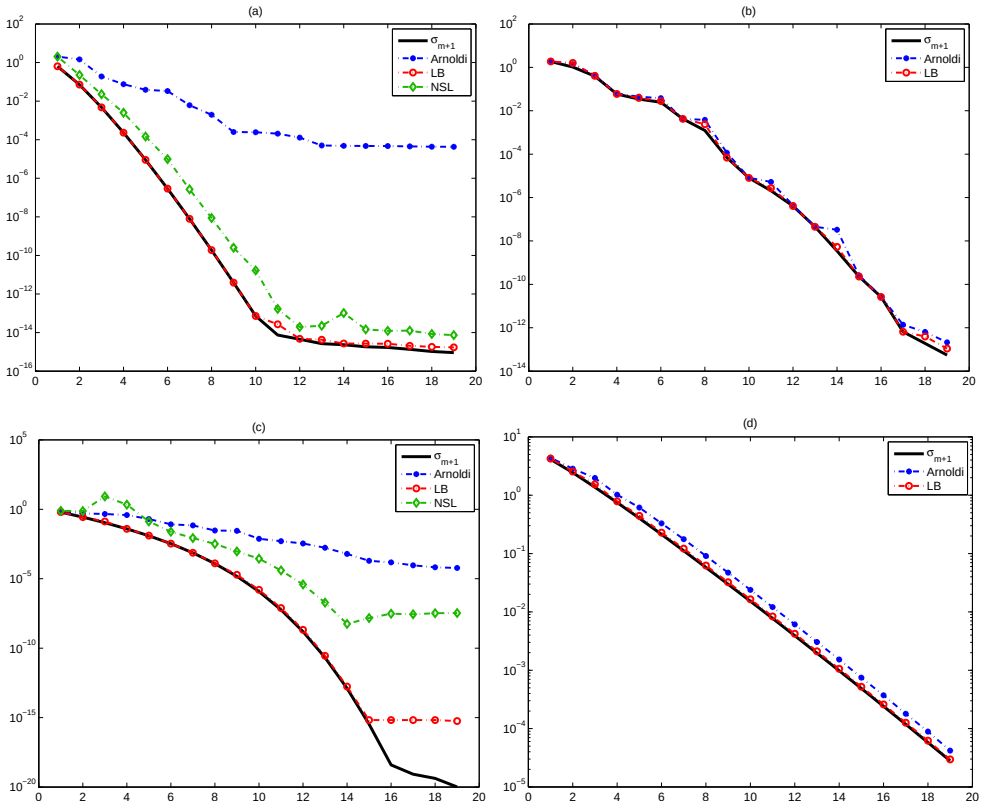

FIG. 3.1. *Plots of* $\left\|A - A_m^{\mathcal{K}}\right\|$ *with respect to the singular values of* A *for the problems* `baart` *(a),* `shaw` *(b),* `i_laplace` *(c), and* `gravity` *(d).*

where, as usual, we have assumed that W_m and K_m have full rank. The solution of the above least squares problem is approximated by taking the solution of the projected least squares problem (3.2). We again underline that in (3.2) equality holds just for LSQR and GMRES. After introducing the matrix

$$A_m^{\mathcal{K}} := Z_{m+1}\bar{D}_m K_m^T, \tag{3.6}$$

which is a sort of regularized matrix associated to the generic Krylov subspace methods defined by the factorization (3.1), we want to compare the approximation and regularization properties of the Krylov methods with the ones of the TSVD method. We do this by plotting the quantity $\left\|A - A_m^{\mathcal{K}}\right\|$ (recall the optimality property $\left\|A - A_m^{\mathcal{S}}\right\| = \sigma_{m+1}$, [30, §2.5.5]). The results are reported in Figure 3.1. The subplots (b) and (d) refer to the problems `shaw` and `gravity`, whose coefficient matrices are symmetric, so that the nonsymmetric Lanczos process (NSL) is equivalent to the Arnoldi algorithm. The Lanczos bidiagonalization process is denoted by LB.

The ability of the projected matrices $\bar{D}_m$ of approximating the dominating singular values of A has been studied in terms of the residuals in Propositions 2.3 and 2.7 for the Lanczos bidiagonalization and the Arnoldi algorithms, respectively. In Figures 3.2 and 3.3 we display graphs of some experiments for all the considered methods. The results illustrate the good approximation properties of these methods and implicitly ensure that all the methods show a very fast initial convergence, which can be measured in terms of the number of approximated singular values greater than the noise level $\|e\|/\|b^{ex}\|$.

ETNA
Kent State University
http://etna.math.kent.edu

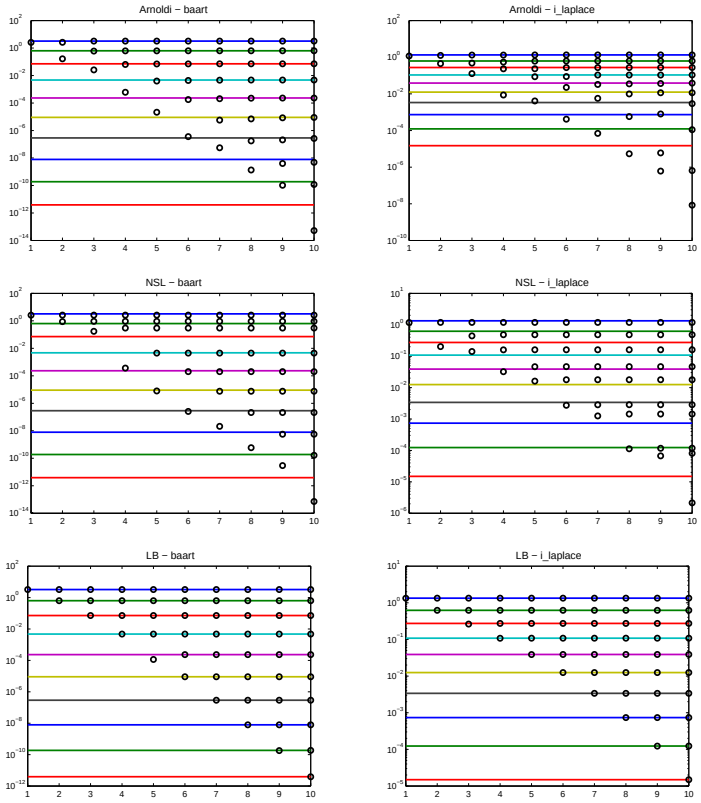

FIG. 3.2. *Approximation of the dominating singular values—the nonsymmetric case. The solid horizontal lines stand for the first singular values of A. The circles display the singular values of the matrix $\bar{D}_m$ in (3.1), where m is varied along the horizontal axis.*

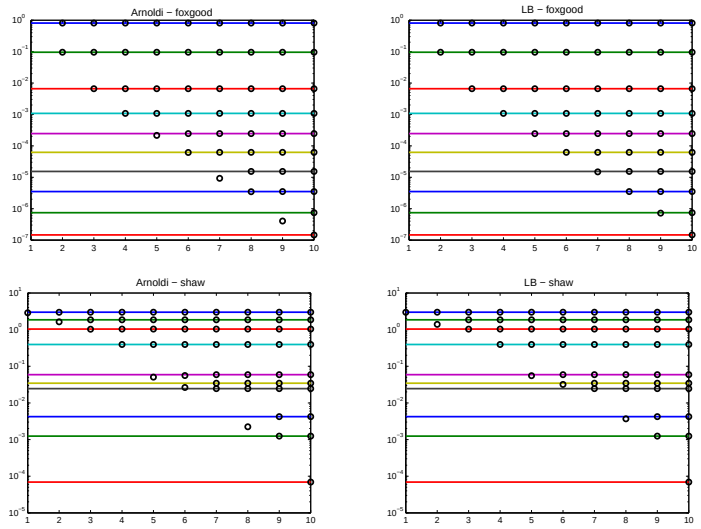

FIG. 3.3. *Approximation of the dominating singular values—the symmetric case. The layout of the plots is as described in Figure 3.2.*

ETNA
Kent State University
http://etna.math.kent.edu

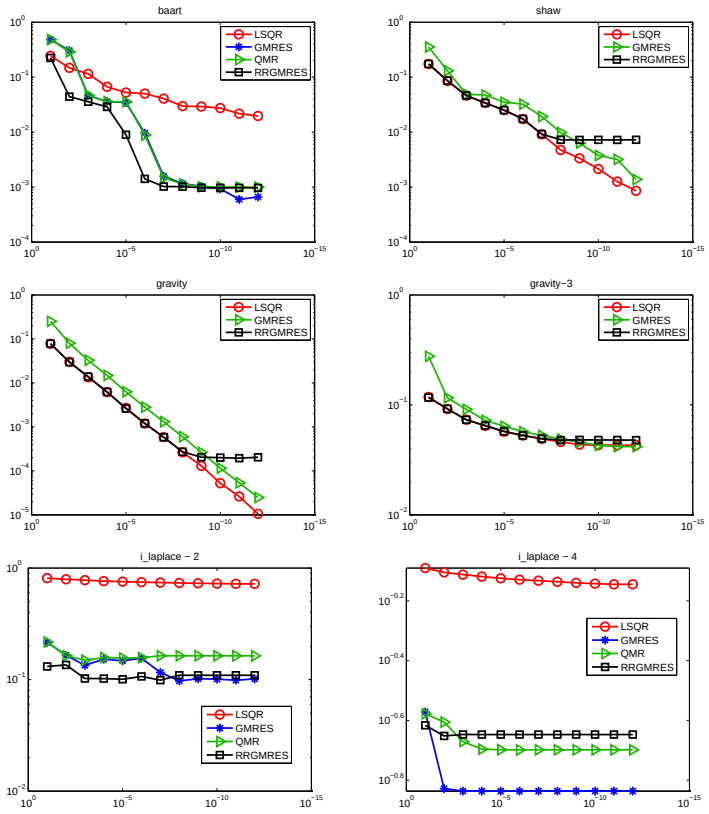

FIG. 3.4. *Optimal attainable accuracy (i.e., minimum relative error with respect to the number of iterations) versus different noise levels (from* 10^{-1} *to* 10^{-12}*). The displayed values are averages over 30 runs of the methods for each level.*

3.2.2. Accuracy analysis for standard test problems. In this section we consider the accuracy of the methods introduced in Section 2 in terms of the minimum relative error attainable (with respect to the number of performed iterations) for different noise levels from 10^{-1} to 10^{-12}. The results, on an average of 30 runs, are reported in Figure 3.4.

Whenever the noise level is relatively high, RRGMRES seems to be the most accurate method. The reason obviously lies in the use of a starting vector Ab, in which most of the noise has been removed. This fact also agrees with the results presented in [54, 55]. The difference is less evident when the noise level is small, and it is interesting to see that the attainable accuracy of RRGMRES typically stagnates around a certain level. This is the downside of the range-restricted approach. It is also interesting to observe that the methods may show little differences in presence of nonsmooth solutions (such as `gravity` - 3 and `i_laplace` - 4, where the solution is piecewise constant).

3.2.3. Stability. In order to understand the practical usefulness of the Krylov methods considered in this paper, we present some results showing how difficult it may be to exploit the potential accuracy of these methods together with their speed. As stopping (or parameter selection) rule we use the discrepancy principle [52], which is one of the most popular techniques to set the regularization parameters when the error e on the right hand side b is assumed to be of Gaussian white type and its norm is known (or well estimated). The

ETNA
Kent State University
http://etna.math.kent.edu

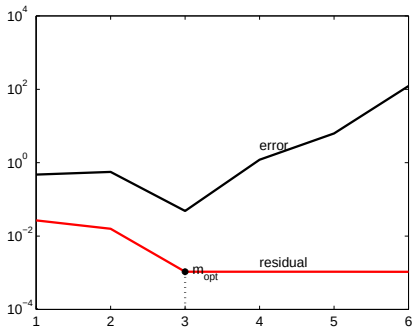

FIG. 3.5. *Problem* `baart`*: example of fast convergence/divergence behavior of GMRES.*

discrepancy principle prescribes to stop the iterations as soon as

$$\|b - Ax_m\| \leq \eta \|e\|, \tag{3.7}$$

where $\eta > 1$ (typically $\eta \approx 1$) is a safety factor. In Table 3.1 we compare the best attainable accuracy (with respect to the number of iterations) with the accuracy attained at the iteration selected by the stopping rule. We consider the average of 100 runs of the methods with different realizations of the random vector e with $\|e\| / \|b^{ex}\| = 10^{-3}$. In particular, denoting by m_{opt} the iteration number corresponding to the optimal accuracy, we also consider the accuracy at the iterations $m_{opt} - 1$ and $m_{opt} + 1$. The differences may be huge and cannot be detected by the residual norm, which is generally flat around m_{opt}; see Figure 3.5.

