# On Krylov projection methods
# and Tikhonov regularization

S. Gazzola, P. Novati, M. R. Russo
Department of Mathematics
University of Padova, Italy

May 27, 2014

## 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 50'. In recent years, the use of these methods in a hybrid fashion or to solve Tikhonov regularized problems has received a great attention especially for problems involving the restoration of digital images. In this paper we review the fundamental Krylov-Tikhonov techniques based on the Lanczos bidiagonalization and the Arnoldi algorithm. 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.

# 1 Introduction

The solution of large-scale linear systems

$$Ax = b, \quad A \in \mathbb{R}^{N \times N}, \quad b, x \in \mathbb{R}^N, \tag{1}$$

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, i.e., $b = b^{ex} + e$, where $b^{ex}$ represents the unknown error-free right-hand side.

Historically, since the introduction of the Conjugate Gradient (CG) method [38], CG-like techniques such as the CGLS and Craig's method (CGNE) [16] in which (1) is solved in terms of a least-squares problem, have been widely studied. After the famous paper [26], in which the authors define the so-called Lanczos bidiagonalization procedure by exploiting the Lanczos algorithm for the tridiagonalization of symmetric matrices [44], in [53] the LSQR method is introduced. This method is mathematically equivalent to the CGLS, but with better stability properties, and it is still widely used to solve least-squares problems.

It is worth to remember that these methods well compare with direct techniques such as the TSVD, especially for severely ill-conditioned problems. Indeed, as pointed out in [28], contrary to the TSVD, the projection attained with the CGLS (LSQR) method is tailored to the specific right-hand side $b$, providing a more rapid convergence. All the iterative methods above mentioned are Krylov subspace methods.

In this framework, the well known basic problem with Krylov iterative methods is the so-called semiconvergence phenomenon, i.e., the solution computed by a Krylov subspace method typically converges quite rapidly at the beginning of the iterative process, but then diverges. 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 to 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 [52], 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 error curve on its minimum. 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 GCV and the L-curve, were then studied in [28, 43, 29, 3, 2].

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), regularizing it by Tikhonov method consists in solving the minimization problem

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

where $\lambda > 0$ is the regularization parameter and $L \in \mathbb{R}^{P \times N}$ is said regularization matrix (see [29, 33] for a background); the norm considered in this paper is always the Euclidean one. Assuming that $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$, we denote the solution of (2) by $x_\lambda$. It is known that a proper choice of $\lambda$ and $L$ is crucial for a meaningful approximation of the solution $x^{ex}$ of the error-free problem $Ax = b^{ex}$. The regularization matrix $L$ can be suitably chosen if some information on the behavior of $x^{ex}$ is available. The simplest regularization consist in taking $L = I_N$, where $I_N$ is the identity matrix of order $N$ (standard form Tikhonov regularization).

The simultaneous use of Krylov methods and Tikhonov regularization for approximating the exact solution of (1) can be formulated in two ways. The first one (hybrid methods) consists in regularizing the projected problem; from now on, the word hybrid will always address the Tikhonov regularization of a projected problem. The second one consists in projecting the regularized problem, i.e., in solving (2) by projections. To the best of our knowledge, the use of hybrid methods has been first suggested in [52] and [53], with the aim of regularizing the Lanczos-bidiagonalization-based methods with the identity matrix. This technique has then been used in a number of paper (e.g., [43, 55, 57, 9, 15, 10, 11, 2, 4, 12]) 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 [62]) have a more recent history: they have been first introduced in [9], and then studied (also with the Range-Restricted implementation) in [40, 45]. We remark that a hybrid Arnoldi method has even been implicitly used in

[47], where the symmetric Lanczos process is used for $A$ symmetric positive semidefinite in connection with the Levrentiev (Franklin) regularization

$$(A + \lambda I_N)x = b, \quad \lambda > 0. \tag{3}$$

Again, in [13] the same algorithm is applied for $A$ symmetric with the standard Tikhonov regularization.

Beyond the hybrid approaches, the use of Krylov projection methods for solving (2) 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 (2) has been used in [42] where an algorithm for the simultaneous bidiagonalization of $A$ and $L$ is introduced, and in [39] where the skinny QR factorization of the penalty term is used to fully project the problem. The Arnoldi algorithm for (2) has been used in [50, 51, 24, 25]; in [23] 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 [56]. We remark that, starting from [2], many authors have suggested to bridge the gap between hybrid methods and solution of (2) by Krylov projection: indeed, after transforming (2) into standard form [18], the smoothing effect of $L$ is incorporated into the hybrid process (see also [40, 45, 34]). Anyway, unless $L$ is invertible, this transformation is not much friendly to use and the arising algorithm would suffer from that; probably, for this reason, this transformation is not implemented and tested in the papers where it is mentioned.

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 (2), with particular attention to the parameter choice rules (for both $\lambda$ 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 BCG, CGS, QMR and BiCGstab, 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 used 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 [32].

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 includes all the methods considered in Section 2; in order to assess the regularizing performances of the considered Krylov subspace methods, in the second part of Section 3 (Section 3.2) we include the results of many 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 be 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: we consider a medical and an astronomical test image, and all the parameter choice strategies described in Section 5 are taken into account. During this paper, we will use the following

**Notation 1** *We denote the SVD of the full-dimensional matrix $A$ by*

$$A = U^{\mathcal{S}} \Sigma^{\mathcal{S}} \left( V^{\mathcal{S}} \right)^T, \tag{4}$$

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

$$A_m^{\mathcal{S}} = U_m^{\mathcal{S}} \Sigma_m^{\mathcal{S}} (V_m^{\mathcal{S}})^T, \tag{5}$$

*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 (4), respectively, and $\Sigma_m^{\mathcal{S}}$ is the leading $m \times m$ submatrix of $\Sigma^{\mathcal{S}}$ in (4). We also denote by $x_m^{\mathcal{S}}$ the TSVD solution of (1), that is,*

$$x_m^{\mathcal{S}} = V_m^{\mathcal{S}} \left( \Sigma_m^{\mathcal{S}} \right)^{-1} (U_m^{\mathcal{S}})^T b. \tag{6}$$

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

$$AX^{\mathcal{G}} = U^{\mathcal{G}} S^{\mathcal{G}}, \quad LX^{\mathcal{G}} = V^{\mathcal{G}} C^{\mathcal{G}}, \tag{7}$$

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

$$\gamma_i = \frac{s_i}{c_i}, \qquad i = 1, ..., N. \tag{8}$$

*To keep the notation simpler, in (7) and (8) we have assumed $L \in \mathbb{R}^{N \times N}$.*

**Notation 2 (Test problems).** *In order to numerically derscribe the properties of the methods considered in the paper, we make use of the test problems available from Hansen's Regularization Toolbox [32]. 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

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

typically, in this paper, $C = A$, $A^T$, $A^T A$, $AA^T$ and $d = b$, $A^T b$, $Ab$. 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

$$
\begin{aligned}
x_m &\in & x_0 + \mathcal{K}'_m, & \qquad (9) \\
b - A x_m &\perp & \mathcal{K}''_m, & \qquad (10)
\end{aligned}
$$

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 $\mathcal{K}'_m$, that is, $\mathcal{R}(W_m) = \mathcal{K}'_m$, we are interested in methods where $x_m = W_m y_m$ (approximatively) solves

$$
\min_{x \in \mathcal{K}'_m} \| b - Ax \| = \min_{y \in \mathbb{R}^m} \| b - A W_m y \| = \| b - A W_m y_m \| . \qquad (11)
$$

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

**Definition 3** *Assume that $b$ 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. [30]) is satisfied if $\left\{ \left| u_m^T b \right| \right\}_{1 \le m \le N}$, 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 [33, p.9]). For discrete problems, the DPC ensures that the TSVD solutions of (1) are uniformly bounded. If we assume to work with severely ill-conditioned problems, that is $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$, (cf. [41]), and that the Fourier coefficients $u_m^T b$, $1 \le m \le N$ satisfy the model

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

(cf. [33, p.88]), then the TSVD solutions (6) are bounded as

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

Similar bounds can be straightforwardly obtained dealing with mildly ill-conditioned problems, in which $\sigma_j = O(j^{-\alpha})$, provided that $\alpha$ 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 Lanczos bidiagonalization algorithm

The Lanczos bidiagonalization algorithm [26] computes two orthonormal basis $\{ w_1, \ldots, w_m \}$ and $\{ z_1, \ldots, z_m \}$ for the Krylov subspaces $\mathcal{K}_m(A^T A, A^T b)$ and $\mathcal{K}_m(AA^T, b)$, respectively. In Algorithm 1 we summarize the main computations involved in the Lanczos bidiagonalization procedure.

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

$$
\begin{aligned}
A W_m &=& Z_{m+1} \bar{B}_m, & \qquad (12) \\
A^T Z_{m+1} &=& W_m \bar{B}_m^T + \mu_{m+1} w_{m+1} e_{m+1}^T, & \qquad (13)
\end{aligned}
$$

---

**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 $w_j = w/\mu_j$.

---

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 $(m+1)$-th 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 (9) and (10), the LSQR method has $\mathcal{K}'_m = \mathcal{K}_m(A^T A, A^T b)$ and $\mathcal{K}''_m = A\mathcal{K}_m(A^T A, A^T b)$. This method consists in computing, at the $m$-th iteration of the Lanczos bidiagonalization algorithm,

$$
y_m = \arg \min_{y \in \mathbb{R}^m} \left\| \|b\| e_1 - \bar{B}_m y \right\|, \tag{14}
$$

and in taking $x_m = W_m y_m$ as approximate solution of (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 subspaces methods to be employed with regularization purposes in a purely iterative fashion (cf. [63]), as hybrid methods (cf. [52]), and to approximate the solution of (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.

6

**Proposition 4** *Assume that (1) is severely ill-conditioned, i.e., $\sigma_j = O(e^{-\alpha j})$, $\alpha > 0$. Assume moreover that $b$ satisfies the DPC. Then, for $m = 1, \ldots, N - 2$,*

$$\mu_{m+1}\nu_{m+1} = O(m\sigma_m^2), \tag{15}$$

$$\mu_{m+1}\nu_{m+2} = O(m\sigma_{m+1}^2), \tag{16}$$

**Proof.** Concerning estimate (15), by (12) and (13) we have that

$$(A^T A)W_m = W_m(\bar{B}_m^T \bar{B}_m) + \mu_{m+1}\nu_{m+1}w_{m+1}e_m^T, \tag{17}$$

so that $W_m$ is the matrix generated by the symmetric Lanczos process applied to the system $A^T A x = A^T b$, and its columns span $\mathcal{K}_m(A^T A, A^T b)$. After recalling that the singular values of $A^T A$ are the scalars $\sigma_i^2$, $i = 1, \ldots, N$, and that the over/under-diagonal elements of the symmetric tridiagonal matrix $\bar{B}_m^T \bar{B}_m \in \mathbb{R}^{m \times m}$ are of the form $\mu_{m+1}\nu_{m+1}$, $m = 1, \ldots, N-1$, (15) directly follows by applying Proposition 8 reported in Section 2.2 (for a proof, see [51, Proposition 3.3] and the refinement given in [21, Theorem 6]).

Concerning estimate (16), using again (13) and (12), we have that

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

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 $AA^T x = b$, and its columns span $\mathcal{K}_{m+1}(AA^T, b)$. Since the singular values of $AA^T$ are the scalars $\sigma_i^2$, $i = 1, \ldots, N$, and the over/under-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 (15) follows again by applying [21, Theorem 6]. ∎

The potential of the Lanczos bidiagonalization as a tool for regularization 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 5** *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*

$$AV_m - U_m \bar{\Sigma}_m = 0, \tag{18}$$

$$\left\| A^T U_m - V_m \bar{\Sigma}_m^T \right\| \leq \mu_{m+1}. \tag{19}$$

**Proof.** Relation (18) immediately follows from (12) and $\bar{B}_m \bar{V}_m = \bar{U}_m \bar{\Sigma}_m$. Moreover, by employing (13),

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

which, since $\|w_{m+1}\| = \|e_{m+1}\| = \|\bar{U}_m\| = 1$, leads to (19). ∎

Provided that $\mu_m \to 0$, relations (18) and (19) 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, 28, 29, 39] and [33, Chapter 6]). Indeed, looking at Algorithm 1, we have that $\mu_j = \|w\|$ where $w \in \mathcal{K}_j(A^T A, A^T b)$ and $w \perp \mathcal{K}_{j-1}(A^T A, A^T b)$. If $A$ represent a compact operator we know that quite rapidly $\mathcal{K}_j(A^T A, A^T b)$ becomes almost $A^T A$-invariant, that is $\mathcal{K}_j(A^T A, A^T b) \approx \mathcal{K}_{j-1}(A^T A, A^T b)$ (see e.g. [46] and the references therein).

