ITERATIVE METHODS FOR ILL-POSED PROBLEMS AND SEMICONVERGENT SEQUENCES

S. MORIGI *, L. REICHEL †, F. SGALLARI ‡, AND F. ZAMA §

Abstract. Iterative schemes, such as LSQR and RRGMRES, are among the most efficient methods for the solution of large-scale ill-posed problems. The iterates generated by these methods form semiconvergent sequences. A meaningful approximation of the desired solution of an ill-posed problem often can be obtained by choosing a suitable member of this sequence. However, it is not always a simple matter to decide which member to choose. Semiconvergent sequences also arise when approximating integrals by asymptotic expansions, and considerable experience and analysis of how to choose a suitable member of a semiconvergent sequence in this context are available. The present note explores how the guidelines developed within the context of asymptotic expansions can be applied to iterative methods for ill-posed problems.

Key words. Ill-posed problem, iterative method, stopping criterion, L-curve.

1. Introduction.

Many problems in science and engineering require the determination of the unknown input of a linear system from known, but generally corrupt, output. Such problems are referred to as inverse problems and are often modeled by Fredholm integral equations of the first kind with a smooth kernel. The task of solving these integral equations is an ill-posed problem, because the solution might not exist, not be unique, or not depend continuously on the data (the right-hand side). The computation of meaningful approximate solutions of inverse problems therefore can be quite challenging.

The discretization of a Fredholm integral equation of the first kind with a smooth kernel gives rise to a linear system of equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{m\times n}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \qquad \boldsymbol{b} \in \mathbb{R}^m, \tag{1.1}$$

with a matrix A of ill-determined rank. In particular, A is severely ill-conditioned and may be singular. The right-hand side vector $\boldsymbol{b}$ in (1.1) represents the available output of the linear system, and is assumed to be contaminated by an error $\boldsymbol{e} \in \mathbb{R}^m$ which, for instance, may stem from measurement errors of the output or from discretization errors. Due to the contamination $\boldsymbol{e}$, the system (1.1) might not be consistent. We will for ease of discussion assume that $m = n$; however, our methods apply, mutatis mutandis, also when $m \neq n$.

Let $\hat{\boldsymbol{b}}$ denote the unknown error-free output of the linear system and let $\hat{\boldsymbol{x}}$ denote the desired solution of the (unavailable) error-free consistent linear system of equations

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}. \tag{1.2}$$

We would like to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of the available linear system (1.1).

*Dipartimento di Matematica, Universitá degli Studi di Bologna, Piazza Porta S. Donato 5, 40127 Bologna, Italy. E-mail: morigi@dm.unibo.it.

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-0107858.

‡Dipartimento di Matematica, Universitá degli Studi di Bologna, Piazza Porta S. Donato 5, 40127 Bologna, Italy. E-mail: sgallari@dm.unibo.it.

§Dipartimento di Matematica, Universitá degli Studi di Bologna, Piazza Porta S. Donato 5, 40127 Bologna, Italy. E-mail: zama@dm.unibo.it.

Straightforward solution of (1.1) typically does not yield a meaningful approximation of $\hat{\boldsymbol{x}}$ due to the error

$$\boldsymbol{e} = \hat{\boldsymbol{b}} - \boldsymbol{b} \tag{1.3}$$

in the right-hand side $\boldsymbol{b}$ and the ill-conditioning of the matrix A. A popular and quite efficient approach to computing an approximation of $\hat{\boldsymbol{x}}$ is to apply a few steps of an iterative method to (1.1). Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$ denote the sequence of iterates generated by a particular iterative method, such as LSQR or RRGMRES.

Let $\|\cdot\|$ denote the Euclidean vector norm and consider the sequence $\eta_k = \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$, $k = 1, 2, 3 \ldots$. Generally, the η_k decrease when k increases and k is fairly small, but, due to the error $\boldsymbol{e}$ and the ill-conditioning of A, the η_k typically increase rapidly with k when k is large. A sequence $\eta_1, \eta_2, \eta_3, \ldots$ with these properties is said to be semiconvergent. The sequence of iterates $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$ also is said to be semiconvergent.

Let $\hat{k} \geq 0$ be the smallest index, such that

$$\|\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}\| = \min_{k \geq 0} \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|. \tag{1.4}$$

For many linear systems of equations (1.1) that arise from the discretization of an ill-posed problem, the approximation error $\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}$ is small enough to make $\boldsymbol{x}_{\hat{k}}$ a meaningful approximation of $\hat{\boldsymbol{x}}$. However, the index $\hat{k}$ in (1.4) generally is not explicitly known. This note considers the problem of determining an approximation of $\hat{k}$. For recent discussions of iterative methods for ill-posed problems and stopping criteria; see [4, 7, 8, 10, 11].