In this view, using the values $\eta_1 = 1.02$, $\eta_2 = 1.05$, $\eta_3 = 1.1$ for the discrepancy rule in (3.7) and denoting by m_{DP} the iteration number selected, in Table 3.1, we report the number of times in which $|m_{\mathrm{DP}} - m_{opt}| = 1$ and $|m_{\mathrm{DP}} - m_{opt}| \geq 2$, denoted by semi-failure and total failure of the stopping rule, respectively.

The results reported in Table 3.1 are rather clear: independently of the choice of the safety factor η, in many cases the stopping rule does not allow to exploit the potentials of these methods. In other words, in practice, the fast convergence/divergence of the methods makes them rather unreliable whenever the singular values of A decay very rapidly. Obviously, the situation is even more pronounced whenever $\|e\|$ is not known, and then other stopping rules such as the GCV or the L-curve need to be used.

4. Krylov methods and Tikhonov regularization. As shown in Section 3.2, the Krylov methods considered in this paper are able to obtain a good accuracy when applied to discrete ill-posed problem, but the fast transition between convergence and divergence, which is not detected by the residual, makes their practical use quite difficult. For this reason, the regularization of the projected subproblems (hybrid methods, cf. the introduction) is generally necessary.

In this setting, the standard form Tikhonov regularization of (3.2) reads

$$\min_{y \in \mathbb{R}^m} \left\{ \left\| Q_{m+1}^T b - \bar{D}_m y \right\|^2 + \lambda^2 \|y\|^2 \right\}.$$

If the regularization parameter λ is defined (at each step) independently of the original problem, i.e., with the only aim of regularizing (3.2), then the corresponding method is traditionally called hybrid; cf. again the introduction. As already addressed, regularization by Krylov methods or their use to solve the Tikhonov minimization problem has a long history dating back to [58]. Regarding GMRES, the hybrid approach called Arnoldi-Tikhonov method,

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 3.1
Stability results. Each test is performed 100 times with different noise realizations.

Method	Average error			Semi-failure			Total failure		
	m_{opt}	$m_{opt}-1$	$m_{opt}+1$	η_1	η_2	η_3	η_1	η_2	η_3
`baart`									
LSQR	0.116	0.160	2.139	56	67	78	21	18	11
GMRES	0.047	0.548	1.054	4	18	2	13	3	1
RRGMRES	0.034	0.384	0.320	25	28	30	18	14	1
QMR	0.046	0.513	0.382	10	5	10	15	10	0
`shaw`									
LSQR	0.047	0.057	0.300	22	25	39	19	66	23
GMRES	0.048	0.107	0.541	3	15	0	4	0	2
RRGMRES	0.046	0.059	0.297	56	30	65	22	65	24
`i_laplace`									
LSQR	0.140	0.145	0.190	30	66	20	70	12	86
GMRES	0.547	0.943	4.748	2	97	1	99	3	97
RRGMRES	0.429	0.891	1.034	4	96	15	85	8	91
QMR	0.048	0.107	0.541	3	15	0	97	2	98
`gravity`									
LSQR	0.138	0.018	0.026	34	55	31	63	26	74
GMRES	0.032	0.041	0.038	58	42	72	28	95	5
RRGMRES	0.014	0.018	0.026	48	42	48	46	47	91

was first considered in [10] with the basic aim of avoiding the matrix-vector multiplications with A^T used by Lanczos-bidiagonalization-type schemes.

Throughout the remainder of the paper we use Krylov methods to iteratively solve (1.2) (i.e., according to the classification given in the introduction, Krylov-Tikhonov methods), and hence we define λ step by step with the aim of regularizing the original problem. In other words, we iteratively solve a sequence of constrained minimization problems of the form

$$\min_{x\in\mathcal{K}_m}\left\{\|b-Ax\|^2+\lambda^2\|Lx\|^2\right\}. \tag{4.1}$$

In the sequel, for theoretical purposes, it will be useful to consider the following expression for the Tikhonov regularized solution

$$x_\lambda=(A^TA+\lambda^2L^TL)^{-1}A^Tb. \tag{4.2}$$

In this sense, at each step we approximate the solution of (1.2) by solving

$$\min_{y\in\mathbb{R}^m}\left\{\left\|Q_{m+1}^Tb-\bar{D}_my\right\|^2+\lambda^2\|LW_my\|^2\right\}; \tag{4.3}$$

cf. Section 3. Minimizing (4.3) is equivalent to solving the following regularized least squares problem

$$\min_{y\in\mathbb{R}^m}\left\|\begin{bmatrix}\bar{D}_m\\ \lambda LW_m\end{bmatrix}y-\begin{bmatrix}Q_{m+1}^Tb\\ 0\end{bmatrix}\right\|^2. \tag{4.4}$$

If we denote by $y_{m,\lambda}$ the solution of (4.3), then $x_{m,\lambda}=W_my_{m,\lambda}$ is the corresponding approximate solution of (1.2) and regularized solution of (1.1). It is well known that, in

ETNA
Kent State University
http://etna.math.kent.edu

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 $L = I_N$. As for the Lanczos bidiagonalization algorithm, the solution of (4.3) with $L \neq I_N$ has been considered, among the others, in [43, 46], whereas the Arnoldi algorithm has been used in [26, 27, 28, 56, 57].

It is important to observe that if $L = I_N$, then the dimension of the problem (4.4) is fully reduced whenever W_m is orthogonal while if $L \in \mathbb{R}^{P\times N}$ is a general matrix having $P \approx N$ rows, then the dimension of (4.4) inherits the dimension of the original problem. In order to fully reduce the dimension of the subproblem (4.4) when $L \neq I_N$, one could consider the "skinny" QR factorization of LW_m (see [43]), i.e.,

$$LW_m = Q_m^L L_m, \tag{4.5}$$

where $Q_m^L \in \mathbb{R}^{P\times m}$ has orthonormal columns and $L_m \in \mathbb{R}^{m\times m}$ is upper triangular. Alternatively, assuming that $P \leq N$, one could also add $N - P$ zero rows to L (which does not alter (4.1)) and consider the projection of L onto $\mathcal{K}_m(A, b)$ (see [57]), i.e.,

$$L_m = K_m^T L W_m \in \mathbb{R}^{m\times m}, \tag{4.6}$$

where K_m depends on the Krylov subspace (cf. (3.1)). In both cases, (4.3) reads

$$\begin{aligned} &\min_{y\in\mathbb{R}^m} \left\{ \left\| Q_{m+1}^T b - \bar{D}_m y \right\|^2 + \lambda^2 \left\| L_m y \right\|^2 \right\} \\ &= \min_{y\in\mathbb{R}^m} \left\| \begin{bmatrix} \bar{D}_m \\ \lambda L_m \end{bmatrix} y - \begin{bmatrix} Q_{m+1}^T b \\ 0 \end{bmatrix} \right\|^2, \qquad \begin{bmatrix} \bar{D}_m \\ \lambda L_m \end{bmatrix} \in \mathbb{R}^{(2m+1)\times m}. \end{aligned} \tag{4.7}$$

For theoretical purposes, it will be useful to consider the following expression

$$y_{m,\lambda} = (\bar{D}_m^T \bar{D}_m + \lambda^2 L_m^T L_m)^{-1} \bar{D}_m^T Q_{m+1}^T b. \tag{4.8}$$

We remark that when we consider the matrix (4.6), problem (4.7) is not equivalent to (4.1) anymore. However, the use of the matrix L_m defined in (4.6) appears natural in this framework: L_m would be the regularization operator of the projection of the Franklin-type regularization [22]

$$(A + \lambda L)x = b, \quad \lambda > 0.$$

According to our experience, employing the upper triangular L_m in (4.5) or considering the projected operator (4.6) perform about the same in terms of convergence rate and accuracy even if the latter approach requires $P \leq N$. Because of this limitation, in what follows, we always tacitly assume to work with the matrix L_m defined in (4.5). In the following we use the acronyms LBT (Lanczos-Bidiagonalization-Tikhonov), AT (Arnoldi-Tikhonov), RRAT (Range- Restricted-Arnoldi-Tikhonov), NSLT (Non-Symmetric-Lanczos-Tikhonov), and RRNSLT (Range-Restricted-Non-Symmetric-Lanczos-Tikhonov) to denote that the matrices in (4.7) have been computed by the Lanczos bidiagonalization, Arnoldi, Range-Restricted Arnoldi, nonsymmetric Lanczos algorithms, and Range-Restricted nonsymmetric Lanczos algorithms, respectively.

Now let $\bar{D}_m = \bar{U}_m \bar{S}_m \bar{X}_m^{-1}$ and $L_m = \bar{V}_m \bar{C}_m \bar{X}_m^{-1}$ be the GSVD decomposition of the matrix pair $\left(\bar{D}_m, L_m\right)$, where $\bar{U}_m \in \mathbb{R}^{(m+1)\times(m+1)}$ and $\bar{V}_m \in \mathbb{R}^{m\times m}$ are orthogonal, $\bar{X}_m \in \mathbb{R}^{m\times m}$ is nonsingular, and

$$\bar{S}_m = \begin{bmatrix} s_1^{(m)} & & \\ & \ddots & \\ & & s_m^{(m)} \\ 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}, \qquad \bar{C}_m = \begin{bmatrix} c_1^{(m)} & & \\ & \ddots & \\ & & c_m^{(m)} \end{bmatrix} \in \mathbb{R}^{m\times m}.$$

ETNA
Kent State University
http://etna.math.kent.edu

The generalized singular values of $(\bar{D}_m, L_m)$ are defined by the ratios

$$\gamma_i^{(m)} = \frac{s_i^{(m)}}{c_i^{(m)}}, \quad i = 1, \ldots, m,$$

and the columns of $\bar{U}_m$ are denoted by $\bar{u}_i^{(m)}$, $i = 1, \ldots, m+1$.

We have the following proposition, which provides an approximation of the residual $\|b - Ax_{m,\lambda}\|$ and, at the same time, can be used in some parameter-choice rules (cf. Section 5).

PROPOSITION 4.1. *Let $y_{m,\lambda}$ be the solution of (4.7). Then the (pseudo)residual satisfies (cf. (3.5))*

$$\left\|\bar{D}_m y_{m,\lambda} - Q_{m+1}^T b\right\|^2 = \sum_{i=1}^{m} \left(\frac{\lambda^2}{\gamma_i^{(m)2} + \lambda^2} (\bar{u}_i^{(m)})^T Q_{m+1}^T b \right)^2 + \left((\bar{u}_{m+1}^{(m)})^T Q_{m+1}^T b \right)^2. \tag{4.9}$$

Proof. This result simply follows by substituting the GSVD of $(\bar{D}_m, L_m)$ into (4.8) in order to obtain

$$y_{m,\lambda} = \bar{X}_m (\bar{S}_m^T \bar{S}_m + \lambda^2 \bar{C}_m^T \bar{C}_m)^{-1} \bar{S}_m^T \bar{U}_m^T Q_{m+1}^T b \tag{4.10}$$

and by replacing the above expression in $\left\|\bar{D}_m y_{m,\lambda} - Q_{m+1}^T b\right\|^2$. ∎

Some numerical experiments. In this section we provide some experiments concerning the method (4.7). We assume that the quantity $\|e\|$ is known quite accurately, and consequently we use the discrepancy principle to simultaneously select the number of iterations (stopping rule) and the value of the regularization parameter λ. Similarly to the discrete case of Section 3.2, when solving regularized problems of the form (4.7), one commonly says that the discrepancy principle is satisfied when

$$\|b - Ax_{m,\lambda}\| \leq \eta \|e\|,$$

where $\eta \gtrsim 1$. Using the same arguments as the ones employed in Section 3 for evaluating the norm of the (pseudo)residuals associated to the projection methods described by the decomposition (3.1), we have that

$$\|b - Ax_{m,\lambda}\| \approx \left\|Q_{m+1}^T b - \bar{D}_m y_{m,\lambda}\right\|,$$

and the discrepancy principle consists in solving, at each iteration m and with respect to the regularization parameter λ, the following nonlinear equation

$$\phi_m(\lambda) := \left\|Q_{m+1}^T b - \bar{D}_m y_{m,\lambda}\right\| = \eta \|e\|, \tag{4.11}$$

where $y_{m,\lambda}$ is the solution of (4.7).

