

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

A stopping criterion for iterative regularization methods

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

A stopping criterion for iterative regularization methods / Landi, G.; Loli Piccolomini, E; Tomba, I.. - In: APPLIED NUMERICAL MATHEMATICS. - ISSN 0168-9274. - STAMPA. - 106:(2016), pp. 53-68.
[10.1016/j.apnum.2016.03.006]

Availability:

This version is available at: <https://hdl.handle.net/11585/589928> since: 2017-05-20

Published:

DOI: <http://doi.org/10.1016/j.apnum.2016.03.006>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Landi, G., E. Loli Piccolomini, and I. Tomba. "A Stopping Criterion for Iterative Regularization Methods." *Applied Numerical Mathematics*, vol. 106, 2016, pp. 53-68.

The final published version is available online at:
<http://dx.doi.org/10.1016/j.apnum.2016.03.006>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

A stopping criterion for iterative regularization methods

G. Landi¹ and E. Loli Piccolomini² and I. Tomba³

Keywords: Iterative regularization; regularization parameter choice; linear discrete ill-posed problems; CGLS; Discrete Picard Condition.

1. Introduction

In this work we consider least squares problems

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2, \quad \mathbf{A} \in \mathbb{R}^{m \times n}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{b} \in \mathbb{R}^m, \quad m \geq n \quad (1)$$

where the coefficient matrix \mathbf{A} is ill-conditioned and derives from the discretization of a continuous ill-posed operator. The data \mathbf{b} is assumed to be corrupted by measurement errors, which we will refer to as noise. In particular, we suppose that $\mathbf{b} = \mathbf{b}^{\text{exact}} + \mathbf{e}$, where $\mathbf{b}^{\text{exact}}$ is the unknown noise-free right-hand side vector and \mathbf{e} is a zero-mean white noise vector. In (1) and in the sequel, $\|\cdot\|$ denotes the Euclidean norm.

Discrete ill-posed problems of the form (1) arise, for example, from the discretization of Fredholm integral equations of the first kind that are commonly used to model instrument distortions. They are often encountered in large-scale image deblurring applications, where \mathbf{A} is typically the matrix representation of a convolution operator. Under periodic boundary conditions, \mathbf{A} is block circulant with circulant blocks (BCCB) and matrix-vector products can be performed using FFTs [28]. This type of observation model may describe, for example, motion blur, atmospheric turbulence blur and out-of-focus blur.

Background. Because of the ill-conditioning of \mathbf{A} , regularization techniques are necessary in order to reduce the sensitivity of the solution of (1) to the noise in \mathbf{b} . Iterative regularization methods are some methods of choice when the dimensions of problem (1) are large and \mathbf{A} cannot be explicitly stored, but matrix-vector products involving \mathbf{A} can be easily computed. Iterative methods have a semiconvergence behaviour when applied to ill-posed problems and they can be used as regularization methods if suitably stopped before the noise enters the computed solution. The stopping iteration plays the role of the regularization parameter, providing a fair balance between data fidelity and solution smoothness.

For example, in image restoration problems, Krylov subspace methods are very important, as pointed out in [22] where an insightful analysis is performed.

Popular iterative regularization methods are, for example, the Conjugate Gradient Least Squares (CGLS), the Preconditioned CGLS (PCGLS), the MINimal RESidual (MINRES), the Generalized Minimal Residual (GMRES) and the Range Restricted GMRES (RRGMRES) methods [16, 18, 31, 43, 10, 35, 32]. When nonnegativity constraints are added to the least squares problem (1), the Scaled Gradient Projection method (SGP), the Iterative Space Reconstruction Algorithm (ISRA), the Projected Landweber (PL) method or the Projected Newton (PN) method may be used as iterative regularization methods [4, 11, 12, 5, 38, 39].

A wise choice of the regularization parameter is a vital issue in applying iterative regularization methods to practical applications, since the quality of the regularized solution crucially depends on this choice. The recent literature on ill-posed inverse problems shows that efficient regularization parameter selection techniques are under active research.

The famous *discrepancy principle* [41] is probably the most widely used parameter choice strategy in the context of regularization. It is an *a-posteriori* criterion choosing the regularization parameter as a function of the data and the noise norm which must be known. Unfortunately, this information may not be available in real-world applications and methods not requiring an estimate of the noise norm are actually desirable. The recent literature shows an increasing interest in *heuristic* (or *noise-level-free*) parameter choice strategies, even if the well-known Bakushinsky veto [1] states that, in Hilbert spaces, all heuristic parameter choice rules, which do not make use of the knowledge about the exact noise level, will never converge in the worst-case scenario analysis. Nevertheless, heuristic parameter selection techniques are used quite frequently in practical applications, often giving good results [46, 3]. The L-curve method [40, 26] and the Generalized Cross Validation method [14] are likely the most popular heuristic parameter selection strategies; they have been deeply investigated by several authors [9, 34, 17, 53]. Some variants of the L-curve criterion have been described, e.g. the residual L-curve criterion [47, 46] and the Regin'ska's method [45]. A number of other choice rules have been proposed in the literature, e.g. the Hanke-Raus rule [19] and the quasi-optimality criterion [52, 51] which have recently received an increasing interest [2, 37, 36, 42]. Other parameter choice techniques, called extrapolation methods [6, 7], are based on suitable *a posteriori* estimates of the error norm in the solution of (1) while another method obtains estimates of the noise level from the Golub-Kahan iterative bidiagonalization [29] (this last method, referred to as quadrature method, is specific for LSQR). Several minimization rules for the selection of the regularization parameter for many iterative regularization methods as the Landweber method and the CGLS method are given in [15], both in case of known and unknown noise norm. A detailed and careful comparison of many parameter choice rules is performed in [3, 46].

Other criteria for choosing the parameter of iterative regularization methods are based on the estimation of the residual norm of a regularized solution. A recent and innovative approach is illustrated in [50, 49, 48] where several diagnostic tools, which are statistically motivated, are presented to evaluate the

suitability of a candidate regularization parameter for the Truncated Singular Value Decomposition (TSVD) and Tikhonov methods. In [50], an automatic procedure, based on the aforementioned diagnostic tools, is presented to select the regularization parameter so that the residual of the corresponding solution resembles white noise. In [21], the authors develop the so-called Normalized Cumulative Periodogram (NCP) method, an automatic procedure that chooses the regularization parameter making the residual as close as possible to white noise. The NCP method can be applied to direct regularization (TSVD and Tikhonov regularization) as well as iterative regularization and requires to calculate the NCP of the residual vector for each choice of the regularization parameter, starting from large values and stopping at the first parameter whose associated residual satisfies the Kolmogorov-Smirnov test. The methods developed in [50, 49] and [21] use the Fourier coefficients to determine the regularization parameter since they all use the periodogram and the cumulative periodogram to judge if the residual resembles white noise.

Finally, even if the Singular Value Decomposition and the Discrete Picard Condition (DPC) [20, 26] are well-known tools for the analysis of ill-posed inverse problems, to the best of the authors' knowledge, very little work has been done in the literature on the development of suitable parameter choice techniques using the DPC. Zama [54] has proposed a fast and efficient method based on the SVD coefficients for the computation of the regularization parameter of the TSVD and Tikhonov methods. Jones [33] has developed a method for automatically estimating the noise norm via the DPC and thus the regularization parameter for Tikhonov method.

Contribution. In this work, we focus on iterative regularization methods such as, for example, Krylov subspace methods and other methods exhibiting a semi-convergence behavior [5, 26]. In the sequel, therefore, the general term *iterative method* will refer to a method with the semiconvergence property.

The main purpose of this work is to propose a criterion for the selection of the stopping iteration of iterative methods. Our criterion is based on an estimate of the residual norm of a suitable regularized solution performed through the TSVD solution of (1). We present the criterion from an algorithmic point of view and we show that it is efficient on large size problems.

As pointed out in [21], almost all the parameter choice methods proposed in the literature involve some information about the norm of the residual vector of a regularized solution close to the exact one. For example, the famous discrepancy principle chooses the regularization parameter so that the residual norm is as close as possible to the noise norm, i.e. to the residual norm of the exact solution. In this work, we propose to choose the regularization parameter of iterative methods so that the residual norm of the corresponding solution is as close as possible to the residual norm of a suitable regularized TSVD solution. It is well known that the TSVD residual norm can be expressed in terms of the SVD coefficients and that a suitable truncation index p can be obtained by the visual inspection of their plot [26]. Therefore, we obtain the residual norm approximation to be used in the proposed stopping criterion via the DPC

[20, 26] using the coefficients of a spectral decomposition of \mathbf{A} . In particular, we propose two algorithmic procedures for the estimation of the TSVD residual norm. The first algorithm requires the SVD of \mathbf{A} and obtains the residual norm approximation by projecting the data \mathbf{b} onto the subspace spanned by the last $m - p$ left singular vectors [26]. The algorithm automatically determines the value p from the solution coefficients. The second algorithm does not require the SVD of \mathbf{A} but needs the noise norm value or an estimate of it which, in practical applications, can be obtained by heuristic strategies as those proposed in [29, 21, 50] or by the measurement instruments. This second algorithm automatically obtains the TSVD residual norm approximation from the Fourier coefficients of the data \mathbf{b} and from the noise norm estimate.