Considerable experience with semiconvergent sequences has been gained in the context of asymptotic approximation of integrals. The following example is from Dahlquist and Björck [5, Section 3.2.5].

Example 1.1. Consider the evaluation of the integral

$$f(t) = e^t \int_t^\infty e^{-u} u^{-1} du$$

for (large) positive values of t. The change of variables $u = t + v$ yields after some manipulations the decomposition $f(t) = s_k(t) + r_k(t)$, where

$$s_k(t) = \frac{1}{t} - \frac{1!}{t^2} + \frac{2!}{t^3} + \ldots + (-1)^{k-1} \frac{(k-1)!}{t^k}$$

and $r_k(t)$ is the remainder. For a fixed positive value of t, we seek to approximate $f(t)$ by a member of the sequence

$$s_1(t), s_2(t), s_3(t), \ldots \; , \tag{1.5}$$

which is easily seen to be semiconvergent; for any fixed value $t > 0$, $\lim_{k \to \infty} |s_k(t)| = \infty$. Thus, we would like to determine an index $\tilde{k}$, which may depend on t, such that $s_{\tilde{k}}(t)$ is an accurate approximation of $f(t)$, or equivalently, such that $|r_{\tilde{k}}(t)|$ is small. This is achieved by bounding $r_k(t)$ for fixed t as follows. Since the signs of the k terms that make up $s_k(t)$ alternate, it follows that for every k the remainder satisfies

$$|r_k(t)| \leq \frac{k!}{t^{k+1}}.$$

Let $\tilde{k}$ minimize this bound, i.e., $\tilde{k}$ is the integer part of t. Then, for all $k \geq 1$,

$$|s_{\tilde{k}+1}(t) - s_{\tilde{k}}(t)| = \frac{\tilde{k}!}{t^{\tilde{k}+1}} \leq \frac{k!}{t^{k+1}} = |s_{k+1}(t) - s_k(t)|.$$

Thus, we can determine $\tilde{k}$ by minimizing the difference $|s_{k+1}(t) - s_k(t)|$ over $k \geq 1$. For instance, $t = 5$ yields $\tilde{k} = 5$, and we obtain $s_5(5) = 0.17408$ and $|f(5) - s_5(5)| < 1.2 \cdot 10^{-2}$. $\square$

Recently, Ferreira et al. [6] presented a nice survey of asymptotic approximation methods for the evaluation of integrals. They consider members of sequences analogous to (1.5) to be truncations of the (divergent) formal series $s_\infty(t)$, and state [6, p. 61] "It usually happens that this optimal truncation occurs at the smallest term (in absolute value) of the series". It is the purpose of this note to discuss this observation in the context of iterative methods for ill-posed problems. Section 2 outlines the RRGMRES and LSQR iterative methods and proposes stopping criteria inspired by the above observation by Ferreira et al. [6]. Computed examples are presented in Section 3, and concluding remarks can be found in Section 4.

2. Iterative methods and stopping criteria. We will apply the stopping criteria of the present note to the Range Restricted GMRES (RRGMRES) and LSQR iterative methods. The initial iterate is chosen to be $\boldsymbol{x}_0 = \boldsymbol{0}$.

The kth iterate, $\boldsymbol{x}_k$, determined by RRGMRES applied to the solution of (1.1), satisfies

$$\|A\boldsymbol{x}_k - \boldsymbol{b}\| = \min_{\boldsymbol{x} \in \mathbb{K}_k(A, A\boldsymbol{b})} \|A\boldsymbol{x} - \boldsymbol{b}\|, \qquad \boldsymbol{x}_k \in \mathbb{K}_k(A, A\boldsymbol{b}),$$

where

$$\mathbb{K}_k(A, A\boldsymbol{b}) = \mathrm{span}\{A\boldsymbol{b}, A^2\boldsymbol{b}, \ldots, A^k\boldsymbol{b}\}$$

is a Krylov subspace. Thus, RRGMRES differs from the better known GMRES method only in that the Krylov subspaces $\mathbb{K}_k(A, \boldsymbol{b})$ used by the latter are replaced by $\mathbb{K}_k(A, A\boldsymbol{b})$, which is in the range of A. RRGMRES typically yields higher accuracy than GMRES when the right-hand side $\boldsymbol{b}$ is contaminated by error and the desired solution $\hat{\boldsymbol{x}}$ is smooth; see [2] for examples. When A is symmetric, the iterates are orthogonal to the null space of A; see Calvetti et al. [3] and Hanke [7] for discussions. For large matrices A, the dominating work for computing $\boldsymbol{x}_k$ is the evaluation of $k+1$ matrix-vector products with A.