Among the existing algorithms that solve (4.11) within a Krylov methods coupled with Tikhonov regularization (see, e.g., [49, 66]), the one proposed in [27] has been shown to be quite efficient and very simple to implement. Denoting by $r_m = Q_{m+1}^T b - \bar{D}_m y_m$ the (pseudo)residual applied to the unregularized linear system (i.e., $\lambda = 0$), then clearly $\phi_m(0) = \|r_m\|$. In this setting, the authors solve (4.11) after considering the linear approximation

$$\phi_m(\lambda) \approx \phi_m(0) + \lambda \chi_m, \tag{4.12}$$

ETNA
Kent State University
http://etna.math.kent.edu

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

$$\chi_m = \frac{\phi_m(\lambda_{m-1}) - \phi_m(0)}{\lambda_{m-1}}. \tag{4.13}$$

In (4.13), $\phi_m(\lambda_{m-1})$ is obtained by solving the m-dimensional problem (4.7) using the parameter $\lambda = \lambda_{m-1}$, which is computed at the previous step. Therefore, to select $\lambda = \lambda_m$ for the next step, we impose

$$\phi_m(\lambda_m) = \eta\|e\|. \tag{4.14}$$

Substituting in the linear approximation (4.12) of $\phi_m(\lambda_m)$ the expression derived in (4.13) and using the condition (4.14), one can easily obtain the following rule for λ_m:

$$\lambda_m^2 = \left| \frac{\eta\|e\| - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \right| \lambda_{m-1}^2.$$

In [27] this scheme was called secant-update method: this is the rule that we employ in the following experiments. Depending on the problem, we use the following classical regularization matrices,

$$L_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(N-1)\times N}, \tag{4.15}$$

$$L_2 = \begin{bmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2)\times N},$$

which represent scaled finite difference approximations of the first and the second derivative operators, respectively. In particular, looking at the quality of the best attainable approximation and at the regularity of the solution, we use L_1 for `shaw`, `i_laplace`, `i_laplace`-4, `gravity`, `gravity`-3, and L_2 for `baart`, `foxgood`, `gravity`-2 (piecewise linear solution). The results are reported in Figure 4.1.

5. Other parameter choice rules. In this section, we discuss some regularization parameter selection techniques that have already been proposed in the literature but have never been coupled with some of the Krylov methods considered in this paper. In the following we assume that no information on $\|e\|$ is available.

5.1. Embedded-based discrepancy principle. This strategy is a generalization of the secant-update approach (see the previous section) first proposed in [28]. This strategy has to be considered different from other well-known techniques since 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. The basic assumption is that, after just a few iterations of each Krylov method described by (3.1), the norm of the (pseudo)residual associated to the purely iterative method lies around the threshold $\|e\|$ (i.e., $\phi_m(0) \approx \|e\|$) and, despite being usually slightly decreasing, stabilizes during the following iterations. This property is rather clear since all the methods of Section 2 are based on the minimization of the (pseudo)residual. This motivates the use of the following update formula to choose the regularization parameter at the m-th iteration

$$\lambda_m^2 = \frac{\eta\phi_{m-1}(0) - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \lambda_{m-1}^2, \quad \eta \gtrsim 1.$$

ETNA
Kent State University
http://etna.math.kent.edu

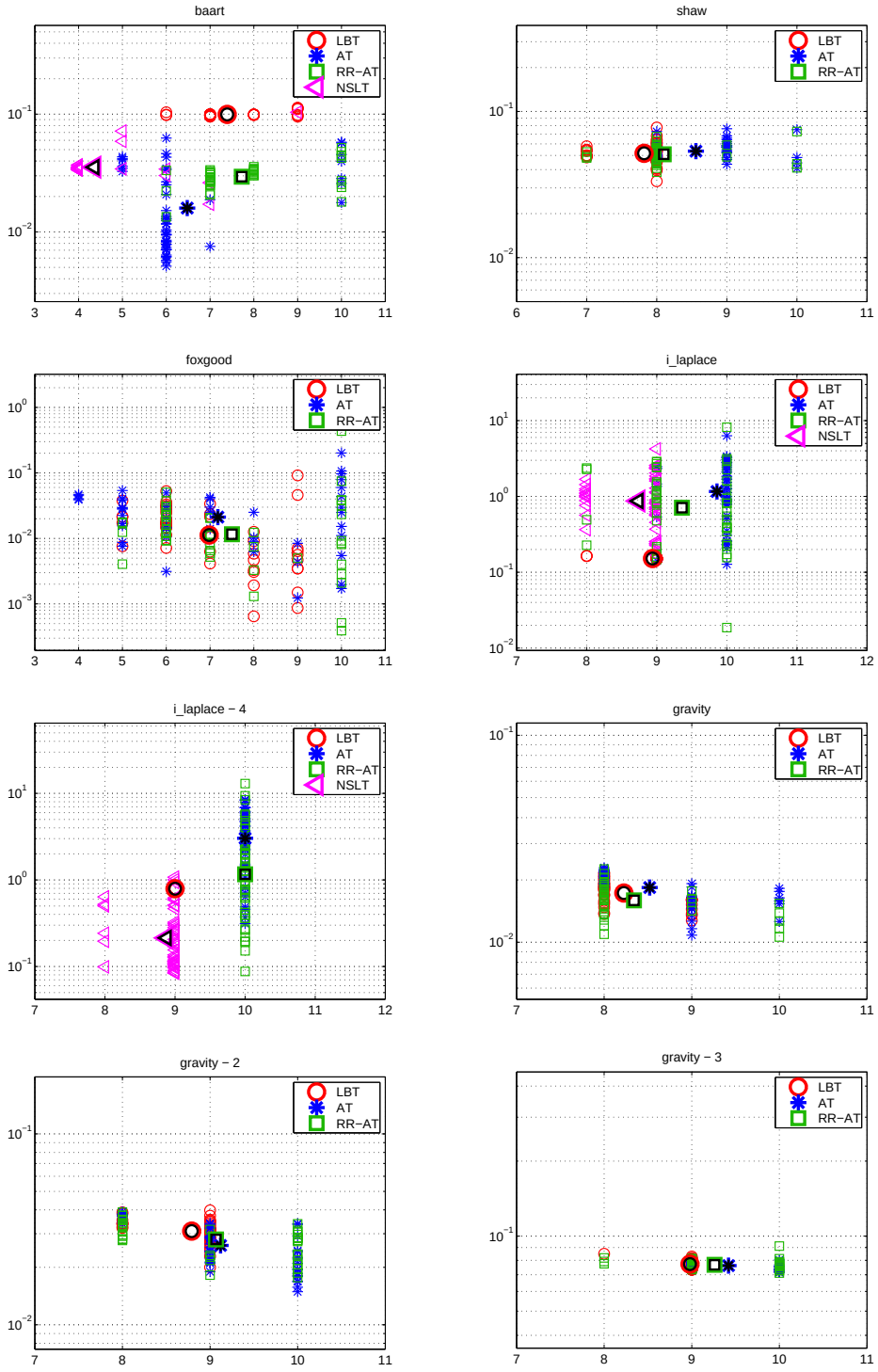

FIG. 4.1. *Accuracy of the automatically selected final approximation versus the number of iterations m. The pictures collect the relative errors $\|x^{ex} - x_{m,\lambda}\|/\|x^{ex}\|$ resulting from 30 runs of each Krylov-Tikhonov method (small markers) and the corresponding mean values (big markers). The dimension of each problem is $N = 200$, and the noise level is 10^{-3}.*

ETNA
Kent State University
http://etna.math.kent.edu

5.2. Generalized Cross Validation (GCV). The GCV parameter choice criterion prescribes to choose as regularization parameter the scalar λ that minimizes the GCV functional

$$G(\lambda) = \frac{\|(I - AA_\lambda^\sharp)b\|^2}{(\text{trace}(I - AA_\lambda^\sharp))^2}, \tag{5.1}$$

where $A_\lambda^\sharp$ stands for the regularized inverse of A associated to Tikhonov regularization (1.2); more precisely, considering the expression (4.2), we derive

$$A_\lambda^\sharp = (A^TA + \lambda^2 L^TL)^{-1}A^T.$$

To obtain an expression of $G(\lambda)$ easy to handle, one considers the GSVD of the matrix pair (A, L), defined by (1.6).

When dealing with the regularized problems (4.7), in order to set λ step by step, i.e., to define the sequence of regularization parameters $\{\lambda_m\}_{m\geq 1}$, we assume that the GSVD decomposition of the matrix pair $\left(\bar{D}_m, L_m\right)$ constitutes an increasingly better approximation of the truncated GSVD of (A, L). Similarly to Section 4, let $\bar{D}_m\bar{X}_m = \bar{U}_m\bar{S}_m$ and $L_m\bar{X}_m = \bar{V}_m\bar{C}_m$ be the GSVD of the matrix pair $(\bar{D}_m, L_m)$.

Following the approach of the recent paper [57], since the numerator of (5.1) is just the squared norm of the residual corresponding to the regularized solution and

$$(\text{trace}(I - AA_\lambda^\sharp))^2 = \sum_{i=1}^{N} \frac{\lambda^2}{\gamma_i^2 + \lambda^2},$$

where γ_i are the generalized singular values of (A, L) (cf. (1.7)), the definition of the sequence of regularization parameters $\{\lambda_m\}_{m\geq 1}$ can be obtained by means of the minimization of the functionals

$$G_m^{\mathcal{K}}(\lambda) := \frac{\sum_{i=1}^{m}\left(\frac{\lambda^2}{\gamma_i^{(m)2}+\lambda^2}(\bar{u}_i^{(m)})^TQ_{m+1}^Tb\right)^2 + \left((\bar{u}_{m+1}^{(m)})^TQ_{m+1}^Tb\right)^2}{\left(N - m + \sum_{i=1}^{m}\frac{\lambda^2}{\gamma_i^{(m)2}+\lambda^2}\right)^2}, \tag{5.2}$$

where we have used the expression of the (pseudo)residual given by Proposition 4.1. In other words, the numerator is defined using the (pseudo)residual and the denominator by replacing γ_i with $\gamma_i^{(m)}$, for $i \leq m$, and γ_i with 0, for $i \geq m+1$. Clearly, the above approximation can be obtained working in reduced dimension, and it is in perfect agreement with the formula commonly used for both Tikhonov-regularized problems and iterative methods; see [36, Chapter 7] and [17].

By considering the expression (4.8), we can immediately state that the Krylov-Tikhonov methods produce a regularized solution given by

$$x_{m,\lambda} = W_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^TL_m)^{-1}\bar{D}_m^TZ_{m+1}^Tb,$$

so that we can derive the following result, which provides a clearer interpretation of (5.2).

ETNA
Kent State University
http://etna.math.kent.edu

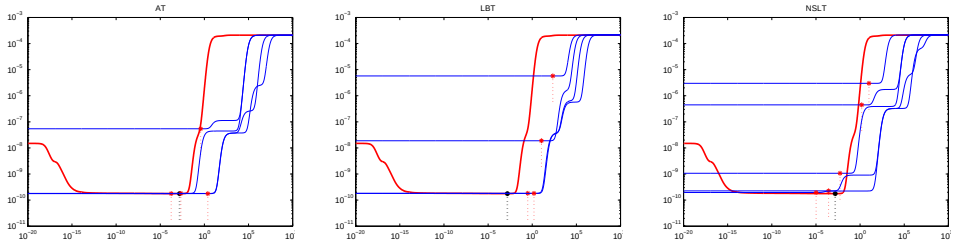

FIG. 5.1. *Problem* `baart` *with regularization matrix* $L = L_2$. *Approximations of* $G(\lambda)$ *obtained by* $G_m^{\mathcal{K}}(\lambda)$ *for some values of* m. *The markers indicate the stationary points, i.e., the selected values for* λ_m.

PROPOSITION 5.1. *For each one of the considered methods, let*

$$\left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp} = W_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^T L_m)^{-1}\bar{D}_m^T Z_{m+1}^T.$$

Then

$$G_m^{\mathcal{K}}(\lambda) \approx \frac{\left\| \left(I - A_m^{\mathcal{K}} \left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp} \right) b \right\|^2}{\left(\text{trace}\left(I - A_m^{\mathcal{K}} \left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp} \right) \right)^2},$$

where $A_m^{\mathcal{K}}$ *is given by* (3.6). *The equal sign just holds for the LBT and the AT methods.*

Proof.

$$\begin{aligned}(I - A_m^{\mathcal{K}} \left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp}) &= \left(I - Z_{m+1}\bar{D}_m K_m^T W_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^T L_m)^{-1}\bar{D}_m^T Z_{m+1}^T\right)\\ &= \left(I - Z_{m+1}\bar{D}_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^T L_m)^{-1}\bar{D}_m^T Z_{m+1}^T\right).\end{aligned}$$

Therefore

$$\begin{aligned}\left\|(I - A_m^{\mathcal{K}} \left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp})b\right\| &= \left\|b - Z_{m+1}\bar{D}_m y_{m,\lambda}\right\|\\ &= \|b - AW_m y_{m,\lambda}\|\\ &\approx \left\|Q_{m+1}^T b - \bar{D}_m y_{m,\lambda}\right\|,\end{aligned}$$

where the equal sign holds for the LBT and the AT methods (recall the discussion in Section 3). ∎

We remark that, since

$$\begin{aligned}\text{trace}(I - A_m^{\mathcal{K}} \left(A_{\lambda,m}^{\mathcal{K}}\right)^{\sharp}) &= N - \text{trace}(Z_{m+1}\bar{D}_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^T L_m)^{-1}\bar{D}_m^T Z_{m+1}^T)\\ &\approx N - \text{trace}(\bar{D}_m(\bar{D}_m^T\bar{D}_m + \lambda^2 L_m^T L_m)^{-1}\bar{D}_m^T),\end{aligned}$$

and since

$$\text{trace}(\bar{D}_m(\bar{D}_m^T\bar{D}_m + \lambda L_m^T L_m)^{-1}\bar{D}_m^T) = m - \sum_{i=1}^{m} \frac{\lambda^2}{\gamma_i^{(m)2} + \lambda^2},$$

we have fully justified the expression (5.2). In Figure 5.1 we plot some of the approximations of $G(\lambda)$ attained by $G_m^{\mathcal{K}}(\lambda)$.

