Comp. Appl. Math.
DOI 10.1007/s40314-014-0174-3

Automatic stopping rule for iterative methods in discrete ill-posed problems

Leonardo S. Borges · Fermín S. Viloche Bazán · Maria C. C. Cunha

Received: 20 December 2013 / Revised: 2 April 2014 / Accepted: 18 July 2014

Abstract The numerical treatment of large-scale discrete ill-posed problems is often accomplished iteratively by projecting the original problem onto a k-dimensional subspace with k acting as regularization parameter. Hence, to filter out the contribution of noise in the computed solution, the iterative process must be stopped early. In this paper, we analyze in detail a stopping rule for LSQR proposed recently by the authors, and show how to extend it to Krylov subspace methods such as GMRES, MINRES, etc. Like the original rule, the extended version works well without requiring a priori knowledge about the error norm and stops automatically after $\widetilde{k}+1$ steps where $\widetilde{k}$ is the computed regularization parameter. The performance of the stopping rule on several test problems is illustrated numerically.

Keywords Iterative methods · Parameter choice rules · Large-scale discrete ill-posed problems

Mathematics Subject Classification 15A06 · 15A18 · 15A23

Communicated by Jinyun Yuan.

Leonardo S. Borges Part of this research is supported by FAPESP, Brazil, grant 2009/52193-1.

Fermín S. Viloche Bazán The work of this author is supported by CNPq, Brazil, grants 308709/2011-0, 477093/2011-6.

L. S. Borges
Federal University of Santa Catarina, 89065-300 Blumenau, SC, Brazil
e-mail: lsbplsb@yahoo.com

F. S. Viloche Bazán (✉)
Department of Mathematics, Federal University of Santa Catarina,
88040-900 Florianópolis, SC, Brazil
e-mail: fermin@mtm.ufsc.br

M. C. C. Cunha
Department of Applied Mathematics, IMECC-UNICAMP,
University of Campinas, CP 6065, 13081-970 Campinas, SP, Brazil

Published online: 09 August 2014

1 Introduction

Linear discrete ill-posed problems of the form

$$\underset{f\in\mathbb{R}^n}{\operatorname{argmin}} \|g - Af\|_2^2, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$, $m \geq n = \operatorname{rank}(A)$, is severely ill-conditioned and the available data $g \in \mathbb{R}^m$ are contaminated with noise, i.e., $g = g_{\text{exact}} + e$, arises naturally in a range of applications in science and engineering. It is well known that the naive least squares solution $f_{\text{LS}} = A^\dagger g$ (in which $A^\dagger$ denotes the Moore–Penrose pseudoinverse of A) has no relation with the desired noise-free solution f_{exact}. To calculate solutions that resemble f_{exact} and are less sensitive to inaccuracies in g, some kind of regularization must be imposed to (1.1). Perhaps, one of the most well-known methods is due to Tikhonov (1963), in which a penalization term is added to (1.1) with a positive scalar controlling how much regularization is applied, see, e.g., (Engl et al. 1996; Hanke and Hansen 1993; Hansen 1998), and the references therein. Another way to construct regularized solutions is through iterative methods for which the iteration index k plays the role of the regularization parameter. The reason is that for many iterative methods, in the early stages of the algorithm, the iterates seem to converge to f_{exact}, then the iterates become worse and finally converge to the solution of (1.1) that is useless; this phenomenon is called *semi-convergence*. Thus, stopping the iteration after a well-chosen number of steps has a regularizing effect. Such methods are attractive for the following reasons:

- Matrix A is only accessed to perform matrix–vector products and no decomposition is required: if A is sparse, any matrix decomposition will, probably, destroy this sparsity or any special structure Björck (1996).
- Usually the performed operations are matrix–vector products, saxpy ($f \leftarrow \alpha f + g$), vector norms and solutions of triangular systems. These operations are easily parallelizable.
- If matrix A can only be accessed by means of a black-box, i.e., A is not known, iterative methods are the only alternative.
- Iterative methods must be interrupted, specially while dealing with large-scale linear discrete ill-posed problems. The iterated solutions as well as the respective residuals can be monitored and used to design stopping criteria.

Iterative methods include Landweber iterations Landweber (1951), Cimmino iterations Cimmino (1983), algebraic reconstruction technique Dold and Eckmann (1986), and Krylov projection methods including CGLS, MINRES, etc (Bunse-Gerstner et al. 2006; Hanke 1995; Hestenes and Stiefel 1952; Neuman et al. 2012; Paige and Saunders 1975, 1982; Saad and Schultz 1986). While most researchers exploit a priori knowledge of the norm of the noise in the data and stop the iterative process using the discrepancy principle (DP) of (Morozov 1984), developing stopping rules for iterative methods that do not exploit this information is still a topic of active research (Bazán et al. 2013; Castellanos et al. 2002; Hansen et al. 2007; Kilmer and O'Leary 2001; Reichel and Rodriguez 2012). A drawback associated with rules that do not exploit information about the noise, referred to as heuristic, is that they can fail, at least for some problems (Bakushinski 1984). Nevertheless, they are important on its own right and proven to be successful in a number of areas. For a discussion and analyses of heuristic rules, the reader is referred to (Engl et al. 1996; Kindermann 2011).

In this paper, we analyze the underlying properties of a stopping rule for LSQR proposed recently by Bazán et al. (2013), and show how to extend it to Krylov subspace projection methods such as GMRES, MINRES, etc., for which regularization is achieved by projecting

the original problem onto a Krylov subspace and where the dimension of this subspace plays the role of the regularization parameter. In addition, because some classical iterative methods (e.g., GMRES) are not always able to calculate useful solutions to some problems, as an attempt to overcome this drawback, we also consider how to apply the rule to certain preconditioned versions of the methods. In the present work, by preconditioning we mean the incorporation of smoothing properties of f_{exact} into the computed iterates, as suggested by Calvetti et al. (2005) and Hansen and Jensen (2006). Our focus is on a stopping rule that works in connection with large-scale problems with and without preconditioning and without a priori knowledge about the norm of the noise.

The rest of the paper is organized as follows. In Sect. 2, some classical projection methods are reviewed. In Sect. 3, the k-th iterated solution generated by the projection method is compared with the TSVD solution, and error bounds that provide insight into the choice of the regularization parameter are derived. The bounds reveal that the closeness of the k-th iterate and the TSVD solution depends on the subspace angle between the Krylov subspace and the corresponding right singular subspace of A, with the angle depending on the distance $\|A - AP_k\|_2$ as a function of k, where P_k is the orthogonal projector onto the Krylov subspace. Section 4 describes the extension of the automatic stopping rule for LSQR proposed in Bazán et al. (2013) to the general Krylov projection method. How to include a priori information about the desired solution into the iterative process can be found in Sect. 5. Section 6 contains some numerical experiments. The paper ends with concluding remarks in Sect. 7.

We close this section by introducing some notation that will be used throughout the paper. As usual, the SVD of A is given by

$$A = \mathsf{U}\begin{pmatrix}\Sigma\\ 0\end{pmatrix}\mathsf{V}^T, \tag{1.2}$$

where $\mathsf{U} = [\mathsf{u}_1, \ldots, \mathsf{u}_m] \in \mathbb{R}^{m\times m}$ and $\mathsf{V} = [\mathsf{v}_1, \ldots, \mathsf{v}_n] \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{n\times n}$ is a diagonal matrix, $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$, with the singular values σ_i ordered as usual, i.e., $\sigma_1 \geq \cdots \geq \sigma_n > 0$. The naive least squares solution is thus given by

$$f_{\text{LS}} = \sum_{i=1}^{n} \frac{\mathsf{u}_i^T g}{\sigma_i}\mathsf{v}_i. \tag{1.3}$$

Furthermore, for future use let us define

$$\mathsf{U}_k = [\mathsf{u}_1, \ldots, \mathsf{u}_k], \quad \mathsf{V}_k = [\mathsf{v}_1, \ldots, \mathsf{v}_k], \quad \mathcal{S}_k = \operatorname{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}. \tag{1.4}$$

2 Review of some projection methods

In this paper, we are interested in projection methods that construct approximations f_k to the solution of (1.1) by solving the constrained problem

$$f_k = \underset{f\in\mathcal{V}_k}{\operatorname{argmin}} \|g - Af\|_2^2, \quad k \geq 1, \tag{2.1}$$

where $\{\mathcal{V}_k\}_{k\geq 1}$ is a family of nested k-dimensional subspaces. If we are given a matrix $V_k = [v_1 \cdots v_k]$ with orthonormal columns such that $\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{V}_k$, then (2.1) can be reformulated by taking $f = V_k d$ for some $d \in \mathbb{R}^k$, and the approximate solution is

$$f_k = V_k d_k, \quad d_k = \underset{d\in\mathbb{R}^k}{\operatorname{argmin}} \|g - (AV_k)d\|_2^2. \tag{2.2}$$

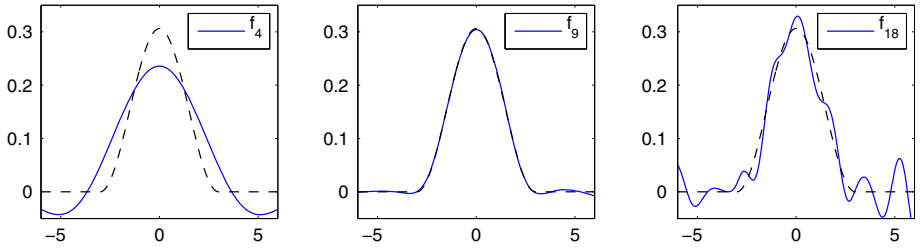

Fig. 1 Exact solution (*dashed*) and TSVD solutions (*solid*) for $k = 4, 9, 18$ for phillips test problem with $n = 512$ and NL $= \|e\|_2/\|g_{\text{exact}}\|_2 = 0.01$

This least squares problem is often referred to as the *projected problem* because it is obtained by projecting the original problem onto the k-dimensional subspace $\mathcal{V}_k$. The purpose of this section is to briefly describe some of the classical projection methods currently used for solving discrete ill-posed problems.