LSQR is an implementation of the conjugate gradient method applied to the normal equations

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

associated with (1.1). The kth iterate, $\boldsymbol{x}_k$, determined by LSQR satisfies

$$\|A\boldsymbol{x}_k - \boldsymbol{b}\| = \min_{\boldsymbol{x} \in \mathbb{K}_k(A^T A, A^T\boldsymbol{b})} \|A\boldsymbol{x} - \boldsymbol{b}\|, \qquad \boldsymbol{x}_k \in \mathbb{K}_k(A^T A, A^T\boldsymbol{b});$$

see, e.g., Björck [1] and Hansen [10] for recent discussions of this method. For large matrices, the dominating work for computing $\boldsymbol{x}_k$ is the evaluation of $2k$ matrix-vector products with A and A^T.

The following result, which follows from the triangle inequality, suggests a criterion for when to terminate the iterations with RRGMRES and LSQR.

PROPOSITION 2.1. *The following inequalities hold:*

$$\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| < \|\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}\| \Rightarrow \|\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}\| > \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\|, \tag{2.1}$$

$$\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| \leq \|\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}\| \Rightarrow \|\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}\| \geq \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|. \tag{2.2}$$

Moreover, let $\rho_{\hat{k}} = \|\boldsymbol{x}_{\hat{k}+1} - \hat{\boldsymbol{x}}\| / \|\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}\|$. *Then*

$$\|\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}\| \geq \frac{1}{1+\rho_{\hat{k}}}\|\boldsymbol{x}_{\hat{k}+1} - \boldsymbol{x}_{\hat{k}}\|. \tag{2.3}$$

Many iterates satisfy (2.1). In particular, the iterate $\boldsymbol{x}_{\hat{k}}$ defined by (1.4) satisfies both (2.1) and (2.2) with $k = \hat{k}$. The constant $\rho_{\hat{k}} \geq 1$ is close to unity for many linear systems of equations (1.1). Therefore $\frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ provides a lower bound, or an estimate thereof, of $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$ for many iterates $\boldsymbol{x}_k$. This suggests that an iterate $\boldsymbol{x}_{k'}$, such that $\|\boldsymbol{x}_{k'+1} - \boldsymbol{x}_{k'}\|$ is small, may furnish an accurate approximation of $\hat{\boldsymbol{x}}$. Specifically, let k' be the smallest index, such that

$$\|\boldsymbol{x}_{k'+1} - \boldsymbol{x}_{k'}\| = \min_{k \geq 0} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|. \tag{2.4}$$

This selection criterion is analogous to the choice of $\tilde{k}$ in Example 1.1. Numerous numerical experiments, some of which are reported in Section 3, show that $\boldsymbol{x}_{k'}$, indeed, often is a quite good approximation of $\hat{\boldsymbol{x}}$.

It is sometimes possible to determine a better approximation of $\hat{\boldsymbol{x}}$ than $\boldsymbol{x}_{k'}$ by modifying the criterion (2.4) to allow the incorporation of auxiliary information. Let $\mathbb{M}$ denote the set of indices k for which the function

$$k \to \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| \tag{2.5}$$

achieves a local minimum. We then choose an iterate $\boldsymbol{x}_k$ with $k \in \mathbb{M}$ with the aid of other criteria, such as the discrepancy principle or the L-curve. For instance, assume that an estimate of $\|\boldsymbol{e}\|$ is available. Then the discrepancy principle suggests that the first iterate $\boldsymbol{x}_k$ that satisfies

$$\|A\boldsymbol{x}_k - \boldsymbol{b}\| \leq \eta \|\boldsymbol{e}\| \tag{2.6}$$

be chosen as an approximation of $\hat{\boldsymbol{x}}$, where η is a parameter larger than or equal to one; see, e.g., [10]. Let k_d denote the index of the iterate determined by the discrepancy principle. We may then select the iterate $\boldsymbol{x}_{k'_d}$, where k'_d an index in $\mathbb{M}$ larger than or equal to k_d, as our approximation of $\hat{\boldsymbol{x}}$. This is illustrated in Section 3.