ETNA
Kent State University
http://etna.math.kent.edu

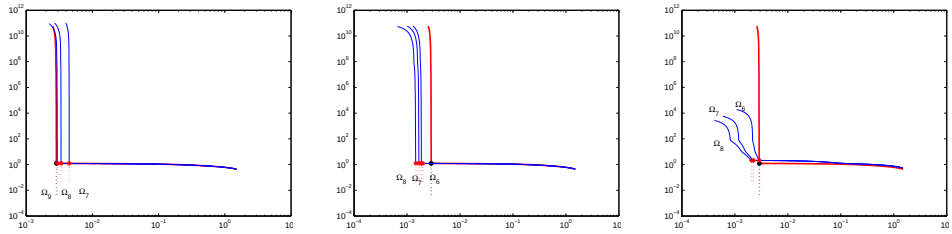

FIG. 5.2. *Problem* `baart` *with regularization matrix* $L = I$. *Approximations of* $\Omega(\lambda)$ *obtained by* $\Omega_m(\lambda)$ *for some values of* m. *The markers indicate the corners selected, i.e., the points corresponding to the chosen values for* λ_m.

5.3. L-curve criterion. The L-curve criterion [34] is based on defining the regularization parameter for (1.2) as the scalar λ that maximizes the curvature of the parametric curve

$$\Omega(\lambda) = \left(\log \|b - Ax_\lambda\|, \log \|Lx_\lambda\|\right).$$

Remarkably, this curve very often has an L-shaped dependence on λ, and its corner (i.e., the point of maximum curvature) represents a good value for λ. Indeed, if we choose the λ corresponding to the corner, we consider a compromise between the minimization of the residual and the minimization of the penalty term.

This criterion has already been used in connection with Krylov-Tikhonov methods (cf. [4, 10]). The basic idea consists in assuming that the curves

$$\Omega_m(\lambda) = \left(\log \left\|Q_{m+1}^T b - \bar{D}_m y_{m,\lambda}\right\|, \log \|L_m y_{m,\lambda}\|\right)$$

are increasingly better approximations of $\Omega(\lambda)$, so that the parameter λ_m corresponding to the corner of $\Omega_m(\lambda)$ should represent a good approximation of a suitable regularization parameter for (1.2).

PROPOSITION 5.2. *Let* $y_{m,\lambda}$ *be the solution of* (4.7). *Using the GSVD of the matrix pair* $(\bar{D}_m, L_m)$, *we obtain*

$$\|L_m y_{m,\lambda}\|^2 = \sum_{i=1}^{m} \left(\frac{\gamma_i^{(m)}}{\gamma_i^{(m)2} + \lambda^2} (\bar{u}_i^{(m)})^T Q_{m+1}^T b \right)^2. \tag{5.3}$$

Proof. Since $L_m = \bar{V}_m \bar{C}_m \bar{X}_m^{-1}$, the proof follows directly from (4.10). ∎

Using the expressions (4.9) and (5.3), the analysis of the "projected" L-curves $\Omega_m(\lambda)$ can be performed quite easily in reduced dimension. Among the existing corner-finding methods (see, e.g., [16, 40, 66]), in our experiments we use the L-curve criterion based on the adaptive algorithm referred to as "pruning algorithm" [39]. In Figure 5.2 we plot some of the approximations of $\Omega(\lambda)$ obtained with $\Omega_m(\lambda)$.

5.4. Regińska criterion. Regińska criterion [60] is a very efficient parameter choice rule for Tikhonov regularization, and it is closely related to the L-curve criterion. The regularization parameter λ is defined as the minimizer of the function

$$\Psi_\mu(\lambda) = \|b - Ax_\lambda\|^2 \|Lx_\lambda\|^{2\mu}, \quad \mu > 0,$$

for a proper μ. Analogously to the L-curve criterion, this rule is motivated by the observation that finding the minimizer of Ψ_μ corresponds to considering a good balance between the size of the regularization term and the size of the residual norm. In [60] the author proves that, if the curvature of the L-curve is maximized at λ^* and if the tangent to the L-curve at

ETNA
Kent State University
http://etna.math.kent.edu

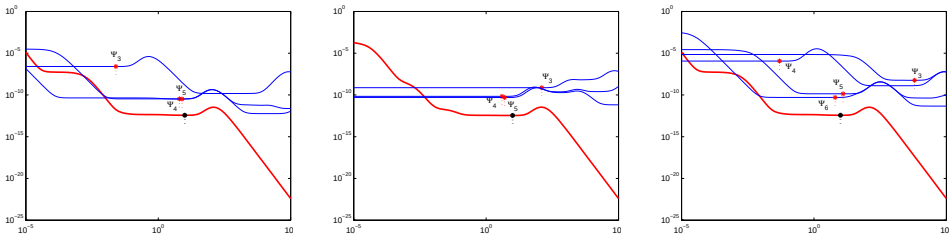

FIG. 5.3. *Problem* `baart` *with regularization matrix* $L = L_2$. *Approximations of* $\Psi_\mu(\lambda)$ *obtained with* $\Psi_{\mu,m}(\lambda)$ *for some values of* m. *The markers indicate the stationary points, i.e., the selected values for* λ_m. *In each picture* $\mu = 1$.

$(\log \|b - Ax_{\lambda^*}\|, \log \|Lx_{\lambda^*}\|)$ has slope $-1/\mu$, then $\Psi_\mu(\lambda)$ is minimized at λ^*. The authors of [61] derive a modification of the basic Regińska criterion in order to improve its performance when dealing with a discrete regularization parameter.

As in the previous cases, in order to set λ_m step by step when dealing with the projected regularized problems (4.7), we consider the function

$$\Psi_{\mu,m}(\lambda) = \left\|Q_{m+1}^T b - \bar{D}_m y_{m,\lambda}\right\|^2 \|L_m y_{m,\lambda}\|^{2\mu}, \quad \mu > 0,$$

which can be written in terms of the generalized singular values of $\left(\bar{D}_m, L_m\right)$ by using again (4.9) and (5.3). In Figure 5.3 we plot some of the approximations of $\Psi_\mu(\lambda)$ obtained with $\Psi_{\mu,m}(\lambda)$; we choose $\mu = 1$.

5.5. Numerical experiments. In order to check the performance of the considered Krylov methods together with the parameter selection strategies just outlined, here we present some experiments in which each method is coupled with the four criteria (Sections 5.1–5.4). In each picture, 50 runs of each method have been executed, considering different realizations of the random noise. The final approximations have been selected by checking the relative residual. In particular each run is stopped whenever

$$\frac{\left|\|r_m\| - \|r_{m-1}\|\right|}{\|r_m\|} \leq \varepsilon, \tag{5.4}$$

where $\varepsilon = 1.05$. The results in terms of relative error versus number of iterations are reported in Figure 5.4. These pictures (together with many other that are not reported) reveal that the four criteria are somehow equivalent when coupled with the Krylov methods considered here since it is not easy to detect the one that clearly overtakes the others.

6. Image deblurring and denoising. As already addressed in the introduction, regularization techniques based on Krylov subspace methods are particularly effective when applied to image restoration problems. Many papers have been devoted to studying the performances of different Krylov methods when applied to the denoising and deblurring of corrupted images: among the most recent ones we cite [1, 7, 25, 38, 56]. In this section we closely follow the approach adopted in [38], and we consider a medical and an astronomical test image of size 256×256 pixels distorted by three different kinds of spatially invariant blurs (isotropic, non-isotropic and experimentally defined). The boundary conditions are set to zero for all the tests (cf. Figure 6.1). Both the isotropic and the non-isotropic blurs are analytically defined starting from a Gaussian point spread function (PSF) $K(s,t)$, i.e.,

$$K(s,t) = \frac{1}{2\pi\sqrt{\alpha_1^2\alpha_2^2 - \rho^4}} \exp\left(-\frac{1}{2}\begin{bmatrix} s & t \end{bmatrix}\begin{bmatrix} \alpha_1^2 & \rho^2 \\ \rho^2 & \alpha_2^2 \end{bmatrix}^{-1}\begin{bmatrix} s \\ t \end{bmatrix}\right).$$

ETNA
Kent State University
http://etna.math.kent.edu

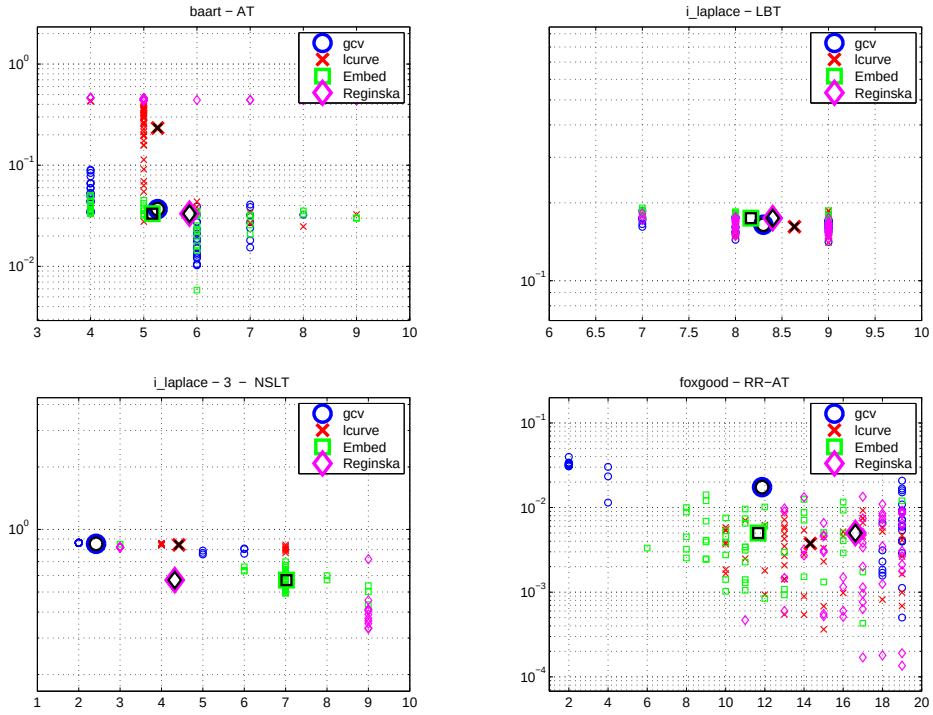

FIG. 5.4. *Relative error of the final approximation selected by* (5.4) *versus the number of iterations. The pictures collect the results of 50 runs of each method (small markers) and the corresponding mean values (big markers). The dimension of each problem is* $N = 200$*, and the noise level is* 10^{-2}.

In the isotropic case, we take $\alpha_1 = \alpha_2 = 3$ and $\rho = 0$; in the non-isotropic case we take $\alpha_1 = 10$, $\alpha_2 = 8$, and $\rho = 4$; in the experimental case we use the data made available in [53], which simulate how an extraterrestrial satellite can be detected from ground-based telescopes. Once the PSFs have been set, the corresponding blurring matrices are generated by employing the *Restore Tools* [53] routines: the matrix associated to the isotropic blur is symmetric, while the matrices associated to the non-isotropic and the experimental blurs are nonsymmetric. We further perturb the blurred images by adding 5% Gaussian white noise in the medical image case and 1% Gaussian white noise in the astronomical image case.