**Proposition 6** *Under the same hypothesis of Proposition 4, we have that* $\nu_m, \mu_m \to 0$ *and*

$$\left\|AA^T z_m - \mu_m^2 z_m\right\| = O(\nu_m), \tag{20}$$
$$\left\|A^T A w_m - \nu_m^2 w_m\right\| = O(\mu_{m-1}). \tag{21}$$

**Proof.** We start by assuming that $\nu_m \nrightarrow 0$. Then, by Proposition 4, we immediately have that $\mu_m = O(m\sigma_m^2)$. Thus, by step 1 of Algorithm 1,

$$z = Aw_{m-1} + d_{m-1}, \quad \|d_{m-1}\| = O(m\sigma_{m-1}^2) \tag{22}$$

for $m$ large enough. Then, by step 4 and by (22),

$$
\begin{aligned}
w &= A^T z_m - \nu_m w_{m-1} \\
&= A^T \left(\frac{Aw_{m-1} + d_{m-1}}{\nu_m}\right) - \nu_m w_{m-1},
\end{aligned}
$$

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

$$\left\|A^T A w_{m-1} - \nu_m^2 w_{m-1}\right\| = O(m\sigma_{m-1}^2). \tag{23}$$

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

$$w = A^T z_m + d_m', \quad \|d_m'\| = \nu_m. \tag{24}$$

Then at the next step 1,

$$
\begin{aligned}
z &= Aw_m - \mu_m z_m \\
&= A\left(\frac{A^T z_m + d_m'}{\mu_m}\right) - \mu_m z_m,
\end{aligned}
$$

so that

$$\nu_{m+1}\mu_m z_{m+1} = AA^T z_m - \mu_m^2 z_m + Ad_m',$$

and hence

$$\left\|AA^T z_m - \mu_m^2 z_m\right\| = O(\nu_m). \tag{25}$$

8

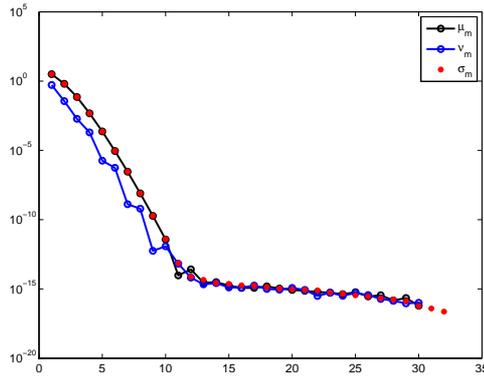Figure 1: Problem baart: decay behavior of the sequences $\{\nu_m\}_m$ and $\{\mu_m\}_m$ with respect to the singular values of $A$.

The above relation means that $\mu_m$ 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 $\mu_m \to 0$. Therefore, rewriting the first part of equality (23) by replacing $\|d_{m-1}\| = \mu_{m-1}$ in (22), we have the result. ∎

Proposition 6 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 closed to the one of the singular values of $A$. An example of this behavior is reported in Figure 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 5).

**Proposition 7** *If the full-dimensional system (1) satisfies the DPC, then the DPC is inherited by the projected problems (14), for $1 \leq m \leq N$.*

**Proof.** Recalling that the LSQR is mathematically equivalent to the CG applied to the normal equations $A^T A x = A^T b$, and thanks to the relations derived in [38, Theorem 6.1] and elaborated in [33, Chapter 6], we can state that

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

Since the DPC holds for the problem (1), $\|y_N\| = \|x_N\| = \|x^{ex}\| = c < \infty$. Moreover, since

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

we can state that

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

which proves the result. ∎

## 2.2 Methods based on Arnoldi algorithm

The Arnoldi algorithm computes an orthonormal basis $\{w_1, \dots, w_m\}$ 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

$$AW_m = W_m H_m + h_{m+1,m} w_{m+1} e_m^T, \tag{26}$$

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 (26) can be written as

$$AW_m = W_{m+1} \bar{H}_m, \tag{27}$$

where

$$\bar{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1) \times m}. \tag{28}$$

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 [61, §6.3], which may considerably improve its accuracy, measured in term of the quantity $\left\| W_m^T W_m - I \right\|$. 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 arithmetics.

### 2.2.1 The GMRES method

The most popular Krylov subspace method based on the Arnoldi algorithm is the GMRES method [62]. Referring to (9) and (10), 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 the LSQR, we have

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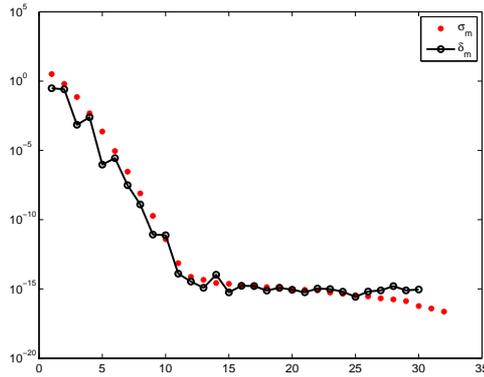\end{aligned}
$$

10

Figure 2: Problem baart: decay behavior of the sequence $\{h_{m+1,m}\}_m$ with respect to the singular values of $A$.

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

$$y_m = \arg \min_{y \in \mathbb{R}^m} \left\| \|b\| \, e_1 - \bar{H}_m y \right\|, \tag{29}$$

and to take $x_m = W_m y_m$ as approximate solution of (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 [8], 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 extendibily of the Krylov subspaces. Moreover, it is also known that the residual of the GMRES can be bounded using the Full Orthogonalization Method (FOM, see e.g. [61, §6.4]) residual as follows

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

In presence of severely ill-conditioned problems, the following result has been proved in [51] (cf. Figure 2).

**Proposition 8** *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*

$$h_{m+1,m} = O\left( m \, \sigma_m \right). \tag{30}$$

The authors of [51] 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 Lanczos bidiagonalization algorithm (cf. Section 2.1 and [1, 28]). More precisely, let us consider the SVD factorization of $\bar{H}_m$, i.e., $\bar{H}_m = \bar{U}_m \bar{\Sigma}_m \bar{V}_m^T$.

11

**Proposition 9** *Let $U_m = W_{m+1}\bar{U}_m$ and $V_m = W_m\bar{V}_m$. Then*

$$AV_m - U_m\bar{\Sigma}_m = 0, \qquad (31)$$
$$W_m^T(A^T U_m - V_m\bar{\Sigma}_m^T) = 0. \qquad (32)$$

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 capabilities of the two algorithms can also be understood by comparing (19) and (32): 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 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

The Range-Restricted GMRES (RRGMRES) has been 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 [17] the 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 (26) still holds, i.e.,

$$AW_m = W_{m+1}\bar{H}_m, \qquad (33)$$

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

$$\begin{aligned}
b &= W_{m+1}W_{m+1}^T b + \left(I - W_{m+1}W_{m+1}^T\right)b \\
&= W_{m+1}W_{m+1}^T b + W_{m+1}^{\perp}\left(W_{m+1}^{\perp}\right)^T b
\end{aligned}$$

we have

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

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 9 is still valid, since it only involves the Arnoldi decomposition (33): this assures that the RRGMRES can still be interpreted as a method able to approximate the singular values of $A$.

## 2.3 Methods based on nonsymmetric Lanczos algorithm

The Nonsymmetric Lanczos algorithm (also referred to as two-sided Lanczos process, or Lanczos biorthogonalization procedure) is employed to compute two basis $\{w_1, \ldots, w_m\}$ and $\{k_1, \ldots, k_m\}$ for the Krylov subspaces $\mathcal{K}_m(A, b)$ and $\mathcal{K}_m(A^T, b)$, respectively, satisfying the biorthogonality condition $w_i^T k_j = \delta_{ij}$, $i, j = 1, \ldots, m$. 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 $j = 1, \ldots, m$

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

$$\begin{align}
AW_m &= W_m T_m + \delta_{m+1} w_{m+1} e_m^T, \tag{34}\\
A^T K_m &= K_m T_m^T + \beta_{m+1} k_{m+1} e_m^T, \tag{35}
\end{align}$$

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 (34) yields to

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

It is well known that, if the matrix $A$ is symmetric, then the 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 to 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 (34) can be written as

$$AW_m = W_{m+1}\bar{T}_m, \tag{36}$$

where

$$\bar{T}_m = \begin{bmatrix} T_m \\ \delta_{m+1,m}e_m^T \end{bmatrix} \in \mathbb{R}^{(m+1)\times m}. \tag{37}$$

We remark that the definition of $\delta_{j+1} = |k^T w|^{1/2}$ 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 the Lanczos biorthogonalization are the BCG and QMR methods (cf. [61, 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 on or before the $m$-th step.

At the $m$-th iteration of the nonsymmetric Lanczos algorithm, the QMR [20] method prescribes to compute

$$y_m = \arg \min_{y \in \mathbb{R}^m} \left\| \|b\| e_1 - \bar{T}_m y \right\|, \tag{38}$$

and to take $x_m = W_m y_m$ as approximate solution of (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 the QMR and the GMRES, it can be proved that (cf. [20])

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

where $r_m^{\text{QMR}}$ and $r_m^{\text{GMRES}}$ are the residuals of the 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 the 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 [60, 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 symmetry, then $\bar{T}_m$ can also be used to approximate its singular values. To study the convergence of the Lanczos unsymmetric process, we recall the following proposition, originally proved in [51].

**Proposition 10** *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*

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

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

$$\widetilde{V}_m = U_m^{\mathcal{S}} C_m + E_m, \quad \|E_m\| = O(\sigma_m), \tag{39}$$
$$U_m^{\mathcal{S}} = \widetilde{V}_m C_m^{-1} + F_m, \quad \left\| F_m \Sigma_m^{\mathcal{S}} \right\| = O(m \sigma_m). \tag{40}$$

Then, we can prove the following result (cf. Figure 3).

**Proposition 11** *Under the same hypothesis of Proposition 10, for $m = 1, \ldots, N-1$*

$$\delta_{m+1} = O(m \sigma_m). \tag{41}$$

**Proof.** Directly from relation (36), 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 [27, §2.5.5], we can write $A = A_m^{\mathcal{S}} + \Delta_m$, where $A_m^{\mathcal{S}}$ is defined in (5) 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 (40). 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

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

Since $\|k_{m+1}\| \|w_m\|$ does not depend on the rate of the decay of $\sigma_m$, we have (41). ∎

As 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 $\delta_{j+1}$ as defined in line 4 of Algorithm 3 vanishes. If this occur 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 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.

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) are analogous to the ones explained in Section 2.2.2 for the RRGMRES.
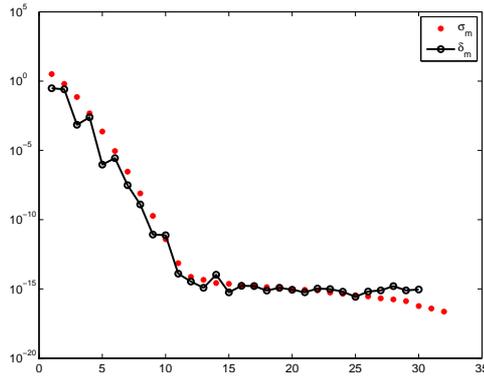
15

Figure 3: Problem baart: decay behavior of the sequence $\{\delta_m\}_m$ with respect to the singular values of $A$.

# 3  General formulation

In this section we provide a general formulation that embraces the Krylov methods considered in this work. We also give the general formulation of an hybrid approach and of the solution of the Tikhonov-regularized problem by projections.