The L-curve is the graph obtained by connecting consecutive points in the sequence

$$\boldsymbol{p}_k = (\ln \|\boldsymbol{x}_k\|, \ln \|\boldsymbol{r}_k\|), \qquad k = 1, 2, 3, \dots, \tag{2.7}$$

by straight lines. This curve often looks roughly like the letter "L". It was first used in the context of iterative methods for ill-posed problems by Hanke and Hansen [8]; see Hansen [10] for a recent discussion. The reason for the L-shape of the curve is that the residual error typically decreases rapidly as k increases when k is small, but decreases slowly as k increases when k is large. Conversely, the iterates $\boldsymbol{x}_k$ typically grow fairly slowly (if at all) with k when k is small, but very rapidly when k is large. The rapid growth for k large is due to contamination of $\boldsymbol{x}_k$ by propagated error. Let k_L denote the index of the point at the "vertex" of the L-curve. The significance of the vertex is that $\|\boldsymbol{r}_{k_L}\|$ is close to $\min_{k\geq 0}\|\boldsymbol{r}_k\|$, and $\boldsymbol{x}_{k_L}$ is, we hope, not contaminated by a significant propagated error. Hanke and Hansen [8] advocate to choose the iterate $\boldsymbol{x}_{k_L}$ as an approximation of $\hat{\boldsymbol{x}}$.

This recipe for choosing an iterate is not always easy to follow, because the L-curve is piecewise linear and the points (2.7) may "cluster" at the vertex. It is therefore not always obvious by visual inspection which point on the curve should be considered the vertex. This is illustrated in Section 3. We propose to use the L-curve to determine a subset of consecutive iterates $\boldsymbol{x}_{k_1}, \boldsymbol{x}_{k_1+1}, \ldots, \boldsymbol{x}_{k_2}$ associated with points in a neighborhood of the vertex of the L-curve, and to choose an iterate $\boldsymbol{x}_{k'_L}$, with an index k'_L that satisfies

$$k_1 \leq k'_L \leq k_2, \qquad k'_L \in \mathbb{M}. \tag{2.8}$$

This approach obviates the need to determine the location of the vertex with high accuracy. We remark that the L-curve is not guaranteed to give a good approximant. A different approach to determine a suitable iterate from the L-curve, based on a geometric considerations of the L-curve, recently has been advocated by Castellanos et al. [4].

3. Computed examples. All computations were carried out using Matlab with unit roundoff $\epsilon \approx 2 \cdot 10^{-16}$. The examples illustrate the different choices of iterates discussed in Section 2. The experiments have been carried out with the RRGMRES and LSQR iterative methods implemented with reorthogonalization. In all examples, the error vector $\boldsymbol{e} \in \mathbb{R}^m$ has normally distributed entries with zero mean, and is normalized so as to achieve a specific noise level

$$\delta = \frac{\|\boldsymbol{e}\|}{\|\hat{\boldsymbol{b}}\|}. \tag{3.1}$$

In most examples the noise level is not assumed to be explicitly known.

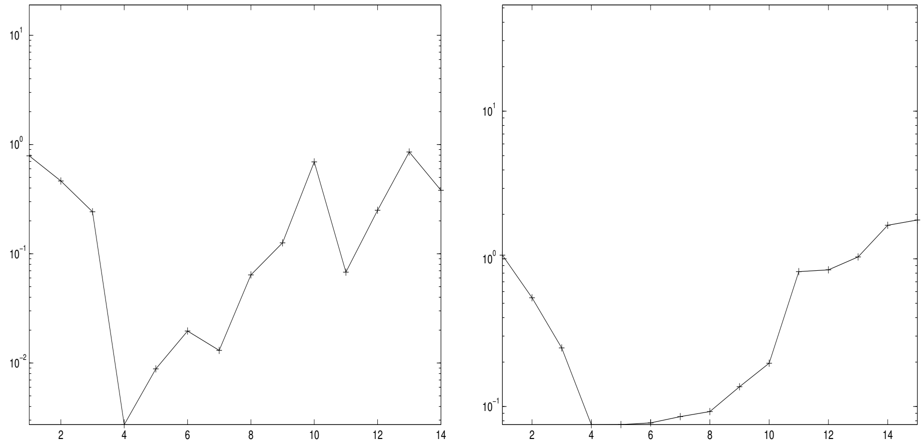
Example 3.1. Consider the solution of the Fredholm integral equation of the first kind

$$\int_{-6}^{6} \kappa(\tau,\sigma)x(\sigma)d\sigma = b(\tau), \qquad -6 \leq \tau \leq 6, \tag{3.2}$$

discussed by Phillips [12]. Its solution, kernel, and right-hand side are given by