Let us focus on the symmetrically blurred medical image. Since in this case both the Arnoldi and the Lanczos nonsymmetric algorithms reduce to the Lanczos tridiagonalization algorithm, we take into account just the AT, the RRAT, and the LBT methods. We extensively test both the standard and the general form regularization, taking

$$L = L_1 \otimes I_N + I_N \otimes L_1, \tag{6.1}$$

where L_1 is defined in (4.15). The regularization matrix L represents the sum of the first derivatives in the horizontal and vertical directions of the two-dimensional image. In Tables 6.1 and 6.2 we collect the results obtained by testing the AT, RRAT, and LBT methods: we run each method 50 times considering different noise realizations, and the average values are displayed. In particular, we are interested in comparing the performances of the various methods and parameter choice strategies (cf. Section 5): for this reason we report the relative errors and the regularization parameters attained when an appropriate stopping criterion is satisfied. Moreover, to evaluate the efficiency of each method, in Table 6.3 we report the average of the minimum relative errors (with respect to the number of iterations) and the

ETNA
Kent State University
http://etna.math.kent.edu

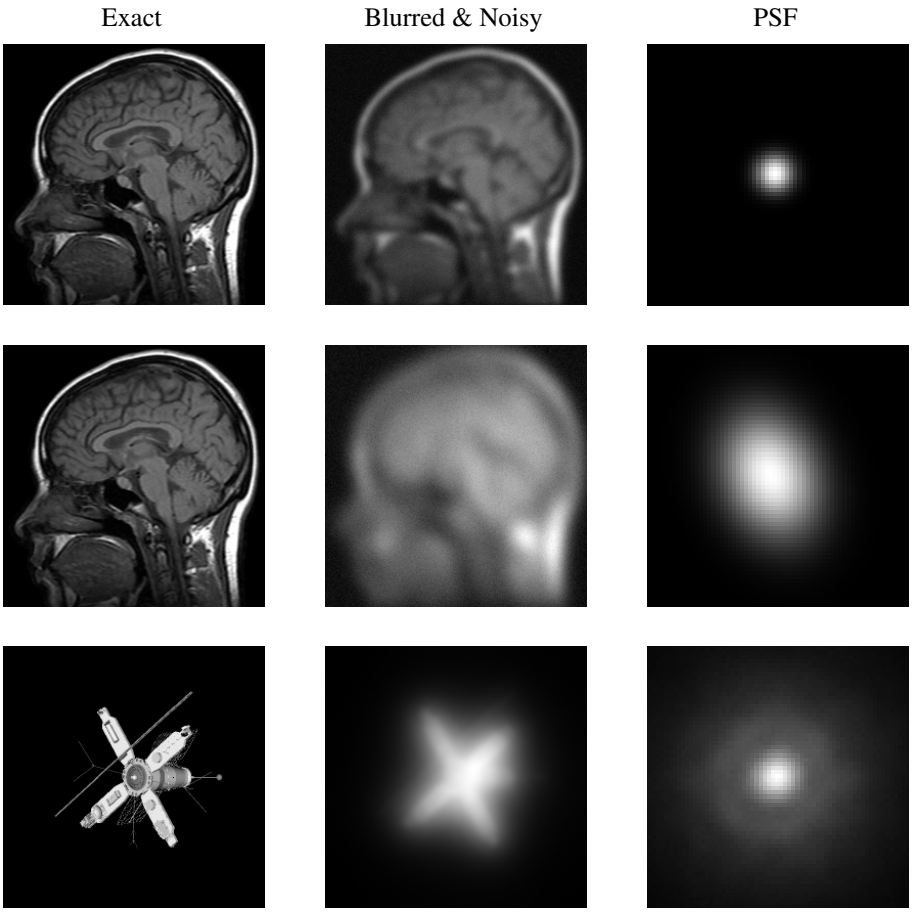

FIG. 6.1. *Test images employed in this section. In the first column we display the ideally exact image, in the second column we display the blurred and noisy image, and in the third column we display a blow-up (400%) of the PSFs.*

average of the iterations at which it is attained. In Figure 6.2 we display the history of the relative errors when a "quasi-optimal" regularization parameter is chosen. In Figure 6.3 we show the best reconstructions obtained by the general-form AT, RRAT, and LBT methods after 12 iterations have been performed. To set the "quasi-optimal" regularization parameter at each iteration, we consider a fixed set of trial regularization parameters, and we choose the one that delivers the minimum relative error. Of course, this strategy is possible only if the exact solution is available. However, we decide to consider it in order to show the regularizing properties of the different Krylov subspace methods applied to the Tikhonov problem independently on the strategy employed to set the parameter.

So far, we can see that the AT and LBT methods seem to outperform the RRAT method both in terms of efficiency and quality of the reconstruction. With respect to the LBT method, the AT method needs considerably less iterations to deliver good approximations of the exact solution, and including a regularization matrix different from the identity leads to more accurate reconstructions. On the downside, while the LBT method seems to be very robust with respect to the choice of the regularization parameters (and often even very tiny regularization parameters are suitable in the LBT case), the performance of the AT methods seems to be much more dependent on an accurate tuning of the regularization parameter.

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 6.1
Averages of the results obtained running 50 times the isotropic deblurring and denoising problem; Tikhonov regularization in standard form is employed. In the first three columns we display the average of the attained relative errors, and the average number of required iterations (between brackets); in the last three columns we display the average of the corresponding regularization parameters.

Relative Error			
	LBT	AT	RRAT
DISCREPANCY	$2.0745 \cdot 10^{-1}$ (9.28)	$3.2355 \cdot 10^{-1}$ (4)	$3.5581 \cdot 10^{-1}$ (30)
EMBEDDED	$2.2291 \cdot 10^{-1}$ (7)	$2.4345 \cdot 10^{-1}$ (6)	$3.0491 \cdot 10^{-1}$ (5)
GCV	$2.2487 \cdot 10^{-1}$ (6)	$3.7878 \cdot 10^{-1}$ (4)	$3.0909 \cdot 10^{-1}$ (3)
L-CURVE	$2.1811 \cdot 10^{-1}$ (14)	$2.1350 \cdot 10^{-1}$ (6)	$3.1318 \cdot 10^{-1}$ (9)
REGIŃSKA	$2.2565 \cdot 10^{-1}$ (6)	$3.1209 \cdot 10^{-1}$ (4)	$3.0434 \cdot 10^{-1}$ (4)
Regularization Parameter			
DISCREPANCY	$7.3730 \cdot 10^{-2}$	$6.3057 \cdot 10^{-2}$	$3.2389 \cdot 10^{-1}$
EMBEDDED	$9.6528 \cdot 10^{-2}$	$2.7193 \cdot 10^{-2}$	$5.2685 \cdot 10^{-2}$
GCV	$6.6069 \cdot 10^{-4}$	$5.0119 \cdot 10^{-4}$	$3.9811 \cdot 10^{-5}$
L-CURVE	$1.4010 \cdot 10^{-1}$	$4.7940 \cdot 10^{-2}$	$1.0251 \cdot 10^{-1}$
REGIŃSKA	$5.0119 \cdot 10^{-2}$	$5.0119 \cdot 10^{-2}$	$1.9953 \cdot 10^{-3}$

TABLE 6.2
Averages of the results obtained running 50 times the isotropic deblurring and denoising problem; Tikhonov regularization in general form is employed. In the first three columns we display the average of the attained relative errors, and the average number of required iterations (between brackets); in the last three columns we display the average of the corresponding regularization parameters.

Relative Error			
	LBT	AT	RRAT
DISCREPANCY	$2.0821 \cdot 10^{-1}$ (9.52)	$2.0047 \cdot 10^{-1}$ (6.98)	$3.6561 \cdot 10^{-1}$ (2)
EMBEDDED	$2.2363 \cdot 10^{-1}$ (8)	$2.1087 \cdot 10^{-1}$ (6)	$3.0545 \cdot 10^{-1}$ (6)
GCV	$2.2487 \cdot 10^{-1}$ (6)	$3.7877 \cdot 10^{-1}$ (4)	$3.0909 \cdot 10^{-1}$ (3)
L-CURVE	$3.6390 \cdot 10^{-1}$ (2)	$2.0161 \cdot 10^{-1}$ (7)	$3.6564 \cdot 10^{-1}$ (2)
REGIŃSKA	$2.4059 \cdot 10^{-1}$ (6)	$2.3710 \cdot 10^{-1}$ (7)	$3.0458 \cdot 10^{-1}$ (4)
Regularization Parameter			
DISCREPANCY	$1.7492 \cdot 10^{0}$	$4.3601 \cdot 10^{-2}$	$1.3591 \cdot 10^{1}$
EMBEDDED	$2.1745 \cdot 10^{-1}$	$4.7113 \cdot 10^{-2}$	$1.4447 \cdot 10^{-1}$
GCV	$1.9953 \cdot 10^{-3}$	$2.5119 \cdot 10^{-4}$	$3.1623 \cdot 10^{-4}$
L-CURVE	$1.4030 \cdot 10^{1}$	$6.3991 \cdot 10^{-2}$	$1.4030 \cdot 10^{1}$
REGIŃSKA	$3.9811 \cdot 10^{-1}$	$3.9811 \cdot 10^{-1}$	$1.9953 \cdot 10^{-3}$

Then, let us consider the non-isotropically blurred medical image. In this case we take into account all the methods previously described, i.e., the AT, RRAT, LBT, NSLT, and RRNSLT methods. Similarly to what we have done in the symmetric case, we consider both the standard and the general form regularization, where we still employ the matrix L defined in (6.1). As before, in Tables 6.4, 6.5, and 6.6 we collect the averages of the results obtained running 50 times each test with different noise realizations. Figure 6.4 displays the history of the relative errors when a "quasi-optimal" regularization parameter is chosen. The remarks just made about the performance of the different Krylov-Tikhonov methods in the symmetric case still hold for the unsymmetric blur. Moreover, looking at the results in Tables 6.4, 6.5, and Figure 6.4, we can clearly see that the performances of the AT and NSLT methods are very similar.

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 6.3
Averages of the minimum relative errors obtained running 50 times the isotropic deblurring and denoising problem; between brackets we display the average number of iterations required to attain such minima.

	LBT	AT	RRAT
Standard Form			
DISCREPANCY	$1.96 \cdot 10^{-1}$ (30)	$1.94 \cdot 10^{-1}$ (9)	$3.56 \cdot 10^{-1}$ (30)
EMBEDDED	$1.88 \cdot 10^{-1}$ (30)	$2.42 \cdot 10^{-1}$ (5)	$3.02 \cdot 10^{-1}$ (10)
GCV	$1.89 \cdot 10^{-1}$ (19.3)	$2.54 \cdot 10^{-1}$ (2)	$2.98 \cdot 10^{-1}$ (23.8)
L-CURVE	$1.88 \cdot 10^{-1}$ (30)	$1.99 \cdot 10^{-1}$ (9)	$2.98 \cdot 10^{-1}$ (30)
REGIŃSKA	$1.87 \cdot 10^{-1}$ (28.2)	$1.88 \cdot 10^{-1}$ (11.4)	$2.98 \cdot 10^{-1}$ (23)
General Form			
DISCREPANCY	$2.07 \cdot 10^{-1}$ (13.3)	$1.90 \cdot 10^{-1}$ (9)	$3.66 \cdot 10^{-1}$ (2)
EMBEDDED	$1.89 \cdot 10^{-1}$ (30)	$1.91 \cdot 10^{-1}$ (9.1)	$3.05 \cdot 10^{-1}$ (23.1)
GCV	$1.89 \cdot 10^{-1}$ (19.3)	$2.54 \cdot 10^{-1}$ (2)	$2.98 \cdot 10^{-1}$ (23.8)
L-CURVE	$3.64 \cdot 10^{-1}$ (2)	$1.87 \cdot 10^{-1}$ (27.3)	$3.66 \cdot 10^{-1}$ (2)
REGIŃSKA	$2.38 \cdot 10^{-1}$ (9)	$2.36 \cdot 10^{-1}$ (6)	$2.98 \cdot 10^{-1}$ (23)

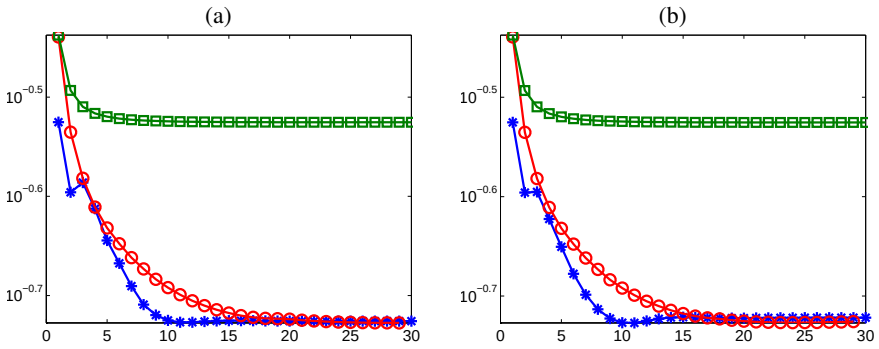

FIG. 6.2. *History of the best relative errors attainable when applying Krylov-Tikhonov regularization in standard form (frame (a)) and general form (frame (b)) to the symmetrically blurred medical image. In both frames, the AT method is denoted by an asterisk, the RRAT method is denoted by a square, and the LBT method is denoted by a circle.*