2.1 Truncated SVD

When dealing with small to moderate size problems, the SVD is the preferred "tool" for analysis and solution of discrete ill-posed problems, see, e.g., Hansen (1998). If the summation (1.3) is interrupted at $k \leq n$, we obtain the method known as Truncated SVD (TSVD), see, e.g., Hansen (1998). Hence, the k-th TSVD solution is defined by

$$\mathsf{f}_k = \sum_{j=1}^{k} \frac{\mathsf{u}_j^T g}{\sigma_j} \mathsf{v}_j, \quad k \leq n. \tag{2.3}$$

Alternatively, f_k is the solution to the following constrained least squares problem

$$\mathsf{f}_k = \underset{f \in \mathcal{S}_k}{\operatorname{argmin}} \|g - Af\|_2^2 = \mathsf{A}_k^{\dagger} g, \quad \mathsf{A}_k = \sum_{j=1}^{k} \sigma_j \mathsf{u}_j \mathsf{v}_j^T. \tag{2.4}$$

The point here is that if k is poorly chosen, the solution f_k either captures not enough information about the problem or the noise in the data dominates the approximate solution. This is illustrated in Fig. 1 where are depicted some TSVD solutions for phillips test problem from Hansen (1994), with $n = 512$ and NL $= \|e\|_2/\|g_{\text{exact}}\|_2 = 0.01$. Throughout the paper, NL stands for Noise Level. For this problem, the first 4 SVD components are not able to capture enough information about the problem, whereas for $k > 9$, the noise begins to dominate. However, it is apparent that for $k = 9$ the subspace $\mathcal{S}_k$ is appropriate. Therefore, the challenge in connection with TSVD for the case when no estimate of the error norm $\|e\|_2$ is available, is how to select a proper truncation parameter.

2.2 CGLS/LSQR method

If the SVD is not available, as it is more likely to occur when treating large-scale problems, iterative methods that only use matrix–vector products are preferable. One of the classical iterative methods to treat symmetric positive definite linear systems is the CG method. The normal equations associated to (1.1) are indeed a symmetric linear system which can be solved by CG, as proposed by Hestenes and Stiefel (1952). This gave rise to what is known

as CGLS method. An interesting property is that the CGLS iterates satisfy the constrained least squares problem Björck (1996)

$$f_k = \underset{f \in \mathcal{K}_k(A^T A, A^T g)}{\operatorname{argmin}} \|g - Af\|_2^2, \tag{2.5}$$

where $\mathcal{K}_k(A^T A, A^T g)$ is the Krylov subspace associated to the pair $(A^T A, A^T g)$. It is well known that in some problems, the first CGLS iterates look like regularized solutions but after a certain number of steps, these solutions approximate the naive least squares solution which is completely dominated by noise. It is out of the scope of this paper to deal with such properties. For more information about the regularizing effects of the CGLS iterations, the reader is referred to Hansen (1998) and references therein.

The LSQR algorithm, due to Paige and Saunders (1982), also constructs a sequence of iterated solutions that satisfies (2.5), but it is obtained in a completely distinct way. LSQR is based on the Golub–Kahan bidiagonalization (GKB) procedure, which, after k steps constructs matrices $U_{k+1} = [u_1, \ldots, u_{k+1}] \in \mathbb{R}^{m \times (k+1)}$ and $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}$ with orthonormal columns, and a lower bidiagonal matrix $B_k \in \mathbb{R}^{(k+1) \times k}$,

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix}, \tag{2.6}$$

such that

$$\beta_1 U_{k+1} e_1 = g = \beta_1 u_1, \tag{2.7}$$

$$A V_k = U_{k+1} B_k, \tag{2.8}$$

$$A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T, \tag{2.9}$$

where e_k denotes the kth unit vector of appropriate dimension. Using Eqs. (2.7)–(2.9), it is possible to show that f_k satisfies

$$f_k = V_k d_k, \quad d_k = \underset{d \in \mathbb{R}^k}{\operatorname{argmin}} \|\beta_1 e_1 - B_k d\|_2^2. \tag{2.10}$$

Paige and Saunders also explain how to efficiently update f_k from f_{k-1} and, hence, avoiding the need of saving all vectors v_j, see Paige and Saunders (1982) for details.

2.3 Minimal residual methods

For square A, two classical minimal residual methods are MINRES (1975) by Paige and Saunders, and GMRES (1986) by Saad and Schultz. The former requires that $A = A^T$ and seeks for a solution f_k that satisfies

$$f_k = \underset{f \in \mathcal{K}_k(A, g)}{\operatorname{argmin}} \|g - Af\|_2^2. \tag{2.11}$$

The latter works with any A and seeks for a solution f_k such that

$$f_k = \underset{f \in \mathcal{K}_k(A, g - Af_0)}{\operatorname{argmin}} \|g - Af\|_2^2, \tag{2.12}$$

where f_0 is an approximation of the desired solution. Note that, if $f_0 = 0$ both methods share the same search space. Algorithmically, MINRES is based on the Lanczos Tridiagonalization (LT) process and GMRES is based on the Arnoldi process. If both are initialized with vector g, after k steps, for the LT process we obtain matrices V_k with orthonormal columns and $T_k \in \mathbb{R}^{k\times k}$ being tridiagonal, while for the Arnoldi process, we obtain matrices $Q_k \in \mathbb{R}^{n\times k}$ with orthonormal columns, and $H_k \in \mathbb{R}^{k\times k}$ being upper Hessenberg matrix:

$$T_k = \begin{pmatrix} \alpha_1 & \beta_2 & & & & \\ \beta_2 & \alpha_2 & \beta_3 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \beta_{k-1} & \alpha_{k-1} & \beta_k \\ & & & \beta_k & \alpha_k \end{pmatrix}, \quad H_k = \begin{pmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & \cdots & h_{2k} \\ & h_{32} & \cdots & \cdots & h_{3k} \\ & & \ddots & \cdots & \vdots \\ & & & h_{k,k-1} & h_{kk} \end{pmatrix}, \tag{2.13}$$

and the following relations hold

$$\begin{array}{c|c} \text{LT Process} & \text{Arnoldi Process} \\ \beta_1 V_k e_1 = g = \beta_1 v_1 & \|g\|_2 Q_k e_1 = g = \|g\|_2 q_1 \\ AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T & AQ_k = Q_k H_k + h_{k+1,k} q_{k+1} e_k^T \end{array} \tag{2.14}$$

Using the above relations, the respective iterated solutions can also be obtained by

$$f_k = V_k d_k, \quad d_k = \underset{d\in\mathbb{R}^k}{\operatorname{argmin}} \|\beta_1 V_k e_1 - (V_k T_k + \beta_{k+1} v_{k+1} e_k^T) d\|_2^2, \tag{2.15}$$

for MINRES, and

$$f_k = Q_k d_k, \quad d_k = \underset{d\in\mathbb{R}^k}{\operatorname{argmin}} \left\| \|g\|_2 e_1 - \widetilde{H}_k d \right\|_2^2, \quad \widetilde{H}_k = \begin{pmatrix} H_k \\ \mathbf{0}\ h_{k+1,k} \end{pmatrix}, \tag{2.16}$$

for GMRES where $\mathbf{0} = [0 \ \cdots \ 0] \in \mathbb{R}^{1\times k-1}$. If we look for a solution in $\mathcal{K}_k(A, Ag)$ and $\mathcal{K}_k(A, A(g - Af_0))$, instead of $\mathcal{K}_k(A, g)$ and $\mathcal{K}_k(A, g - Af_0)$, respectively, we obtain methods known as MR-II (Hanke 1995) and RRGMRES (Calvetti et al. 2000). As shown in Jensen and Hansen (2007), MINRES and MR-II have regularization properties by "killing" the large SVD components to reduce the influence of the noise. On the other hand, GMRES and RRGMRES mix the SVD components in each iteration, so it is possible that for some problems neither GMRES nor RRGMRES is able to produce regularized solutions (Jensen and Hansen 2007).

3 Iterated solution f_k vs. TSVD solution f_k

Probably due to the fact that the noise-free solution of (1.1) is expressed in terms of the right singular vectors, the subspace $\mathcal{S}_k = \operatorname{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$ is sometimes referred to as the "best" subspace, see, e.g., Hansen (2010). Thus, a natural question that arises is how close the iterated solution f_k defined in (2.1) is to the TSVD solution. The goal of this section is to determine bounds on $\|\mathsf{f}_k - f_k\|_2$. Intuitively, for the kth iterated solution f_k to be a good approximation to the kth TSVD solution, the subspace $\mathcal{V}_k$ should be a good approximation to $\mathcal{S}_k$, and a way to compare these subspaces is by assessing the angle between them. For this, let P_k be the orthogonal projector onto $\mathcal{V}_k$ and let $\gamma_k = \|A - AP_k\|_2$. This number

is important since it measures how well the operator AP_k approximates the operator A. In addition, we shall show that it is fundamental to estimate the quality of f_k compared to f_k. The following Lemma gives a simple expression for γ_k.

Lemma 3.1 *Let $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ be a matrix with orthonormal columns such that span$\{v_1, \ldots, v_k\} = \mathcal{V}_k$. Thus, there is a matrix $M = [M_1 \;\; M_2] \in \mathbb{R}^{m\times n}$ where $M_1 \in \mathbb{R}^{m\times k}$ and $M_2 \in \mathbb{R}^{m\times(n-k)}$ such that*

$$\gamma_k = \|M_2\|_2. \tag{3.1}$$

Proof Since V_k has k columns, there is a matrix $V_\perp$ such that $V = [V_k \;\; V_\perp]$ is orthogonal. Define $\overline{U} = AV$. We can decompose $\overline{U}$ in $\overline{U} = UM$ where $U \in \mathbb{R}^{m\times m}$ is orthogonal and $M = [M_1 \;\; M_2] \in \mathbb{R}^{m\times n}$. Thereby, $AV_k = UM_1$ and

$$\gamma_k = \|A - AP_k\|_2 = \|A - AV_kV_k^T\|_2 = \|M - M_1V_k^T[V_k \;\; V_\perp]\|_2 = \|M_2\|_2. \tag{3.2}$$

□

The theorem below gives an upper bound for the angle between the subspaces $\mathcal{V}_k$ and $\mathcal{S}_k$.

Theorem 3.2 *Let $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ be a matrix with orthonormal columns such that span$\{v_1, \ldots, v_k\} = \mathcal{V}_k$ and let Ω_k denotes the angle between the subspaces $\mathcal{V}_k$ and $\mathcal{S}_k$. Thus,*

$$\sin(\Omega_k) \le \frac{\gamma_k}{\sigma_k}, \tag{3.3}$$

where γ_k is from Lemma 3.1.

Proof From proof of Lemma 3.1, we have $AV_\perp = UM_2$. Multiplying this equation by U_k^T and using the fact that $\mathsf{U}_k^T A = \Sigma_k\mathsf{V}_k^T$ where Σ_k contains the first k singular values of A, we have $\mathsf{V}_k^T V_\perp = \Sigma_k^{-1}\mathsf{U}_k UM_2$. Taking norms leads to $\sin(\Omega_k) \le \gamma_k/\sigma_k$, where we used the fact that $\sin(\Omega_k) = \|\mathsf{V}_k^T V_\perp\|_2$, see Golub and Van Loan (1996). □

If, in particular, the subspace $\mathcal{V}_k$ is generated by the Golub–Kahan process, we have the following upper bound for $\sin(\Omega_k)$.

