Decomposition methods for large linear discrete ill-posed problems

James Baglama [a,1], Lothar Reichel [b,2,*],

[a] *Department of Mathematics, University of Rhode Island, Kingston, RI 02881, USA.*

[b] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.*

Abstract

The solution of large linear discrete ill-posed problems by iterative methods continues to receive considerable attention. This paper presents decomposition methods that split the solution space into a Krylov subspace that is determined by the iterative method and an auxiliary subspace that can be chosen to help represent pertinent features of the solution. Decomposition is well suited for use with the GMRES, RRGMRES, and LSQR iterative schemes.

Key words: GMRES, RRGMRES, LSQR, iterative method, ill-posed problem, inverse problem, decomposition, augmentation

1 Introduction

This paper is concerned with the iterative solution of large linear systems of equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{m\times n}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \qquad \boldsymbol{b} \in \mathbb{R}^m, \tag{1}$$

with a matrix A of ill-determined rank, i.e., A has many singular values of different orders of magnitude close to the origin. In particular, A is severely

* Corresponding author.
Email addresses: `jbaglama@math.uri.edu` (James Baglama), `reichel@math.kent.edu` (Lothar Reichel).

[1] Research supported in part by NSF grant DMS-0311786.
[2] Research supported in part by NSF grant DMS-0107858.

ill-conditioned and may be singular. Matrices of ill-determined rank arise from the discretization of ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. Linear systems of equations with such a matrix are often referred to as linear discrete ill-posed problems. If the linear system (1) is inconsistent, e.g., when $m > n$, then we consider the system a least-squares minimization problem.

The right-hand side $\boldsymbol{b}$ in linear discrete ill-posed problems that arise in applications typically is contaminated by an error $\boldsymbol{e} \in \mathbb{R}^m$, which may stem from measurement or discretization errors. Let $\hat{\boldsymbol{b}}$ denote the unknown error-free vector associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}, \tag{2}$$

and assume that the linear system of equations with the unknown error-free right-hand side,

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \tag{3}$$

is consistent. The available linear system (1) is not required to be consistent.

We would like to determine a solution $\hat{\boldsymbol{x}}$ of (3), e.g., the least-squares solution of minimal Euclidean norm. Since the right-hand side $\hat{\boldsymbol{b}}$ is not available, we seek to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of the available linear system of equations (1). A popular approach to determining an approximation of $\hat{\boldsymbol{x}}$ for large-scale linear discrete ill-posed problems is to apply a few, say j, steps of an iterative method to (1). Denote the approximate solution so obtained by $\boldsymbol{x}_j$ and let $\|\cdot\|$ denote the Euclidean vector norm or the associated induced matrix norm. For many linear discrete ill-posed problems, the optimal number of iterations, denoted $j_{\rm opt}$, and defined as the smallest integer, such that

$$\|\boldsymbol{x}_{j_{\rm opt}} - \hat{\boldsymbol{x}}\| = \min_{j \geq 0} \|\boldsymbol{x}_j - \hat{\boldsymbol{x}}\|, \tag{4}$$

is quite small. This depends on that iterates $\boldsymbol{x}_j$ for j large generally are severely contaminated by propagated errors due to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ and round-off errors introduced during the computation of $\boldsymbol{x}_j$; see, e.g., [5,6,10,12] for discussions and illustrative computed examples. It is therefore important that the subspaces in which the iterates $\boldsymbol{x}_j$ live allow the representation of pertinent features of $\hat{\boldsymbol{x}}$ already for small values of j. These features may be jumps, spikes, or just linear increase.

This paper proposes decomposition of the linear system of equations (1) that corresponds to a decomposition of the solution space into a Krylov subspace determined by a standard iterative method, such as GMRES, RRGMRES or LSQR, and a user-supplied subspace. The latter can be chosen to allow the representation of desirable features of $\hat{\boldsymbol{x}}$ that may be difficult to represent by a vector in a low-dimensional Krylov subspace.