The presented criterion and the two algorithms, proposed for the residual norm computation, are deeply investigated for some large-scale image restoration problems. Numerous numerical results and comparisons with some state-of-the-art techniques are presented to validate the proposed criterion and to illustrate the numerical effectiveness of the two algorithms.

We remark that, even if we focus here on iterative regularization, the presented criterion can be easily extended to Tikhonov-like regularization formulated as a constrained optimization problem.

Outline of the paper. In section 2 we give insights into the proposed criterion motivation. In section 3, we formulate the new parameter choice criterion and we discuss two algorithms for its implementation. Section 4 presents the numerical results and comparisons with some state-of-the-art criteria for the parameter selection in iterative regularization methods. Conclusions are presented in Section 5.

2. Motivation

As already stated in the previous section, almost all criteria for the selection of the stopping index in iterative regularization methods employ some a priori information about the residual norm of a regularized (or exact) solution. The rationale behind our criterion is therefore based on the observation that the SVD and the DPC, which are effective tools for the analysis of the properties of discrete ill-conditioned least squares problems, can also be used as effective tools for the computation of a residual norm approximation which can be used in a discrepancy-like stopping criterion.

The Singular Value Decomposition and the Discrete Picard Condition. The state-of-the-art tool for the analysis of the features of the ill-posed problem (1) is the SVD of \mathbf{A} [26]:

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices, $\mathbf{0} \in \mathbb{R}^{(m-n) \times n}$ is a zero matrix and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose entries are the *singular*

values of \mathbf{A} appearing in nonincreasing order:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

The SVD solution to (1) is

$$\mathbf{x}^{\text{SVD}} = \sum_{i=1}^n \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i.$$

The terms $\mathbf{u}_i^T \mathbf{b}$ are referred to as *SVD coefficients* and the terms $\mathbf{u}_i^T \mathbf{b} / \sigma_i$ as *solution coefficients*.

The TSVD is a well-known regularization method [26], with the truncation index p playing the role of the regularization parameter. The TSVD solution is

$$\mathbf{x}_p^{\text{TSVD}} = \sum_{i=1}^p \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i, \quad p < n$$

and the corresponding residual norm r_p^{TSVD} is

$$r_p^{\text{TSVD}} = \sqrt{\sum_{i=p+1}^m |\mathbf{u}_i^T \mathbf{b}|^2}. \quad (2)$$

Obviously, p should be chosen so that noise is suppressed, as much as possible, in the approximated solution $\mathbf{x}_p^{\text{TSVD}}$ and is only left in the residual vector. Therefore, a suitable value of the truncation index p should provide a good regularized solution $\mathbf{x}_p^{\text{TSVD}}$. The DPC, deeply discussed by Hansen in [26, 20], provides insight into the influence of noise in the SVD solution. It can be stated as follows.

Discrete Picard condition. *On average, the absolute values $|\mathbf{u}_i^T \mathbf{b}^{\text{exact}}|$ of the SVD coefficients of the true right-hand side $\mathbf{b}^{\text{exact}}$ decay faster than the corresponding singular values.*

As claimed in [20], the DPC is “a necessary condition for obtaining a good regularized solution”. The DPC is usually not satisfied by least squares problems (1) where the data \mathbf{b} is assumed to be contaminated with measurement errors. In fact, even if $\mathbf{b}^{\text{exact}}$ meets the DPC, the perturbed vector \mathbf{b} usually does not. This is due to the considerable influence of the terms $\mathbf{u}_i^T \mathbf{e}$, corresponding to small singular values, on the solution norm. In the following, we will suppose that $\mathbf{b}^{\text{exact}}$ satisfies the DPC. If this is the case, and if the noise \mathbf{e} is white noise, the coefficients $|\mathbf{u}_i^T \mathbf{b}^{\text{exact}}|$ decrease with i faster than σ_i and approach zero for increasing i . Moreover, since the noise \mathbf{e} is supposed to be white noise and \mathbf{U} is orthogonal, the terms $|\mathbf{u}_i^T \mathbf{e}|$ are roughly constant for all i and the following approximation holds:

$$\|\mathbf{e}\| = \|\mathbf{U}^T \mathbf{e}\| \approx \sqrt{m} |\mathbf{u}_i^T \mathbf{e}|.$$

Therefore, the TSVD residual norm can be approximated as

$$r_p^{\text{TSVD}} = \sqrt{\sum_{i=p+1}^m |\mathbf{u}_i^T (\mathbf{b}^{\text{exact}} + \mathbf{e})|^2} \approx \sqrt{\sum_{i=p+1}^m |\mathbf{u}_i^T \mathbf{e}|^2} \approx \sqrt{\frac{m-p}{m}} \|\mathbf{e}\|. \quad (3)$$

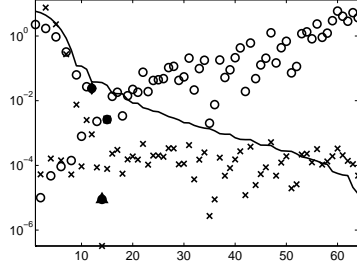


Figure 1: **phillips** test problem, Picard plot. The line denotes the singular values, the crosses indicate the SVD coefficients while the circles represent the solution coefficients. The diamond, the triangle and the square respectively denote the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$ for $i = 12$ and $i = 14, 15$.

The SVD and the DPC as tools for the analysis of ill-posed problems. The DPC is a powerful diagnostic tool able to identify the components of $\mathbf{U}^T \mathbf{b}$ that are completely dominated by noise. A useful truncation index p for TSVD regularization may be chosen via the DPC by discarding those SVD coefficients $|\mathbf{u}_i^T \mathbf{b}|$ where this condition is violated.

As a test problem, let us consider the famous problem **phillips** [44] from Hansen's Regularization Tools [25]. This test problem is widely used in the literature to model instruments distortion. The problem size is $m = n = 64$ and the noise level δ in \mathbf{b} is

$$\delta = \frac{\|\mathbf{e}\|}{\|\mathbf{b}_{\text{exact}}\|} = 10^{-4}. \quad (4)$$

From the Picard plot for this problem (figure 1) it is evident that the SVD coefficients, represented by crosses, initially decay and then level off at a plateau determined by the noise level. At the same time, the solution coefficients, represented by circles, initially decrease and then start to increase again. This means that the projections $|\mathbf{u}_i^T \mathbf{b}|$ of the right-hand side \mathbf{b} onto the left singular vector subspace of the matrix \mathbf{A} corresponding to small singular values are completely dominated by noise. A visual inspection of the Picard plot suggests that a suitable choice for the SVD truncation index is $p = 14$. The best truncation parameter, i.e. the truncation index minimizing the TSVD relative error, is $p = 15$. The truncation parameter satisfying the discrepancy principle is $p = 12$. We remark that, for the TSVD method, the discrepancy principle chooses the smallest p such that $\|\mathbf{A}\mathbf{x}_p - \mathbf{b}\| \leq \tau\|\mathbf{e}\|$, with $\tau > 1$. In the following, the TSVD solution truncated at the index p will be denoted by $\mathbf{x}_p^{\text{TSVD}}$, the corresponding relative error and residual norm values by $\text{err}_p^{\text{TSVD}}$ and r_p^{TSVD} . The relative error values of the TSVD solution satisfying the discrepancy principle ($p = 12$) and those with $p = 14, 15$, respectively, are

$$\text{err}_{12}^{\text{TSVD}} = 4.565238 \cdot 10^{-3}, \quad \text{err}_{14}^{\text{TSVD}} = 4.629029 \cdot 10^{-3}, \quad \text{err}_{15}^{\text{TSVD}} = 4.468127 \cdot 10^{-3}.$$

The corresponding residual norms are

$$r_{11}^{\text{TSVD}} = 1.477459 \cdot 10^{-3}, \quad r_{14}^{\text{TSVD}} = 1.474911 \cdot 10^{-3}, \quad r_{15}^{\text{TSVD}} = 1.472855 \cdot 10^{-3}$$

while the noise norm is $\|\mathbf{e}\| = 1.528649 \cdot 10^{-3}$.