Theorem 3.3 *Let Ω_k be the angle between the subspace $\mathcal{V}_k$ generated by the GKB process and the right singular subspace $\mathcal{S}_k$. Assume that the smallest singular value of B_k given in (2.6), denoted by $\sigma_{\min}(B_k)$, satisfies, $\sigma_{\min}(B_k) > \sigma_{k+1}$. Then,*

$$\sin(\Omega_k) \le \frac{\sigma_{\min}(B_k)\alpha_{k+1}}{\sigma_{\min}(B_k)^2 - \sigma_{k+1}^2}. \tag{3.4}$$

Proof The SVD of A can be rewritten as

$$A = [\mathsf{U}_k \;\; \mathsf{U}_0 \;\; \mathsf{U}_\perp]\begin{bmatrix}\Sigma_k & 0\\ 0 & \Sigma_0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathsf{V}_k^T\\ \mathsf{V}_0^T\end{bmatrix} = \mathsf{U}_k\Sigma_k\mathsf{V}_k^T + \mathsf{U}_0\Sigma_0\mathsf{V}_0^T. \tag{3.5}$$

Since A is full rank, after n GKB steps, matrix A can be written as

$$A = [U_{k+1} \;\; U_0 \;\; U_\perp]\begin{bmatrix}B_k & C_k\\ 0 & F_k\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_k^T\\ V_0^T\end{bmatrix}, \tag{3.6}$$

where $C_k = \alpha_{k+1} e_{k+1} e_1^T \in \mathbb{R}^{(k+1)\times(n-k)}$,

$$F_k = \begin{bmatrix} \beta_{k+2} & \alpha_{k+2} & & \\ & \ddots & \ddots & \\ & & \beta_n & \alpha_n \\ & & & \beta_{n+1} \end{bmatrix} \in \mathbb{R}^{(n-k)\times(n-k)}. \tag{3.7}$$

Let $B_k = P_k \begin{pmatrix} D_k \\ 0 \end{pmatrix} Q_k^T$ be the SVD of B_k. Then, (3.6) can be rewritten as

$$A = [\widetilde{U}_k \;\; \widetilde{U}_0 \;\; \widetilde{U}_\perp] \begin{bmatrix} D_k & \widetilde{C}_k \\ 0 & \widetilde{F}_k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{V}_k^T \\ \widetilde{V}_0^T \end{bmatrix} = \widetilde{U}_k D_k \widetilde{V}_k^T + \widetilde{U}_k \widetilde{C}_k \widetilde{V}_0^T + \widetilde{U}_0 \widetilde{F}_k \widetilde{V}_0^T, \tag{3.8}$$

where $\widetilde{U}_k \in \mathbb{R}^{m\times k}$ contains the first k columns of matrix $U_{k+1}P_k$, $\widetilde{U}_0 = [\widetilde{p} \;\; U_0] \in \mathbb{R}^{m\times(n-k+1)}$ with $\widetilde{p}$ being the last column of $U_{k+1}P_k$, $\widetilde{U}_\perp = U_\perp P_k \in \mathbb{R}^{m\times(m-n-1)}$, $\widetilde{C}_k \in \mathbb{R}^{k\times(n-k)}$ contains the first k rows of matrix $P_k^T C_k$, $\widetilde{F}_k = [\widetilde{f}^T \;\; F_k^T]^T \in \mathbb{R}^{(n-k+1)\times(n-k)}$ with $\widetilde{f}$ being the last row of $P_k^T C_k$, $\widetilde{V}_k = V_k Q_k$ and $\widetilde{V}_0 = V_0$. Thus,

$$\begin{aligned} \widetilde{V}_k^T \mathsf{V}_0 &= D_k^{-1} \left(\widetilde{U}_k^T A - \widetilde{C}_k \widetilde{V}_0^T \right) \mathsf{V}_0 = D_k^{-1} \left(\widetilde{U}_k^T A \mathsf{V}_0 - \widetilde{C}_k \widetilde{V}_0^T \mathsf{V}_0 \right) \\ &= D_k^{-1} \left(\widetilde{U}_k^T \mathsf{U}_0 \Sigma_0 - \widetilde{C}_k \widetilde{V}_0^T \mathsf{V}_0 \right). \end{aligned} \tag{3.9}$$

We also have

$$A\widetilde{V}_k = \widetilde{U}_k D_k \Rightarrow \widetilde{V}_k^T A^T = D_k^T \widetilde{U}_k^T \Rightarrow \widetilde{U}_k^T = D_k^{-1} \widetilde{V}_k^T A^T = D_k^{-1} \widetilde{V}_k^T A^T, \tag{3.10}$$

and $\widetilde{U}_k^T \mathsf{U}_0 = D_k^{-1} \widetilde{V}_k^T A^T \mathsf{U}_0 = D_k^{-1} \widetilde{V}_k^T \mathsf{V}_0 \Sigma_0$. Thus,

$$\begin{aligned} \widetilde{V}_k^T \mathsf{V}_0 &= D_k^{-1} \left(D_k^{-1} \widetilde{V}_k^T \mathsf{V}_0 \Sigma_0^2 - \widetilde{C}_k \widetilde{V}_0^T \mathsf{V}_0 \right) \\ &= D_k^{-2} \widetilde{V}_k^T \mathsf{V}_0 \Sigma_0^2 - D_k^{-1} \widetilde{C}_k \widetilde{V}_0^T \mathsf{V}_0 \end{aligned} \tag{3.11}$$

Taking norms, we obtain

$$\sin(\Omega_k) \leq \frac{\sigma_{k+1}^2}{\sigma_{\min}(B_k)^2} \sin(\Omega_k) + \frac{\|\widetilde{C}_k\|_2}{\sigma_{\min}(B_k)}. \tag{3.12}$$

Using the assumption $\sigma_{\min}(B_k) > \sigma_{k+1}$ and the fact that $\|\widetilde{C}_k\|_2 \leq \alpha_{k+1}$, we have

$$\sin(\Omega_k) \leq \frac{\sigma_{\min}(B_k)\alpha_{k+1}}{\sigma_{\min}(B_k)^2 - \sigma_{k+1}^2}. \tag{3.13}$$

□

This theorem is very similar in spirit to one due to Fierro and Bunch (1995, Thm. 2.2) who considered subspaces generated by rank-revealing decompositions. The results of these authors are not applicable here since B_k is not square. Bounds (3.3) and (3.4) denoted, respectively, by "Bound A" and "Bound B", as well as $\sin(\Omega_k)$ for the first 20 iterations of GKB process applied to i_laplace and shaw test problems with $n = 512$ and NL = 0.01, are depicted in Fig 2.

The main result of this section gives upper bounds on the distance between f_k and f_k.

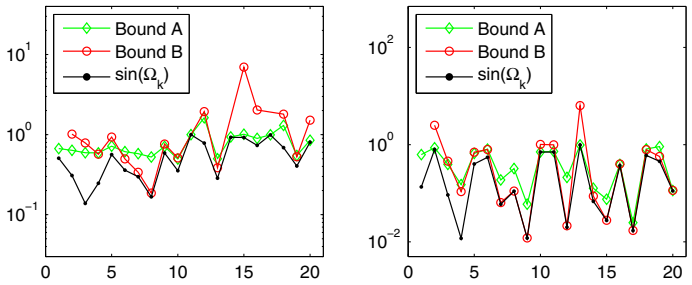

Fig. 2 Bounds (3.3) and (3.4) for i_laplace (*left*) and shaw (*right*) test problems with $n = 512$ and NL = 0.01

Theorem 3.4 *Let the columns of $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ form a basis of subspace $\mathcal{V}_k$. The distance between f_k given by (2.1) and the TSVD solution f_k can be bounded as*

$$\frac{\|\mathsf{f}_k - f_k\|_2}{\|f_k\|_2} \leq \frac{1}{\sigma_k}(\phi_k + \gamma_k), \tag{3.14}$$

where $\phi_k = \|g - Af_k\|_2/\|f_k\|_2$ and γ_k is as (3.1). In addition, under assumptions given in Theorem (3.3), we have

$$\frac{\|\mathsf{f}_k - f_k\|_2}{\|f_k\|_2} \leq \frac{\phi_k}{\sigma_k} + \frac{\sigma_{\min}(B_k)\alpha_{k+1}}{\sigma_{\min}(B_k)^2 - \sigma_{k+1}^2}. \tag{3.15}$$

Proof Let $r_k = g - Af_k$ and $f_k = V_k d_k$ where $d_k = (AV_k)^\dagger g$. Thus,

$$\mathsf{f}_k - f_k = \mathsf{A}_k^\dagger g - V_k(AV_k)^\dagger g = \mathsf{A}_k^\dagger r_k + \mathsf{A}_k^\dagger A f_k - V_k(AV_k)^\dagger A f_k, \tag{3.16}$$

where we used the fact that $V_k(AV_k)^\dagger r_k = 0$. Notice that

$$V_k(AV_k)^\dagger A f_k = V_k(AV_k)^\dagger AV_k V_k^T f_k = V_k V_k^T f_k = f_k, \tag{3.17}$$

and since $\mathsf{A}_k^\dagger A = \mathsf{A}_k^\dagger \mathsf{A}_k$, it follows that

$$\mathsf{f}_k - f_k = \mathsf{A}_k^\dagger r_k + (\mathsf{A}_k^\dagger \mathsf{A}_k - I_n) f_k = \mathsf{A}_k^\dagger r_k - (I_n - \mathsf{P}_k) P_k f_k, \tag{3.18}$$

where $P_k = V_k V_k^T$ and $\mathsf{P}_k = \mathsf{A}_k^\dagger \mathsf{A}_k$ are, respectively, the orthogonal projectors onto $\mathcal{V}_k$ and $\mathcal{S}_k$. Therefore,

$$\begin{aligned} \|\mathsf{f}_k - f_k\|_2 &\leq \|\mathsf{A}_k^\dagger\|_2 \|r_k\|_2 + \|(I_n - \mathsf{P}_k)P_k\|_2 \|f_k\|_2, \\ &\leq \|f_k\|_2 \phi_k/\sigma_k + \|f_k\|_2 \gamma_k/\sigma_k, \end{aligned} \tag{3.19}$$

where we used the fact that $\|(I_n - \mathsf{P}_k)P_k\|_2 = \sin(\Omega_k)$ and Theorem 3.2. Inequality (3.15) follows from equation (3.19) and Theorem 3.3. □