Let the span of the orthonormal columns of $W \in \mathbb{R}^{n\times\ell}$ represent the user-chosen linear space and introduce the orthogonal projectors $P_W = WW^T$ and $P_W^{\perp} = I - P_W$. We use these projectors to split the computed approximate solutions $\boldsymbol{x}_j$ according to

$$\boldsymbol{x}_j = \boldsymbol{x}'_j + \boldsymbol{x}''_j, \qquad \boldsymbol{x}'_j = P_W \boldsymbol{x}_j, \qquad \boldsymbol{x}''_j = P_W^{\perp} \boldsymbol{x}_j, \qquad j = 1, 2, 3, \dots . \quad (5)$$

The component $\boldsymbol{x}''_j$ of $\boldsymbol{x}_j$ is computed by an iterative method. Since ℓ generally is chosen quite small, e.g., $1 \leq \ell \leq 3$, we can determine $\boldsymbol{x}'_j$ by solving a small linear system of equations by a direct method.

Section 2 describes decomposition methods for linear systems of equations (1) with $m = n$, and discusses application of the GMRES and RRGMRES iterative methods to the computation of an approximation of $P_W^{\perp}\hat{\boldsymbol{x}}$. A decomposition method for linear systems with $m \neq n$, in which an approximation of $P_W^{\perp}\hat{\boldsymbol{x}}$ is computed by the LSQR iterative method, is presented in Section 3. Section 4 shows a few computed examples and Section 5 contains concluding remarks.

2 Decomposition and GMRES-type methods

This section discusses iterative schemes based on the decomposition (5) and the application of iterative methods of GMRES-type. We assume throughout this section that $m = n$ in (1).

Our solution method uses the QR-factorization

$$AW = QR, \quad (6)$$

i.e., $Q \in \mathbb{R}^{n\times\ell}$ has orthonormal columns and $R \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Since in our applications ℓ is small, the factorization (6) can be computed quite rapidly. We will assume that R is nonsingular. This can be secured by choosing ℓ sufficiently small. Let P_Q be the orthogonal projector onto the range of AW, i.e., $P_Q = QQ^T$, and let $P_Q^{\perp} = I - P_Q$. By using these projectors and by splitting $\boldsymbol{x}$ into $P_W\boldsymbol{x}$ and $P_W^{\perp}\boldsymbol{x}$ analogously to (5), we obtain the decomposition

$$P_Q A P_W \boldsymbol{x} + P_Q A P_W^{\perp} \boldsymbol{x} = P_Q \boldsymbol{b}, \tag{7}$$
$$P_Q^{\perp} A P_W^{\perp} \boldsymbol{x} = P_Q^{\perp} \boldsymbol{b} \tag{8}$$

of the linear system (1), where we have used the fact that

$$P_Q^{\perp} A P_W = 0. \tag{9}$$

Remark 2.1. Introduce the A-weighted pseudo-inverse of P_W,

$$P_{W,A}^{\dagger} = (I - (AP_W)^{\dagger} A) P_W^{\perp},$$

where $(AP_W)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of AP_W; see Eldén [8] or Hansen [12, Section 2.3] for the definition of the A-weighted pseudo-inverse of a general matrix. Then

$$AP_{W,A}^{\dagger} = P_Q^{\perp} A P_W^{\perp},$$

i.e., the system of equations (8) formally may be considered preconditioned by the right preconditioner $P_{W,A}^{\dagger}$. Of course, $P_{W,A}^{\dagger}$ does not approximate the inverse or Moore-Penrose pseudo-inverse of A in any meaningful way and therefore $P_{W,A}^{\dagger}$ is not a preconditioner in a traditional sense. □

We solve the large linear system (8) by the GMRES or RRGMRES iterative methods. A brief description of these methods is provided below. It follows from (9) that $P_Q^{\perp} A P_W^{\perp} = P_Q^{\perp} A$, and therefore we apply the iterative methods to

$$P_Q^{\perp} A \boldsymbol{z}'' = P_Q^{\perp} \boldsymbol{b}, \tag{10}$$

i.e., we do not require the computed iterates $\boldsymbol{z}_j''$ to be in the range of $P_W^{\perp}$.

Let $\boldsymbol{z}_j''$ be a satisfactory approximate solution of (10). We then determine

$$\boldsymbol{x}_j'' = P_W^{\perp} \boldsymbol{z}_j'' \tag{11}$$

and solve (7) for the component $\boldsymbol{x}_j'$ in the range of P_W of the approximate solution $\boldsymbol{x}_j$ of (1), cf. (5). Thus, $\boldsymbol{x}_j' = P_W \boldsymbol{x}_j$ satisfies

$$P_Q A \boldsymbol{x}_j' = P_Q \boldsymbol{b} - P_Q A \boldsymbol{x}_j''. \tag{12}$$

This system is equivalent to the small linear system of equations

$$R\boldsymbol{z}' = Q^T (\boldsymbol{b} - A\boldsymbol{x}_j''), \tag{13}$$

whose solution we denote by $\boldsymbol{z}'_j$. Then

$$\boldsymbol{x}'_j = W\boldsymbol{z}'_j \tag{14}$$

and we obtain $\boldsymbol{x}_j$ from (5).