## 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

$$AW_m = Z_{m+1}\bar{D}_m, \quad K_m^T W_m = I_m, \tag{42}$$

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

$$\min_{y \in \mathbb{R}^m} \left\| d - \bar{D}_m y \right\| \approx \min_{y \in \mathbb{R}^m} \left\| b - AW_m y \right\|, \tag{43}$$

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

$$Z_{m+1} = Q_{m+1}R_{m+1}, \quad Q_{m+1} \in \mathbb{R}^{N \times (m+1)}, \ R_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}, \tag{44}$$

we can state the following general result.

**Proposition 12** *Given a Krylov subspace method based on the decomposition (42), for each $y \in \mathbb{R}^m$ we have*

$$\|b - AW_m y\|^2 = \left\| Q_{m+1}^T b - R_{m+1}\bar{D}_m y \right\|^2 + \left\| (Q_{m+1}^\perp)^T b \right\|^2. \tag{45}$$

**Proof.** Considering the factorizations (42) and (44), and writing

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

16

we have

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

Thanks to the orthonormality of the columns of $Q_{m+1}$ and $Q_{m+1}^\perp$, we immediately have (45). ∎

Depending on the properties of the considered Krylov method, expression (45) can assume simpler forms. In particular:

- For the 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 (43), $d = \|b\|e_1$.

- For the GMRES we have $\bar{D}_m = \bar{H}_m$. Moreover, $Q_m = Z_m = W_m = K_m$ has 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 (43), $d = \|b\|e_1$.

- For the 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 (43), $d = Q_{m+1}^T b$.

- For the QMR we have $\bar{D}_m = \bar{T}_m$ and $Z_m = W_m$. Unless $A$ is symmetric, the QR factorization (44) 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 (43), $d = \|b\|e_1$. Moreover, the matrix $Q_m$ is just the orthogonal matrix $W_m$ generated by the Arnoldi algorithm. By comparing (43) with (45) it is clear that in the 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 (43) with $d = Q_{m+1}^T b$. Observe, however, that none of them makes use of the QR decomposition (44), because, except for the RRGMRES, we have $Q_{m+1}^T b = \|b\|\, e_1$, and, for the 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

$$\|b - Ax_m\| \approx \left\|Q_{m+1}^T b - \bar{D}_m y_m\right\|. \tag{46}$$

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

**Proposition 13** *Let $y_m$ be the solution of (43), and let $x_m = W_m y_m$ be the corresponding approximate solution of (11). 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

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., [33, Chapter 2]), and the TSVD is a commonly used tool for regularization. Denoting by $A_m^{\mathcal{S}}$ the truncated singular value decomposition of $A$ (5), 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 (42), we have that the least-square solution of (1) is approximated by the solution of

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

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 (43); we again underline that in (43) the equality holds just for the LSQR and the GMRES. After introducing the matrix

$$A_m^{\mathcal{K}} := Z_{m+1} \bar{D}_m K_m^T, \tag{47}$$

which is a sort of regularized matrix associated to the generic Krylov subspace methods defined by the factorization (42), we want to compare the regularization properties of the Krylov methods with the ones of the TSVD. We do this by plotting the quantity $\left\| A - A_m^{\mathcal{K}} \right\|$ (recall the optimality property $\left\| A - A_m^{\mathcal{S}} \right\| = \sigma_{m+1}$, [27, §2.5.5]). The results are reported in Figure 4. 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 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 Propositions 5 and 9 in terms of residual, for the Lanczos bidiagonalization and the Arnoldi algorithms, respectively. In Figures 5 and 6 we show some experiments for all the considered methods. The results show 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 $\|e\|/\|b^{ex}\|$.

### 3.2.2 Accuracy analysis on standard test problems

In this section we consider the accuracy of the methods, in terms of the minimum relative error attainable, for different noise levels, from $10^{-1}$ to $10^{-12}$. The results, on an average of 30 runs, are reported in Figure 7.
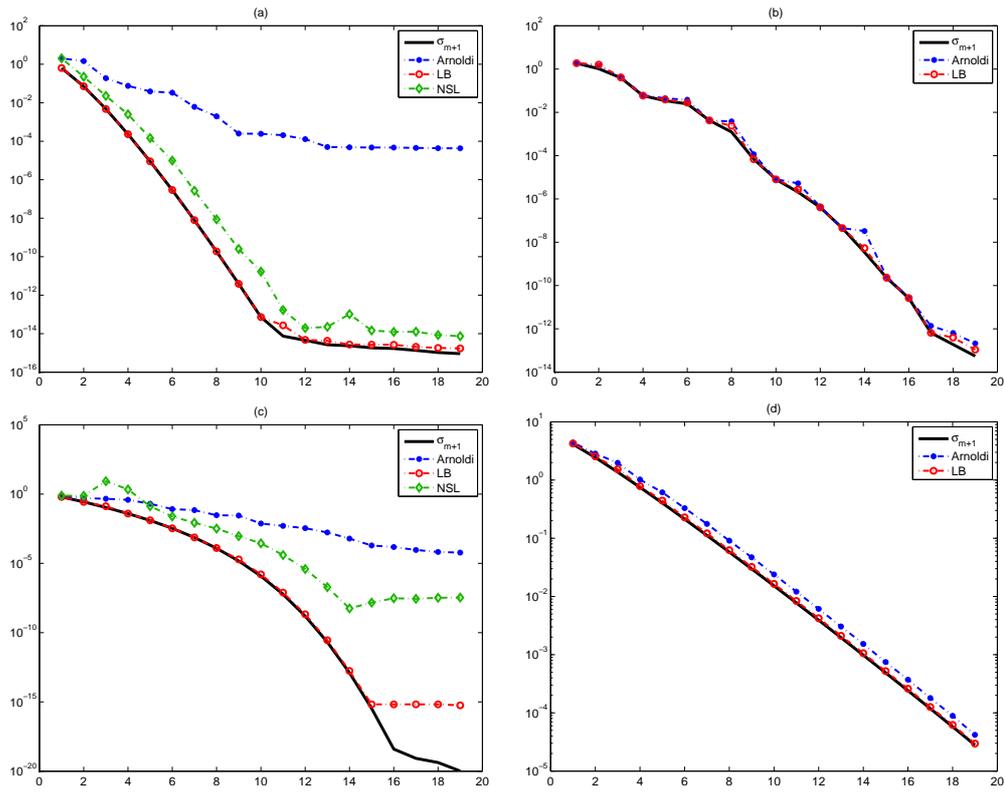
Figure 4: Plots of $\left\| A - A_m^{\mathcal{K}} \right\|$ with respect to the singular values of $A$ for the problems baart (a), shaw (b), i_laplace (c) and gravity (d).
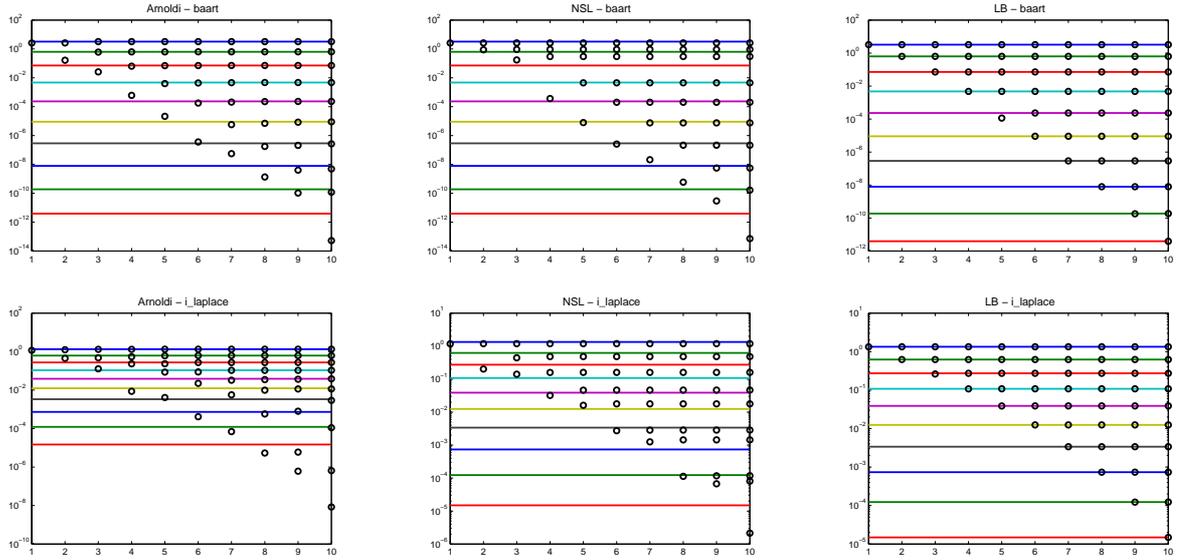
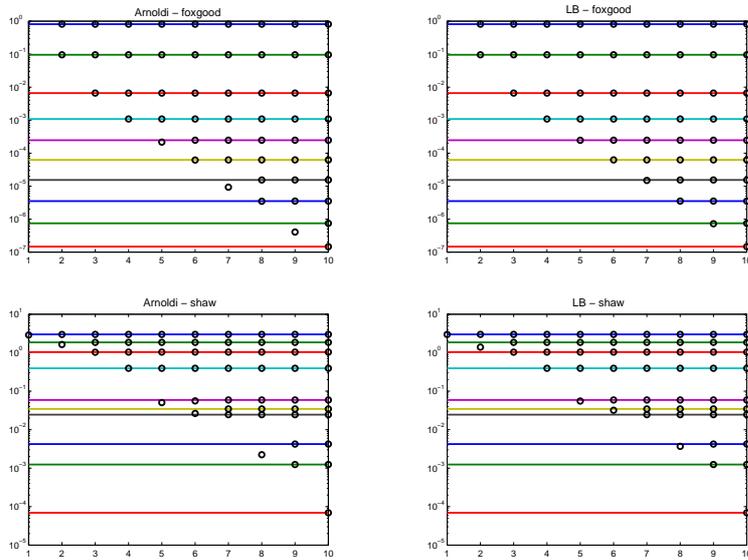Figure 5: Approximation of the dominating singular values - the nonsymmetric case.



Figure 6: Approximation of the dominating singular values - the symmetric case.
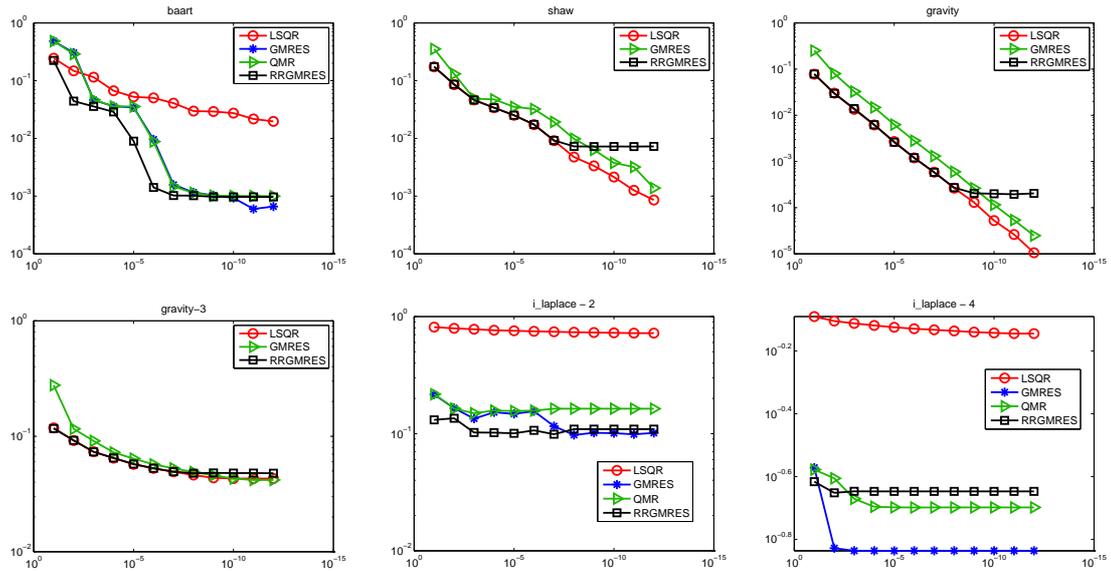
Figure 7: Optimal attainable accuracy (minimum relative error) versus different noise levels (from $10^{-1}$ to $10^{-12}$). The displayed values are averages over 30 runs of the methods for each level.