As the error bound becomes useless when ϕ_k/σ_k>1, Theorem 3.4 suggests that the quality of the iterates f_k with respect to the TSVD solutions will probably deteriorate after this inequality is met. We shall see that the inequality ϕ_k/σ_k>1 marks a kind of transition between the size of the solution norm and the size of the residual norm, which is crucial for the proper functioning of the parameter choice rule we are going to propose. To illustrate how the behavior of ϕ_k/σ_k relates to the error $E_k = \|f_k - \mathsf{f}_k\|_2/\|f_k\|_2$, these quantities are displayed in Fig. 3 for LSQR and i_laplace test problem with $n = 512$ and NL=0.01. As we can see, the range of values of k for which the bounds (3.14) and (3.15) start to deteriorate occurs

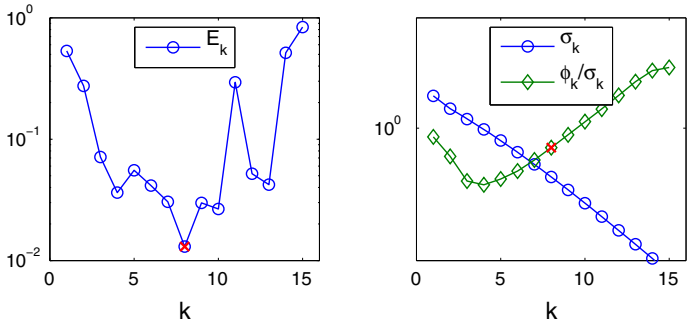

Fig. 3 Relative error E_k (*left*), and $\sigma_k, \phi_k/\sigma_k$ for the first 15 iterated solutions obtained by LSQR for i_laplace test problem with $n = 512$ and NL = 0.01. Integer k that minimizes E_k is marked in *red* with *times*

when $k > 8$ which is precisely when the inequality $\phi_k/\sigma_k > 1$ holds. This suggests that the iteration that minimizes the error should occur near the first k such that $\phi_k/\sigma_k > 1$.

If $\mathcal{V}_k$ is the Krylov subspace generated by the GKB process, we have the following result.

Corollary 3.5 *Let $\mathcal{V}_k$ be the Krylov subspace $\mathcal{K}_k(A^TA, A^Tg)$, and $A_k = U_{k+1}B_kV_k^T$. Then,*

$$\frac{\|\mathsf{f}_k - f_k\|_2}{\|f_k\|_2} \leq \sin(\Omega_k)\left(\frac{\sigma_1}{\sigma_k}\frac{\|r_k\|_2}{\|g\|_2}\frac{\sigma_{k+1}}{\tau_k\sigma_{min}(R_k)} + 1\right), \quad \tau_k = \sqrt{1 - \frac{\|r_k\|_2^2}{\|g\|_2^2}}. \tag{3.20}$$

Proof From Eq. (3.8), we have $A_k = \widetilde{U}_k D_k \widetilde{V}_k^T$, thus

$$\begin{aligned}
\mathsf{f}_k - f_k &= (\mathsf{A}_k^\dagger - A_k^\dagger)g \\
&= (\mathsf{A}_k^\dagger - A_k^\dagger)g - (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k + (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k \\
&= (\mathsf{A}_k^\dagger - A_k^\dagger)g - (\mathsf{A}_k^\dagger A - A_k^\dagger A + \widetilde{V}_k D_k^{-1}\widetilde{C}_k\widetilde{V}_0^T)f_k + (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k \\
&= (\mathsf{A}_k^\dagger - A_k^\dagger)r_k + (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k \\
&= \mathsf{A}_k^\dagger r_k + (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k
\end{aligned} \tag{3.21}$$

where we used that $\mathsf{A}_k^\dagger\mathsf{A}_k = \mathsf{A}_k^\dagger A$, $A = A_k + \widetilde{U}_k\widetilde{C}_k\widetilde{V}_0^T + \widetilde{U}_0\widetilde{F}_k\widetilde{V}_0^T$, $r_k = g - Af_k$, $\widetilde{V}_0^T f_k = 0$ and $A_k^\dagger r_k = 0$. Since $\mathsf{A}_k^\dagger = \mathsf{A}_k^\dagger\mathsf{U}_k\mathsf{U}_k^T$ and $r_k = \widetilde{U}_k^\perp\widetilde{U}_k^{\perp T}r_k$, it follows

$$\mathsf{f}_k - f_k = \mathsf{A}_k^\dagger\mathsf{U}_k\mathsf{U}_k^T\widetilde{U}_k^\perp\widetilde{U}_k^{\perp T}r_k + (\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)f_k. \tag{3.22}$$

Thus,

$$\|\mathsf{f}_k - f_k\|_2 \leq \|\mathsf{A}_k^\dagger\|_2\|\mathsf{U}_k^T\widetilde{U}_k^\perp\|_2\|r_k\|_2 + \|(\mathsf{A}_k^\dagger\mathsf{A}_k - A_k^\dagger A_k)\|_2\|f_k\|_2. \tag{3.23}$$

Using the inequality $\frac{1}{\|f_k\|_2}\sqrt{1 - \frac{\|r_k\|_2^2}{\|g\|_2^2}} \leq \frac{\|A\|_2}{\|g\|_2}$ and denoting by Π_k the angle between the subspaces $\mathcal{R}(\mathsf{U}_k)$ and $\mathcal{R}(\widetilde{U}_k)$, we have

$$\frac{\|\mathsf{f}_k - f_k\|_2}{\|f_k\|_2} \leq \frac{\sigma_1}{\sigma_k}\frac{\|r_k\|_2}{\|g\|_2}\frac{\sin(\Pi_k)}{\tau_k} + \sin(\Omega_k). \tag{3.24}$$

Combining Eqs. (3.5) and (3.8), it can be proved that $\sin(\Pi_k) \leq \frac{\sigma_{k+1}}{\sigma_{\min}(B_k)} \sin(\Omega_k)$ (Fierro and Bunch 1995), thus

$$\frac{\|\mathsf{f}_k - f_k\|_2}{\|f_k\|_2} \leq \sin(\Omega_k) \left(\frac{\sigma_1}{\sigma_k} \frac{\|r_k\|_2}{\|g\|_2} \frac{\sigma_{k+1}}{\tau_k \sigma_{\min}(R_k)} + 1 \right), \quad \tau_k = \sqrt{1 - \frac{\|r_k\|_2^2}{\|g\|_2^2}}. \tag{3.25}$$

□

This corollary confirms that the iterated solutions will be close to the TSVD solutions provided the sine of the angle between the subspaces $\mathcal{V}_k$ and $\mathcal{S}_k$ is small.

4 Stopping rule

As already mentioned, projection methods must be suitably stopped to produce good approximations to the noise-free solution of (1.1). If a good estimate of $\|e\|_2$ is available, say $\|e\|_2 \leq \delta$, the discrepancy principle (DP) is probably the most appropriate criterion to be used. However, in this paper, we consider that estimates of $\|e\|_2$ are not available, hence DP will not be used. One stopping rule that is derived within a statistical framework and that does not require any estimate of the error norm is the generalized cross-validation (GCV). The main idea is that the regularization parameter should be able to predict data in vector g that is missing. If the iterated solutions are carried out by TSVD method, the stopping index is chosen to be the minimizer of $G(k) = \|r_k\|_2^2/(m-k)^2$, see (Björck 1996; Hansen 1998; Bauer and Lukas 2011; Lukas 2006) for robust versions of this method for small to moderate size problems. Since the SVD is computationally expensive for large A, GCV will not be considered in this paper. Hence, the development of simple and efficient stopping rules for Krylov projection methods is required.

4.1 Automatic stopping rule

To describe our automatic stopping rule, we start as in (Bazán et al. 2013, Sect. 4.1) and discuss a truncation criterion for the TSVD method. We shall assume that the noise-free data g_{exact} satisfy the discrete Picard condition (DPC) (Hansen 1990), i.e., the coefficients $|\mathsf{u}_i^T g_{\text{exact}}|$, on average, decay to zero faster than the singular values, as well as the noise contains zero-mean Gaussian random numbers. These assumptions imply that there exists a integer k^* such that

$$|\mathsf{u}_i^T g| = |\mathsf{u}_i^T g_{\text{exact}} + \mathsf{u}_i^T e| \approx |\mathsf{u}_i^T e| \approx \text{constant}, \ \text{for } i > k^*. \tag{4.1}$$

The error in the kth TSVD solution can be written as

$$\mathcal{E}_k = \|\mathsf{f}_k - f_{\text{exact}}\|_2^2 = \sum_{i=1}^{k} \frac{|\mathsf{u}_i^T e|^2}{\sigma_i^2} + \sum_{i=k+1}^{n} \frac{|\mathsf{u}_i^T g_{\text{exact}}|^2}{\sigma_i^2} \equiv E_1(k) + E_2(k). \tag{4.2}$$

The first term measures the error caused by the noise; it increases with k and can be large for $\sigma_i \approx 0$. The second term, called the regularization error, decreases with k and can be small when k is large. Thus, to choose a good truncation parameter, it is required that these errors balance each other to make the global error small. A closer look to these errors reveals that for $k > k^*$, the error $E_1(k)$ increases dramatically while $E_2(k)$ remains under control, hence the error should not be minimized for $k > k^*$. Also, for $k < k^*$, $E_2(k)$ increases with k while $E_1(k)$ stays under control since σ_i dominates $|\mathsf{u}_i^T e|$, and hence, the error should not be minimized. This suggests that the error should be minimized at $k = k^*$.