The SVD and the DPC as tools for the selection of the regularization parameter of CGLS. The CGLS method is an iterative regularization method with a well-known semiconvergence behavior: in order to get an acceptable solution, CGLS must be stopped before the noise enters the solution. Given the value of the noise norm $\|\mathbf{e}\|$, the discrepancy principle stops the CGLS iterations when the residual norm becomes less than $\tau\|\mathbf{e}\|$ where $\tau > 1$ [26]. Let us now consider again the test problem **phillips** of the previous section. We will show that a good regularized solution can be obtained by stopping the CGLS iterations as soon as the residual norm is less than the residual norm of a suitable TSVD solution. Let now $\mathbf{x}_k^{\text{CGLS}}$ denote the k -th CGLS iterate and let $\text{err}_k^{\text{CGLS}}$ and r_k^{CGLS} indicate the corresponding error and residual norm. The iteration index $k = 13$ is the first integer satisfying

$$r_{13}^{\text{CGLS}} \leq \|\mathbf{e}\|, \quad r_{13}^{\text{CGLS}} \leq r_{14}^{\text{TSVD}}, \quad r_{13}^{\text{CGLS}} \leq r_{15}^{\text{TSVD}}$$

where r_{14}^{TSVD} and r_{15}^{TSVD} are, respectively, the residual norm values associated to the TSVD solutions satisfying the DPC and minimizing the relative error. The optimal CGLS solution is obtained for $k = 15$. The relative error values of the regularized solutions obtained for $k = 13$ and $k = 15$ are

$$\text{err}_{13}^{\text{CGLS}} = 3.894432 \cdot 10^{-3}, \quad \text{err}_{15}^{\text{CGLS}} = 3.894081 \cdot 10^{-3}.$$

The SVD and CGLS solutions. From the previous example, based on the **phillips** test problem, we can infer that *the residual norm estimate of a suitable TSVD solution, obtained via the DPC, can be successfully used in a discrepancy-like criterion for the CGLS regularization method.* However, the need to compute the TSVD residual norm in order to obtain a stopping criterion for iterative methods may seem contradictory.

Let us now consider again the previous **phillips** test problem. Figure 2 shows, in logarithmic scale for the y-axis, the relative error histories for the TSVD and CGLS methods versus the integers from 1 to n . Observe that these integers represent the values of the TSVD and CGLS regularization parameters. From the graphs, we observe that the CGLS error decreases more rapidly; furthermore, as the integer grows, the CGLS error grows more slowly.

We have already shown that, when the SVD truncation index is chosen to minimize the relative error ($p = 15$), the TSVD solution $\mathbf{x}_{15}^{\text{TSVD}}$ and the CGLS solution $\mathbf{x}_{13}^{\text{CGLS}}$, satisfying the stopping condition $r_{13}^{\text{CGLS}} \leq r_{15}^{\text{TSVD}}$, are both good approximations to the exact solution with relative error

$$\text{err}_{15}^{\text{TSVD}} = 4.468127 \cdot 10^{-3}, \quad \text{err}_{13}^{\text{CGLS}} = 3.894432 \cdot 10^{-3}.$$

However, in practical applications, an automatic procedure for the selection of the SVD truncation index using the DPC may fail in providing the optimal

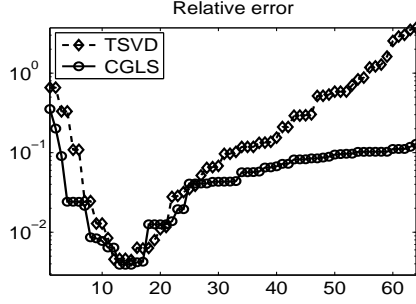


Figure 2: phillips test problem. Relative error histories for TSVD and CGLS.

value of truncation index. The following example shows that, in this case, if we choose a wrong value of p (lower or upper the best one) the CGLS solutions obtained are better than the corresponding TSVD solution, thus motivating our algorithm. Figure 3 compares the TSVD solution $\mathbf{x}_{10}^{\text{TSVD}}$ and $\mathbf{x}_{22}^{\text{TSVD}}$ with the CGLS solution $\mathbf{x}_8^{\text{CGLS}}$ and $\mathbf{x}_{23}^{\text{CGLS}}$ obtained with the stopping condition $r_k^{\text{CGLS}} \leq r_p^{\text{TSVD}}$ respectively for $p = 10$ and $p = 22$. The graphs shows that, in terms of visual evaluation, the CGLS solution are better approximations to the exact one. This qualitative evaluation is confirmed by the relative error values of the solutions:

$$\begin{aligned} \text{err}_{10}^{\text{TSVD}} &= 1.285097 \cdot 10^{-2}, & \text{err}_8^{\text{CGLS}} &= 8.629706 \cdot 10^{-3} \\ \text{err}_{22}^{\text{TSVD}} &= 2.780182 \cdot 10^{-2}, & \text{err}_{23}^{\text{CGLS}} &= 1.949018 \cdot 10^{-2}. \end{aligned}$$

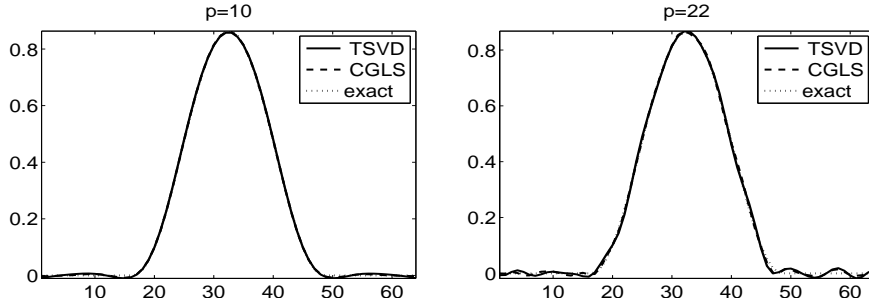


Figure 3: phillips test problem. TSVD and CGLS solutions obtained for $p = 10$ (right) and $p = 22$ (left).

These examples show that, as pointed out in [27, p. 124], we can expect the CGLS solution to be a better approximation to the exact one than the TSVD solution. This is the reason why CGLS is preferred to TSVD even when it is possible to cheaply compute the SVD decomposition.

The SVD and the DPC as tools for the selection of the regularization parameter of iterative methods. The previous example shows the well-known fact that the

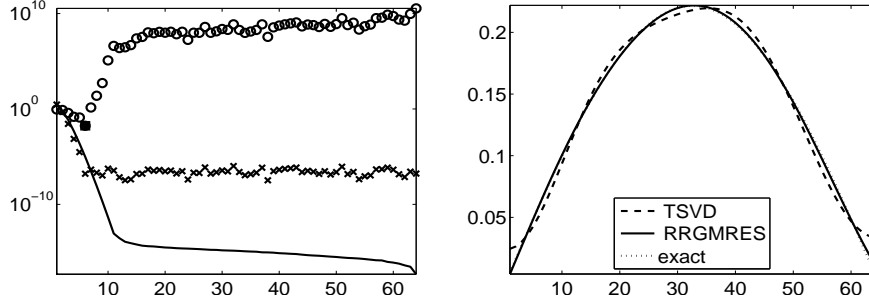


Figure 4: **baart** test problem. Left: Picard plot. The line denotes the singular values, the crosses indicate the SVD coefficients while the circles represent the solution coefficients. The square denotes the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$ for $i = 6$. Right: TSVD and RRGMRRES solutions for $p = 6$ and $k = 6$.

TSVD solution is qualitatively comparable to the CGLS solution. Let us now consider the problem **baart** [25] of size $n = m = 64$ and Gaussian noise added of level 10^{-6} .

The RRGMRRES [8] is an iterative methods whose regularizing properties are deeply analyzed in [32] where it is shown that RRGMRRES does not provide a filtered SVD solution. It is well known that the basis vectors of the Krylov subspace on which RRGMRRES is based are well suited for the problem **baart**. The visual inspection of the Picard plot (figure 4) suggests that a suitable value of the SVD truncation parameter is $p = 6$. The iteration index $k = 6$ gives the RRGMRRES solution satisfying the stopping condition $r_6^{\text{RRGMRRES}} < r_6^{\text{TSVD}}$. Figure 4 shows the solutions $\mathbf{x}_6^{\text{TSVD}}$ and $\mathbf{x}_6^{\text{RRGMRRES}}$ whose relative errors respectively are

$$\text{err}_6^{\text{TSVD}} = 4.850912 \cdot 10^{-2}, \quad \text{err}_6^{\text{RRGMRRES}} = 9.722830 \cdot 10^{-3}.$$

This example shows that the residual norm estimate of a good TSVD solution could be effectively used in a discrepancy-like criterion suited not only for CGLS but also for other iterative regularization methods.