We conclude this section with a brief review of the GMRES and RRGMRES iterative methods for the solution of (10) and comment on when to terminate the iterations. The iterative methods are applied with initial approximate solution $\boldsymbol{z}''_0 = \boldsymbol{0}$. The jth iterate, $\boldsymbol{z}''_j$, determined by GMRES satisfies

$$\begin{cases} \|P_Q^\perp A\boldsymbol{z}''_j - P_Q^\perp \boldsymbol{b}\| = \displaystyle\min_{\boldsymbol{z}''\in\mathbb{K}_j(P_Q^\perp A, P_Q^\perp \boldsymbol{b})} \|P_Q^\perp A\boldsymbol{z}'' - P_Q^\perp \boldsymbol{b}\|, \\ \\ \qquad\qquad\qquad \boldsymbol{z}''_j \in \mathbb{K}_j(P_Q^\perp A, P_Q^\perp \boldsymbol{b}), \end{cases} \tag{15}$$

where

$$\mathbb{K}_j(P_Q^\perp A, P_Q^\perp \boldsymbol{b}) = \mathrm{span}\{P_Q^\perp \boldsymbol{b}, P_Q^\perp A P_Q^\perp \boldsymbol{b}, \dots, (P_Q^\perp A)^{j-1} P_Q^\perp \boldsymbol{b}\} \tag{16}$$

is a Krylov subspace. The standard GMRES implementation determines the Arnoldi decomposition

$$P_Q^\perp A V_j = V_{j+1} H_{j+1,j}, \tag{17}$$

where $V_i \in \mathbb{R}^{n\times i}$, $V_i^T V_i = I_i$, $\mathrm{range}(V_i) = \mathbb{K}_i(P_Q^\perp A, P_Q^\perp \boldsymbol{b})$, and $V_i\boldsymbol{e}_1 = P_Q^\perp \boldsymbol{b}/\|P_Q^\perp \boldsymbol{b}\|$ for $i = j$ and $i = j+1$. Here I_i denotes the identity matrix of order i and $\boldsymbol{e}_i$ the ith axis vector. Moreover, the matrix $H_{j+1,j} \in \mathbb{R}^{(j+1)\times j}$ is of upper Hessenberg form with positive subdiagonal entries; see, e.g., Saad and Schultz [16] for details. We assume that j is small enough so that the decomposition (17) with the stated properties exists.

Substituting $\boldsymbol{z}'' = V_j\boldsymbol{y}''$ into the minimization problem in (15) and using the Arnoldi decomposition (17) yields

$$\min_{\boldsymbol{y}''\in\mathbb{R}^j} \|P_Q^\perp A V_j \boldsymbol{y}'' - P_Q^\perp \boldsymbol{b}\| = \min_{\boldsymbol{y}''\in\mathbb{R}^j} \|H_{j+1,j}\boldsymbol{y}'' - \boldsymbol{e}_1\|P_Q^\perp \boldsymbol{b}\|\,\|. \tag{18}$$

Introduce the QR-factorization $H_{j+1,j} = Q_{j+1}R_{j+1,j}$, where $Q_{j+1} \in \mathbb{R}^{(j+1)\times(j+1)}$ is orthogonal and $R_{j+1,j} \in \mathbb{R}^{(j+1)\times j}$ has a leading $j\times j$ upper triangular submatrix and a vanishing last row. Then (18) gives

$$\min_{\boldsymbol{y}''\in\mathbb{R}^j} \|P_Q^\perp A V_j \boldsymbol{y}'' - P_Q^\perp \boldsymbol{b}\| = |\boldsymbol{e}_{j+1}^T Q_{j+1}^T \boldsymbol{e}_1|\,\|P_Q^\perp \boldsymbol{b}\|. \tag{19}$$

Therefore the residual error is inexpensive to determine.