Another way to perform the above analysis, which we introduce here by the first time and we use in the general framework of Krylov projection methods, is by analyzing the finite forward differences

$$\nabla\mathcal{E}_k = \mathcal{E}_{k+1} - \mathcal{E}_k = \frac{|\mathsf{u}_{k+1}^T e|^2}{\sigma_{k+1}^2} - \frac{|\mathsf{u}_{k+1}^T g_{\text{exact}}|^2}{\sigma_{k+1}^2}. \tag{4.3}$$

Since the DPC is satisfied, for $k \geq k^*$ we have $|\mathsf{u}_{k+1}^T e| > |\mathsf{u}_{k+1}^T g_{\text{exact}}|$. Thus, $\mathcal{E}_k$ increases with k and it is not minimized for these values of k. On the other hand, for $k < k^*$, we have $|\mathsf{u}_{k+1}^T e| < |\mathsf{u}_{k+1}^T g_{\text{exact}}|$, and $\mathcal{E}_k$ is decreasing. Therefore, the index $k = k^*$ points out a change in $\mathcal{E}_k$ such that $\nabla\mathcal{E}_{k^*-1} < 0$ and $\nabla\mathcal{E}_{k^*} > 0$, thus indicating that $\mathcal{E}_k$ should be minimized at $k = k^*$. Nevertheless, in practical applications, neither the error vector e nor the exact data vector g_{exact} is available and the best we can do is to minimize a model for $\mathcal{E}_k$. Since the sequences $\|\mathsf{f}_k\|_2$ and $\|g - A\mathsf{f}_k\|_2$ are, respectively, increasing and decreasing, a truncation index for the TSVD method can be chosen as the minimizer of

$$\Psi_k(\text{SVD}) = \|g - A\mathsf{f}_k\|_2\|\mathsf{f}_k\|_2, \quad k \geq 1, \tag{4.4}$$

as minimizing $\log\Psi_k(\text{SVD})$ is equivalent to minimizing a sum of competing terms, one increasing and other decreasing. A similar analysis lead us to conclude that the minimizer of $\Psi_k(\text{SVD})$ can serve as a good estimate for $k = k^*$. Unfortunately, the SVD is infeasible for large-scale problems and the TSVD method may not be of practical utility. To overcome this possible difficulty, our proposal is to use iterative methods such as LSQR, GMRES, etc., and stop the iterations at the first integer k such that

$$\widetilde{k} = \text{argmin}\Psi_k, \quad \Psi_k = \|g - Af_k\|_2\|f_k\|_2, \tag{4.5}$$

where f_k denotes the kth iterate computed by the chosen method. Figure 4 depicts three sequences Ψ_k calculated by three different methods. Notice that criterion (4.5) is nothing more than a discrete counterpart of Regińska's rule (1996) which looks for a corner of the continuous Tikhonov L-curve; for details the reader is referred to (Bazán 2008; Bazán and Francisco 2009). We mention that in (Bazán et al. 2013) the rule (4.5) was analyzed and illustrated in connection with LSQR method. Here, we extend its use to a more general framework along with a detailed analysis to understand its behavior. Notice also that from the practical point of view, the stopping criterion (4.5) can be implemented by monitoring the finite forward differences

$$\nabla\Psi_k = \Psi_{k+1} - \Psi_k, \quad k \geq 1, \tag{4.6}$$

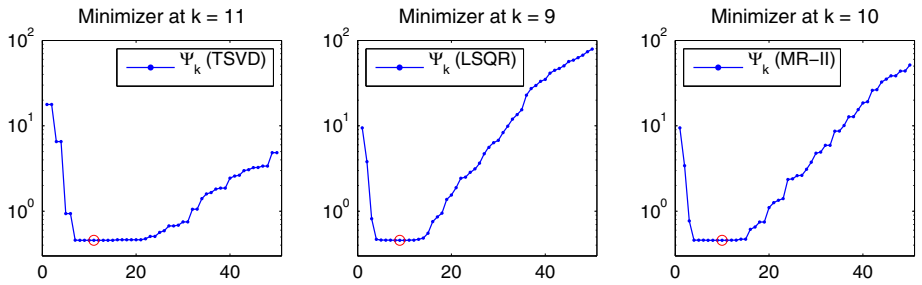

Fig. 4 Sequences Ψ_k obtained by TSVD, LSQR and MR-II for phillips test problem with $n = 512$ and NL $= 0.01$

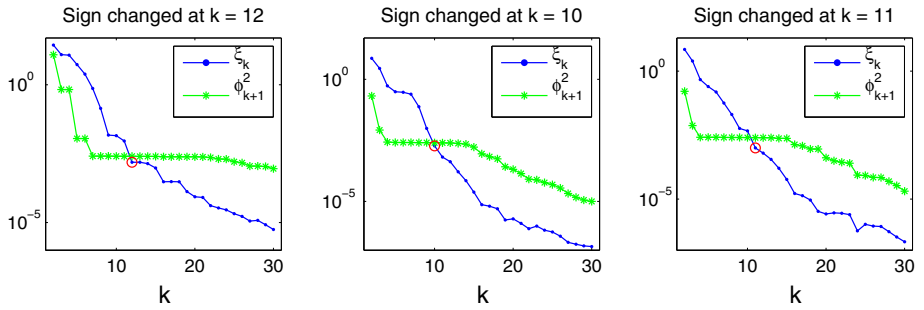

Fig. 5 Sequences ϕ_{k+1}^2 and ξ_k calculated by TSVD (*left*), LSQR (*middle*) and MR-II (*right*) using the same data as in previous figure

and thus selecting the first index k such that $\nabla\Psi_{k-1} < 0$ and $\nabla\Psi_k > 0$. However, to know more about the rule, motivated by the above analysis for TSVD based on forward differences, we will consider the sequence Ψ_k^2 (which shares minimum with Ψ_k) and look for extreme values by identifying sign changes of $\nabla\Psi_k^2$. In fact, letting $x_k = \|g - Af_k\|_2^2$, $y_k = \|f_k\|_2^2$, $\nabla x_k = x_{k+1} - x_k$, and $\nabla y_k = y_{k+1} - y_k$, after some calculation, for $\nabla y_k \neq 0$, we have

$$\nabla\Psi_k^2 = \Psi_{k+1}^2 - \Psi_k^2 = y_{k+1}\nabla y_k \left(\phi_{k+1}^2 - \xi_k\right), \tag{4.7}$$

where

$$\phi_k^2 = x_k/y_k, \quad \xi_k = -\frac{y_k}{y_{k+1}}\frac{\nabla x_k}{\nabla y_k}, \tag{4.8}$$

while for $\nabla y_k = 0$, we have

$$\nabla\Psi_k^2 = y_k(x_{k+1} - x_k). \tag{4.9}$$

To analyze the sign of $\nabla\Psi_k^2$, three cases are considered.

- If $\nabla y_k = 0$, which is unlikely to happen in practical applications, from (4.9) we conclude that $\nabla\Psi_k^2 \leq 0$ as the residual norm sequence x_k is not increasing, and Ψ_k cannot be minimized at this k.
- If $\nabla y_k < 0$, it follows that $\xi_k \leq 0$ and therefore $\phi_{k+1}^2 - \xi_k \geq 0$. Hence, $\nabla\Psi_k^2$ and ∇y_k share the same sign and Ψ_k cannot be minimized at this k .
- If $\nabla y_k > 0$ we have that $\xi_k \geq 0$, and the sign of $\nabla\Psi_k^2$ depends on the sign of $\phi_{k+1}^2 - \xi_k$. Therefore, $\nabla\Psi_k^2 \leq 0$ if $\xi_k/\phi_{k+1}^2 \geq 1$ and $\nabla\Psi_k \geq 0$ if $\xi_k/\phi_{k+1}^2 \leq 1$.

Thus, minimization of Ψ_k depends on sign changes of $\phi_{k+1}^2 - \xi_k$ and occurs at the first k such that $\nabla\Psi_{k-1}^2 \leq 0$ and $\nabla\Psi_k^2 \geq 0$. Figure 5 displays sequences ϕ_{k+1}^2 and ξ_k, where x_k and y_k are calculated by TSVD, LSQR and MR-II using the same data of the previous figure. In this case, sign changes occur at $k = 12$ for TSVD, at $k = 10$ for LSQR and at $k = 11$ for MR-II.

As a particular case, if the sequence f_k is calculated by TSVD it is easy to show that

$$\frac{\nabla x_k}{\nabla y_k} = -\sigma_{k+1}^2, \tag{4.10}$$

hence

$$\operatorname{sign}\left(\nabla\Psi_k^2\right) = \begin{cases} 1, \text{ if } \frac{\|g - Af_{k+1}\|_2}{\|f_k\|_2} > \sigma_{k+1}, \\ -1, \text{ if } \frac{\|g - Af_{k+1}\|_2}{\|f_k\|_2} < \sigma_{k+1}. \end{cases} \tag{4.11}$$

We end this subsection with a few observations.

- If the sequence $\|f_k\|_2$ is nondecreasing, as happening with TSVD and LSQR, we always have $\nabla y_k \geq 0$ and the sequence ϕ_k is non-increasing.
- The sequence ϕ_k may be non-increasing even if $\|f_k\|_2 > \|f_{k+1}\|_2$ for some k, since ϕ_k is a quotient between $\|f_k\|_2$ and $\|g - Af_k\|_2$.
- Although the proposed rule has already been used in connection with LSQR in Bazán et al. (2013), we emphasize that the discussion and analysis on the behavior of Ψ_k given here do not appear elsewhere.
- A key feature of this stopping rule is that there is no need to run a predetermined amount of steps, as required, e.g., by the L-Curve criterion (Hansen et al. 2007; Kilmer et al. 2007). More precisely, it stops automatically in at most $\widetilde{k} + 1$ steps where $\widetilde{k}$ is the selected regularization parameter.

5 Iterative methods with preconditioning via smoothing norm

It is possible that for a certain class of problems, certain projection methods cannot always capture some intrinsic feature of the solution probably because the search space is not well suited. One way to circumvent this difficulty is by changing the search space with the incorporation of prior information via regularization matrices. For instance, if the required solution is known to be smooth, we can use general-form Tikhonov regularization to construct regularized solutions defined as

$$f_\lambda = \underset{f \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \|g - Af\|_2^2 + \lambda^2 \|Lf\|_2^2 \right\}, \tag{5.1}$$

where $\lambda > 0$ is the regularization parameter and L is chosen so as to incorporate into the minimization process (hence into the computed f_λ) prior information such as smoothness. If $L \neq I_n$ (the identity matrix of order n), Eq. (5.1) can be transformed into Hansen (1998)

$$f_\lambda = \underset{\overline{f} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \|\overline{g} - \overline{Af}\|_2^2 + \lambda^2 \|\overline{f}\|_2^2 \right\}, \tag{5.2}$$

thus incorporating the properties of matrix L into $\overline{A}$, the transformation being given by

$$\overline{A} = AL_A^\dagger, \quad \overline{g} = g - Af_N, \quad f_\lambda = L_A^\dagger \overline{f}_\lambda + f_N, \tag{5.3}$$

where $L_A^\dagger = \left(I_n - \left(A \left(I_n - L^\dagger L \right) \right)^\dagger A \right) L^\dagger$ is known as the A-weighted generalized inverse of L (Eldén 1982), and f_N is the component of the solutions that belongs to the null space of L, $\mathcal{N}(L)$. Nevertheless, on large-scale problems, the computation of the regularization parameter λ can be a burdensome task and a way to overcome this issue is to replace Tikhonov regularization by iterative regularization via projection methods. As far as iterative regularization in this new context is concerned, the idea is to incorporate prior information into the computed iterates by changing the original search subspace in such a way that the original problem (1.1) is replaced by the following minimization problem

$$\underset{\overline{f} \in \mathbb{R}^p}{\operatorname{argmin}} \|\overline{g} - \overline{Af}\|_2^2, \tag{5.4}$$

and then compute regularized solution by early truncation of the iterative process. To exploit the stopping rule described in the previous section in this new context, our proposal is to apply Krylov projection methods (e.g., MINRES/MR-II, GMRES/RRGMRES, etc) to (5.4)

and stop the iterations according to the rule described in the previous section, i.e., stop the iterations at the first integer k such that

$$\widetilde{k} = \text{argmin} \|\overline{g} - \overline{Af}_k\|_2 \|\overline{f}_k\|_2. \tag{5.5}$$

Once the stopping index $\widetilde{k}$ is determined, the computed approximation $\overline{f}_{\widetilde{k}}$ is transformed back to obtain

$$f_{\widetilde{k}} = L_A^\dagger \overline{f}_{\widetilde{k}} + f_N$$

Two approaches can be used, one approach due to (Calvetti et al. 2005) which requires L to be square, and a more general approach by Hansen and Jensen (2006) where such condition is not needed. As we only consider the latter, we describe it briefly.