The Fourier coefficients. The analysis of the features of the ill-posed problem by means of the solution coefficients requires the SVD of \mathbf{A} . However, for large scale problems the SVD computation is often impracticable. Let \mathcal{F} be the unitary matrix representing the Discrete Fourier Transform (DFT) such that, for a vector $\mathbf{x} \in \mathbb{R}^m$, the vector $\mathcal{F}^H \mathbf{x}$ is the DFT of \mathbf{x} and let \mathbf{f}_i denote a column of \mathcal{F} (the superscript H is used to indicate the complex conjugate transpose operation). The strict relationship between the SVD coefficients $|\mathbf{u}_i^T \mathbf{b}|$ and the so-called *Fourier coefficients* $|\mathbf{f}_i^H \mathbf{b}|$ has been deeply investigated by several authors. In [18], the authors use the stagnation of the Fourier coefficients of the data to find an appropriate truncation parameter to be used in their “modified preconditioner”. The Fourier coefficients are also employed in the regularization parameter evaluation tools developed in [50, 49] and [21]. Moreover, in [21], the overall behavior of the Fourier coefficients is shown to be similar to the

behavior of the SVD coefficients. In fact, the Fourier coefficients, as well as the SVD coefficients, first decrease and then level off at the noise level. The stagnation of the Fourier coefficients of the noisy data \mathbf{b} can be explained as follows [21, 18]. Since $\mathbf{b}^{\text{exact}}$ is the result of the action of an integral operator having a smoothing effect, its Fourier coefficients tend to have small values at high frequencies. On the other hand, the Fourier coefficients of the white noise \mathbf{e} have the same magnitude for all frequencies and, at high frequencies, they dominate the Fourier coefficients of the noisy data \mathbf{b} . This behavior of the Fourier coefficients of \mathbf{b} suggests that they could be used to individuate the SVD coefficients not satisfying the DPC. Figure 5 shows the Fourier and SVD coefficients of the `phillips` test problem.

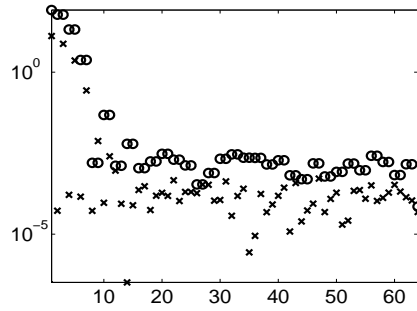


Figure 5: `phillips` test problem. Fourier coefficients (circles) and SVD coefficients (crosses).

Notation. For clarity of exposition, the following notation is used henceforth: p is a TSVD truncation parameter; k is the iteration index of a general iterative method, \mathbf{x}_k and r_k are respectively the k th iterate and the corresponding residual norm.

3. The stopping criterion

In the previous section, we have shown that the TSVD residual norm can be used in a discrepancy-like criterion for the selection of the stopping iteration of iterative methods. Moreover, we have observed that the Fourier coefficients are strictly connected to the SVD coefficients.

On the basis of these observations, we propose the following criterion for the selection of the stopping iteration of iterative regularization methods.

Definition 3.1 (Stopping Criterion). *Let $1 \leq p \leq n$ be the first index such that, for $i \geq p$, the DPC is not satisfied. Let r_p^{TSVD} denote the residual norm of the associated TSVD solution. Then, stop the iterations of the iterative regularization method at the first index k for which*

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| \leq \tau r_p^{\text{TSVD}}, \quad (5)$$

where $\tau > 1$.

As in the discrepancy principle, the parameter τ is a “safety factor”. In the sequel, we will refer to the proposed criterion as Residual-based Discrepancy-like Criterion (RDC) since it is a discrepancy-like criterion based on an estimate of the TSVD residual norm. The practical implementation of the stopping criterion (5) requires an estimate of the TSVD residual norm r_p which can be obtained from equation (2), if the left singular vectors are available, or from equation (3), if the computation of the left singular vectors is impracticable but the noise norm, or its approximated value, is known. Observe that formula (2), as well as formula (3), needs the value of the truncation index p .

In the following, we describe two algorithms computing r_p required for the practical implementation of the stopping criterion (5); the first algorithm uses the solution coefficients for the computation of the truncation index p and employs the formula (2) for obtaining the TSVD residual norm. The second algorithm uses the Fourier coefficients for the computation of p and employs the formula (3) for the TSVD residual norm. Both algorithms employ an automated choice of the truncation index p and do not require any human inspection of the Picard plot.

3.1. Residual norm computation: Algorithm 1

Truncation index computation. In the first algorithm, henceforth referred to as RDC1, the truncation index is computed as the index p such that the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$, on the average, decay for $i < p$ and start to increase for $i \geq p$. Such truncation index p is determined as follows.

Let h be a natural number such that $1 \leq h \leq n - 1$ and let $M = \lceil n/h \rceil$ be the nearest integer greater than or equal to the real number n/h . For $j = 1, \dots, M$, let ϕ_j be the sequence containing local maxima of the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$ defined as:

$$\begin{aligned} \phi_j &= \max \left\{ \frac{|\mathbf{u}_i^T \mathbf{b}|}{\sigma_i}, i = 1 + (j-1)h, \dots, jh \right\}, \quad j = 1, \dots, M-1 \\ \phi_M &= \max \left\{ \frac{|\mathbf{u}_i^T \mathbf{b}|}{\sigma_i}, i = 1 + (M-1)h, \dots, n \right\} \end{aligned} \quad (6)$$

and let α_j be the sequence of the corresponding indices in $\{1, \dots, n\}$.

Finally, let $\bar{\phi}_j$ be the sequence of the local mean values:

$$\bar{\phi}_j = \frac{\phi_{2j-1} + \phi_{2j}}{2}, \quad j = 1, \dots, \lfloor \frac{M}{2} \rfloor. \quad (7)$$

Let \bar{j} be the point where the sequence $\bar{\phi}_j$ attains its minimum value, then the truncation index p is chosen as the integer which corresponds to $2\bar{j} - 1$ in $\{1, \dots, n\}$, i.e:

$$\bar{j} = \operatorname{argmin}_j \{\bar{\phi}_j\}, \quad p = \alpha_{2\bar{j}-1}.$$

Residual norm computation. The TSVD residual norm is approximated by:

$$r_p^{\text{TSVD}} \approx \sqrt{\sum_{i=p+1}^m |\mathbf{u}_i^T \mathbf{b}|^2}. \quad (8)$$

Comments. The sequence $\bar{\phi}_j$ is basically the subsequence of the mean values of the local maxima of the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$; this subsequence is able to “capture” the overall behavior of the solution coefficients and thus it gives usable information on the index where, on average, the solution coefficients start to increase. The entire sequence of the solution coefficients can not be used in an automatic procedure for locating the value p since it is very oscillating.

As an example, figure 6 depicts the sequence of the solution coefficients $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$ (circles) and the subsequences ϕ_j (dashed line) and $\bar{\phi}_j$ (solid line) obtained for the problem **phillips** introduced in Section 2, with $h = 2$. The computed truncation parameter is $p = 15$ and is denoted in the plot by the square.

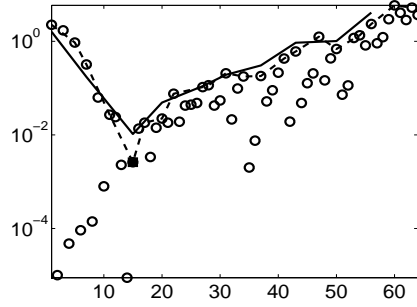


Figure 6: Solution coefficients (circles) and subsequences ϕ_j (dashed line) and $\bar{\phi}_j$ (solid line) with $h = 2$ for the **phillips** test problem. The square indicates the computed truncation parameter $p = 15$.

Finally, we observe that this algorithm requires the SVD of \mathbf{A} and therefore it is suited for linear least squares problems where the SVD is computationally inexpensive as small-size problems or image restoration problems, where \mathbf{A} is BCCB and the SVD is computed by means of FFTs [28].

3.2. Residual norm computation: Algorithm 2

Truncation index computation. In the second algorithm, henceforth referred to as RDC2, the truncation index is computed as the index p such that the Fourier coefficients $|\mathbf{f}_i^H \mathbf{b}|$ start to level off for $i \geq p$. Such truncation index p is determined as follows.

Since \mathbf{b} has real values, from the conjugate symmetry of the DFT it follows that

$$\mathbf{f}_{i+1}^H \mathbf{b} = \text{conj}(\mathbf{f}_{m-i+1}^H \mathbf{b}), \quad i = 1, \dots, q-1,$$

where $q = \lfloor m/2 \rfloor + 1$, and thus

$$|\mathbf{f}_{i+1}^H \mathbf{b}| = |\mathbf{f}_{m-i+1}^H \mathbf{b}|, \quad i = 1, \dots, q-1.$$

Therefore, it is sufficient to consider only the first q Fourier coefficients. Let ψ_j be the sequence of the Fourier coefficients $|\mathbf{f}_i^H \mathbf{b}|$, $i = 1, \dots, q$, sorted into non increasing order. Let \bar{j} be the point where the sequence ψ_j , $i = 1, \dots, q$, begins to level off, i.e: \bar{j} is the first index such that

$$\frac{|\psi_{\bar{j}+1} - \psi_{\bar{j}}|}{|\psi_2 - \psi_1|} < \text{tol}_\psi$$

where tol_ψ is a positive tolerance. Then the truncation index p is chosen as

$$p = 2\bar{j} - 1. \quad (9)$$

Residual norm computation. Given an estimate ϵ of the noise norm $\|\mathbf{e}\|$, the TSVD residual norm is computed as

$$r_p = \sqrt{\frac{m-p}{m}} \epsilon. \quad (10)$$

Comments. This algorithm needs an estimate ϵ of the noise norm $\|\mathbf{e}\|$. Some strategies have been recently proposed in the literature for the computation of a noise norm approximation [21, 50, 29, 30]. Among others, the Normalized Cumulative Periodogram (NCP) method [21] chooses the stopping iteration of an iterative method so that the residual is white noise-like; the corresponding residual norm could be used as a fair value of ϵ . Moreover, in some practical applications, a measurement error bound can be calculated from the measurement instruments and used as value of ϵ .