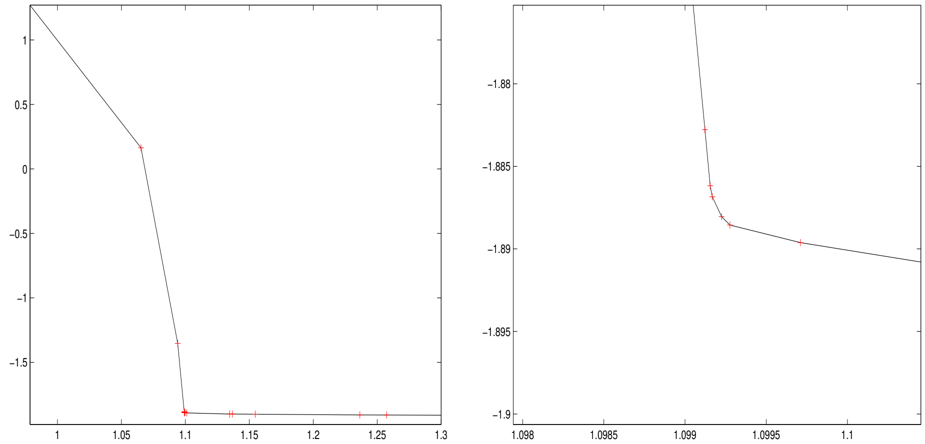
$$\begin{aligned}
x(\sigma) &= \begin{cases} 1+\cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma|<3, \\ 0, & \text{otherwise,} \end{cases} \\
\kappa(\tau,\sigma) &= x(\tau-\sigma), \\
b(\tau) &= (6-|\tau|)(1+\frac{1}{2}\cos(\frac{\pi}{3}\tau)) + \frac{9}{2\pi}\sin(\frac{\pi}{3}|\tau|).
\end{aligned}$$

We use the code phillips from the Matlab package Regularization Tools [9] to discretize (3.2) by a Galerkin method with orthonormal box functions as test and trial functions to obtain the symmetric indefinite matrix $A \in \mathbb{R}^{300\times 300}$ and the solution $\hat{\boldsymbol{x}}$ of the error-free linear system (1.2), whose right-hand side we compute as $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$. The



(a) (b)

FIG. 3.1. *Example 3.1: (a) Graph of* $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ *for* $1 \leq k \leq 14$. *(b) Error history* $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$, $1 \leq k \leq 15$.



(a) (b)

FIG. 3.2. *Example 3.1: (a) L-curve with points* $\boldsymbol{p}_k$, $1 \leq k \leq 15$, *cf. (2.7), marked by (red) crosses. (b) Blow-up of L-curve near the "vertex" showing the points* $\boldsymbol{p}_k$, $4 \leq k \leq 9$.

error vector $\boldsymbol{e}$ has noise level $\delta = 5 \cdot 10^{-2}$ and the right-hand side of (1.1) is given by $\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}$; cf. (1.3).

Figure 3.1(a) displays $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ for $1 \leq k \leq 14$. The figure shows that the index k', defined by (2.4), has the value 4. The error history $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$, $1 \leq k \leq 15$, is displayed in Figure 3.1(b), which shows $\boldsymbol{x}_4$ to be the iterate with the smallest error, i.e., $\hat{k} = 4$; cf. (1.4). We have $\|\boldsymbol{x}_4 - \hat{\boldsymbol{x}}\| = 7.5 \cdot 10^{-2}$.

Figures 3.2(a) and 3.2(b) illustrate how the L-curve can be used to support the selection of the index k'. Figure 3.2(a) displays the L-curve (2.7) and Figure 3.2(b) shows a blow-up of the region around the vertex of the L-curve. The clustering of nodes at the vertex makes it difficult to decide which node should be considered the vertex of the L-curve. However, the L-curve plots show that norm of the residual error $\boldsymbol{r}_k$ does not decrease significantly for $k \geq k'$. This suggests that $\boldsymbol{x}_{k'}$ may be a