5.1 The smoothing norm (SN) approach

As our interest is to use MINRES/MR-II and GMRES/RRGMRES, we assume that A is square and we deal with consistent linear systems. We shall also note that the iterated solution can be written as (we will suppress the index k) $f = L_A^\dagger \overline{f} + f_N = L_A^\dagger \overline{f} + Wz$, with the columns of W being a basis to $\mathcal{N}(L)$. Thus, the equation $Af = g$ can be written as

$$A(L_A^\dagger, W) \begin{pmatrix} \overline{f} \\ z \end{pmatrix} = g. \tag{5.6}$$

Multiplying the above equation by $(L_A^\dagger \;\; W)^T$ leads to

$$\begin{pmatrix} L_A^{\dagger T} A L_A^\dagger & L_A^{\dagger T} AW \\ W^T A L_A^\dagger & W^T AW \end{pmatrix} \begin{pmatrix} \overline{f} \\ z \end{pmatrix} = \begin{pmatrix} L_A^{\dagger T} g \\ W^T g \end{pmatrix}. \tag{5.7}$$

Using the Schur complement, it follows that $S\overline{f} = \overline{d}$ where

$$\begin{aligned} S &= L_A^{\dagger T} A L_A^\dagger - L_A^{\dagger T} AW (W^T AW)^{-1} W^T A L_A^\dagger = L_A^{\dagger T} P A L_A^\dagger, \\ \overline{d} &= L_A^{\dagger T} g - L_A^{\dagger T} AW (W^T AW)^{-1} W^T g = L_A^{\dagger T} P g, \end{aligned} \tag{5.8}$$

with $P = I - AW(W^T AW)^{-1} W^T$. The SN approach looks for approximations to f_{exact} from the linear system $S\overline{f} = \overline{d}$. As demonstrated in Hansen and Jensen (2006), based on the fact that the subspaces $\mathcal{R}(L^T)$ and $\mathcal{R}(AW)$ are complementary, the above system reduces to

$$L^{\dagger T} P A L^\dagger \overline{f} = L^{\dagger T} P g, \tag{5.9}$$

which means $L_A^\dagger$ is replaced by $L^\dagger$. Table 1 summarizes the basic ideas of the algorithm SN-X, where X is any method, e.g., MINRES, MR-II, GMRES and RRGMRES.

Table 1 Algorithm SN-X where X = MINRES/MR-II or X = GMRES/RRGMRES

Input: A, L, g
Output: Iterated solution f_k.
1. Apply the method X to the system $S\overline{f} = \overline{d}$ until some stopping rule is satisfied to give $\overline{f}_k$.
2. Compute $f_k = L_A^\dagger \overline{f}_k + f_N$ whit f_N and $L_A^\dagger$ according to a (5.3).

5.2 Smoothing preconditioning for LSQR

If A is not square, methods such as MINRES/MR-II and GMRES/RRGMRES cannot be used, and in such cases LSQR is probably the most appropriate method to be chosen. Thus, the proposal is to apply LSQR to the problem (5.4), as suggested by Hanke and Hansen (1993), and stop the iterations using the parameter choice rule described before. Then, the regularization parameter solution f_k is computed as in step 2 of algorithm SN-X. As in Bazán et al. (2013), for simplicity this preconditioned version of LSQR will be referred to as P-LSQR. Obviously, for P-LSQR to be computationally viable, the dimension of $\mathcal{N}(L)$ should be small and the matrix–vector products involving both $L^{\dagger}$ and $L^{\dagger^T}$ must be performed as efficiently as possible. For details and numerical results involving P-LSQR, the reader is referred to Bazán et al. (2013).

6 Numerical experiments

To illustrate that our stopping rule is efficient, we shall apply it to a number of problems and compare the results with those obtained by the L-Curve criterion. As in the continuous case, the L-curve criterion promotes choosing the regularization parameter as the index associated to the corner of the discrete L-curve defined by

$$\mathcal{L}_q = \{(\log \|g - Af_k\|_2, \log \|f_k\|_2), \quad k = 1, \ldots, q\}. \tag{6.1}$$

Methods to find such "corner" include the use of splines due to Hansen and O'Leary (1993), the triangle method of Castellanos et al. (2002), a method developed by Rodriguez and Theis (2005) and an adaptive pruning algorithm due to Hansen et al. (2007). However, finding the "corner" using a finite number of points is not an easy task and the above algorithms are not without difficulties, see, e.g., (Hansen 1998, p. 190) for some discussions, Hansen et al. (2007) for a case with multiple "corners", and more recently, Bazán et al. (2013) for a comparison of results obtained by LSQR coupled with the proposed stopping criterion and L-curve. In our numerical experiments, we use an implementation of L-curve based on the pruning algorithm (Hansen et al. 2007).

The following notation will be used:

- k_{Ψ}, k_{LC}, k_{opt}: stopping index determined by (4.5), by the L-Curve criterion and the optimum one, respectively;
- $\bar{E}_{\Psi}$, $\bar{E}_{\mathrm{LC}}$, $\bar{E}_{\mathrm{opt}}$: average values of relative error in $f_{k_{\Psi}}$, $f_{k_{\mathrm{LC}}}$, $f_{k_{\mathrm{opt}}}$, respectively.

The optimal regularization parameter is defined as $k_{\mathrm{opt}} = \mathrm{argmin}_k \|f_k - f_{\mathrm{exact}}\|_2 / \|f_{\mathrm{exact}}\|_2$.

6.1 Methods without preconditioning

The purpose of this section is to illustrate the potential of the stopping rule (4.5) in connection with algorithms LSQR, RRGMRES (if $A \neq A^T$), MR-II (if $A = A^T$) and TSVD. The choice of RRGMRES and MR-II obeys the well-known fact that they outperform their counterparts GMRES and MINRES, respectively Jensen and Hansen (2007). To this end, we selected six test problems from Hansen (1994), namely, foxgood, shaw, deriv2, phillips, heat and baart, and an image deblurring problem.

6.1.1 Results for test problems from Hansen (1994)

As it is well known, Hansen's Toolbox provides triplets $[A, g_{\text{exact}}, f_{\text{exact}}]$ with A $n \times n$ such that $A\, f_{\text{exact}} = g_{\text{exact}}$. For our experiment, we choose $n = 800$ and for each problem we consider 20 data vectors of the form $g = g_{\text{exact}} + e$ where the noise vector e contains zero-mean Gaussian random numbers scaled such that $\text{NL} = \|e\|_2/\|g_{\text{exact}}\|_2 = 0.001,\ 0.01,\ 0.025$. To ensure that the overall behavior of the discrete L-Curve is captured, we take $q = 200$. This choice seems reasonable compared to the dimension of the problem and the numerical rank of each test problem. Numerical rank as well as the condition number of each test problem, all computed by Matlab, are reported in Table 2.

Average results are reported in Tables 3, 4, 5, 6, 7, 8. Number inside the parenthesis corresponds to the maximum stopping index of all realizations.

As we can see, except for the poor performance of RRGMRES on heat test problem, see Table 7, which is nothing new because, as we know, this method is not always able to produce regularized solutions Jensen and Hansen (2007), all projection methods tested in this study were able to produce very good results. In particular, the results indicate: (a) that both L-curve and the new parameter choice rule work well in connection with projection methods

Table 2 Numerical rank and condition number of 6 test problems

	foxgood	shaw	deriv2	phillips	heat	baart
rank	30	20	800	800	596	13
$\kappa(A)$	1.1×10^{20}	2.4×10^{20}	7.8×10^{5}	1.1×10^{10}	7.1×10^{188}	3.0×10^{18}

Table 3 Results obtained by LSQR, MR-II and TSVD for foxgood test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD
k_Ψ	3 (4)	3 (4)	3 (4)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)
k_{LC}	3 (4)	3 (4)	4 (4)	2 (2)	2 (2)	2 (2)	2(2)	2(2)	2(2)
k_{opt}	3 (4)	3 (4)	3 (4)	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)
$\bar{E}_\Psi$	0.0217	0.0211	0.0193	0.0311	0.0312	0.0312	0.0319	0.0319	0.0320
$\bar{E}_{\text{LC}}$	0.0212	0.0211	0.0193	0.0311	0.0312	0.0312	0.0319	0.0319	0.0320
$\bar{E}_{\text{opt}}$	0.0080	0.0080	0.0078	0.0256	0.0256	0.0249	0.0308	0.0308	0.0296

Table 4 Results obtained by LSQR, MR-II and TSVD for shaw test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD
k_Ψ	7 (8)	7 (8)	7 (8)	5 (6)	5 (7)	7 (7)	4 (4)	4 (4)	4 (5)
k_{LC}	7 (8)	7 (8)	8 (8)	4 (7)	5 (7)	7 (7)	4 (4)	4 (4)	4 (5)
k_{opt}	7 (9)	7 (9)	7 (10)	5 (8)	5 (8)	6 (7)	5 (7)	5 (7)	5 (7)
$\bar{E}_\Psi$	0.0498	0.0500	0.0500	0.0775	0.0898	0.0670	0.1683	0.1688	0.1679
$\bar{E}_{\text{LC}}$	0.0496	0.0493	0.0501	0.1082	0.0811	0.0670	0.1683	0.1688	0.1690
$\bar{E}_{\text{opt}}$	0.0427	0.0430	0.0440	0.0637	0.0637	0.0665	0.0969	0.0952	0.1036