The computation of $\boldsymbol{z}_j''$ requires the evaluation of j matrix-vector products with A. For large-scale problems this is the dominant computational work when j is not very large, as is typically the case in our applications.

The matrix A in discrete ill-posed problems generally represents a smoothing operator, such as a convolution with a Gaussian. If the desired solution $\hat{\boldsymbol{x}}$ of (3) is known to have non-smooth features, such as jumps, then it is generally advantageous to choose the matrix W, so that these features can be represented by a linear combination of its columns. Moreover, we illustrate in Section 4 that significant increase in accuracy sometimes also can be achieved by letting the columns of W represent smooth functions, such as constants and linear functions.

Several criteria for when to terminate the iterations with GMRES are available, the most reliable of which is the discrepancy principle. It requires that $\|\boldsymbol{e}\|$, or an estimate thereof, be known, and prescribes that the iterations be terminated as soon as an approximate solution $\boldsymbol{x}_j$, such that

$$\|\boldsymbol{b} - A\boldsymbol{x}_j\| \leq \eta\|\boldsymbol{e}\|, \tag{20}$$

has been found, where $\eta \geq 1$ is a user-specified constant; see [5] for a justification of this stopping criterion.

Assume that the linear system of equations (12) is solved exactly. This assumption is reasonable, because the solution $\boldsymbol{x}_j'$ is determined by solving the linear system of equations (13) with a small and generally not very ill-conditioned matrix. Then the following theorem shows that

$$\|\boldsymbol{b} - A\boldsymbol{x}_j\| = \|P_Q^\perp \boldsymbol{b} - P_Q^\perp A\boldsymbol{z}_j''\|, \tag{21}$$

where $\boldsymbol{z}_j''$ is defined by (15). Thus, the left-hand side of (20) can be computed inexpensively by evaluating the right-hand side of (19) during the GMRES iterations.

Theorem 2.1 *Let the approximate solution $\boldsymbol{x}_j$ be given by (5) and assume that (12) holds. Let $\boldsymbol{z}_j''$ satisfy (15). Then (21) is valid.*

Proof. It follows from (5) and $I = P_Q + P_Q^\perp$ that

$$\|\boldsymbol{b} - A\boldsymbol{x}_j\| = \|\boldsymbol{b} - A(\boldsymbol{x}_j' + \boldsymbol{x}_j'')\| = \|P_Q\boldsymbol{b} + P_Q^\perp\boldsymbol{b} - (P_QA + P_Q^\perp A)(\boldsymbol{x}_j' + \boldsymbol{x}_j'')\|.$$

The right-hand side can be simplified using (12) and $P_Q^\perp A\boldsymbol{x}_j' = 0$, which follows from (5) and (9). Thus, we obtain

$$\|\boldsymbol{b} - A\boldsymbol{x}_j\| = \|P_Q^{\perp}\boldsymbol{b} - P_Q^{\perp}A\boldsymbol{x}_j''\|.$$

Application of (11) and (9) yields $P_Q^{\perp}A\boldsymbol{x}_j'' = P_Q^{\perp}A\boldsymbol{z}_j''$ and the theorem follows. $\square$

We remark that no specific properties of GMRES are used in the proof of Theorem 2.1. Therefore (21) also holds for iterates $\boldsymbol{z}_j''$ and the corresponding approximate solutions $\boldsymbol{x}_j$ determined by other iterative methods. When no estimate of $\|\boldsymbol{e}\|$ is available, one can use an L-curve to decide when to terminate the iterations; see [6].

The computation of the approximate solution $\boldsymbol{x}_j$ of (1) as described demands the evaluation of $\ell+j+1$ matrix-vector products with the matrix A: ℓ evaluations are required to determine the QR-factorization (6), j by GMRES, and 1 for computing the right-hand side of (13).

The linear system of equations (10) may be inconsistent and the recursion formulas for GMRES might break down before a sufficient number of iterations have been carried out. This can be remedied by using a breakdown-free variant of GMRES described in [15]. However, breakdown is rare and in our experience standard GMRES performs well.