In Figure 6.5, we also perform a Discrete Cosine Transform (DCT) analysis: we display just the results relative to the general form regularization since the results relative to the standard form regularization are very similar. The regularization parameter is chosen according to the "quasi-optimal" strategy. This analysis is based on splitting the "exact" and the perturbed components. Recalling that $b = b^{ex} + e$ and adopting the general notation of Proposition 5.1, we write

$$x_m^{ex} = (A_{\lambda,m}^{\mathcal{K}})^{\sharp} b^{ex} \quad \text{and} \quad x_m^{e} = (A_{\lambda,m}^{\mathcal{K}})^{\sharp} e.$$

We remark that x_m^{ex} is still dependent on the noise component in b since the considered Krylov subspaces are generated taking b, $A^T b$, or Ab as starting vector. The goal of this analysis is to understand how the noise propagates during the Krylov-Tikhonov iterations and to assess the quality of the restoration that can be achieved performing different Krylov-Tikhonov methods. In this way we extend the analysis performed in [38] for the purely iterative methods. Looking at Figure 6.5 we can clearly see that, although the basis of the Krylov subspaces associated to the considered methods are quite different (cf. again the analysis in [37], or [38]), the properties of the reconstructed solutions obtained by all the Krylov-Tikhonov methods are similar. As

ETNA
Kent State University
http://etna.math.kent.edu

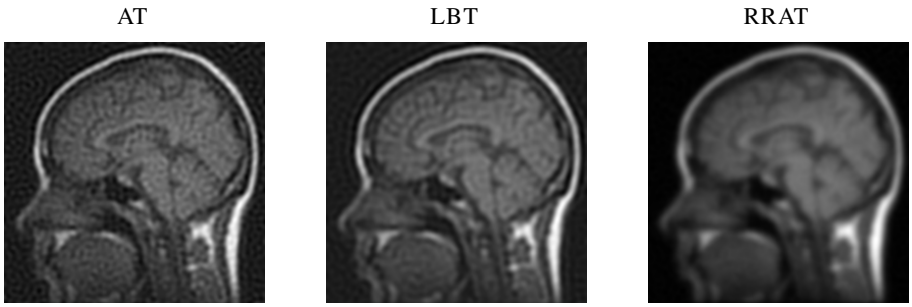

FIG. 6.3. *Best reconstructions obtained at the 12-th iteration of Krylov-Tikhonov regularization methods in general form applied to the symmetrically blurred medical image.*

TABLE 6.4
Averages of the relative errors obtained running 50 times the non-isotropic deblurring and denoising problem; Tikhonov regularization in standard form is employed. Between brackets we report the average number of iterations required to satisfy the stopping criterion.

	LBT	AT	NSLT	RRAT	RRNSLT
DISCREPANCY	$4.05 \cdot 10^{-1}$ (17.6)	$6.64 \cdot 10^{-1}$ (6)	$6.64 \cdot 10^{-1}$ (6)	$5.19 \cdot 10^{-1}$ (30)	$5.19 \cdot 10^{-1}$ (30)
EMBEDDED	$4.45 \cdot 10^{-1}$ (7)	$4.45 \cdot 10^{-1}$ (6)	$4.45 \cdot 10^{-1}$ (6)	$4.94 \cdot 10^{-1}$ (5)	$4.94 \cdot 10^{-1}$ (5)
GCV	$4.38 \cdot 10^{-1}$ (7)	$5.26 \cdot 10^{-1}$ (4)	$5.26 \cdot 10^{-1}$ (4)	$4.98 \cdot 10^{-1}$ (3)	$4.98 \cdot 10^{-1}$ (3)
L-CURVE	$4.42 \cdot 10^{-1}$ (13)	$4.32 \cdot 10^{-1}$ (6)	$4.32 \cdot 10^{-1}$ (6)	$4.99 \cdot 10^{-1}$ (9)	$4.99 \cdot 10^{-1}$ (9)
REGIŃSKA	$4.39 \cdot 10^{-1}$ (7)	$4.83 \cdot 10^{-1}$ (4)	$4.83 \cdot 10^{-1}$ (4)	$4.94 \cdot 10^{-1}$ (4)	$4.94 \cdot 10^{-1}$ (4)

clearly shown in column (b), all the images in column (a) are dominated by low-frequency components (i.e., their spectral components are mostly located in the upper left corner). In particular, the image x_m^{ex} obtained by the RRAT method seems to be the most low-frequency one. This behavior is quite natural if we consider the slow performance of the RRAT method (cf. Figure 6.4, frame (b)). Among the other images x_m^{ex}, the one produced by the LBT method is slightly more low-frequent than the one produced by the AT method. All the images in column (c) appear to be dominated by bandpass-filtered noise in the form of freckles, which are in connection with the image contours. This description is coherent with the information displayed in column (d), where the dominating frequencies appear inside a bandlimited ring. In the RRAT and LBT cases, the freckles in (c) are slightly less enhanced, and the ring in (d) is slightly narrower. The same analysis has been performed for the previously considered symmetric case, and the results are very similar.

Finally, let us consider the satellite test image. Due to the well-marked edges of the ideally exact image (cf. Figure 6.1), we exclusively consider general-form Tikhonov regularization equipped with particular regularization matrices that are adaptively defined in order to increasingly better approximate a TV-like regularization method [67]. Basically, after a suitable number of iterations has been performed (according to a fixed parameter choice strategy), we restart the underlying Arnoldi, Lanczos bidiagonalization, or nonsymmetric Lanczos algorithms, and we define a new regularization matrix of the form $L = \widetilde{D}_m L_1^{hv}$, where $\widetilde{D}_m$ is a suitable diagonal weighting matrix dependent on the last computed solution and

$$L_1^{hv} = \begin{bmatrix} L_1 \otimes I_N \\ I_N \otimes L_1 \end{bmatrix}, \quad L_1 \text{ given in (4.15)}.$$

This approach was first derived in [25], where the authors consider the AT and the discrepancy principle. We refer to this paper for some additional details about the choice of the matrix $\widetilde{D}_m$ and the number of restarts to be performed. In Figure 6.6 we display the history of

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 6.5
Averages of the relative errors obtained running 50 times the non-isotropic deblurring and denoising problem; Tikhonov regularization in general form is employed. Between brackets we report the average number of iterations required to satisfy the stopping criterion.

	LBT	AT	NSLT	RRAT	RRNSLT
DISCREPANCY	$4.05 \cdot 10^{-1}$ (17.6)	$4.24 \cdot 10^{-1}$ (9)	$4.24 \cdot 10^{-1}$ (9)	$5.24 \cdot 10^{-1}$ (2)	$8.73 \cdot 10^{-1}$ (2)
EMBEDDED	$4.44 \cdot 10^{-1}$ (8)	$4.19 \cdot 10^{-1}$ (7)	$4.19 \cdot 10^{-1}$ (7)	$4.95 \cdot 10^{-1}$ (6)	$8.25 \cdot 10^{-1}$ (6)
GCV	$4.38 \cdot 10^{-1}$ (7)	$5.26 \cdot 10^{-1}$ (4)	$5.26 \cdot 10^{-1}$ (4)	$4.98 \cdot 10^{-1}$ (3)	$8.30 \cdot 10^{-1}$ (3)
L-CURVE	$5.23 \cdot 10^{-1}$ (2)	$4.22 \cdot 10^{-1}$ (7)	$4.22 \cdot 10^{-1}$ (7)	$5.08 \cdot 10^{-1}$ (6)	$8.46 \cdot 10^{-1}$ (6)
REGIŃSKA	$4.92 \cdot 10^{-1}$ (5)	$4.95 \cdot 10^{-1}$ (6.4)	$4.95 \cdot 10^{-1}$ (6.4)	$4.94 \cdot 10^{-1}$ (4)	$4.94 \cdot 10^{-1}$ (4)

TABLE 6.6
Averages of the minimum relative errors obtained running 50 times the non-isotropic deblurring and denoising problem; between brackets we display the average number of iterations required to attain such minima.

	LBT	AT	NSLT	RRAT	RRNSLT
Standard Form					
DISCREPANCY	$3.99 \cdot 10^{-1}$ (28.8)	$3.91 \cdot 10^{-1}$ (12)	$3.91 \cdot 10^{-1}$ (12)	$5.19 \cdot 10^{-1}$ (30)	$5.19 \cdot 10^{-1}$ (30)
EMBEDDED	$3.99 \cdot 10^{-1}$ (30)	$3.86 \cdot 10^{-1}$ (13.1)	$3.86 \cdot 10^{-1}$ (13.1)	$4.91 \cdot 10^{-1}$ (11)	$4.91 \cdot 10^{-1}$ (11)
GCV	$3.866 \cdot 10^{-1}$ (30)	$4.56 \cdot 10^{-1}$ (2)	$4.56 \cdot 10^{-1}$ (2)	$4.87 \cdot 10^{-1}$ (30)	$4.87 \cdot 10^{-1}$ (30)
L-CURVE	$4.05 \cdot 10^{-1}$ (30)	$3.76 \cdot 10^{-1}$ (19.2)	$3.76 \cdot 10^{-1}$ (19.2)	$4.88 \cdot 10^{-1}$ (30)	$4.88 \cdot 10^{-1}$ (30)
REGIŃSKA	$3.99 \cdot 10^{-1}$ (30)	$3.96 \cdot 10^{-1}$ (16)	$3.96 \cdot 10^{-1}$ (16)	$4.87 \cdot 10^{-1}$ (30)	$4.87 \cdot 10^{-1}$ (30)
General Form					
DISCREPANCY	$4.04 \cdot 10^{-1}$ (24.1)	$3.82 \cdot 10^{-1}$ (14)	$3.82 \cdot 10^{-1}$ (14)	$5.26 \cdot 10^{-1}$ (2)	$5.24 \cdot 10^{-1}$ (2)
EMBEDDED	$4.03 \cdot 10^{-1}$ (30)	$3.81 \cdot 10^{-1}$ (14.9)	$3.81 \cdot 10^{-1}$ (14.9)	$4.94 \cdot 10^{-1}$ (8)	$4.94 \cdot 10^{-1}$ (8)
GCV	$3.87 \cdot 10^{-1}$ (30)	$4.56 \cdot 10^{-1}$ (3)	$4.56 \cdot 10^{-1}$ (3)	$4.87 \cdot 10^{-1}$ (30)	$4.87 \cdot 10^{-1}$ (30)
L-CURVE	$5.23 \cdot 10^{-1}$ (2)	$3.74 \cdot 10^{-1}$ (19.2)	$3.74 \cdot 10^{-1}$ (19.2)	$4.87 \cdot 10^{-1}$ (30)	$4.87 \cdot 10^{-1}$ (30)
REGIŃSKA	$4.90 \cdot 10^{-1}$ (4)	$4.88 \cdot 10^{-1}$ (2)	$4.88 \cdot 10^{-1}$ (2)	$4.87 \cdot 10^{-1}$ (30)	$4.87 \cdot 10^{-1}$ (30)

the relative errors obtained by projecting the Tikhonov-regularized problem into the Krylov subspaces associated to the Arnoldi, Lanczos bidiagonalization, and nonsymmetric Lanczos algorithms; different parameter choice strategies are taken into account.

For all the methods we allow at most 40 inner iterations and 20 restarts. We can state that, typically, many steps are performed during the first set of iterations and then, as soon as the first restart happens, the stopping criterion is almost immediately fulfilled and a few iterations are considered (i.e., the number of iterations at each cycle decreases as the number of restarts increases). This is due to the fact that, when more restarts are considered, an increasingly more accurate initial guess for the solution is available. Of course the performance of the method depends on the particular Krylov subspace taken into account. As previously remarked, the AT method is the fastest one, while the RRAT method is the slowest one; the LBT method is the most stable one.