As an example, figure 7 shows the sequence of the Fourier coefficients $|\mathbf{f}_i^H \mathbf{b}|$ (circles) and the sequence ψ_j (line) obtained for the test problem **phillips** of section 2; the truncation parameter, computed for $\text{tol}_\psi = 10^{-5}$, is $p = 11$ and is denoted by the square.

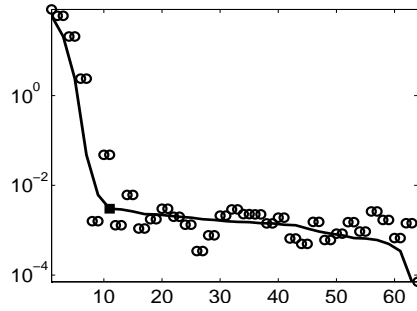


Figure 7: Fourier coefficients (circles) and sequence ψ_j (line) for the **phillips** test problem. The square indicates the computed truncation parameter $p = 11$ ($\text{tol}_\psi = 10^{-5}$).

Problem	cond	Problem	cond
1) phillips	$2.4 \cdot 10^9$	2) baart	$5.2 \cdot 10^{18}$
3) shaw	$2.4 \cdot 10^{21}$	4) foxgood	$3.2 \cdot 10^{20}$
5) heat(1)	$7.7 \cdot 10^{270}$	6) deriv2(2)	$3.7 \cdot 10^{05}$
7) gravity(1)	$1.4 \cdot 10^{20}$	8) spikes(1)	$2.2 \cdot 10^{23}$
9) wing	$2.8 \cdot 10^{22}$	10) hilbert	$1.0 \cdot 10^{21}$
11) lotkin	$4.5 \cdot 10^{21}$	12) moler	$2.2 \cdot 10^{305}$
13) random	$2.1 \cdot 10^{05}$	14) prolate	$3.9 \cdot 10^{20}$

Table 1: Test problems and condition numbers for $n = 500$, obtained with the `condtest` Matlab function.

Finally, we observe that this algorithm does not require the SVD of \mathbf{A} and therefore it can be used for large-scale problems where the SVD computation is impracticable.

4. Numerical results

In this section we present some numerical results obtained with the RDC1 and RDC2 proposed in section 3.

The aim is to evaluate, on image deblurring problems, the proposed criteria for stopping an iterative regularization method. As iterative method, we mainly consider the CGLS method since it is the most widely used method for linear iterative regularization and it has been proved to be efficient in image deblurring [13]. Moreover, we present a few tests on the Landweber and the RRGMR methods. Finally, we compare the results obtained with RDC1 and RDC2 with other existing criteria for stopping iterative regularization methods; the results are evaluated in terms of relative error between the exact solution \mathbf{x} and the computed solution $\tilde{\mathbf{x}}$:

$$\text{err} = \frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|}{\|\mathbf{x}\|}.$$

We remark that RDC1 is completely heuristic, in the sense that it needs only the data vector \mathbf{b} , while RDC2 also requires in input the noise norm (or its estimate). In particular, in all the presented tests, we use the exact value of the noise norm when implementing RDC2.

The tests have been performed in Matlab 2015A, on a Windows 7 platform.

4.1. One-dimensional test problems

The aim of this experiment is to test the performance of RDC1 and RDC2 on some linear discrete ill-posed problems from Hansen's regularization Tools [25] and from the `gallery` function of Matlab, for different noise level values. The considered test problems, whose size is $n = m = 500$ and $n = m = 1000$, are listed in table 1 with their condition number (estimated with the `condtest` Matlab function) for $n = 500$. The matrix \mathbf{A} of the test problem

“random” is obtained by the `random` Matlab function. For each problem from [25] the exact solution \mathbf{x} of the noise-free linear system $\mathbf{Ax} = \mathbf{b}^{\text{exact}}$ is given. For the test problems from the Matlab `gallery` function, the exact solution of the test problem `shaw` from [25] is used. The noisy right-hand side \mathbf{b} is obtained by adding a Gaussian white noise vector \mathbf{e} to $\mathbf{b}^{\text{exact}}$, with noise level $\delta = 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$. For each value of δ , 10 realizations of \mathbf{e} are generated by using the Matlab `randn` function. This results in 1800 test problems.

Parameters setting. For this experiment, the parameters involved in the proposed algorithms have been set to the following fixed values:

- $\tau = 1 + 10\varsigma$ in (5) and in the DP criterion (where ς is the machine precision);
- $h = 2$ in (6), if the matrix \mathbf{A} is symmetric, $h = 1$ otherwise;
- $tol_\psi = 0.99$ in (9).

For each test problem, we compute the percentage of error values satisfying:

$$\frac{\text{err}_{\text{meth}}}{\text{err}_{\text{best}}} > \eta \quad (11)$$

where err_{meth} is the relative error obtained with the considered stopping method and err_{best} is the relative error obtained stopping CGLS at best.

Table 2 compares the results obtained with $\eta = 5, 10, 100$ by RDC1 and RDC2 with those provided by the Discrepancy Principle (DP), the L-curve (LC) criterion [24, 23], the residual L-curve (RLC) criterion [47, 46], the Hanke-Raus (HR) rule [19] and the Rule HRmC [15]. In the L-curve and in the residual L-curve methods, the corner of the L-curve has been computed with the `lcorner` Matlab function of the Regularization Tools package [25]. Table 2 also shows the results for DP and RDC2 obtained with $\tau = 1.3$ in order to simulate the case of overestimated noise norm. In order to make the results of RDC1 and RDC2 comparable with those reported in [15] for the Hanke-Raus rule and for Rule HRmC, table 3 reports the averages of error ratios $\text{err}_{\text{meth}}/\text{err}_{\text{best}}$ for $n = m = 100$ over 10 realizations of noise at noise norm values $\|\mathbf{e}\| = 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$. (We remind that the noise norm $\|\mathbf{e}\|$ and the noise level δ are proportional since $\|\mathbf{e}\| = \delta\|\mathbf{b}^{\text{exact}}\|$.) The numerical results in the tables show that, on average, RDC1, RLC and HRmC perform quite comparably. It is worth mentioning that these three rules are all heuristic rules but RDC1 needs the SVD of \mathbf{A} while RLC and HRmC require to perform many CGLS iterations to a posteriori select the stopping iteration. Therefore, RDC1 may be computationally preferable when the SVD of \mathbf{A} can be computed using inexpensive techniques such as in image deblurring applications, for example. For the one-dimensional test problems, DP performs very well and outperforms RDC2. Indeed, in authors’ experience, RDC2 tends to perform favourably compared to DP in large scale application when the value of n is very large.

Criterion	$\eta = 5$	$\eta = 10$	$\eta = 100$
DP	1%	0%	0%
DP ($\tau = 1.3$)	4%	0%	0%
LC	24%	18%	7%
RLC	6%	1%	0%
RDC1	9%	5%	0%
RDC2	13%	10%	7%
RDC2 ($\tau = 1.3$)	3%	0%	0%
HR	23%	11%	1%
HRmC	11%	9%	4%

Table 2: Comparison between different rules on CGLS for different values of η in (11).

Probl.	DP	LC	RLC	RDC1	RDC2	HR	HRmC
1	1.06	39.95	1.84	1.51	1.06	8.96	2.76
2	1.12	1.29	1.43	1.61	107.37	2.72	1.28
3	1.26	2.12	1.88	2.30	1.26	3.09	2.05
4	1.44	45.80	3.04	16.17	45.04	9.03	4.40
5	1.19	5.25	2.15	1.25	1.17	2.42	2.43
6	1.13	1.32	120.73	1.19	1.60	2.39	1.71
7	1.46	33.38	3.55	1.90	1.38	4.92	2.85
8	1.00	1.00	1.00	1.04	1.00	1.03	1.50
9	1.10	1.06	82.59	1.32	439199.39	1.31	1.19
10	1.09	1.07	1.18	1.41	7.23	1.29	1.27
11	1.12	1.08	1.21	1.46	1.20	1.48	1.44
12	1.00	1.00	1.42	1.00	1.00	1.00	1.00
13	2.44	1.53	572.91	81743.69	2.36	1.53	9342.25
14	1.85	519.02	5.89	4.64	1.78	3.11	3.21

Table 3: Averages of error ratios for CGLS ($n = 100$).