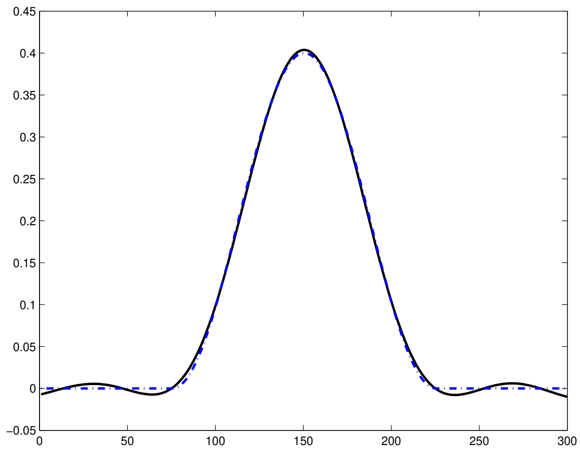
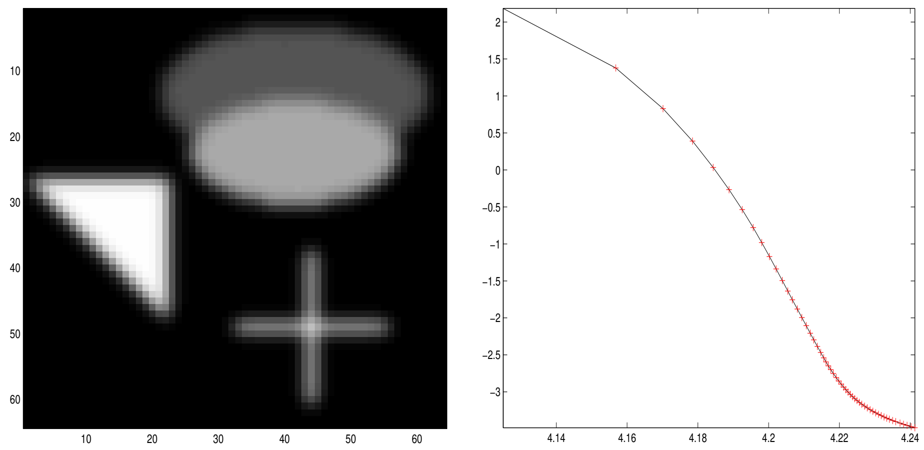


FIG. 3.3. *Example 3.1: Computed approximate solution* $\boldsymbol{x}_4$ *(black continuous curve) and solution* $\hat{\boldsymbol{x}}$ *of error-free linear system (1.2) (blue dashed curve).*

good approximation of $\hat{\boldsymbol{x}}$ and provides additional support for the choice of $\boldsymbol{x}_{k'}$ as an approximant of $\hat{\boldsymbol{x}}$.

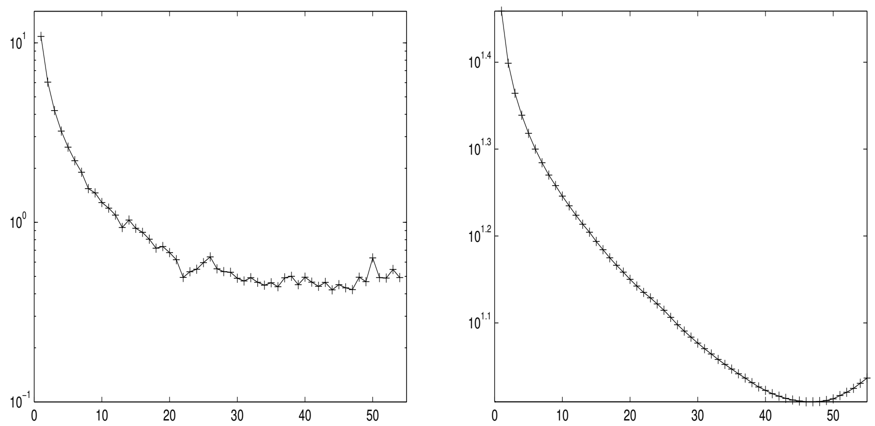
When RRGMRES is replaced by LSQR, we obtain $k' = 5$ and $\|\boldsymbol{x}_5 - \hat{\boldsymbol{x}}\| = 7.5 \cdot 10^{-2}$. Also for LSQR, we have that $k' = \hat{k}$. Note that LSQR and RRGMRES yield approximations of $\hat{\boldsymbol{x}}$ of the same accuracy, but the computation of $\boldsymbol{x}_5$ by LSQR requires the evaluation of twice as many matrix-vector products as the computation of $\boldsymbol{x}_4$ by RRGMRES. □



(a) (b)