Table 5 Results obtained by LSQR, MR-II and TSVD for deriv2 test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD
k_Ψ	16 (17)	13 (13)	17 (34)	7 (7)	6 (6)	9 (10)	5 (5)	4 (5)	5 (6)
k_{LC}	16 (16)	9 (14)	32 (34)	6 (7)	6 (6)	9 (10)	5 (5)	4 (5)	6 (6)
k_{opt}	13 (16)	11 (12)	25 (35)	7 (9)	6 (8)	10 (17)	5 (7)	5 (6)	8 (14)
$\bar{E}_\Psi$	0.1474	0.1523	0.1525	0.2145	0.2161	0.2323	0.2656	0.2689	0.2949
$\bar{E}_{\mathrm{LC}}$	0.1471	0.1537	0.1562	0.2151	0.2161	0.2325	0.2656	0.2603	0.2949
$\bar{E}_{\mathrm{opt}}$	0.1392	0.1399	0.1451	0.2019	0.2027	0.2063	0.2344	0.2364	0.2406

Table 6 Results obtained by LSQR, MR-II and TSVD for phillips test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD	LSQR	MR-II	TSVD
k_Ψ	11 (15)	7 (17)	12 (25)	9 (10)	8 (10)	11 (11)	7 (8)	7 (8)	8 (8)
k_{LC}	14 (16)	11 (18)	21 (26)	9 (10)	8 (10)	11 (11)	7 (8)	6 (8)	8 (8)
k_{opt}	9 (10)	9 (10)	11 (12)	5 (9)	4 (9)	7 (11)	5 (8)	4 (7)	7 (9)
$\bar{E}_\Psi$	0.0617	0.0343	0.0497	0.0374	0.0347	0.0280	0.0327	0.0335	0.0276
$\bar{E}_{\mathrm{LC}}$	0.0706	0.0677	0.0721	0.0374	0.0431	0.0280	0.0327	0.0288	0.0276
$\bar{E}_{\mathrm{opt}}$	0.0072	0.0077	0.0065	0.0223	0.0196	0.0165	0.0255	0.0246	0.0224

Table 7 Results obtained by LSQR, RRGMRES and TSVD for heat test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	RRGMRES	TSVD	LSQR	RRGMRES	TSVD	LSQR	RRGMRES	TSVD
k_Ψ	28 (29)	18 (24)	17 (33)	16 (16)	17 (25)	17 (24)	11 (11)	14 (24)	16 (18)
k_{LC}	27 (31)	9 (69)	45 (46)	15 (18)	11 (80)	23 (24)	10 (12)	12 (67)	17 (18)
k_{opt}	20 (22)	1 (1)	29 (34)	13 (15)	1 (1)	21 (28)	11 (13)	1 (1)	16 (23)
$\bar{E}_\Psi$	0.0812	1.3×10^{6}	0.0660	0.0798	8.8×10^{5}	0.0762	0.1091	7.4×10^{5}	0.1129
$\bar{E}_{\mathrm{LC}}$	0.0925	3.8×10^{8}	0.0809	0.0778	1.4×10^{8}	0.0747	0.1021	5.5×10^{7}	0.1108
$\bar{E}_{\mathrm{opt}}$	0.0253	1.0756	0.0233	0.0711	1.0756	0.0681	0.1021	1.0755	0.1069

Table 8 Results obtained by LSQR, RRGMRES and TSVD for baart test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	LSQR	RRGMRES	TSVD	LSQR	RRGMRES	TSVD	LSQR	RRGMRES	TSVD
k_Ψ	4 (4)	3 (5)	4 (4)	3 (3)	3 (4)	3 (3)	3 (3)	3 (4)	3 (3)
k_{LC}	4 (4)	3 (6)	4 (4)	3 (3)	3 (6)	3 (3)	3 (3)	3 (6)	3 (3)
k_{opt}	4 (5)	3 (4)	4 (5)	3 (4)	3 (4)	3 (4)	3 (4)	3 (4)	3 (4)
$\bar{E}_\Psi$	0.1159	0.0402	0.1160	0.1662	0.0569	0.668	0.1684	0.0655	0.1691
$\bar{E}_{\mathrm{LC}}$	0.1159	0.8584	0.1160	0.1662	17.7101	0.1668	0.1684	6.9093	0.1691
$\bar{E}_{\mathrm{opt}}$	0.1027	0.0337	0.1028	0.1459	0.0385	0.1461	0.1614	0.0483	0.1629

Fig. 6 Exact and blurred and noisy images for NL = 0.001 and NL = 0.05

Table 9 Results obtained by LSQR and MR-II for image deblurring test problem

	NL = 0.001		NL = 0.01		NL = 0.025		NL = 0.05	
	LSQR	MR-II	LSQR	MR-II	LSQR	MR-II	LSQR	MR-II
k_Ψ	535 (545)	69 (70)	90 (93)	22 (22)	39 (40)	13 (14)	20 (21)	9 (9)
k_{LC}	495 (577)	62 (77)	80 (103)	21 (23)	36 (40)	13 (13)	18 (23)	9 (9)
k_{opt}	247 (272)	40 (43)	43 (47)	14 (15)	22 (23)	9 (10)	13 (15)	7 (7)
$\bar{E}_\Psi$	0.1618	0.1616	0.1672	0.1673	0.1702	0.1698	0.1733	0.1722
$\bar{E}_{\text{LC}}$	0.1598	0.1607	0.1662	0.1649	0.1683	0.1680	0.1724	0.1722
$\bar{E}_{\text{opt}}$	0.1494	0.1492	0.1577	0.1575	0.1629	0.1629	0.1689	0.1688

and TSVD, and (b), that the rules are able to produce results with relative errors close to the optimal, with the observation that the new rule is cheaper.

6.1.2 Results for deblurring test problem

We consider a part of the image Barbara of size 175×175 which was used in Hansen and Jensen (2008). This implies that we deal with $n = 30625$ unknowns and a linear system with coefficient matrix given by $A = T \otimes T$ where $T \in \mathbb{R}^{175\times 175}$ is symmetric and Toeplitz Hansen and Jensen (2008). The condition number of A is $\kappa(A) \approx 2.1 \times 10^{33}$ and the numerical rank is 8648. In deblurring problems, A acts as blurring operator and the right hand side $g = g_{\text{exact}} + e$ represents the blurred and noisy image in vector form. Since this problem is much larger than those from the previous section, for L-curve we take $q = 800$. The exact image and two blurred and noisy images (one with NL = 0.001 and other with NL = 0.05) are depicted in Fig. 6.

In addition to the same noise levels considered in the previous examples, we also consider data with NL = 0.05 since this value was used in Hansen and Jensen (2008). It is worth mentioning that in Hansen and Jensen (2008) the main concern was to study the behavior of the iterations, not to discuss stopping rules. We report results obtained by LSQR and MR-II. Notice that, except for the fact that LSQR spends a considerably large number of iterations compared to MR-II, both methods produced similar results, see Table 9. For this, only reconstructed images obtained by LSQR for each noise level are reported, see Fig. 7.

From the results, we see that the reconstructed images show artifacts in the form of circular freckles. This is in accordance with observations made in Hansen and Jensen (2008). For a detailed analysis of the appearance of freckles in the reconstructed image, the reader is referred to Hansen and Jensen (2008).

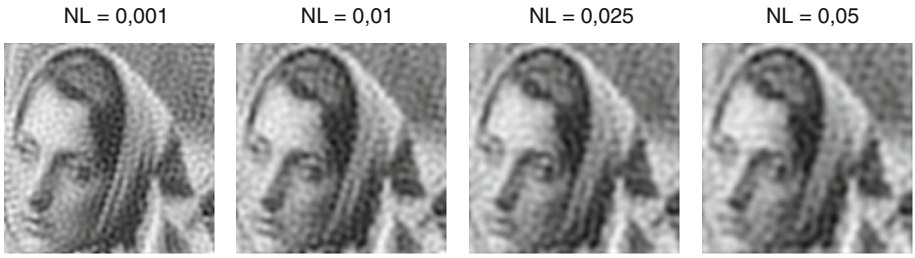

Fig. 7 Solutions obtained via LSQR coupled with stopping rule (4.5)

Table 10 Results obtained by SN-RRGMRES, P-LSQR and TGSVD for deriv2 test problem

	NL = 0.001			NL = 0.01			NL = 0.025		
	SN-R	P-L	TGSVD	SN-R	P-L	TGSVD	SN-R	P-L	TGSVD
k_Ψ	9 (11)	5 (5)	5 (6)	5 (6)	2 (2)	2 (2)	4 (5)	1 (1)	1 (1)
k_{LC}	9 (11)	5 (5)	6 (6)	5 (6)	2 (2)	2 (2)	4 (5)	2 (2)	2 (2)
k_{opt}	7 (10)	7 (9)	9 (15)	3 (5)	3 (5)	3 (7)	3 (5)	3 (4)	3 (5)
$\bar{E}_\Psi$	0.0122	0.0162	0.0176	0.0291	0.0543	0.0609	0.0405	0.0689	0.0701
$\bar{E}_{\text{LC}}$	0.0122	0.0162	0.0176	0.0297	0.0543	0.0609	0.0405	0.0541	0.0601
$\bar{E}_{\text{opt}}$	0.0087	0.0091	0.0092	0.0210	0.0219	0.0221	0.0299	0.0306	0.0307

6.2 Methods with preconditioning via smoothing norm

To improve the quality of the solution to certain problems using a priori knowledge, we will illustrate the effectiveness of some of the preconditioned methods described in the previous section. In all the following examples, the regularization parameter is determined by the stopping rule (4.5) and L-curve.

Our first example involves deriv2 test problem considered in the previous section, and is motivated by the fact LSQR, MR-II and TSVD produced solutions with not so small errors, see Table 5. Smoothing preconditioning is incorporated by considering the minimization problem (5.4) with the regularization matrix defined by

$$L = \mathcal{L}_1 = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}. \tag{6.2}$$

Table 10 shows results obtained by P-LSQR (P-L for short), SN-RRGMRES (SN-R for short) and TGSVD. As in the previous section, number inside the parenthesis corresponds to the maximum stopping index of all realizations.

The results show that both L-curve and stopping rule (4.5) produced high-quality solutions with relative errors of one order of magnitude smaller than the relative errors obtained with methods without preconditioning.