In Figure 6.7 we display an example of the reconstructions obtained at the end of the iterative process when the embedded parameter choice strategy is employed. As in the previous examples, in Table 6.7 we collect the results obtained by extensively testing the AT, RRAT, LBT, and NSLT methods: we run each method 50 times considering different noise realizations, and the average values are displayed. In order to compare the performances of the different methods and the different parameter choice strategies, we report the relative errors attained at the end of the outer iteration cycle and the total number iterations (i.e., the sum of the iterations performed during each restart). Moreover, to assess the quality of the restoration achieved by each Krylov-Tikhonov method, in Table 6.8 we report the best attainable relative error (with respect to the number of steps) and the number of required iterations. The worst reconstructions are associated to the RRAT method. This is due to the fact that many iterations are required to deliver a suitable reconstruction and therefore, when a restart happens (i.e.,

ETNA
Kent State University
http://etna.math.kent.edu

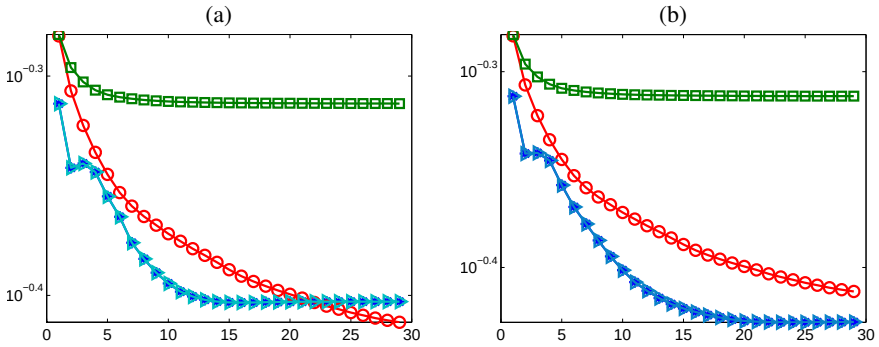

FIG. 6.4. *History of the best relative errors attainable when applying Krylov-Tikhonov regularization in standard form (frame (a)) and general form (frame (b)) to the unsymmetrically blurred medical image. In both frames, the AT method is denoted by an asterisk, the RRAT method is denoted by a square, the LBT method is denoted by a circle, and the NSLT method is denoted by a triangle; since the relative errors associated to the RRNSLT method basically coincide with the RRAT ones, they are omitted.*

TABLE 6.7

Averages of the relative errors obtained running 50 times the test problem associated to the satellite image; the values displayed are obtained after all the prescribed restarts have been performed. Between brackets we display the average of the total number of required iterations.

	LBT	AT	NSLT
DISCREPANCY	$3.6679 \cdot 10^{-1}$ (89.03)	$3.4162 \cdot 10^{-1}$ (66)	$3.4162 \cdot 10^{-1}$ (66)
EMBEDDED	$6.2061 \cdot 10^{-1}$ (64)	$3.3932 \cdot 10^{-1}$ (70)	$3.3928 \cdot 10^{-1}$ (70)
GCV	$4.0588 \cdot 10^{-1}$ (73)	$6.9823 \cdot 10^{-1}$ (81.13)	$6.8135 \cdot 10^{-1}$ (79.27)
L-CURVE	$3.5585 \cdot 10^{-1}$ (86.60)	$3.0717 \cdot 10^{-1}$ (79.17)	$3.0274 \cdot 10^{-1}$ (83.97)
REGIŃSKA	$4.4106 \cdot 10^{-1}$ (65.50)	$3.5297 \cdot 10^{-1}$ (71)	$3.5199 \cdot 10^{-1}$ (71)

		RRAT	RRNSLT
DISCREPANCY		$7.6822 \cdot 10^{-1}$ (181)	$7.6761 \cdot 10^{-1}$ (181)
EMBEDDED		$5.6985 \cdot 10^{-1}$ (60)	$5.6986 \cdot 10^{-1}$ (60)
GCV		$5.4931 \cdot 10^{-1}$ (30)	$5.4930 \cdot 10^{-1}$ (30)
L-CURVE		$6.0955 \cdot 10^{-1}$ (44)	$6.0955 \cdot 10^{-1}$ (44)
REGIŃSKA		$5.5428 \cdot 10^{-1}$ (33)	$5.4327 \cdot 10^{-1}$ (37)

TABLE 6.8

Averages (over 50 runs) of the regularization parameters obtained when the stopping criterion is satisfied. The satellite test problem is considered (as in Table 6.7).

	LBT	AT	NSLT	RRAT	RRNSLT
DISCREPANCY	$1.1682 \cdot 10^{1}$	$3.9632 \cdot 10^{1}$	$3.9673 \cdot 10^{1}$	$3.9209 \cdot 10^{0}$	$7.2011 \cdot 10^{0}$
EMBEDDED	$5.0156 \cdot 10^{-1}$	$6.2059 \cdot 10^{-1}$	$6.2464 \cdot 10^{-1}$	$1.8579 \cdot 10^{-3}$	$1.8582 \cdot 10^{-3}$
GCV	$1.6596 \cdot 10^{-4}$	$7.8632 \cdot 10^{1}$	$4.3102 \cdot 10^{1}$	$5.0119 \cdot 10^{-5}$	$5.0372 \cdot 10^{-5}$
L-CURVE	$1.6685 \cdot 10^{-3}$	$3.2662 \cdot 10^{-3}$	$3.4124 \cdot 10^{-3}$	$2.6954 \cdot 10^{-3}$	$2.6971 \cdot 10^{-3}$
REGIŃSKA	$5.8916 \cdot 10^{-3}$	$3.1623 \cdot 10^{-3}$	$1.9140 \cdot 10^{-3}$	$1.5849 \cdot 10^{-3}$	$1.0212 \cdot 10^{-3}$

when the fixed maximum number of iterations per restarts is performed) the exact solution is poorly approximated. The AT method delivers good reconstructions, except when a parameter choice based on the GCV is employed: this is due to the fact that the AT method is very sensitive to the value of the regularization parameter and the solution rapidly deteriorates. The performances of the NSLT and RRNSLT methods are very similar to the AT and RRAT ones, respectively.

ETNA
Kent State University
http://etna.math.kent.edu

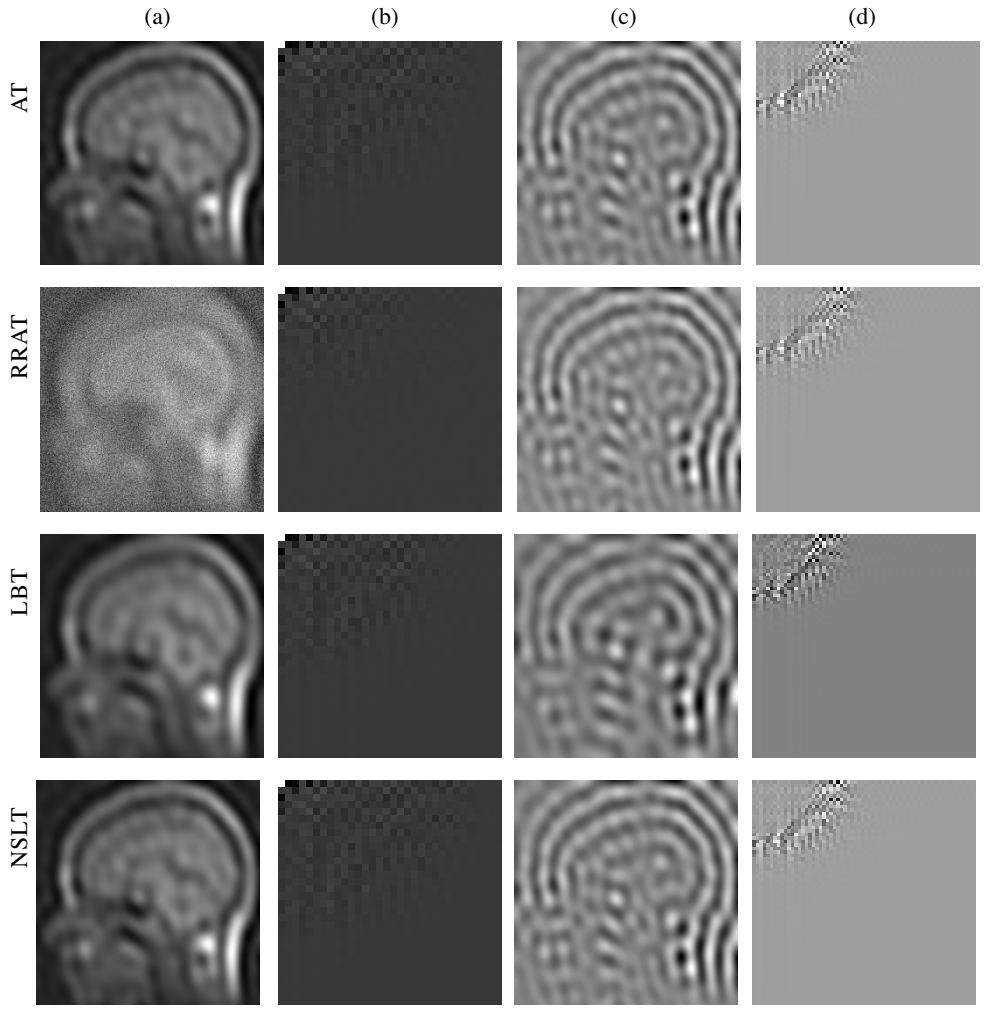

FIG. 6.5. *Spectral analysis of different Krylov subspace methods employed to project the general-form Tikhonov-regularized problem. More precisely, we show: (a) x_m^{ex} , (b) a blow-up (800%) of the DCT of x_m^{ex}, (c), x_m^e, and (d) a blow-up (400%) of the DCT of x_m^e after 10 iterations have been performed (i.e., $m = 10$).*

To summarize the results of all the performed numerical experiments, we remark that many strategies based on the projection of the Tikhonov-regularized problems are very efficient when one has to deal with image restoration problems. Contrarily to what is stated in [38] for the purely iterative methods, we conclude that the most effective Krylov-Tikhonov methods seem to be the ones based on the standard Arnoldi algorithm and the Lanczos bidiagonalization algorithm. In general, AT is faster and cheaper (as far as matrix-vectors multiplications are concerned) than LBT. However, LBT is more reliable than AT when different parameter choice strategies and stopping criteria are considered. Moreover, the performances of the Arnoldi and nonsymmetric Lanczos based methods are very similar. The reason behind this is that since the regularized solutions typically belong to Krylov subspaces of low dimension, the projected problems have a similar behavior (cf. the remarks in Section 2.3). Therefore we propose the NSLT method as a valid alternative to the AT method for the regularization of nonsymmetric problems. Finally, since the starting point of the Krylov-Tikhonov methods

ETNA
Kent State University
http://etna.math.kent.edu

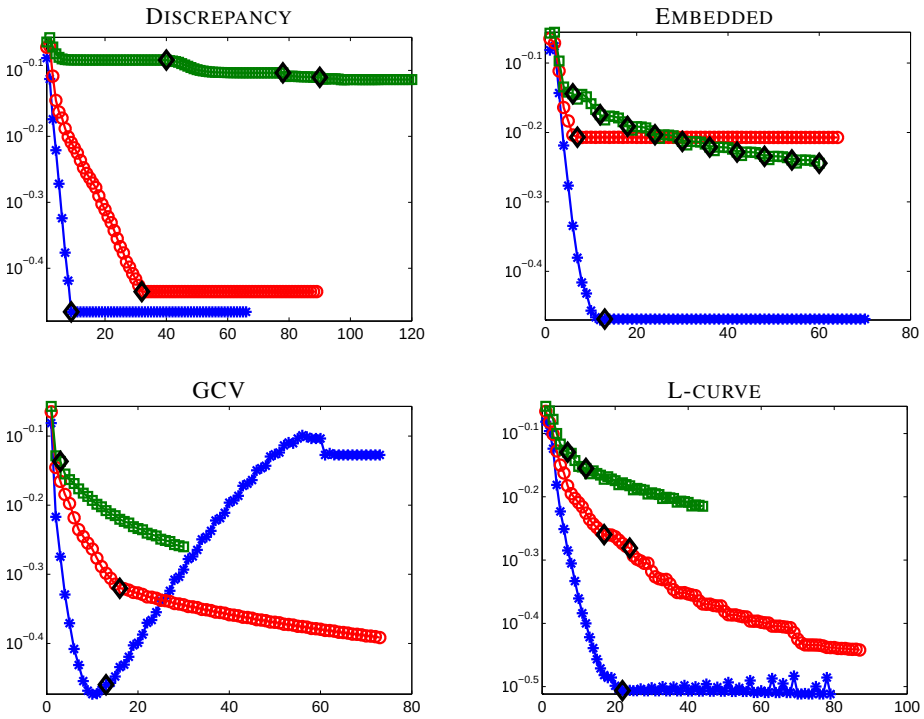

FIG. 6.6. *History of the relative errors obtained when approximating the TV regularization by means of suitably restarted Krylov-Tikhonov methods. The AT method is denoted by an asterisk, the RRAT method is denoted by a square, the LBT method is denoted by a circle. We use a big diamond to highlight the iterations at which a restart happens: after the last diamond is displayed, the restarts happen almost immediately (typically, after 3 iterations have been performed) and, not to overload the plots, we decide to omit them.*

is a regularized problem, a certain amount of regularization is added as the problem and projected onto Krylov subspaces of increasing dimensions. For this reason the noise, which is potentially more present in the Krylov subspaces generated by the standard Arnoldi algorithm than in the subspaces generated by the Lanczos bidiagonalization and range-restriced Arnoldi algorithms [38], is filtered out. Moreover, even if the SVD components of the matrix A are mixed in $K_m(A,b)$ (cf. [37]), the SVD of the projected matrices quickly approximate the SVD of the full-dimensional original matrix (cf. the arguments in Sections 2.2 and 2.3). For this reason, one can obtain good results when solving the projected regularized problems.