4.2. Image restoration test problems

In image restoration, the least squares problem (1) is obtained as the discretization of a Fredholm integral equation of the first kind on the image domain $\Omega \subset \mathbb{R}^2$:

$$b(v) = \int_{\Omega} A(u, v)x(u)du \quad (12)$$

where the matrix \mathbf{A} is the discretization of the Point Spread Function (PSF) $A(u, v)$, \mathbf{b} is the discretization of the data function $b(v)$ and represents the recorded image and \mathbf{x} is the discretization of the object function $x(u)$ and represents the exact image to be recovered. In the case of image deblurring, the vectors \mathbf{x} and \mathbf{b} are obtained by lexicographically ordering the $n_x \times n_y$ ($n_x = n_y$) image arrays and have the same size n . We consider, in particular, convolution PSF operators with periodic boundary conditions, where the matrix \mathbf{A} has a

PSF	δ	best	DP	RDC1	RDC2
g	0.001	$8.73 \cdot 10^{-2}(178)$	$9.26 \cdot 10^{-2}(67)$	$8.95 \cdot 10^{-2}(103)$	$9.07 \cdot 10^{-2}(86)$
	0.005	$9.96 \cdot 10^{-2}(45)$	$1.05 \cdot 10^{-1}(20)$	$1.00 \cdot 10^{-1}(33)$	$1.03 \cdot 10^{-1}(24)$
d	0.001	$1.19 \cdot 10^{-1}(64)$	$1.22 \cdot 10^{-1}(53)$	$1.20 \cdot 10^{-1}(59)$	$1.22 \cdot 10^{-1}(53)$
	0.005	$1.60 \cdot 10^{-1}(28)$	$1.63 \cdot 10^{-1}(24)$	$1.61 \cdot 10^{-1}(27)$	$1.63 \cdot 10^{-1}(24)$
m	0.001	$4.08 \cdot 10^{-2}(107)$	$4.99 \cdot 10^{-2}(41)$	$6.16 \cdot 10^{-2}(26)$	$4.92 \cdot 10^{-2}(42)$
	0.005	$7.75 \cdot 10^{-2}(24)$	$8.44 \cdot 10^{-2}(15)$	$7.76 \cdot 10^{-2}(23)$	$8.26 \cdot 10^{-2}(16)$

Table 4: Results on the CGLS method.

Block Toeplitz with Toeplitz Blocks (BCCB) structure [28]. In this case, since its spectral decomposition can be obtained by means of Fourier transforms, the SVD decomposition is computationally performed by FFTs and it is not expensive even for large size problems. We choose three images for deblurring experiments: the *cameraman*, the *Lena* and the *satellite* images that are widely used in literature for imaging deblurring. We test a Gaussian (g), a defocus (d) and a motion (m) PSF [28].

In the test problems, the noisy blurred image is obtained by adding a Gaussian noise vector to the blurred version of the exact image.

Parameters setting. For all the image restoration test problems, the parameters involved in the proposed algorithms have been set to the following fixed values:

- $\tau = 1 + 10\varsigma$ in (5) and in the DP criterion (where ς is the machine precision);
- $h = 2 * n_x$ for Gaussian PSF and $h = 3 * n_x$ for motion and defocus PSFs (in (6));
- $tol_\psi = 10^{-10}$ in (9).

4.2.1. Results on CGLS

We start by presenting the results obtained on the CGLS method. In this case we use two different levels of noise, $\delta_1 = 10^{-3}$ and $\delta_2 = 5 \cdot 10^{-3}$. Table 4 shows the relative errors and the number of performed iterations (in brackets) for the three PSFs (g,d,m) combined with the two noise levels δ_1 and δ_2 , (on a single noise realization obtained with the `randn` Matlab function with seed 0) with the *cameraman* image. With **best** we intend, in all this section, the best result, in terms of relative error, that we can obtain with the considered method. In the table we compare RDC1 and RDC2 with the DP. The numerical results show that, on average, our algorithms performs comparably to the DP.

Then we tested the criteria on a larger set of 180 executions obtained from the three images, blurred with the three considered PSFs and noise of levels δ_1 and δ_2 , with 10 realizations of noise for each noise level.

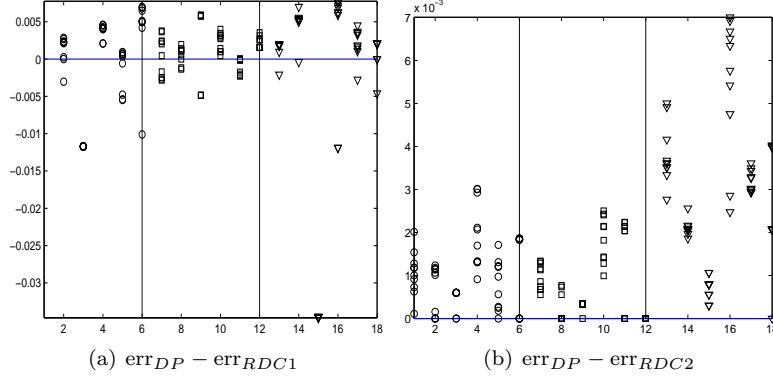


Figure 8: Graphical representation of the difference between the DP, RDC1 and RDC2 errors.

In figure 8 we plot the values of $\text{err}_{DP} - \text{err}_{RDC1}$ (figure (a)) and $\text{err}_{DP} - \text{err}_{RDC2}$ (figure (b)) where err_{DP} , err_{RDC1} and err_{RDC2} are the relative errors obtained by stopping CGLS with the DP, the RDC1 and the RDC2 criteria. In the plot, the circles are related to the *Lena*, the squares to the *cameraman* and the triangles to the *satellite* images. In figure 8(a) there are 122 nonnegative values with mean equal to 0.0031 and 78 negative values with mean equal to -0.01 . We note that there are only few executions where the RDC1 error is consistently greater than the DP error. From figure 8(b), we see that in the case of RDC2 the errors are all positive with the mean equal to 0.0019; hence, RDC2 is never overcome by DP.

In figure 9 we report the values of $k_{best} - k_{RDC1}$ (figure (a)), $k_{best} - k_{RDC2}$ (figure (b)), $k_{best} - k_{DP}$ (figure (c)), where k is the number of performed iterations. We remark that, even if RDC2 always performs more iterations than DP, it never goes further the best iteration. Moreover, we observe that test number 15 is critical for all the three methods, since they stop far before the best iteration.

Finally, we compare the RDC1 and RDC2 algorithms with other well known heuristic stopping criteria. Basing on the results presented in [46], where a wide experimentation has been performed to compare different stopping rules on iterative Krylov regularization methods, we consider the following criteria (in addition to the DP) to compare with RDC1 and RDC2 on CGLS: the L-curve, the residual L-curve and the NCP criterium [21]. For all the considered algorithms, we used, when available, the Matlab functions from the authors.

We experimented the criteria on the 180 tests as before and we report in table 5 the percentage of results satisfying (11). We reported in the table the results for the values $\eta = 1.05$ and $\eta = 1.01$. The RDC1 criterium is the best for $\eta = 1.05$ and it is second only to NCP in the case of $\eta = 1.01$.

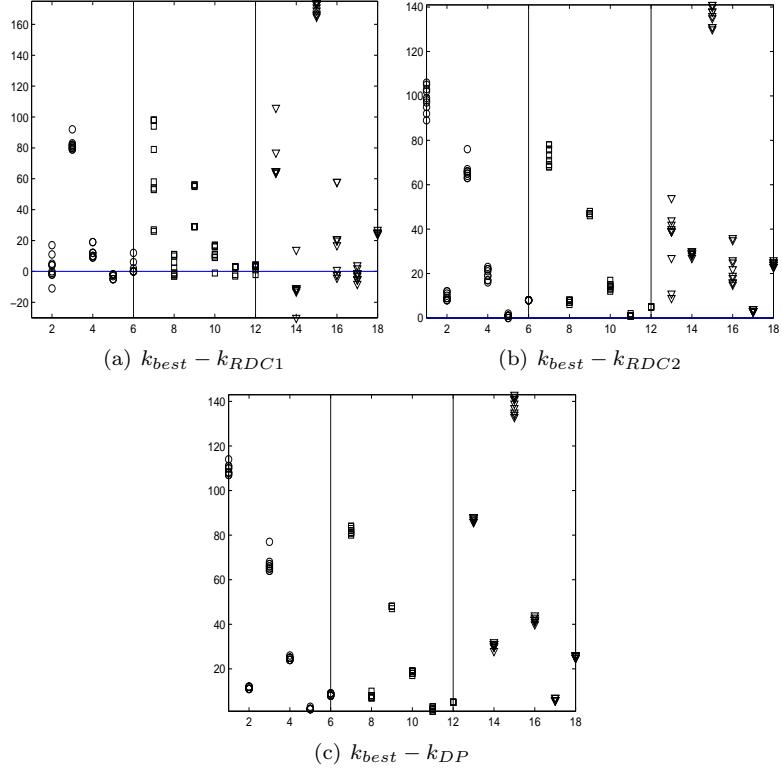


Figure 9: Graphical representation of the difference between the **best** stopping index and the stopping iteration computed by RDC1 (a), RDC2 (b) and Discrepancy (c).