As a second example, we choose again the image deblurring test problem considered in the previous section. In this case, for the regularization matrix, we consider a bidimensional first order differential operator defined by

Table 11 Results obtained by P-LSQR and SN-RRGMRES for image deblurring test problem

	NL = 0.001		NL = 0.01		NL = 0.025		NL = 0.05	
	P-L	SN-R	P-L	SN-R	P-L	SN-R	P-L	SN-R
k_Ψ	285 (290)	182 (183)	67 (68)	85 (86)	33 (33)	59 (60)	17 (17)	45 (46)
k_{LC}	3 (308)	175 (177)	3 (71)	77 (79)	35 (36)	55 (64)	16 (16)	43 (47)
k_{opt}	479 (507)	208 (216)	160 (173)	96 (100)	98 (106)	69 (73)	67 (73)	55 (58)
$\bar{E}_\Psi$	0.1518	0.1498	0.1633	0.1576	0.1731	0.1627	0.1875	0.1684
$\bar{E}_{\text{LC}}$	0.2729	0.1500	0.1691	0.1584	0.1721	0.1629	0.1899	0.1689
$\bar{E}_{\text{opt}}$	0.1495	0.1493	0.1572	0.1571	0.1618	0.1617	0.1669	0.1667

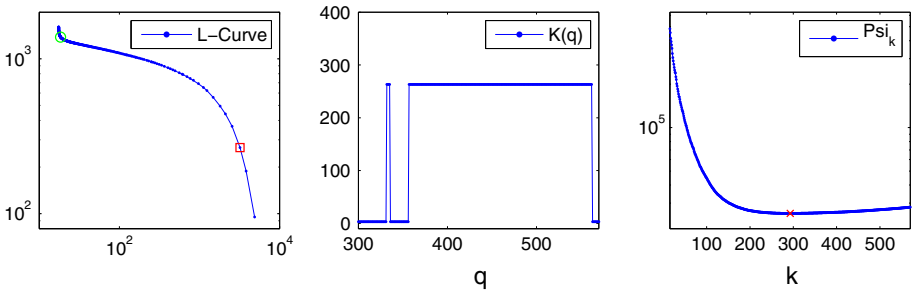

Fig. 8 Discrete L-curve (*left*), corner index as function of q points (*middle*) and sequence Ψ_k for image deblurring test problem for NL = 0.001. In this example, the "well" distinctive corner (marked with a *small circle*) produces oversmoothed solutions, while the "false" corner (marked with a *small square*) produces acceptable solutions

$$L = \begin{bmatrix} \mathcal{L}_1 \otimes I \\ I \otimes \mathcal{L}_1 \end{bmatrix} \tag{6.3}$$

where $\otimes$ denotes Kronecker product. Table 11 shows results obtained by P-LSQR and SN-RRGMRES. As before, for L-curve, we take $q = 800$.

This test problem provides an excellent way to illustrate that the regularization parameter determined by L-curve can be very sensitive to changes in the number of points q used to perform the L-curve analysis. A clear evidence appears in Table 11 for NL $= 0.001$: the corner index varies so much taking two values, $k_{\text{LC}} = 3$ and $k_{\text{LC}} = 263$. To reinforce this, Fig. 8 displays the corner index returned by the pruning algorithm (as implemented by the Matlab function `corner.m` in Hansen (1994)) as a function of q for NL $= 0.001$ and q ranging from 300 to 600. This example also illustrates that even if the L-curve displays a distinctive L-corner, the corner index does not necessarily produce a good regularized solution, as seen in this example for $k_{\text{LC}} = 3$. Notice that, contrary to the behavior of k_{LC}, the parameter determined by minimizing Ψ_k does not suffer from large variations. Thus, the conclusion that can be made here is that L-curve should be carefully used in connection with iterative methods.

Figure 9 displays reconstructed images obtained by P-LSQR and SN-RRGMRES. The benefit from using smoothing preconditioning is apparent.

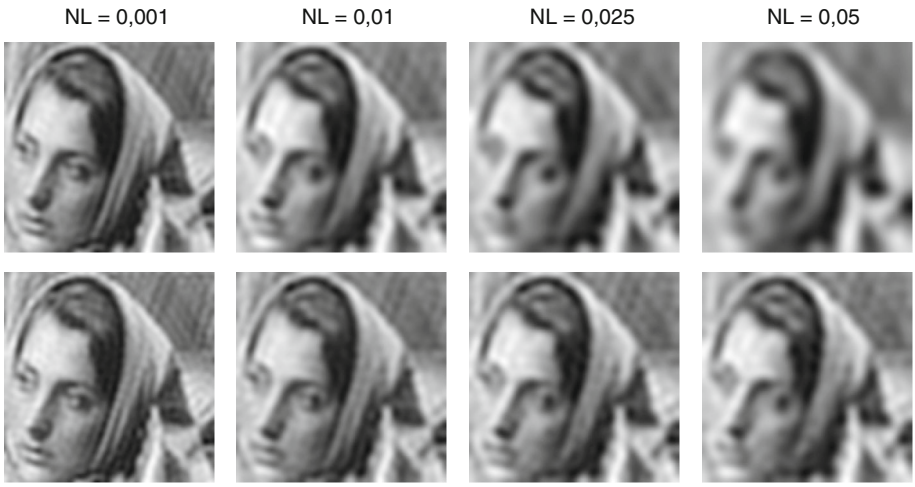

Fig. 9 Solutions obtained by P-LSQR (*top*) and SN-RRGMRES (*bottom*)

7 Conclusions

We extended a stopping rule for LSQR proposed recently in Bazán et al. (2013) to well-known Krylov projection methods. The rule does not require any estimate of the error norm and stops automatically after $\widetilde{k}+1$ steps where $\widetilde{k}$ is the computed regularization parameter. Numerical results show that the rule works well in conjunction with classical projection methods and its smoothing norm preconditioned versions. In particular, the numerical examples show that the proposed stopping rule is cheaper than L-curve, that the regularization parameter determined by L-curve strongly depends on the number of points used to perform the L-curve analysis, and that the proposed rule is able to produce solutions that are as good as those obtained by L-curve when the latter performs well.

References

Bakushinski AB (1984) Remarks on choosing a regularization parameter using quasi-optimality and ratio criterion. USSR Comp Math Phys 24:181–182

Bauer F, Lukas MA (2011) Comparing parameter choice methods for regularization of ill-posed problem. Math Comput Simul 81:1795–1841

Bazán FSV (2008) Fixed-point iterations in determining the Tikhonov regularization parameter. Inverse Problems 24 doi:10.1088/0266-5611/24/3/035001

Bazán FSV, Francisco JB (2009) An improved fixed-point algorithm for determining the Tikhonov regularization parameter. Inverse Problems 25 doi:10.1088/0266-5611/25/4/045007

Bazán FSV, Cunha MCC, Borges LS (2013) Extension of GKB-FP algorithm to large-scale general-form Tikhonov regularization. Numer Lin Alg. doi:10.1002/nla.1874

Björck Å (1996) Numerical Methods for Least Squares Problems. SIAM

Bunse-Gerstner A, Guerra-Ones V (2006) An improved preconditioned LSQR for discrete ill-posed problems. Math Comput in Simul 73:65–75

Calvetti D, Lewis B, Reichel L (2000) GMRES-type methods for inconsistent systems. Lin Alg and its Appl 316:157–169

Calvetti D, Reichel L, Shuibi A (2005) Invertible smoothing preconditioners for linear discrete ill-posed problems. Appl Num Math 54:135–149

Castellanos JJ, Gómez S, Guerra V (2002) The triangle method for finding the corner of the L-curve. Appl Numer Math 43:359–373

Cimmino G (1983) Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. La Ricerca Scientifica II 9:326–333
Dold A, Eckmann B (1986) Lecture Notes in Mathematics. Inverse Problems
Eldén L (1982) A weighted pseudoinverse, generalized singular values, and constrained least square problems. BIT 22:487–502
Engl HW, Hanke M, Neubauer A (1996) Mathematics and its applications., Regularization of inverse problemsKluwer Academic, Dordrecht
Fierro RD, Bunch JR (1995) Bounding the subspaces from rank revealing two-sided orthogonal decomposition. SIAM J Matrix Anal Appl 16:743–759
Golub GH, Van Loan CF (1996) Matrix computations. The Johns Hopkins University Press, Baltimore
Hanke H (1995) Conjugate gradient type methods for Ill-posed problems. Longman, Harlow
Hanke M, Hansen PC (1993) Regularization methods for large-scale problems. Surv Math Ind 3:253–315
Hansen PC (1990) The discrete picard condition for discrete Ill-posed problems. BIT 30:658–672
Hansen PC (1994) Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems. Numer Algebra 6:1–35
Hansen PC (1998) Rank-deficient and discrete ill-posed problems. SIAM, Philadelphia
Hansen PC (2010) Discrete inverse problems: insight and algorithms. SIAM, Philadelphia
Hansen PC, O'Leary DP (1993) The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J Sci Comput 14:1487–1503
Hansen PC, Jensen TK (2006) Smoothing-norm preconditioning for regularizing minimum-residual methods. SIAM J Matrix Anal Appl 29:1–14
Hansen PC, Jensen TK (2008) Noise propagation in regularizing iterations for image deblurring. Elec Trans Numer Anal 31:204–220
Hansen PC, Jensen TK, Rodriguez G (2007) An adaptive pruning algorithm for the discrete L-curve criterion. J Comput Appl Math 198:483–492
Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. J Res Nat Bur Stand 49:409–436
Jensen TK, Hansen PC (2007) Iterative regularization with minimum-residual methods. BIT 47:103–120
Kilmer ME, Hansen PC, Español MI (2007) A projection-based approach to general-form Tikhonov regularization. SIAM J Sci Comput 29:315–330
Kilmer ME, O'Leary DP (2001) Choosing regularization parameters in iterative methods for ill-posed problems. SIAM J Matrix Anal Appl 22:1204–1221
Kindermann S (2011) Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. Electron Trans Numer Anal 38:233–257
Landweber L (1951) An iteration formula for Fredholm integral equations of the first kind. Am J Math 73:615–624
Lukas MA (2006) Robust generalized cross-validation for choosing the regularization parameter. Inverse Probl 22:1883–1902
Neuman A, Reichel L, Sadok H (2012) Implementations of range restricted iterative methods for linear discrete ill-posed problems. Linear Algebra Appl 436:3974–3990
Paige CC, Saunders MA (1975) Solution of sparse indefinite systems of linear equations. SIAM J Numer Anal 12:617–629
Morozov VA (1984) Regularization methods for solving incorrectly posed problems. Springer, New York
Paige CC, Saunders MA (1982) LSQR: An algorithm for sparse linear equations and sparse least squares. ACM Trans Math Softw 8:43–71
Regińska T (1996) A regularization parameter in discrete ill-posed problems. SIAM J Sci Comput 3:740–749
Reichel L, Rodriguez G (2012) Old and new parameter choice rules for discrete ill-posed problems. Numer Algorithms. doi:10.1007/s11075-012-9612-8
Rodriguez G, Theis D (2005) An algorithm for estimating the optimal regularization parameter by the L-curve. Rend Mat 25:69–84
Saad Y, Schultz MH (1986) GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J Sci Statist Comput 7:856–869
Tikhonov AN (1963) Solution of incorrectly formulated problems and the regularization method. Soviet Math Dokl 4:1035–1038