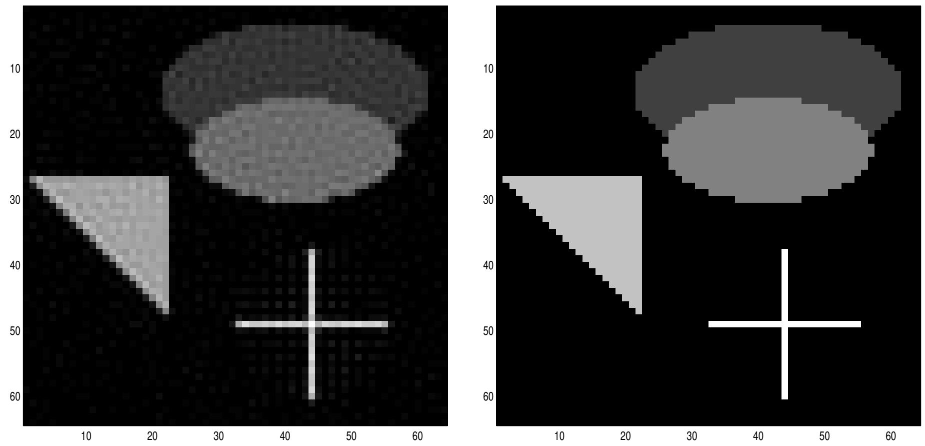
FIG. 3.4. *Example 3.2: (a) Blurred and noisy image. (b) L-curve with points* $\boldsymbol{p}_k$, $1 \leq k \leq 55$, *cf. (2.7), marked by crosses.*

Example 3.2. We would like to restore an image that has been contaminated by blur and noise. The blur- and noise-free image is represented by 64×64 pixels and generated by the Matlab code blur from [9]. It is stored in the vector $\hat{\boldsymbol{x}} \in \mathbb{R}^{4096}$. The code blur also determines a symmetric block Toeplitz matrix with Toeplitz blocks $A \in \mathbb{R}^{4096 \times 4096}$, which represents a blurring operator, and can be thought of as being



(a) (b)

FIG. 3.5. *Example 3.2: (a) Graph of* $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ *for* $1 \leq k \leq 55$. *(b) Error history* $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$, $1 \leq k \leq 55$.



(a) (b)

FIG. 3.6. *Example 3.2: (a) Restored image represented by* x_{44}. *(b) Original image represented by* $\hat{\boldsymbol{x}}$.

determined by a Gaussian point spread function with variance $\sigma = 1.2$. The vector $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ represents the blurred, but noise-free, image associated with $\hat{\boldsymbol{x}}$. The error vector $\boldsymbol{e} \in \mathbb{R}^{4096}$ models noise and is scaled to yield the noise level $\delta = 1 \cdot 10^{-3}$; cf. (3.1). The right-hand side $\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}$ of (1.1) represents the blurred and noisy image, shown in Figure 3.4(a), associated with $\hat{\boldsymbol{x}}$.

Assume first that neither the vector $\hat{\boldsymbol{x}}$ nor the noise level are known. Given A and $\boldsymbol{b}$, we would like to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of the linear system of equations (1.1).

We solve (1.1) by RRGMRES. Figure 3.5(a) displays $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ for $1 \leq k \leq 55$. In particular, the figure shows that $\mathbb{M} = \{13, 22, 31, 36, 39, 42, 44, \dots\}$ and $k' = 44$. The error history $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$ for $1 \leq k \leq 55$, which is displayed in Figure 3.5(b), shows

TABLE 3.1
Performance of the selection criterion (2.4). The last column tabulates the relative error of the error in $\boldsymbol{x}_{k'}$ given by (3.3).

problem	dimension	noise level	$\hat{k}$	k'	relative error
shaw(10)	300	$1 \cdot 10^{-2}$	6	6	0
shaw(10)	300	$1 \cdot 10^{-3}$	7	6	$1.41 \cdot 10^{-4}$
shaw(10)	300	$1 \cdot 10^{-4}$	9	7	$4.62 \cdot 10^{-1}$
phillips(10)	300	$1 \cdot 10^{-2}$	6	4	$7.81 \cdot 10^{-1}$
phillips(30)	300	$1 \cdot 10^{-3}$	10	7	$1.32 \cdot 10^{-1}$
phillips(30)	300	$1 \cdot 10^{-4}$	11	11	0
blur(30)	4096	$1 \cdot 10^{-1}$	4	3	$1.89 \cdot 10^{-2}$
blur(55)	4096	$1 \cdot 10^{-2}$	14	13	$6.31 \cdot 10^{-3}$
blur(55)	4096	$1 \cdot 10^{-3}$	47	44	$4.90 \cdot 10^{-3}$
deriv2(10)	300	$1 \cdot 10^{-2}$	7	6	$8.62 \cdot 10^{-3}$
deriv2(20)	300	$1 \cdot 10^{-3}$	12	10	$6.09 \cdot 10^{-2}$
deriv2(20)	300	$1 \cdot 10^{-4}$	18	15	$7.00 \cdot 10^{-2}$
baart(10)	300	$1 \cdot 10^{-2}$	3	3	0
baart(10)	300	$1 \cdot 10^{-3}$	3	3	0

that $\boldsymbol{x}_{47}$ furnishes the best approximation of $\hat{\boldsymbol{x}}$.