4.2.2. Results on iterative methods different from CGLS

Here we apply the RDC1 and RDC2 criteria to other iterative methods than CGLS. We first consider the Landweber method, whose iterates have an expression similar to a TSVD solution, presenting a semiconvergence behavior [27]. Then we test the criteria on the RRGMRRES iterative method, that is preferred to the standard GMRES in imaging applications, because it produces more smoothed and less noisy restored images [32]. Its iterates do not provide a filtered SVD solution, hence they behave differently from CGLS and Landweber iterates [27].

In the tests, we blurred the *cameraman* image with the Gaussian and motion PSFs; since Landweber and RRGMRRES have different convergence behaviors, in order to have semiconvergence for both methods we use different noise levels: 0.05 for Landweber and 0.005 for RRGMRRES. In table 6 we report the errors obtained with one noise realization (with the Matlab function `randn` and `seed=0`) and the correspondent number of iterations. We compare RDC1 and RDC2 rules with the discrepancy (DP) and with a modified discrepancy (MDP) proposed in [15], where the stopping iteration index is computed as the nearest

Method	$\eta = 1.05$	$\eta = 1.01$
DP	52.22%	93.90%
LC	73.33%	87.22%
RLC	61.66%	87.77%
NCP	37.22%	66.11%
RDC1	28.33%	66.67%
RDC2	37.77%	77.22%

Table 5: Comparison between different stopping methods on CGLS; percentage of tests satisfying (11)

Method	Criterium	g	m
RRGMRES	best	$9.94 \cdot 10^{-2}(14)$	$7.75 \cdot 10^{-2}(21)$
	DP	$1.02 \cdot 10^{-1}(10)$	$8.21 \cdot 10^{-2}(14)$
	RDC1	$9.94 \cdot 10^{-2}(14)$	$7.83 \cdot 10^{-2}(18)$
	RDC2	$1.01 \cdot 10^{-2}(15)$	$8.21 \cdot 10^{-2}(14)$
Landweber	best	$1.26 \cdot 10^{-1}(29)$	$1.42 \cdot 10^{-1}(16)$
	DP	$1.31 \cdot 10^{-1}(10)$	$1.48 \cdot 10^{-1}(7)$
	MDP	$1.26 \cdot 10^{-1}(23)$	$1.42 \cdot 10^{-1}(16)$
	RDC1	$1.29 \cdot 10^{-1}(12)$	$1.42 \cdot 10^{-1}(13)$
	RDC2	$1.28 \cdot 10^{-1}(13)$	$1.45 \cdot 10^{-1}(9)$

Table 6: Errors on the RRGMRRES (noise level $\delta = 0.005$) and the Landweber method (noise level $\delta = 0.05$) on the *cameraman* image blurred with Gaussian (g) and motion (m) PSFs.

Method	RRGMRES $\eta = 1.07$	RRGMRES $\eta = 1.05$	Landweber $\eta = 1.07$	Landweber $\eta = 1.05$
DP	33.3%	100%	15%	88.3%
MDP	/	/	0%	0%
LC	95%	100%	40%	46.6%
RDC1	26.6%	36.6%	13.3%	21.6%
RDC2	0%	8.3%	0%	11.6%

Table 7: Comparison between different stopping criteria on RRGMRRES and Landweber methods. Percentage of executions satisfying (11).

integer to $2.3 \cdot kD$, where kD is the discrepancy iteration.

Finally, we performed tests with the three images used before blurred by Gaussian and motion PSFs with 10 noise realizations for each case. Table 7 reports the percentage of tests satisfying (11) for different values of η . It is evident that both the proposed criteria give very good results; RDC2 better behaves when applied to the RRGMRRES method, while RDC1 when applied to the Landweber method. The MDP rule works very well; we remind that it needs an error estimate in order to apply the discrepancy principle.

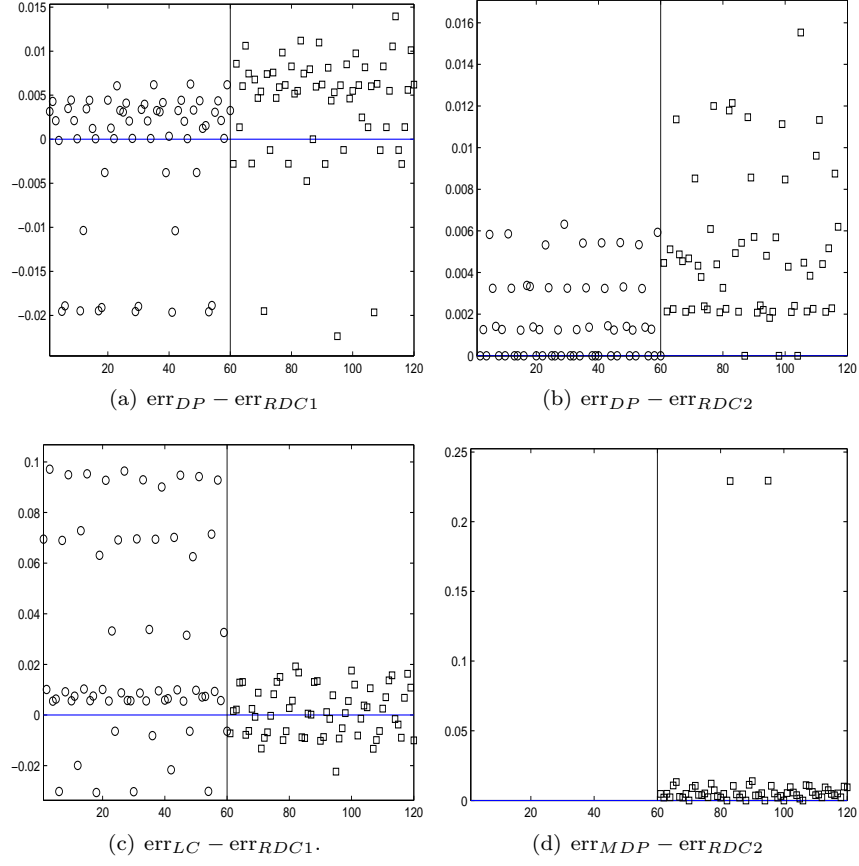


Figure 10: Errors obtained on the iterative methods RRGMRES and Landweber (only Landweber method in figure (d))

Figures 10(a), 10(b), 10(c) and 10(d) report the values: $\text{err}_{DP} - \text{err}_{RDC1}$, $\text{err}_{DP} - \text{err}_{RDC2}$, $\text{err}_{LC} - \text{err}_{RDC1}$ and $\text{err}_{MDP} - \text{err}_{RDC2}$, respectively. Since MDP is a good rule only for Landweber method, only Landweber results are considered in figure 10(d). In each plot the circles are relative to the RRGMRES method and the squares to the Landweber method. The results are very similar to those obtained on CGLS: RDC2 always performs better or equal to DP (figure 10(b)), while if we compare the totally heuristic criteria RDC1 and L-curve, we can see that the proposed RDC1 gives better results in most cases (84 positive values over 120) (figure 10(c)).

From the results we can conclude that the proposed RDC1 and RDC2 criteria work well even when applied to the Landweber and the RRGMRES methods.

4.3. Comments

At the end of this section, some comments about the behavior of RDC1 and RDC2 are worth.

RDC1 is an heuristic criterion since it only requires the coefficient matrix \mathbf{A} and the noisy data vector \mathbf{b} . It performs favorably, in terms of relative error reduction, compared with other well-known heuristic criteria (see table 2 and 5, for the CGLS method, and table 7, for the RRGMRRES and Landweber methods). Surprisingly, on many image deblurring test problems, it performs comparably to the Discrepancy Principle which is indeed an *a-posteriori* criterium (see figure 8). The numerical results show that, in some test problems, the RDC1 solution is under-regularized (see figure 9) since a too large value of the SVD truncation index is found by Algorithm 1. However, we have verified that, when the optimal SVD truncation parameter is used in (8), the corresponding solution is a good approximation to the exact one. This means that the failure of RDC1 is due to the failure of Algorithm 1 in computing a suitable value of the SVD truncation index.

RDC2 is an *a-posteriori* criterium since it needs an estimate of the noise norm. As the discrepancy principle it is sensitive to the quality of this estimate and often gives over-regularized solutions (see figure 9); however, the numerical results indicate that it can be competitive with the discrepancy principle since the value of the truncation index computed by Algorithm 2 allows the iterative method to perform a number of iterations larger than the DP stopping iteration but smaller than the optimal one.

Finally, we remark that our results indicate that the computation of a suitable value of the TSVD index is the critical issue in order to obtain usable regularized solutions. To this end, we have described two algorithms but we intend to investigate the use of the methods recently proposed in [50, 49, 48, 54, 30] to compute a suitable value of the SVD truncation index and improve the efficiency and effectiveness of the proposed stopping criterium (5).

5. Conclusions

In this paper we have proposed a stopping criterion for iterative regularization methods in the solution of linear discrete ill-posed problems. The proposed criterion terminates the iterative procedure when the current residual norm becomes less than the residual norm of a suitable TSVD solution.