The GMRES-based decomposition method described may be considered an augmentation method; the Krylov subspace generated by GMRES is augmented by the space $\mathbb{W} = \mathrm{span}(W)$. A numerical method based on the latter approach is discussed in [2]. Specifically, an approximate solution $\boldsymbol{x}_j$ of (1) that solves the least-squares problem

$$\|A\boldsymbol{x}_j - \boldsymbol{b}\| = \min_{\boldsymbol{x}\in\mathbb{K}_j(A,\boldsymbol{b})\cup\mathbb{W}} \|A\boldsymbol{x} - \boldsymbol{b}\|, \qquad \boldsymbol{x}_j \in \mathbb{K}_j(A,\boldsymbol{b})\cup\mathbb{W}, \tag{22}$$

is computed by using the modified Arnoldi decomposition

$$A[W\,\widetilde{V}_{\ell+1:\ell+j}] = \widetilde{V}_{\ell+j+1}\widetilde{H}_{\ell+j+1,\ell+j}, \tag{23}$$

where $\widetilde{V}_{\ell+j+1} = [Q\,\widetilde{V}_{\ell+1:\ell+j+1}] \in \mathbb{R}^{n\times(\ell+j+1)}$ has orthonormal columns, $Q \in \mathbb{R}^{n\times\ell}$ is determined by the QR-factorization (6), $\widetilde{V}_{i:k}$ denotes the (sub)matrix made up of columns i through k of $\widetilde{V}_{\ell+j+1}$, and $\widetilde{V}_{\ell+1:\ell+1} = P_Q^{\perp}\boldsymbol{b}/\|P_Q^{\perp}\boldsymbol{b}\|$. Further,

$$\widetilde{H}_{\ell+j+1,\ell+j} = \begin{bmatrix} R & S \\ & \\ O & H \end{bmatrix} \in \mathbb{R}^{(\ell+j+1)\times(\ell+j)}, \tag{24}$$

where $R \in \mathbb{R}^{\ell\times\ell}$ is the upper triangular matrix in the QR-factorization (6), $S \in \mathbb{R}^{\ell\times j}$, and $H \in \mathbb{R}^{(j+1)\times j}$ is of upper Hessenberg form with positive subdiagonal entries. Then $\boldsymbol{x}_j = [W\,\widetilde{V}_{\ell+1:\ell+j}]\boldsymbol{y}_j$, where $\boldsymbol{y}_j \in \mathbb{R}^{\ell+j}$ solves the least-squares problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^{\ell+j}} \|\widetilde{H}_{\ell+j+1,\ell+j}\boldsymbol{y} - \widetilde{V}^T_{\ell+j+1}\boldsymbol{b}\|. \tag{25}$$

Theorem 2.2 *In the absence of round-off errors, the iterate $\boldsymbol{x}_j$ determined by the augmented GMRES method described in [2], and outlined above, is identical to the iterate $\boldsymbol{x}_j$ determined by the GMRES-based decomposition method of the present paper.*

Proof. Split the solution $\boldsymbol{y}_j$ of (25) into subvectors, commensurate with the splitting of the matrix (24),

$$\boldsymbol{y}_j = \begin{bmatrix} \boldsymbol{y}'_j \\ \boldsymbol{y}''_j \end{bmatrix}, \qquad \boldsymbol{y}'_j \in \mathbb{R}^{\ell}, \qquad \boldsymbol{y}''_j \in \mathbb{R}^{j}.$$

It follows from (24) and (25) that $\boldsymbol{y}''_j$ solves the least-squares problem

$$\min_{\boldsymbol{y}''\in\mathbb{R}^{j}} \|H\boldsymbol{y}'' - \boldsymbol{e}_1\|P_Q^{\perp}\boldsymbol{b}\|\,\| \tag{26}$$

and $\boldsymbol{y}'_j$ solves the linear system of equations

$$R\boldsymbol{y}' = Q^T\boldsymbol{b} - S\boldsymbol{y}''. \tag{27}$$

Identifying the last j columns of (23) and using (24) yields

$$A\widetilde{V}_{\ell+1:\ell+j} = QS + \widetilde{V}_{\ell+1:\ell+j+1}H,$$

and we obtain

$$P_Q^{\perp}A\widetilde{V}_{\ell+1:\ell+j} = P_Q^{\perp}\widetilde{V}_{\ell+1:\ell+j+1}H = \widetilde{V}_{\ell+1:\ell+j+1}H, \tag{28}$$

where the last equality follows from the fact that the columns of $\widetilde{V}_{\ell+1:\ell+j+1}$ are orthogonal to Q. Thus, (17) and (28) are two Arnoldi decompositions of $P_Q^{\perp}A$ with the same initial vectors and with positive subdiagonal entries in the Hessenberg matrices. It follows from the Implicit Q-Theorem, see [9, Theorem 7.4.2], that the decompositions are identical. In particular, the least-squares problems (18) and (26) have the same solution.