Whenever the noise level is relatively high, the RRGMRES seem to be the most accurate methods. The reason obviously lies in the use of a starting vector, $Ab$, in which most of the noise has been removed. The difference is less evident when the noise level il small, and it is interesting to see that the attainable accuracy of the RRGMRES typically stagnates around a certain level. This is the negative counterpart 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 [48], 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 type, and its norm is known (or well estimated). The discrepancy principle prescribes to stop the iterations as soon as

$$\|b - Ax_m\| \leq \eta \|e\|, \tag{48}$$

where $\eta > 1$ (typically $\eta \approx 1$) is a safety factor. In Table 1 we compare the best attainable accuracy 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
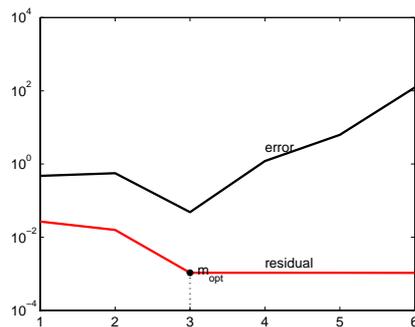
Figure 8: Problem baart: example of fast convergence/divergence behavior of GMRES.

$e$, with $\|e\| / \|b\| = 10^{-3}$. 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 quite huge, and cannot be detected by the residual norm, which is generally flat around $m_{opt}$ (see Figure 8).

In this view, using the values $\eta_1 = 1.02$, $\eta_2 = 1.05$, $\eta_3 = 1.1$ for the discrepancy rule in (48), and denoting by $m_{\mathrm{DP}}$ the iteration number selected, in Table 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 1 are rather clear: independently on the choice of the safety factor $\eta$, 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) is generally necessary.

In this setting, the standard form Tikhonov regularization of (43) reads

$$\min_{y \in \mathbb{R}^m} \left\{ \left\| Q_{m+1}^T b - \bar{D}_m y \right\|^2 + \lambda^2 \|y\|^2 \right\}. \tag{49}$$

If the regularization parameter $\lambda$ is defined (at each step) independently of the original problem, that is, with the only aim of regularizing (43), then the corresponding method is traditionally called hybrid. As already addressed in the Introduction, regularization by Krylov methods, or their use to solve the Tikhonov minimization problem, has a long history,

22

| Method | Average error | | | Semi-failure | | | Total failure | | |
|---|---|---|---|---|---|---|---|---|---|
| | $m_{opt}$ | $m_{opt}-1$ | $m_{opt}+1$ | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\eta_1$ | $\eta_2$ | $\eta_3$ |
| **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 |

Table 1: Stability results.

dating back to [52]. Regarding the GMRES, the hybrid approach, called Arnoldi-Tikhonov method, was first considered in [9] 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 (2), and hence we define $\lambda$ 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

$$\min_{x \in \mathcal{K}_m} \left\{ \|b - Ax\| + \lambda^2 \|Lx\| \right\} ; \tag{50}$$

in the following, for theoretical purposes, it could be useful to consider the following expression for the Tikhonov regularized solution

$$x_\lambda = (A^T A + \lambda^2 L^T L)^{-1} A^T b. \tag{51}$$

In this sense, at each step we approximate the solution of (2) by solving

$$\min_{y \in \mathbb{R}^m} \left\{ \left\| Q_{m+1}^T b - \bar{D}_m y \right\|^2 + \lambda^2 \|L W_m y\|^2 \right\} \tag{52}$$

(cf. Section 3). Minimizing (52) is equivalent to solving the following regularized least squares problem

$$\min_{y_m \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{D}_m \\ \lambda L W_m \end{pmatrix} y_m - \begin{pmatrix} Q_{m+1}^T b \\ 0 \end{pmatrix} \right\|^2 . \tag{53}$$

If we denote by $y_{m,\lambda}$ the solution of (52), then $x_{m,\lambda} = W_m y_{m,\lambda}$ is the corresponding approximate solution of (2) and regularized solution of (1). It is well known that, in many applications, the use of a suitable regularization operator $L \neq I_N$, may substantially improve the quality of the approximate solution with respect to the choice of $L = I_N$. As for the Lanczos bidiagonalization algorithm, the solution of (52) with $L \neq I_N$ has been considered, among the others, in [42] and [39], whereas the Arnoldi algorithm has been used in [24, 23, 25, 50, 51].

It is important to observe that, if $L = I_N$, then the dimension of the problem (53) 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 (53) inherits the dimension of the original problem. In order to fully reduce the dimension of the subproblem (53) when $L \neq I_N$, one could consider the "skinny" QR factorization of $L W_m$ (see [39]), i.e.,

$$Q_m^L L_m = L W_m, \tag{54}$$

where $Q_m^L \in \mathbb{R}^{P \times m}$ and $L_m \in \mathbb{R}^{m \times m}$. Alternatively, assuming that $P \leq N$, one could also add $N - P$ zero rows to $L$ (which does not alter (50)) and consider the projection of $L$ onto $\mathcal{K}_m(A, b)$ (see [51]), i.e.,

$$L_m = K_m^T L W_m \in \mathbb{R}^{m \times m}, \tag{55}$$

where $K_m$ depends on the Krylov subspace (cf. (42)). In both cases, (52) reads

$$\min_{y \in \mathbb{R}^m} \left\{ \left\| Q_{m+1}^T b - \bar{D}_m y \right\|^2 + \lambda^2 \|L_m y\|^2 \right\} \tag{56}$$

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

24

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

$$y_{m,\lambda} = (\bar{D}_m^T \bar{D}_m + \lambda^2 L_m^T L_m)^{-1} \bar{D}_m^T Q_{m+1}^T b. \tag{57}$$

We remark that, when we consider the matrix (55), problem (56) is not equivalent to (50), anymore. However, the use of the matrix $L_m$ defined in (55) appears natural in this framework: $L_m$ would be the regularization operator of the projection of the Franklin-type regularization [19]

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

According to our experience, employing the upper triangular $L_m$ in (54), or considering the projected operator (55) performs 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 (54). 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 (56) have been computed by the Lanczos bidiagonalization, Arnoldi, Range-Restricted Arnoldi, nonsymmetric Lanczos algorithms, and Range-Restricted non-symmetric 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 $(\bar{D}_m, L_m)$, 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

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

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 $\bar{u}_i^{(m)}$, $i = 1, ..., m+1$.

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

**Proposition 14** *Let $y_{m,\lambda}$ be the solution of (56). Then the pseudo-residual satisfies (cf. (46))*

$$\left\| \bar{D}_m y_{m,\lambda} - Q_{m+1}^T b \right\|^2 = \sum_{i=1}^{m} \left( \frac{\lambda^2}{\gamma_i^{(m)2} + \lambda^2} (\bar{u}_i^{(m)})^T Q_{m+1}^T b \right)^2 + \left( (\bar{u}_{m+1}^{(m)})^T Q_{m+1}^T b \right)^2. \tag{58}$$

**Proof.** This result simply follows by substituting the GSVD of $(\bar{D}_m, L_m)$ into (57):

$$y_{m,\lambda} = \bar{X}_m (\bar{S}_m^T \bar{S}_m + \lambda^2 \bar{C}_m^T \bar{C}_m)^{-1} \bar{S}_m^T \bar{U}_m^T Q_{m+1}^T b, \tag{59}$$

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 (56). We assume that the quantity $\|e\|$ is quite accurately known, and consequently we use the discrepancy principle to simultaneously select the number of iterations (stopping rule) and the value of the regularization parameter $\lambda$. Similarly to the discrete case of Section 3.2, when solving regularized problems of the form (56), one commonly says that the discrepancy principle is satisfied when

$$\|b - Ax_{m,\lambda}\| \leq \eta \|e\|, \tag{60}$$

where $\eta \gtrsim 1$. Using the same arguments as the ones employed in Section 2 for evaluating the norm of the residuals associated to the projection methods described by the decomposition (42), we have that

$$\|b - Ax_{m,\lambda}\| \approx \left\| Q_{m+1}^T b - \bar{D}_m y_{m,\lambda} \right\|, \tag{61}$$

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

$$\phi_m(\lambda) := \left\| Q_{m+1}^T b - \bar{D}_m y_{m,\lambda} \right\| = \eta \|e\|, \tag{62}$$

where $y_{m,\lambda}$ is the solution of (56).

Among the existing algorithms that solve (62) inside a Krylov methods coupled with Tikhonov regularization (see e.g. [45] and [58]), the one proposed in [24] 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., $\lambda = 0$), clearly $\phi_m(0) = \|r_m\|$. In this setting, the authors solve (62) after considering the linear approximation

$$\phi_m(\lambda) \approx \phi_m(0) + \lambda \varphi_m, \tag{63}$$

where, at each iteration, the scalar $\varphi_m$ is defined by the ratio

$$\varphi_m = \frac{\phi_m(\lambda_{m-1}) - \phi_m(0)}{\lambda_{m-1}}. \tag{64}$$

In (64), $\phi_m(\lambda_{m-1})$ is obtained by solving the $m$-dimensional problem (56) 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

$$\phi_m(\lambda_m) = \eta \|e\|. \tag{65}$$

Substituting in the linear approximation (63) of $\phi_m(\lambda_m)$ the expression derived in (64), and using the condition (65), one can easily obtain the following rule for $\lambda_m > 0$

$$\lambda_m = \left| \frac{\eta \|e\| - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \right| \lambda_{m-1}. \tag{66}$$

In [24] this scheme has been called secant-update method: this is the rule that we employ in the following experiments. Depending on the problem, we have used the following classical

regularization matrices,

$$L_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(N-1) \times N}, \tag{67}$$

$$L_2 = \begin{bmatrix} 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2) \times N}, \tag{68}$$

which represent scaled finite difference approximations of the first and the second derivative, 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 9.

# 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 [25]. 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 the Krylov method described by (42), the norm of the residual associated to the 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 residual (pseudoresidual). This motivates the use of the following update formula to choose the regularization parameter at the $m$-th iteration

$$\lambda_m = \frac{\eta \phi_{m-1}(0) - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)} \lambda_{m-1}, \quad \eta \gtrsim 1. \tag{69}$$

## 5.2   Generalized Cross Validation (GCV)

The GCV parameter choice criterion prescribes to choose as regularization parameter the scalar $\lambda$ that minimizes the GCV functional