Two algorithms have been described for the practical implementation of the proposed criterion. The first algorithm can be used when the SVD of the coefficient matrix is easily computable. This is the case, for example, of one-dimensional problems that usually have a not so large size and of BCCB matrices arising in image deblurring with periodic boundary conditions on the PSF. The second algorithm, requiring an estimate of the noise norm, can be employed when the SVD of the coefficient matrix is impracticable since it uses the Fourier coefficients.

The proposed algorithms have been tested on many different one-dimensional test problems as well as on large size problems in image deblurring. Many experiments, mainly with the CGLS regularization method, show that the two algorithms often outperform the widely used discrepancy principle and are competitive with the most efficient criteria reported in literature, with comparable computational cost. In particular, the second algorithm tends to perform better than the first one since it uses a noise norm estimate.

Even if the proposed criterion is applied here to iterative regularization, it is indeed quite general and can be extended to direct regularization formulated as a noise-constrained optimization problem. In this case in fact, the objective is a regularization function (Tikhonov or Total Variation regularization, for example) and the constraint, imposing data fidelity, compares the residual norm of the solution with the residual norm of a suitable regularized solution.

References

- [1] A. B. Bakushinski, Remarks on the choice of regularization parameter from quasi-optimality and relation tests, *USSR Comput. Math. Math. Phys.* 24 (4) (1984) 181–182.
- [2] F. Bauer, S. Kindermann, Recent results on the quasi-optimality principle., *J. Inverse Ill-Posed Probl.* 17 (2009) 5–18.
- [3] F. Bauer, M. Lukas, Comparing parameter choice methods for regularization of ill-posed problems, *Mathematics and Computers in Simulation* 81 (9) (2011) 1795–1841.
- [4] F. Benvenuto, R. Zanella, L. Zanni, M. Bertero, Nonnegative least-squares image deblurring: improved gradient projection approaches, *Inv. Probl.* 26 (2) (2010) 025004.
- [5] M. Bertero, P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing, Bristol, 1998.
- [6] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for linear systems with applications to regularization, *Numer. Algorithms* 49 (2008) 85–104.
- [7] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for the regularization of least squares problems, *Numer. Algorithms* 51 (2009) 61–76.
- [8] D. Calvetti, B. Lewis, L. Reichel, GMRES-type methods for inconsistent systems, *Lin. Alg. Appl.* 316 (2000) 157–169.
- [9] D. Calvetti, B. Lewis, L. Reichel, GMRES, L-curves, and discrete ill-posed problems, *BIT* 42 (2002) 44–65.
- [10] D. Calvetti, B. Lewis, L. Reichel, On the regularizing properties of the GMRES method, *Numer. Math.* 91 (2002) 605–625.

- [11] M. E. Daube-Witherspoon, G. Muehllehner, An iterative image space reconstruction algorithm suitable for volume ct, *IEEE Trans. Med. Imaging* 5 (1986) 61–66.
- [12] B. Eicke, Iteration methods for convexly constrained ill-posed problems in Hilbert space, *Num. Funct. Anal. Optim.* 13 (1992) 413–429.
- [13] C. E. F. Di Benedetto, S. S. Capizzano, Superoptimal preconditioned conjugate gradient iteration for image deblurring, *SIAM Sci. Comp.* 26 (3) (2005) 1012–1035.
- [14] G. Golub, M. Heath, G. Wahba, Generalized Cross-Validation as a method for choosing a good ridge parameter, *Technometrics* 21 (1979) 215–223.
- [15] U. Hämarik, R. Palm, T. Raus, On minimization strategies for choice of the regularization parameter in ill-posed problems, *Numer. Funct. Anal. Optim.* 30 (9-10) (2009) 924–950.
- [16] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Pitman Research Notes in Mathematics, Longman House, Harlow, Essex, 1995.
- [17] M. Hanke, Limitations of the L-curve method in ill-posed problems, *BIT* 36 (2) (1996) 287–301.
- [18] M. Hanke, J. Nagy, R. Plemmons, Preconditioned iterative regularization for ill-posed problems, chap. *Numerical Linear Algebra*, De Gruyter (Walter), Berlin, 1993, pp. 141–163.
- [19] M. Hanke, T. Raus, A general heuristic for choosing the regularization parameter in ill-posed problems, *SIAM J. Sci. Computing* 17 (4) (1996) 956–972.
- [20] P. Hansen, The discrete Picard condition for discrete ill-posed problems, *BIT* 30 (4) (1990) 658–672.
- [21] P. Hansen, M. E. Kilmer, R. H. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, *BIT* 46 (2006) 41–59.
- [22] P. Hansen, T. Koldborg, Noise propagation in regularizing iterations for image deblurring, *ETNA* 31 (2008) 204–220.
- [23] P. Hansen, D. O’Leary, The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.* 14 (1993) 1487–1503.
- [24] P. C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Rev.* 34 (4) (1992) 561–580.
- [25] P. C. Hansen, Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems, *Numer. Alg.* 6 (1994) 1–35.

- [26] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, 1998.
- [27] P. C. Hansen, Discrete Inverse Problems: Insight and Algorithms, SIAM, 2010.
- [28] P. C. Hansen, J. Nagy, D. P. O’Leary, Deblurring images. Matrices, Spectra and Filtering, SIAM, 2006.
- [29] I. Hnětynková, M. Plešinger, Z. Strakoš, The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data, BIT 49 (2009) 669–696.
- [30] M. Hochstenbach, L. Reichel, G. Rodriguez, Regularization parameter determination for discrete ill-posed problems, JCAM 273 (2015) 132 – 149.
- [31] T. Huckle, M. Sedlacek, Data based regularization for discrete deconvolution problems, BIT 53 (2013) 459–473.
- [32] T. K. Jensen, P. C. Hansen, Iterative regularization with minimum-residual methods, BIT 43 (1) (2007) 103–120.
- [33] R. E. Jones, Automatically regularized nonnegative solutions for ill-conditioned linear systems, IPES2006.
- [34] M. E. Kilmer, D. P. O’Leary, Choosing regularization parameters in iterative methods for ill-posed problems, SIAM J. Matrix Anal. Appl. 22 (2001) 1204–1221.
- [35] M. E. Kilmer, G. W. Stewart, Iterative regularization and MINRES, SIAM J. Matrix Anal. Appl. 21 (1999) 613–628.
- [36] S. Kindermann, On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization, Inverse Probl. Imaging 2 (2008) 291–299.
- [37] S. Kindermann, Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems, ETNA 38 (2011) 233–257.
- [38] G. Landi, E. L. Piccolomini, A projected Newton-CG method for nonnegative astronomical image deblurring, Num. Alg. 48 (4) (2008) 279–300.
- [39] G. Landi, E. L. Piccolomini, Quasi-Newton projection methods and the discrepancy principle in image restoration, Appl. Math. Comp. 218 (2011) 2091–2107.
- [40] C. L. Lawson, R. J. Hanson, Solving least squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [41] V. A. Morozov, Methods for solving incorrectly posed problems, Springer-Verlag, New York, 1984.

- [42] A. Neubauer, The convergence of a new heuristic parameter selection criterion for general regularization methods, *Inverse Problems* 24 (2008) 055005 (10 pages).
- [43] C. C. Paige, M. A. Saunders, Solution of sparse indefinite systems of linear equations, *SIAM J. Numer. Anal.* 12 (1975) 617–629.
- [44] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Mach.* 9 (1962) 84–97.
- [45] T. Regińska, A regularization parameter in discrete ill-posed problems, *IAMJ. Sci. Comput.* 17 (1996) 740–749.
- [46] L. Reichel, G. Rodriguez, Old and new parameter choice rules for discrete ill-posed problems, *Num. Alg.* 63 (1) (2013) 65–87.
- [47] L. Reichel, H. Sadok, A new L-curve for ill-posed problems, *J. Comput. Appl. Math.* 219 (2) (2008) 493–508.
- [48] B. W. Rust, Truncating the singular value decomposition for ill-posed problems, Tech. rep., Tech. Rep. NISTIR 6131 (1998).
- [49] B. W. Rust, Parameter selection for constrained solutions to ill-posed problems, *Comput. Sci. Stat.* 32 (2000) 333–347.
- [50] B. W. Rust, D. P. O’Leary, Residual periodograms for choosing regularization parameters for ill-posed problems, *Inv. Prob.* 24 (2008) 034005 (30 pages).
- [51] A. N. Tikhonov, V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley, New York, 1977.
- [52] A. N. Tikhonov, V. Glasko, Y. Kriksin, On the question of quasi-optimal choice of a regularized approximation, *Soviet Mathematics Doklady* 20 (5) (1979) 1036–1040.
- [53] C. R. Vogel, Non-convergence of the L-curve regularization parameter selection method, *Inv. Prob.* 12 (1996) 535–547.
- [54] F. Zama, Computation of regularization parameters using the Fourier coefficients, *Almae Matris Studiorum Acta*, http://amsacta.cib.unibo.it/2574/1/FZ_reg.pdf (2009).