Using $\widetilde{V}_{\ell+1:\ell+j+1} = V_j$ and $S = Q^T AV_j$, as well as equation (27) to eliminate $\boldsymbol{y}'_j$, we can express the solution $\boldsymbol{x}_j$ of (22) in the form

$$\boldsymbol{x}_j = W\boldsymbol{y}'_j + V_j\boldsymbol{y}''_j = WR^{-1}(Q^T\boldsymbol{b} - Q^T AV_j\boldsymbol{y}''_j) + V_j\boldsymbol{y}''_j. \tag{29}$$

The iterate $\boldsymbol{x}_j$ determined by the decomposition method of the present paper can be written as

$$\boldsymbol{x}_j = W\boldsymbol{z}'_j + (I - WW^T)\boldsymbol{z}''_j, \tag{30}$$

where we have used (11) and (14). Using (13) to eliminate $\boldsymbol{z}'_j$ and substituting $\boldsymbol{z}''_j = V_j\boldsymbol{y}''_j$ and (6) into (30) yields

$$\begin{aligned}\boldsymbol{x}_j &= WR^{-1}Q^T(\boldsymbol{b} - A(I - WW^T)V_j\boldsymbol{y}''_j) + (I - WW^T)V_j\boldsymbol{y}''_j \\ &= WR^{-1}Q^T(\boldsymbol{b} - AV_j\boldsymbol{y}''_j) + WR^{-1}Q^T AWW^T V_j\boldsymbol{y}''_j + (I - WW^T)V_j\boldsymbol{y}''_j \\ &= WR^{-1}Q^T(\boldsymbol{b} - AV_j\boldsymbol{y}''_j) + WW^T V_j\boldsymbol{y}''_j + (I - WW^T)V_j\boldsymbol{y}''_j. \end{aligned} \tag{31}$$

Hence, the expressions (29) and (31) are the same, which shows the theorem.
□

Thus, the augmented GMRES method described in [2] and the decomposition method based on GMRES of the present paper are equivalent. An advantage of the approach of the present paper is that it easily can be adapted to many iterative solution methods.

Range Restricted GMRES (RRGMRES) differs from GMRES in that the minimization problem (15) is replaced by

$$\begin{cases} \|P_Q^\perp A\boldsymbol{z}''_j - P_Q^\perp\boldsymbol{b}\| = \displaystyle\min_{\boldsymbol{z}''\in\mathbb{K}_j(P_Q^\perp A, P_Q^\perp AP_Q^\perp\boldsymbol{b})} \|P_Q^\perp A\boldsymbol{z}'' - P_Q^\perp\boldsymbol{b}\|, \\ \\ \qquad\qquad \boldsymbol{z}''_j \in \mathbb{K}_j(P_Q^\perp A, P_Q^\perp AP_Q^\perp\boldsymbol{b}); \end{cases}$$

in particular, the computed iterate $\boldsymbol{z}''_j$ lives in the range of $P_Q^\perp A$. This often makes the iterates determined by RRGMRES less sensitive to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ than iterates computed by GMRES. We have found that RRGMRES typically is able to determine a more accurate approximation of $\hat{\boldsymbol{x}}$ than GMRES when the desired solution $\hat{\boldsymbol{x}}$ is smooth; see Section 4 as well as [4] for examples.

The computation of the approximate solution $\boldsymbol{x}_j$ by the RRGMRES-based decomposition method requires the evaluation of $\ell+j+2$ matrix-vector products with the matrix A, of which RRGMRES needs $j + 2$.

The decomposition method for RRGMRES may be considered an augmentation method in which the Krylov subspace generated by RRGMRES is augmented by the space $\mathbb{W}$; cf. the related discussion on GMRES above. In the augmented RRGMRES method discussed in [2], one seeks to compute an approximate solution $\boldsymbol{x}_j$ of (1) that solves the least-squares problem

$$\|A\boldsymbol{x}_j - \boldsymbol{b}\| = \min_{\boldsymbol{x}\in\mathbb{K}_j(A,A\boldsymbol{b})\cup\mathbb{W}} \|A\boldsymbol{x} - \boldsymbol{b}\|, \qquad \boldsymbol{x}_j \in \mathbb{K}_j(A, A\boldsymbol{b}) \cup \mathbb{W}, \tag{32}$$

where $\mathbb{W}$ is the same as in (22).