$$G(\lambda) = \frac{\|(I - AA_\lambda^\sharp)b\|^2}{(\text{trace}(I - AA_\lambda^\sharp))^2}, \tag{70}$$
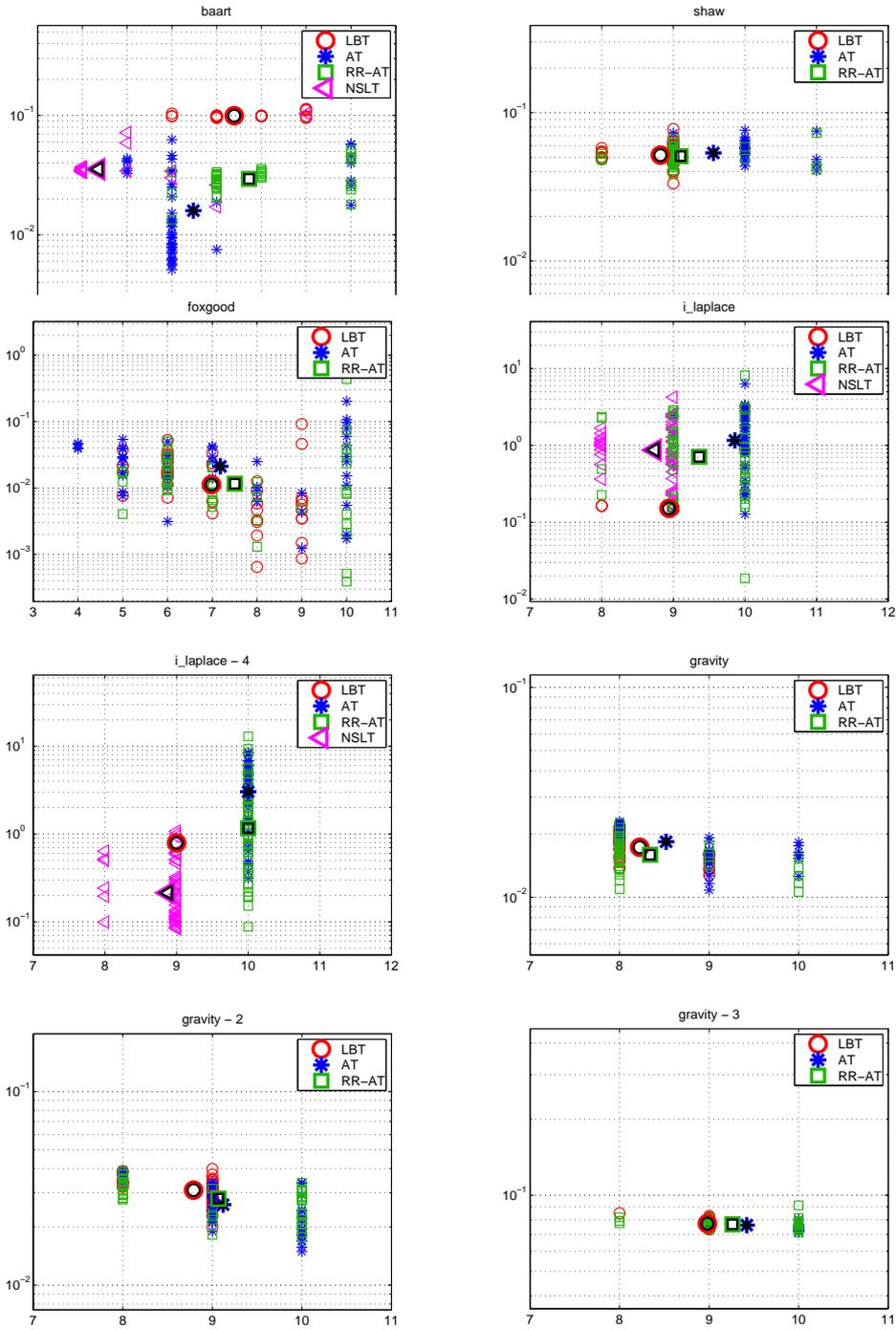
27

Figure 9: Accuracy of the automatically selected final approximation versus the number of iterations. The pictures collect the relative errors resulting from 30 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 $10^{-3}$.

where $A_\lambda^\sharp$ stands for the regularized inverse of $A$ associated to Tikhonov regularization (2); more precisely, considering the expression (51), we derive $A_\lambda^\sharp = (A^T A + \lambda^2 L^T L)^{-1} A^T$. To obtain an expression of $G(\lambda)$ easy to handle, one considers the GSVD of the matrix pair $(A, L)$, defined by (7).

When dealing with the regularized problems (56), in order to set $\lambda$ 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 $(\bar{D}_m, L_m)$ 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 [51], since the numerator of (70) 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 $\gamma_i$ are the generalized singular values of $(A, L)$ (cf. **??**GenSinVals)), the definition of the sequence of regularization parameters $\{\lambda_m\}_{m \geq 1}$ can be obtained by means of the minimization of the functionals

$$G_m^{\mathcal{K}}(\lambda) := \frac{\sum_{i=1}^{m} \left( \frac{\lambda^2}{\gamma_i^{(m)2} + \lambda^2} (\bar{u}_i^{(m)})^T Q_{m+1}^T b \right)^2 + \left( (\bar{u}_{m+1}^{(m)})^T Q_{m+1}^T b \right)^2}{\left( N - m + \sum_{i=1}^{m} \frac{\lambda^2}{\gamma_i^{(m)2} + \lambda^2} \right)^2}, \tag{71}$$

where we have used the expression of the (pseudo)residual given by Proposition 14. In other words, the numerator is defined using the (pseudo)residual, and the denominator by replacing $\gamma_i$ with $\gamma_i^{(m)}$ for $i \leq m$, and $\gamma_i$ 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 [33, Chapter 7]).

By considering the expression (57), we can immediately state that the Krylov-Tikhonov methods produce a regularized solution given by

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

so that, we can state the following result, which provides a clearer interpretation of (71).

**Proposition 15** *For each one of the method considered, 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*

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

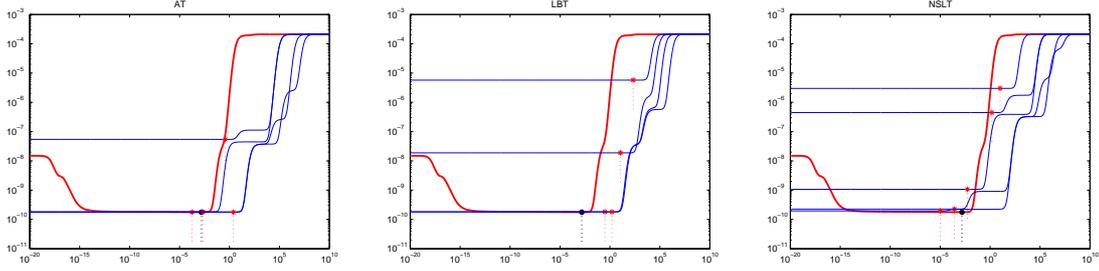*where $A_m^{\mathcal{K}}$ is given by (47). The equal sign just holds for the LBT and the AT methods.*

29

Figure 10: 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 $\lambda_m$.

**Proof.**

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

Therefore

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

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

We remark that, since

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

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 (71).

In Figure 10 we plot some of the approximations of $G(\lambda)$ attained by $G_m^{\mathcal{K}}(\lambda)$.

## 5.3   L-curve

The L-curve criterion [31] is based on defining the regularization parameter for (2) as the scalar $\lambda$ that maximizes the curvature of the parametric curve

$$
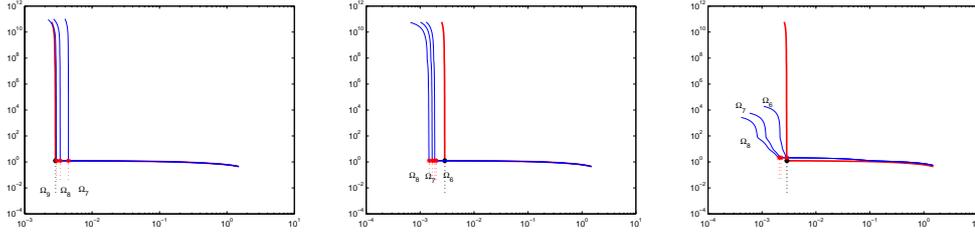\Omega(\lambda) = (\log \|b - A x_\lambda\|, \log \|L x_\lambda\|).
$$

30

Figure 11: 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.

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

This criterion has already been used in connection with Krylov-Tikhonov methods (cf. [4, 9]). The basic idea consists in assuming that the curves

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

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

**Proposition 16** Let $y_{m,\lambda}$ be the solution of (56). Using the GSVD of the matrix pair $(\bar{D}_m, L_m)$, we obtain

$$\| L_m y_{m,\lambda} \|^2 = \sum_{i=1}^{m} \left( \frac{\gamma_i^{(m)}}{\gamma_i^{(m)2} + \lambda^2} (\bar{u}_i^{(m)})^T Q_{m+1}^T b \right)^2. \tag{72}$$

**Proof.** Since $L_m = \bar{V}_m \bar{C}_m \bar{X}_m^{-1}$, the proof follows directly from (59). ∎

Using the expressions (58) and (72), 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. [37, 14, 58]), in our experiments we use the L-curve criterion based on the adaptive algorithm referred to as the "pruning algorithm" [36]. In Figure 11 we plot some of the approximations of $\Omega(\lambda)$ obtained with $\Omega_m(\lambda)$.

## 5.4   Regińska

Regińska criterion [54] is a very efficient parameter choice rule for Tikhonov regularization, and it is closely related to the L-curve criterion. The regularization parameter $\lambda$ is defined as the minimum of the function

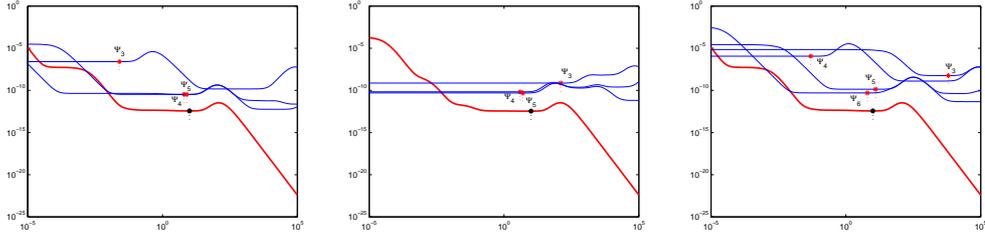$$\Psi_\mu(\lambda) = \| b - A x_\lambda \|^2 \| L x_\lambda \|^{2\mu}, \quad \mu > 0,$$

Figure 12: 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 $\lambda_m$. In each picture $\mu = 1$.

for proper a $\mu$. Analogously to the L-curve criterion, this rule is motivated by the observation that finding the minimizer of $\Psi_\mu$ corresponds to considering a good balance between the size of the regularization term and the size of the residual norm. In [54] the author proves that, if the curvature of the L-curve is maximized at $\lambda^*$ and if the tangent to L-curve at $(\log \|b - Ax_{\lambda^*}\|, \log \|Lx_{\lambda^*}\|)$ has slope $-1/\mu$, then $\Psi_\mu(\lambda)$ is minimized at $\lambda^*$.

As in the previous cases, in order to set $\lambda_m$ step by step when dealing with the projected regularized problems (56), 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 of $(\bar{D}_m, L_m)$ by using again (58) and (72). In Figure 12 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 to the four criteria (Sections 5.1–5.4). In each picture, 50 runs of the method have been executed. The final approximations have been selected by checking the relative residual. In particular each run is stopped whenever

$$\frac{|\|r_m\| - \|r_{m-1}\||}{\|r_m\|} \leq \varepsilon, \tag{73}$$

where $\varepsilon = 1.05$. The results, in terms of relative error versus number of iterations, are reported in Figure 13. These pictures (together with many other that are not reported) reveal that the four criteria are somehow equivalent when coupled with the Krylov methods here considered, since it is not easy to detect the one that clearly overtakes the others.
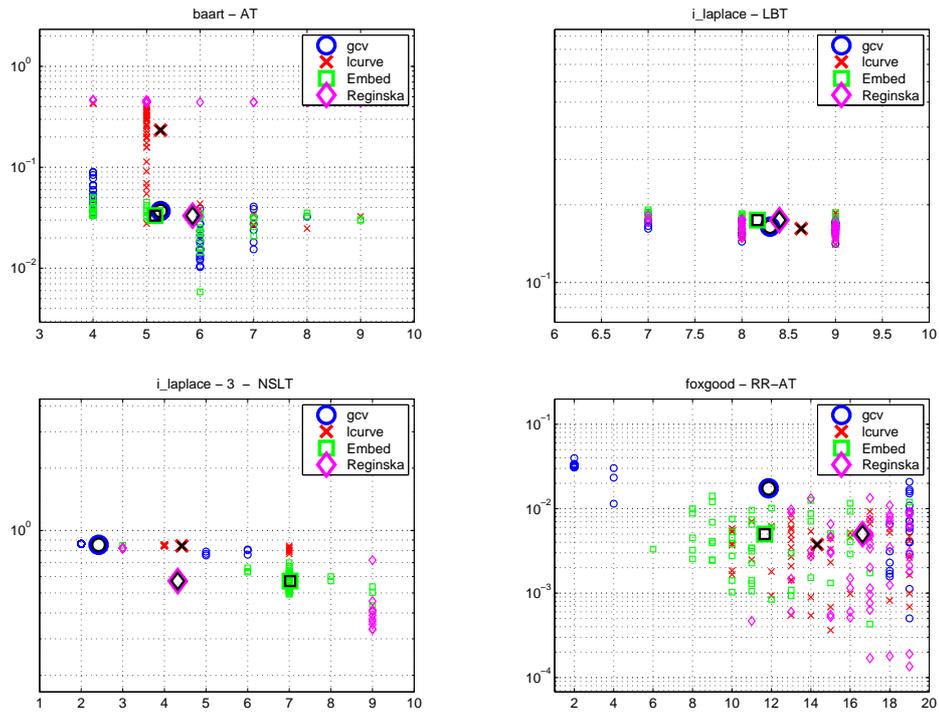
32

Figure 13: Relative error of the final approximation selected by (73) 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 $10^{-2}$.