7. Conclusions. In this paper we have collected many old and new results concerning the use of some well-known Krylov methods for solving the Tikhonov minimization problem. The analysis has been focused on linear discrete ill-posed problems, which include applications in image restoration. We have shown that the projected problem associated to each one of the considered methods rapidly inherits the basic spectral properties of the original problem, so that these methods can be efficiently used in connection with some of the most important parameter choice rules. This property makes these methods particularly attractive for large-scale problems. The performed numerical experiments have revealed that it is difficult to detect which method outperforms the others in terms of accuracy and speed (i.e., number of iterations). However, it should be emphasized that the Arnoldi-based methods do not require computations with the matrix transpose, so the cost per iteration is lower than the iteration-wise cost of the Lanczos-based methods.

ETNA
Kent State University
http://etna.math.kent.edu

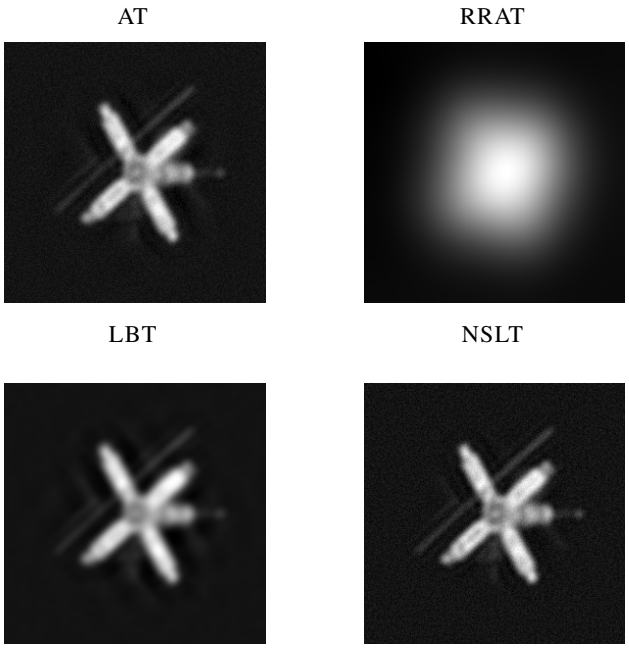

FIG. 6.7. *Reconstructions obtained by different Krylov-Tikhonov regularization methods at the end of the restarting scheme.*

Acknowledgements. We are very grateful to James Nagy for reading an early draft of this work and for his insightful remarks. We wish to thank the referees too: their suggestions helped to improve our presentation.

REFERENCES

[1] S. BERISHA AND J. G. NAGY, *Iterative image restoration*, in Academic Press Library in Signal Processing, Vol. 4, R. Chellappa and S. Theodoridis, eds., Elsevier, Oxford, 2014, pp. 193–243.
[2] Å. BJÖRCK, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
[3] Å. BJÖRCK, E. GRIMME, AND P. VAN DOOREN, *An implicit shift bidiagonalization algorithm for ill-posed systems*, BIT, 34 (1994), pp. 510–534.
[4] D. CALVETTI, G. H. GOLUB, AND L. REICHEL, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
[5] D. CALVETTI, B. LEWIS, AND L. REICHEL, *GMRES-type methods for inconsistent systems*, Linear Algebra Appl., 316 (2000), pp. 157–169.
[6] ———, *Krylov subspace iterative methods for nonsymmetric discrete ill-posed problems in image restoration*, in Advanced Signal Processing Algorithms, Architectures, and Implementations XI, F. T. Luk, ed., vol. 4474 of Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), SPIE, Bellingham, 2001, pp. 224–233.
[7] ———, *On the choice of subspace for iterative methods for linear discrete ill-posed problems*, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
[8] ———, *GMRES, L-curves, and discrete ill-posed problems*, BIT, 42 (2002), pp. 44–65.
[9] ———, *On the regularizing properties of the GMRES method*, Numer. Math., 91 (2002), pp. 605–625.
[10] 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 (2000), pp. 423–446.
[11] D. CALVETTI AND L. REICHEL, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
[12] ———, *Tikhonov regularization with a solution constraint*, SIAM J. Sci. Comput., 26 (2004), pp. 224–239.
[13] D. CALVETTI, L. REICHEL, F. SGALLARI, AND G. SPALETTA, *A regularizing Lanczos iteration method for*

ETNA
Kent State University
http://etna.math.kent.edu

underdetermined linear systems, J. Comput. Appl. Math., 115 (2000), no. 1-2, pp. 101–120.
[14] D. CALVETTI, L. REICHEL, AND A. SHUIBI, *Invertible smoothing preconditioners for linear discrete ill-posed problems*, Appl. Numer. Math., 54 (2005), pp. 135–149.
[15] ———, *Tikhonov regularization of large symmetric problems*, Numer. Linear Algebra Appl., 12 (2005), pp. 127–139.
[16] J. L. CASTELLANOS, S. GÓMEZ, AND V. GUERRA, *The triangle method for finding the corner of the L-curve*, Appl. Numer. Math., 43 (2002), pp. 359–373.
[17] J. CHUNG, J. G. NAGY, AND D. P. O'LEARY, *A weighted-GCV method for Lanczos-hybrid regularization*, Electron. Trans. Numer. Anal., 28 (2007/08), pp. 149–167.
http://etna.mcs.kent.edu/volumes/2001-2010/vol28/abstract.php?vol=28&pages=149-167
[18] E. J. CRAIG, *The N-step iteration procedures*, J. Math. and Phys., 34 (1955), pp. 64–73.
[19] M. DONATELLI, A. NEUMAN, AND L. REICHEL, *Square regularization matrices for large linear discrete ill-posed problems*, Numer. Linear Algebra Appl., 19 (2012), pp. 896–913.
[20] L. DYKES AND L. REICHEL, *A family of range restricted iterative methods for linear discrete ill-posed problems*, Dolomites Res. Notes Approx., 6 (2013), pp. 27–36.
[21] L. ELDÉN, *Algorithms for the regularization of ill-conditioned least squares problems*, Nordisk Tidskr. Informationsbehandling (BIT), 17 (1977), pp. 134–145.
[22] J. N. FRANKLIN, *Minimum principles for ill-posed problems*, SIAM J. Math. Anal., 9 (1978), pp. 638–650.
[23] R. W. FREUND AND N. M. NACHTIGAL, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.
[24] S. GAZZOLA, *Regularization Techniques Based on Krylov Methods for Ill-Posed Linear Systems*, Ph.D. Thesis, Dept. of Mathematics, University of Padua, Italy, 2014.
[25] S. GAZZOLA AND J. G. NAGY, *Generalized Arnoldi-Tikhonov method for sparse reconstruction*, SIAM J. Sci. Comput., 36 (2014), pp. B225–B247.
[26] S. GAZZOLA AND P. NOVATI, *Multi-parameter Arnoldi-Tikhonov methods*, Electron. Trans. Numer. Anal., 40 (2013), pp. 452–475.
http://etna.mcs.kent.edu/volumes/2011-2020/vol40/abstract.php?vol=40&pages=452-475
[27] ———, *Automatic parameter setting for Arnoldi-Tikhonov methods*, J. Comput. Appl. Math., 256 (2014), pp. 180–195.
[28] S. GAZZOLA, P. NOVATI, AND M. R. RUSSO, *Embedded techniques for choosing the parameter in Tikhonov regularization*, Numer. Linear Algebra Appl., 21 (2014), pp. 796–812.
[29] G. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, J. Soc. Indust. Appl. Math. Ser. B Numer. Anal., 2 (1965), pp. 205–224.
[30] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.
[31] M. HANKE, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT, 41 (2001), pp. 1008–1018.
[32] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3 (1993), pp. 253–315.
[33] P. C. HANSEN, *The discrete Picard condition for discrete ill-posed problems*, BIT, 30 (1990), pp. 658–672.
[34] ———, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[35] ———, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
[36] ———, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[37] P. C. HANSEN AND T. K. JENSEN, *Smoothing-norm preconditioning for regularizing minimum-residual methods*, SIAM J. Matrix Anal. Appl., 29 (2006/07), pp. 1–14.
[38] ———, *Noise propagation in regularizing iterations for image deblurring*, Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220.
http://etna.mcs.kent.edu/volumes/2001-2010/vol31/abstract.php?vol=31&pages=204-220
[39] P. C. HANSEN, T. K. JENSEN, AND G. RODRIGUEZ, *An adaptive pruning algorithm for discrete L-curve criterion*, J. Comput. Appl. Math., 198 (2007), pp. 483–492.
[40] P. C. HANSEN AND D. P. O'LEARY, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
[41] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).
[42] M. E. HOCHSTENBACH, N. MCNINCH, AND L. REICHEL, *Discrete ill-posed least-squares problems with a solution norm constraint*, Linear Algebra Appl., 436 (2012), pp. 3801–3818.
[43] M. E. HOCHSTENBACH AND L. REICHEL, *An iterative method for Tikhonov regularization with a general linear regularization operator*, J. Integral Equations Appl., 22 (2010), pp. 465–482.

ETNA
Kent State University
http://etna.math.kent.edu

[44] B. HOFMAN, *Regularization of nonlinear problems and the degree of ill-posedness*, in Inverse Problems: Principle and Applications in Geophysics, Technology and Medicine, G. Anger, R. Gorenflo, H. Jochmann, H. Moritz, and W. Webers, eds., Akademie Verlag, Berlin, 1993, pp. 174–188.
[45] T. K. JENSEN AND P. C. HANSEN, *Iterative regularization with minimum-residual methods*, BIT, 47 (2007), pp. 103–120.
[46] M. E. KILMER, P. C. HANSEN, AND M. I. ESPAÑOL, *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comput., 29 (2007), pp. 315–330.
[47] M. E. KILMER AND D. P. O'LEARY, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.
[48] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
[49] B. LEWIS AND L. REICHEL, *Arnoldi-Tikhonov regularization methods*, J. Comput. Appl. Math., 226 (2009), pp. 92–102.
[50] I. MORET, *A note on the superlinear convergence of GMRES*, SIAM J. Numer. Anal., 34 (1997), pp. 513–516.
[51] S. MORIGI, L. REICHEL, AND F. SGALLARI, *An iterative Lavrentiev regularization method*, BIT, 46 (2006), pp. 589–606.
[52] V. A. MOROZOV, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.
[53] J. G. NAGY, K. PALMER, AND L. PERRONE, *Iterative methods for image deblurring: a Matlab object-oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93.
[54] A. NEUMAN, L. REICHEL, AND H. SADOK, *Algorithms for range restricted iterative methods for linear discrete ill-posed problems*, Numer. Algorithms, 59 (2012), pp. 325–331.
[55] ———, *Implementations of range restricted iterative methods for linear discrete ill-posed problems*, Linear Algebra Appl., 436 (2012), pp. 3974–3990.
[56] P. NOVATI AND M. R. RUSSO, *Adaptive Arnoldi-Tikhonov regularization for image restoration*, Numer. Algorithms, 65 (2014), pp. 745–757.
[57] ———, *A GCV based Arnoldi-Tikhonov regularization method*, BIT, 54 (2014), pp. 501–521.
[58] D. P. O'LEARY AND J. A. SIMMONS, *A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 474–489.
[59] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: an algorithm for sparse linear equations and and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.
[60] T. REGIŃSKA, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 740–749.
[61] L. REICHEL AND G. RODRIGUEZ, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms, 63 (2013), pp. 65–87.
[62] L. REICHEL, G. RODRIGUEZ, AND S. SEATZU, *Error estimates for large-scale ill-posed problems*, Numer. Algorithms, 51 (2009), pp. 341–361.
[63] L. REICHEL, F. SGALLARI, AND Q. YE, *Tikhonov regularization based on generalized Krylov subspace methods*, Appl. Numer. Math., 62 (2012), pp. 1215–1228.
[64] L. REICHEL AND A. SHYSHKOV, *A new zero-finder for Tikhonov regularization*, BIT, 48 (2008), pp. 627–643.
[65] L. REICHEL AND Q. YE, *Simple square smoothing regularization operators*, Electron. Trans. Numer. Anal., 33 (2008/09), pp. 63–83.
http://etna.mcs.kent.edu/volumes/2001-2010/vol33/abstract.php?vol=33&pages=63-83
[66] G. RODRIGUEZ AND D. THEIS, *An algorithm for estimating the optimal regularization parameter by the L-curve*, Rend. Mat. Appl. (7), 25 (2005), pp. 69–84.
[67] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[68] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Halsted Press, New York, 1992.
[69] ———, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[70] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual method for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[71] W. SQUIRE, *The solution of ill-conditioned linear systems arising from Fredholm equations of the first kind by steepest descents and conjugate gradients*, Internat. J. Numer. Methods Engrg., 10 (1976), pp. 607–617.