The L-curve, shown in Figure 3.4(b), does not have an obvious vertex, but indicates that it may be worthwhile to carry out at least about 40 iterations. The L-curve together with knowledge of $k' = 44$ and the set $\mathbb{M}$ suggest that $\boldsymbol{x}_{44}$ be chosen as an approximant of $\hat{\boldsymbol{x}}$. The restored image represented by $\boldsymbol{x}_{44}$ is shown in Figure 3.6(a). For comparison, the original image represented by $\hat{\boldsymbol{x}}$ is displayed by Figure 3.6(b); we have $\|\boldsymbol{x}_{44} - \hat{\boldsymbol{x}}\| = 10.26$. We remark that the images represented by $\boldsymbol{x}_{47}$ and $\boldsymbol{x}_{44}$ cannot be distinguished by visual inspection.

We turn to the situation when the noise level (3.1) is known. The discrepancy principle with $\eta = 1$ yields the iterate $\boldsymbol{x}_{27}$. We have $\|\boldsymbol{x}_{27} - \hat{\boldsymbol{x}}\| = 12.512$; thus the L-curve combined with the criterion (2.8) yields a significantly more accurate approximation of $\hat{\boldsymbol{x}}$ than the discrepancy principle. □

Example 3.3. This example shows our computational experience for several problems from Regularization Tools [9]. The problems yield linear systems of equations with symmetric or nonsymmetric severely ill-conditioned matrices. The "problem" column of Table 3.1 shows the name of the problem (from [9]) and displays in parentheses the number of iterations carried out. All examples have square matrices of size $n \times n$. The column labeled "dimension" tabulates n. The column "noise level" shows the quotient δ; cf. (3.1). Several noise levels are used for each problem. The table displays the indices $\hat{k}$ and k', defined by (1.4) and (2.4), respectively. For a few problems and noise levels, we obtain $k' = \hat{k}$; however, also when $k' \neq \hat{k}$, the iterate $\boldsymbol{x}_{k'}$ often approximates $\hat{\boldsymbol{x}}$ almost as well as $\boldsymbol{x}_{\hat{k}}$. To see this, we tabulate the quotient of the relative error of the error in $\boldsymbol{x}_{k'}$,

$$\frac{\|\boldsymbol{x}_{k'} - \hat{\boldsymbol{x}}\| - \|\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}_{\hat{k}} - \hat{\boldsymbol{x}}\|}, \tag{3.3}$$

which in Table 3.1 is referred to as the "relative error". Note that we did not use the L-curve and the set $\mathbb{M}$ to help us select an iterate in the present example. □

4. Conclusion. This note shows that the norm of the difference between consecutive iterates can help in the determination of a suitable iterate. Example 3.3 shows the performance of the selection criterion (2.4) based on the norm of the difference between consecutive iterates, and Examples 3.1 and 3.2 illustrate how the norm of the difference between consecutive iterates can be used together with the discrepancy principle and the L-curve to select a suitable iterate.

REFERENCES

[1] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[2] D. Calvetti, B. Lewis, and L. Reichel, *On the choice of subspace for iterative methods for linear discrete ill-posed problems*, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
[3] D. Calvetti, L. Reichel, and Q. Zhang, *Conjugate gradient algorithms for symmetric inconsistent linear systems*, in Proceedings of the Cornelius Lanczos International Centenary Conference, eds. J. D. Brown, M. T. Chu, D. C. Ellison, and R. J. Plemmons, SIAM, Philadelphia, 1994, pp. 267–272.
[4] J. L. Castellanos, S. Gómez, and V. Guerra, *The triangle method for finding the corner of the L-curve*, Appl. Numer. Math., 43 (2002), pp. 359–373.
[5] G. Dahlquist and Å. Björck, *Numerical Methods*, Dover, Mineola, 2003.
[6] C. Ferreira, J. L. López, and N. M. Temme, *Asymptotic approximations of integrals: an introduction, with recent developments and applications to orthogonal polynomials*, Elec. Trans. Numer. Anal., 19 (2005), pp. 58–83.
[7] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman, Essex, 1995.
[8] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315.
[9] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algor., 6 (1994), pp. 1–35. Software is available in Netlib at `http://www.netlib.org`.
[10] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[11] L. Kaufman and A. Neumaier, *Regularization of ill-posed problems by envelope guided conjugate gradients*, J. Comput. Graph. Stat., 6 (1997), pp. 451–463.
[12] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.