33

# 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 applied to the denoising and deblurring of corrupted images: among the most recent ones we cite [1, 35, 22, 7, 50]. In this section we closely follow the approach adopted in [35], and we consider a medical and an astronomical test image of size $256 \times 256$ 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 14). Both the

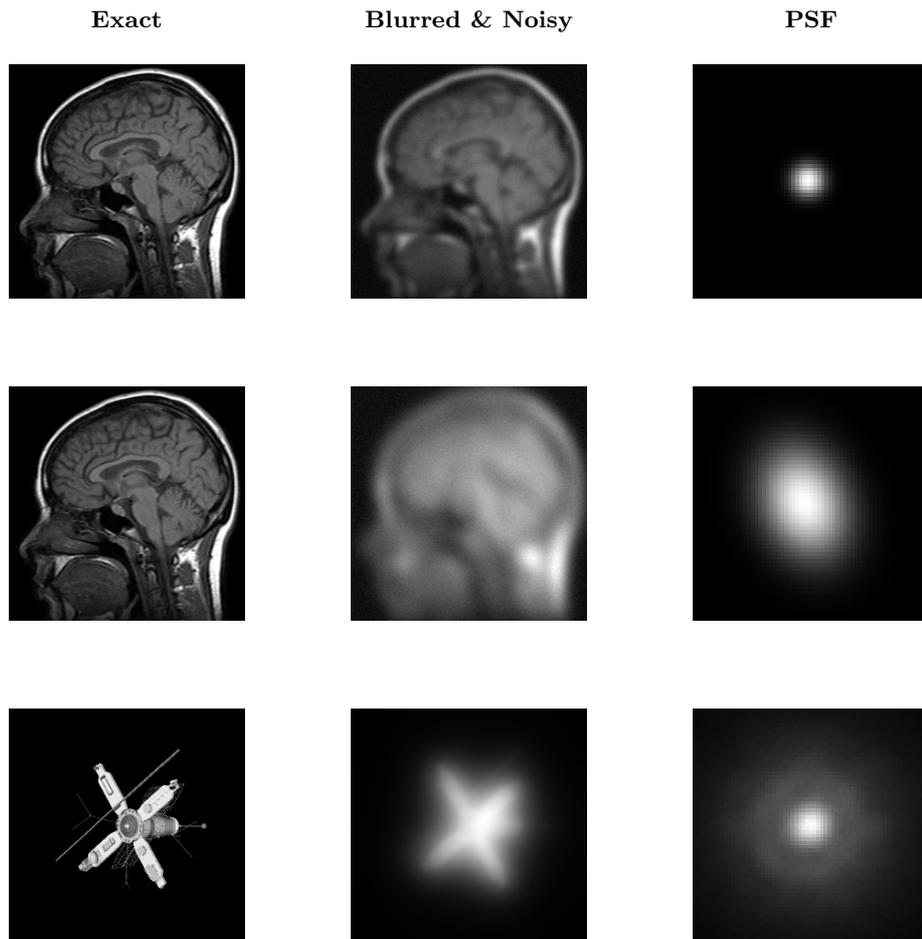| Exact | Blurred & Noisy | PSF |
|:-----:|:---------------:|:---:|



Figure 14: 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.

34

isotropic and the non-isotropic blurs are analytically defined starting from a Gaussian PSF $K(s,t)$, i.e.,

$$K(s,t) = \frac{1}{2\pi\sqrt{\alpha_1^2\alpha_2^2 - \rho^4}} \exp\left(-\frac{1}{2}\begin{bmatrix} s & t \end{bmatrix}\begin{bmatrix} \alpha_1^2 & \rho^2 \\ \rho^2 & \alpha_2^2 \end{bmatrix}^{-1}\begin{bmatrix} s \\ t \end{bmatrix}\right). \tag{74}$$

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 [49], 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* [49] 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 Arnoldi-Tikhonov and Lanczos-Tikhonov methods. We extensively test both the standard and the general form regularization, taking

$$L = L_1 \otimes I_N + I_N \otimes L_1, \tag{75}$$

where $L_1$ is defined in (67); the regularization matrix $L$ represents the sum of first derivatives in the horizontal and vertical directions of the two-dimensional image. In Table 2 and 3 we collect the results obtained by extensively 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, we report the average of the minimum relative errors and the average of the iterations at which it is attained. In Figure 15 we display the history of the relative errors when a "quasi-optimal" regularization parameter is chosen; in Figure 16 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.

35

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

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

Table 3: 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.

| | Standard Form | | | General Form | | |
|---|---|---|---|---|---|---|
| | LBT | AT | RRAT | LBT | AT | RRAT |
| Discrepancy | $1.9583 \cdot 10^{-1}$ (30) | $1.9446 \cdot 10^{-1}$ (9) | $3.5581 \cdot 10^{-1}$ (30) | $2.0674 \cdot 10^{-1}$ (13.32) | $1.9000 \cdot 10^{-1}$ (9) | $3.6561 \cdot 10^{-1}$ (2) |
| Embedded | $1.8773 \cdot 10^{-1}$ (30) | $2.4204 \cdot 10^{-1}$ (5) | $3.0173 \cdot 10^{-1}$ (10) | $1.8897 \cdot 10^{-1}$ (30) | $1.9073 \cdot 10^{-1}$ (9.04) | $3.0468 \cdot 10^{-1}$ (23.12) |
| GCV | $1.8930 \cdot 10^{-1}$ (19.28) | $2.5366 \cdot 10^{-1}$ (2) | $2.9817 \cdot 10^{-1}$ (23.76) | $1.8930 \cdot 10^{-1}$ (19.28) | $2.5366 \cdot 10^{-1}$ (2) | $2.9817 \cdot 10^{-1}$ (23.76) |
| L-curve | $1.8777 \cdot 10^{-1}$ (30) | $1.9845 \cdot 10^{-1}$ (9) | $2.9836 \cdot 10^{-1}$ (30) | $3.6390 \cdot 10^{-1}$ (2) | $1.8681 \cdot 10^{-1}$ (27.27) | $3.6564 \cdot 10^{-1}$ (2) |
| Regińska | $1.8741 \cdot 10^{-1}$ (28.19) | $1.8779 \cdot 10^{-1}$ (11.37) | $2.9817 \cdot 10^{-1}$ (23) | $2.3784 \cdot 10^{-1}$ (9) | $2.3600 \cdot 10^{-1}$ (6) | $2.9817 \cdot 10^{-1}$ (23) |

Table 4: 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.
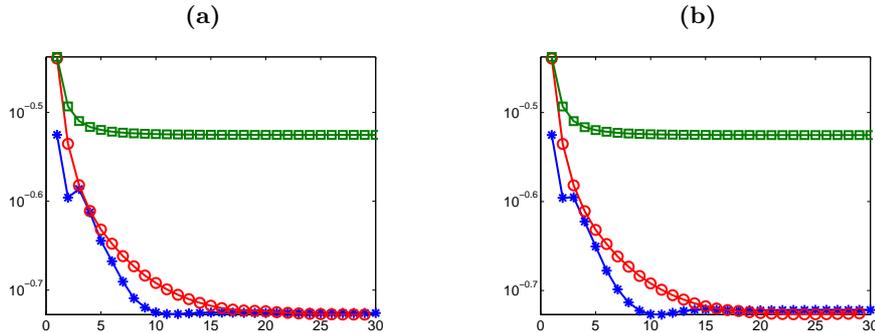
Figure 15: History of the best relative errors attainable when applying a Krylov-Tikhonov regularization in standard form (frame **(a)**) and general form (frame **(b)**) to the symmetrically blurred medical image. In both frames, the AT method is denoted by an asterisk, the RRAT method is denoted by a square, and the LBT method is denoted by a circle.
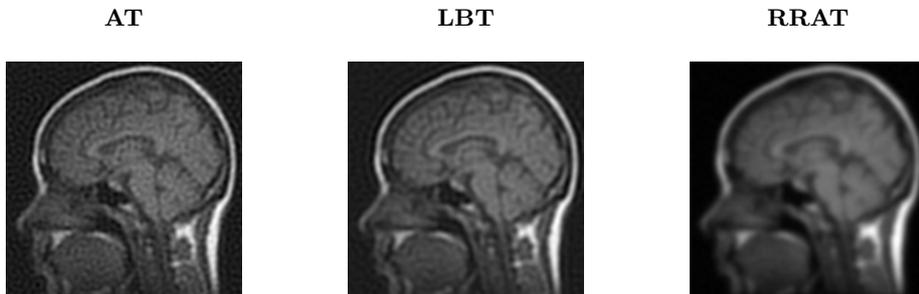


Figure 16: Best reconstructions obtained at the 12-th iteration of a Krylov-Tikhonov regularization method in general form applied to the symmetrically blurred medical image.

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, the LBT, and the NSLT, 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 (75). As before, in Table 5, 6 and 7 we collect the averages of the results obtained running 50 times each test, with different noise realizations; Figure 17 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 Table 5, 6 and Figure 17, we can clearly see that the performances of the AT and NSLT methods are very similar.

In Figure 18, we also perform a Discrete Cosine Transform (DCT) analysis: we display just the results relative to 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 15, 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 [35] for the purely iterative methods. Looking at Figure 18 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 [35], or [34]), the properties of the reconstructed solutions obtained by projecting the Tikhonov-regularized problem by all the Krylov methods are similar. As 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 17, 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 14), we exclusively consider general-form Tikhonov regularization, equipped with particular regularization matrices that are adaptively defined in order to increasingly better approximate the TV regularization [59]. 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

$$L_1^{hv} = \left( \begin{array}{c} L_1 \otimes I_N \\ I_N \otimes L_1 \end{array} \right), \quad L_1 \text{ given in (67)}.$$

This approach was first derived in [22], 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 19 we display the history of 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.

|  | LBT | AT | NSLT | RRAT | RRNSLT |
|---|---|---|---|---|---|
| **Discrepancy** | $4.0452 \cdot 10^{-1}$ (17.58) | $6.6419 \cdot 10^{-1}$ (6) | $6.6419 \cdot 10^{-1}$ (6) | $5.1898 \cdot 10^{-1}$ (30) | $5.1898 \cdot 10^{-1}$ (30) |
| **Embedded** | $4.4454 \cdot 10^{-1}$ (7) | $4.4540 \cdot 10^{-1}$ (6) | $4.4540 \cdot 10^{-1}$ (6) | $4.9395 \cdot 10^{-1}$ (5) | $4.9394 \cdot 10^{-1}$ (5) |
| **GCV** | $4.3772 \cdot 10^{-1}$ (7) | $5.2590 \cdot 10^{-1}$ (4) | $5.2590 \cdot 10^{-1}$ (4) | $4.9814 \cdot 10^{-1}$ (3) | $4.9815 \cdot 10^{-1}$ (3) |
| **L-curve** | $4.4238 \cdot 10^{-1}$ (13) | $4.3188 \cdot 10^{-1}$ (6) | $4.3188 \cdot 10^{-1}$ (6) | $4.9931 \cdot 10^{-1}$ (9) | $4.9931 \cdot 10^{-1}$ (9) |
| **Regińska** | $4.3883 \cdot 10^{-1}$ (7) | $4.8324 \cdot 10^{-1}$ (4) | $4.8324 \cdot 10^{-1}$ (4) | $4.9390 \cdot 10^{-1}$ (4) | $4.9391 \cdot 10^{-1}$ (4) |

Table 5: 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** | $4.0484 \cdot 10^{-1}$ (17.58) | $4.2381 \cdot 10^{-1}$ (9) | $4.2381 \cdot 10^{-1}$ (9) | $5.2356 \cdot 10^{-1}$ (2) | $8.7260 \cdot 10^{-1}$ (2) |
| **Embedded** | $4.4383 \cdot 10^{-1}$ (8) | $4.1973 \cdot 10^{-1}$ (7) | $4.1973 \cdot 10^{-1}$ (7) | $4.9527 \cdot 10^{-1}$ (6) | $8.2545 \cdot 10^{-1}$ (6) |
| **GCV** | $4.3772 \cdot 10^{-1}$ (7) | $5.2589 \cdot 10^{-1}$ (4) | $5.2589 \cdot 10^{-1}$ (4) | $4.9814 \cdot 10^{-1}$ (3) | $8.3023 \cdot 10^{-1}$ (3) |
| **L-curve** | $5.2278 \cdot 10^{-1}$ (2) | $4.2212 \cdot 10^{-1}$ (7) | $4.2212 \cdot 10^{-1}$ (7) | $5.0784 \cdot 10^{-1}$ (6) | $8.4640 \cdot 10^{-1}$ (6) |
| **Regińska** | $4.9180 \cdot 10^{-1}$ (5) | $4.9459 \cdot 10^{-1}$ (6.4) | $4.9459 \cdot 10^{-1}$ (6.4) | $4.9419 \cdot 10^{-1}$ (4) | $4.9419 \cdot 10^{-1}$ (4) |

Table 6: 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 |
|---|---|---|---|---|---|
| | | **Standard Form** | | | |
| **Discrepancy** | $3.9973 \cdot 10^{-1}$ (28.84) | $3.9089 \cdot 10^{-1}$ (12) | $3.9089 \cdot 10^{-1}$ (12) | $5.1898 \cdot 10^{-1}$ (30) | $5.1898 \cdot 10^{-1}$ (30) |
| **Embedded** | $3.9904 \cdot 10^{-1}$ (30) | $3.8631 \cdot 10^{-1}$ (13.08) | $3.8631 \cdot 10^{-1}$ (13.08) | $4.9058 \cdot 10^{-1}$ (11) | $4.9053 \cdot 10^{-1}$ (11) |
| **GCV** | $3.8655 \cdot 10^{-1}$ (30) | $4.5633 \cdot 10^{-1}$ (2) | $4.5633 \cdot 10^{-1}$ (2) | $4.8706 \cdot 10^{-1}$ (30) | $4.8711 \cdot 10^{-1}$ (30) |
| **L-curve** | $4.0481 \cdot 10^{-1}$ (30) | $3.7649 \cdot 10^{-1}$ (19.22) | $3.7649 \cdot 10^{-1}$ (19.22) | $4.8755 \cdot 10^{-1}$ (30) | $4.8757 \cdot 10^{-1}$ (30) |
| **Regińska** | $3.9883 \cdot 10^{-1}$ (30) | $3.9578 \cdot 10^{-1}$ (16) | $3.9578 \cdot 10^{-1}$ (16) | $4.8708 \cdot 10^{-1}$ (30) | $4.8705 \cdot 10^{-1}$ (30) |
| | | **General Form** | | | |
| **Discrepancy** | $4.0375 \cdot 10^{-1}$ (24.14) | $3.8183 \cdot 10^{-1}$ (14) | $3.8183 \cdot 10^{-1}$ (14) | $5.2356 \cdot 10^{-1}$ (2) | $5.2356 \cdot 10^{-1}$ (2) |
| **Embedded** | $4.0315 \cdot 10^{-1}$ (30) | $3.8095 \cdot 10^{-1}$ (14.96) | $3.8095 \cdot 10^{-1}$ (14.96) | $4.9396 \cdot 10^{-1}$ (8) | $4.9396 \cdot 10^{-1}$ (8) |
| **GCV** | $3.8655 \cdot 10^{-1}$ (30) | $4.5633 \cdot 10^{-1}$ (3) | $4.5633 \cdot 10^{-1}$ (3) | $4.8706 \cdot 10^{-1}$ (30) | $4.8706 \cdot 10^{-1}$ (30) |
| **L-curve** | $5.2278 \cdot 10^{-1}$ (2) | $3.7392 \cdot 10^{-1}$ (19.16) | $3.7392 \cdot 10^{-1}$ (19.16) | $4.8748 \cdot 10^{-1}$ (30) | $4.8748 \cdot 10^{-1}$ (30) |
| **Regińska** | $4.9030 \cdot 10^{-1}$ (4) | $4.8789 \cdot 10^{-1}$ (2) | $4.8789 \cdot 10^{-1}$ (2) | $4.8708 \cdot 10^{-1}$ (30) | $4.8708 \cdot 10^{-1}$ (30) |

Table 7: 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.
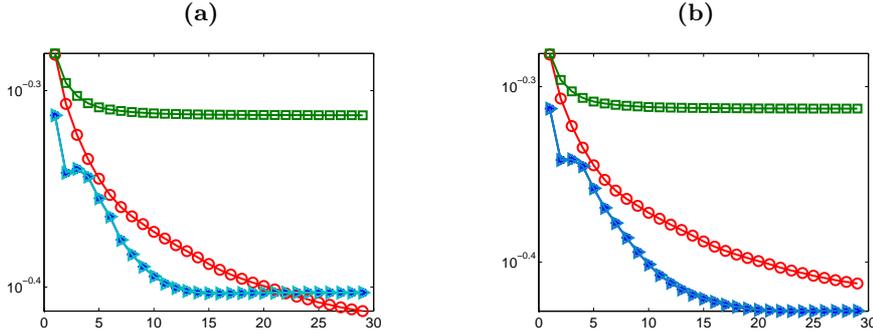
Figure 17: History of the best relative errors attainable when applying a Krylov-Tikhonov regularization in standard form (frame **(a)**) and general form (frame **(b)**) to the unsymmetrically blurred medical image. In both frames, the AT method is denoted by an asterisk, the RRAT method is denoted 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.

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 more stable one.

In Figure 20 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 8 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 9 we report the best attainable relative error 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., when the fixed maximum number of iterations per restarts is performed) the exact solution is poorly approximated and the method fails. 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.
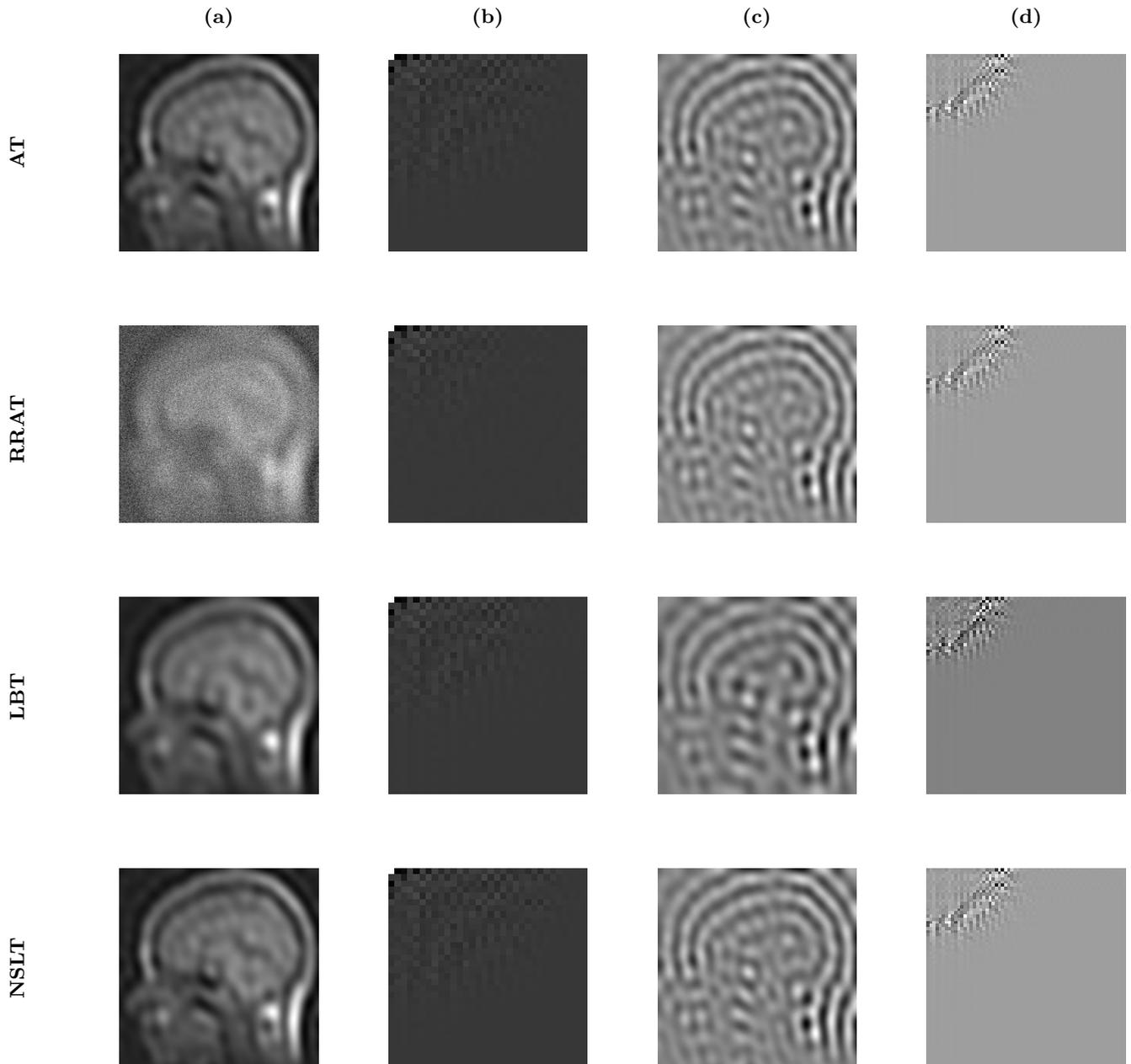
40

Figure 18: Spectral analysis of different Krylov subspace methods employed to project the general-form Tikhonov-regularized problem. More precisely, we show $x_m^{ex}$ **(a)**, a blow-up (800%) of the DCT of $x_m^{ex}$ **(b)**, $x_m^e$ **(c)**, and a blow-up (400%) of the DCT of $x_m^e$ **(d)** after 10 iterations have been performed (i.e., $m = 10$).
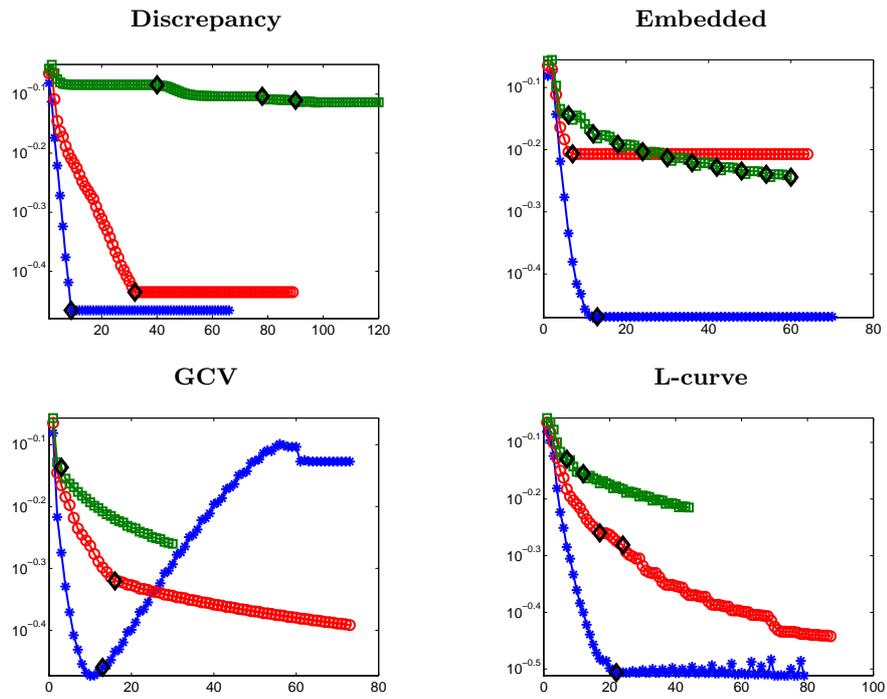
Figure 19: History of the relative errors obtained when approximating the TV regularization by means of suitably restarted Krylov-Tikhonov methods. The AT method is denoted by an asterisk, the RRAT method is denoted by a square, the LBT method is denoted by a circle. We use a big diamond to highlight 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.
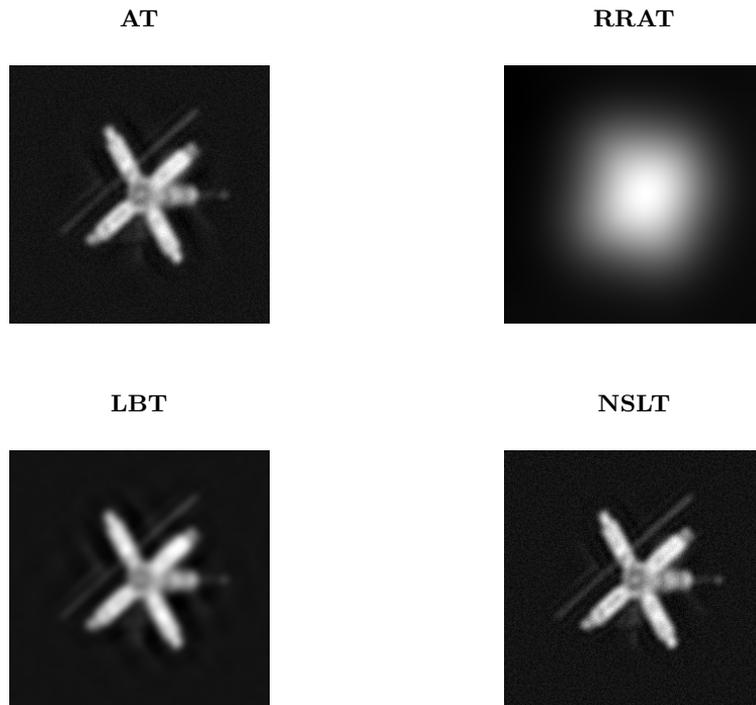
**AT**



**RRAT**



**LBT**



**NSLT**



Figure 20: Reconstructions obtained by different Krylov-Tikhonov regularization methods at the end of the restarting scheme.

43

| | LBT | AT | NSLT | RRAT | RRNSLT |
|---|---|---|---|---|---|
| **Discrepancy** | $3.6679 \cdot 10^{-1}$ (89.03) | $3.4162 \cdot 10^{-1}$ (66) | $3.4162 \cdot 10^{-1}$ (66) | $7.6822 \cdot 10^{-1}$ (181) | $7.6761 \cdot 10^{-1}$ (181) |
| **Embedded** | $6.2061 \cdot 10^{-1}$ (64) | $3.3932 \cdot 10^{-1}$ (70) | $3.3928 \cdot 10^{-1}$ (70) | $5.6985 \cdot 10^{-1}$ (60) | $5.6986 \cdot 10^{-1}$ (60) |
| **GCV** | $4.0588 \cdot 10^{-1}$ (73) | $6.9823 \cdot 10^{-1}$ (81.13) | $6.8135 \cdot 10^{-1}$ (79.27) | $5.4931 \cdot 10^{-1}$ (30) | $5.4930 \cdot 10^{-1}$ (30) |
| **L-curve** | $3.5585 \cdot 10^{-1}$ (86.60) | $3.0717 \cdot 10^{-1}$ (79.17) | $3.0274 \cdot 10^{-1}$ (83.97) | $6.0955 \cdot 10^{-1}$ (44) | $6.0955 \cdot 10^{-1}$ (44) |
| **Regińska** | $4.4106 \cdot 10^{-1}$ (65.50) | $3.5297 \cdot 10^{-1}$ (71) | $3.5199 \cdot 10^{-1}$ (71) | $5.5428 \cdot 10^{-1}$ (33) | $5.4327 \cdot 10^{-1}$ (37) |

Table 8: 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 | 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** | $5.8916 \cdot 10^{-3}$ | $3.1623 \cdot 10^{-3}$ | $1.9140 \cdot 10^{-3}$ | $1.5849 \cdot 10^{-3}$ | $1.0212 \cdot 10^{-3}$ |

Table 9: Averages (over 50 runs) of the regularization parameters obtained when the stopping criterion is satisfied.

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 [35] for the purely iterative methods, we conclude that the most effective Krylov-Tikhonov methods seem 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 is a regularized problem, a certain amount of regularization is added as the problem is 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 [35], is filtered out. Moreover, even if the SVD components of the matrix $A$ are mixed in $K_m(A, b)$ [34], 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 (number of iterations). Anyway, for what concerns the computational cost, we underline that the Arnoldi-based methods do not work with the transpose nor need the reorthogonalization of the Krylov vectors, so that they are basically cheaper than the Lanczos based ones.

# References

[1] S. Berisha and J. G. Nagy. Iterative image restoration. In R. Chellappa and S. Theodoridis, editors, *Academic Press Library in Signal Processing*, volume 4, chapter 7, pages 193–243. Elsevier, 2014.

[2] Å. Björck. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT*, 28(3):659–670, 1988.

[3] Å. Björck, E. Grimme, and P. van Dooren. An implicit shift bidiagonalization algorithm for ill-posed systems. *BIT*, 34(4):510–534, 1994.

[4] D. Calvetti, G. H. Golub, and L. Reichel. Estimation of the L-curve via Lanczos bidiagonalization. *BIT*, 39(4):603–619, 1999.

[5] D. Calvetti, B. Lewis, and L. Reichel. GMRES-type methods for inconsistent systems. *Lin. Alg. Appl.*, 316:157–169, 2000.

[6] D. Calvetti, B. Lewis, and L. Reichel. Krylov subspace iterative methods for nonsymmetric discrete ill-posed problems in image restoration. In F. T. Luk, editor, *Advanced Signal Processing Algorithms, Architectures, and Implementations XI*, volume 4474 of *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, pages 224–233, Bellingham, WA, 2001. The International Society for Optical Engineering.

[7] D. Calvetti, B. Lewis, and L. Reichel. On the choice of subspace for iterative methods for linear discrete ill-posed problems. *Int. J. Appl. Math. Comput. Sci.*, 11(5):1069–1092, 2001.

[8] D. Calvetti, B. Lewis, and L. Reichel. On the regularizing properties of the GMRES method. *Numer. Math.*, 91:605–625, 2002.

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

[10] D. Calvetti and L. Reichel. Tikhonov regularization with a solution constraint. *SIAM J. Sci. Comput.*, 26(1):224–239, 2001.

[11] D. Calvetti and L. Reichel. Tikhonov regularization of large linear problems. *BIT*, 43(2):263–283, 2003.

[12] D. Calvetti, L. Reichel, F. Sgallari, and G. Spaletta. A regularizing Lanczos iteration method for underdetermined linear systems. *J. Comput. Appl. Math.*, 115(1-2):101–120, 2000.

[13] D. Calvetti, L. Reichel, and A. Shuibi. Tikhonov regularization of large symmetric problems. *Numer. Linear Algebra Appl.*, 12(2-3):127–139, 2005.

[14] J. J. Castellanos, S. Gómez, and V. Guerra. The triangle method for finding the corner of the L-curve. *Applied Numerical Mathematics*, 43:359–373, 2002.

[15] J. Chung, J. G. Nagy, and D. P. O'Leary. A weighted GCV method for Lanczos hybrid regularization. *Elec. Trans. Numer. Anal.*, 28:149–167, 2008.

[16] E. J. Craig. The n-step iteration procedures. *J. Math. Phys.*, 34:64–73, 1955.

[17] L. Dykes and L. Reichel. A family of range restricted iterative methods for linear discrete ill-posed problems. *Dolomites Research Notes on Approximation*, 6:27–36, 2013. Proceedings of DWCAA12.

[18] L. Eldén. Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, 17:134–145, 1977.

[19] J. N. Franklin. Minimum principles for ill-posed problems. *SIAM J. Math. Anal.*, 9:638–650, 1978.

[20] R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.

[21] S. Gazzola. *Regularization techniques based on Krylov methods for ill-posed linear systems*. PhD thesis, Dept. of Mathematics, University of Padua, Italy, 2014.

[22] S. Gazzola and J. G. Nagy. Generalized Arnoldi-Tikhonov method for sparse reconstruction. *SIAM J. Sci. Comput.*, 36(2), 2014. DOI: 10.1137/130917673.

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

[24] S. Gazzola and P. Novati. Automatic parameter setting for Arnoldi-Tikhonov methods. *J. Comput. Appl. Math.*, 256:180–195, 2014. DOI: 10.1016/j.cam.2013.07.023.

[25] S. Gazzola, P. Novati, and M. R. Russo. Embedded techniques for choosing the parameter in Tikhonov regularization. *Numer. Linear Algebra Appl.*, 2014. DOI: 10.1002/nla.1934.

[26] G. H. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Numer. Anal.*, 2:205–224, 1965.

[27] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[28] M. Hanke. On Lanczos based methods for the regularization of discrete ill-posed problems. *BIT*, 41(5):1008–1018, 2001.

[29] M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surv. Math. Ind.*, 3:253–315, 1993.

[30] P. C. Hansen. The discrete Picard condition for discrete ill-posed problems. *BIT*, 30(4):658–672, 1990.

[31] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

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

[33] P. C. Hansen. *Rank-deficient and discrete ill-posed problems*. SIAM, Philadelphia, PA, 1998.

[34] P. C. Hansen and T. K. Jensen. Smoothing-norm preconditioning for regularizing minimum-residual methods. *SIAM J. Matrix Anal. Appl.*, 29:1–14, 2006.

[35] P. C. Hansen and T. K. Jensen. Noise propagation in regularizing iterations for image deblurring. *Electron. Trans. Numer. Anal.*, 31:204–220, 2008.

[36] P. C. Hansen, T. K. Jensen, and G. Rodriguez. An adaptive pruning algorithm for discrete L-curve criterion. *J. Comp. Appl. Math.*, 198:483–492, 2007.

[37] 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.*, 14:1487–1503, 1993.

[38] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436, 1952.

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

[40] M. E. Hochstenbach, N. McNinch, and L. Reichel. Discrete ill-posed least-squares problems with a solution norm constraint. *Linear Algebra Appl.*, 436(10):3801–3818, 2012.

[41] B. Hofman. Regularization of nonlinear problems and the degree of ill-posedness. In G. Anger, R. Gorenflo, H. Jochmann, H. Moritz, and W. Webers, editors, *Inverse Problems: Principle and Applications in Geophysics, Technology and Medicine.* Akademie Verlag, Berlin, 1993.

[42] M. E. Kilmer, P. C. Hansen, and M. I. Español. A projection-based approach to general-form Tikhonov regularization. *SIAM J. Sci. Comput.*, 29(1):315–330, 2007.

[43] M. E. Kilmer and D.P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl.*, 22(4):1204–1221, 2001.

[44] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45:255–282, 1950.

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

[46] I. Moret. A note on the superlinear convergence of GMRES. *SIAM J. Numer. Anal.*, 34(2):513–516, 1997.

[47] S. Morigi, L. Reichel, and F. Sgallari. An iterative Lavrentiev regularization method. *BIT*, 46(3):589–606, 2006.

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

[49] J. G. Nagy, K. M. Palmer, and L. Perrone. Iterative methods for image deblurring: a Matlab object oriented approach. *Numer. Algorithms*, 36:73–93, 2004. See also: http://www.mathcs.emory.edu/~nagy/RestoreTools.

[50] P. Novati and M. R. Russo. Adaptive Arnoldi-Tikhonov regularization for image restoration. *Numer. Algorithms*, 65(4):745–757, 2013.

[51] P. Novati and M. R. Russo. A GCV based Arnoldi-Tikhonov regularization method. *BIT*, 2013. DOI:10.1007/s10543-013-0447-z.

[52] 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(4):474–489, 1981.

[53] C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.

[54] T. Regińska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17:740–749, 1996.

[55] L. Reichel, G. Rodriguez, and S. Seatzu. Error estimates for large-scale ill-posed problems. *Numer. Algorithms*, 51(3):341–361, 2009.

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

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

[58] G. Rodriguez and D. Theis. An algorithm for estimating the optimal regularization parameter by the L-curve. *Rendiconti di Matematica*, 25:69–84, 2005.

[59] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[60] Y. Saad. *Numerical methods for large eigenvalue problems. Algorithms and Architectures for Advanced Scientific Computing.* Halsted Press, New York, NY, 1992.

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

[62] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual method for solving nonsymmetric linear systems. *SIAM J. Sci. Comput.*, 7:856–869, 1986.

[63] W. Squire. The solution of ill-conditioned linear systems arising from Fredholm equations of the first kind by steepest descents and conjugate gradients. *Int. J. Numer. Meth. Eng.*, 10:607–617, 1976.