Theorem 2.3 *In the absence of round-off errors, the solution $\boldsymbol{x}_j$ of (32) is identical to the iterate $\boldsymbol{x}_j$ determined by the RRGMRES-based decomposition method of the present paper.*

Proof. The result can be established similarly as Theorem 2.2. □

3 Decomposition and LSQR

In this section we allow $m \neq n$ in (1) and discuss how decomposition can be combined with the LSQR method by Paige and Saunders [13,14]. LSQR is an implementation of the conjugate gradient method applied to the normal equations

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b} \tag{33}$$

associated with (1); see Björck [3, Section 7.6] for a recent discussion of the method.

We would like to determine an approximate solution, $\boldsymbol{x}_j$, of the least-squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x} - \boldsymbol{b}\|, \tag{34}$$

such that $\boldsymbol{x}_j$ is an accurate approximation of $\hat{\boldsymbol{x}}$. Let W be the same matrix as in Section 2 and determine its QR-factorization (6). Similarly as we derived the linear systems of equations (7) and (8) from (1), we split the minimization problem (34) to obtain

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|P_Q A P_W \boldsymbol{x} + P_Q A P_W^{\perp}\boldsymbol{x} - P_Q\boldsymbol{b}\|, \tag{35}$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|P_Q^{\perp} A P_W^{\perp}\boldsymbol{x} - P_Q^{\perp}\boldsymbol{b}\|. \tag{36}$$

Note that the problem (35) is equivalent to the small linear system of equations (13). This suggests that (34) be solved as follows. We first determine an approximate solution $\boldsymbol{z}_j''$ of

$$\min_{\boldsymbol{z}''\in\mathbb{R}^n} \|P_Q^\perp A\boldsymbol{z}'' - P_Q^\perp \boldsymbol{b}\| \tag{37}$$

by j iterations with LSQR, where we use the initial approximate solution $\boldsymbol{z}_0'' = \boldsymbol{0}$. The minimization problem (37) differs from (36) only in that the computed solution is not required to be in the range of $P_W^\perp$. Compute $\boldsymbol{x}_j''$ using (11), solve (13) for $\boldsymbol{z}_j'$, determine $\boldsymbol{x}_j'$ by (14), and finally form the approximate solution $\boldsymbol{x}_j = \boldsymbol{x}_j' + \boldsymbol{x}_j''$ of (34).

LSQR allows inexpensive computation of the norm of the residual error $\|P_Q^\perp \boldsymbol{b} - P_Q^\perp A\boldsymbol{z}_j''\|$ for each iterate $\boldsymbol{z}_j''$. The equality (21) is valid also for the iterates $\boldsymbol{z}_j''$ and $\boldsymbol{x}_j$ determined by the method of this section. Therefore the norms $\|P_Q^\perp \boldsymbol{b} - P_Q^\perp A\boldsymbol{z}_j''\|$ can be used to decide when to terminate the iterations in the same way as for GMRES and RRGMRES in Section 2.

The iterate $\boldsymbol{z}_j''$ lives in the Krylov subspace $\mathbb{K}_j(A^T P_Q^\perp A, P_Q^\perp A^T \boldsymbol{b})$. Its computation requires the evaluation of j matrix-vector product with each one of the matrices A and A^T. Therefore the computation of $\boldsymbol{x}_j$ requires a total of $2j + \ell + 1$ matrix-vector product evaluations.

An augmented conjugate gradient method for the solution of (33) based on the CGLS algorithm has been described in [7]; see Björck [3, Section 7.4] for a discussion of CGLS. This section and Section 2 illustrate that the decomposition framework of the present paper is quite versatile and provides a unified approach to augmentation of Krylov subspace iterative methods.

4 Numerical examples

We present two computed examples that illustrate the performance of the methods described. The desired solution $\hat{\boldsymbol{x}}$ is available for both examples, and we use it to determine the error-free right-hand side

$$\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}} \tag{38}$$

of the linear system (3). The error vector $\boldsymbol{e}$ has normally distributed entries with zero mean and is scaled so that the contaminated right-hand side $\boldsymbol{b}$, defined by (2), has relative error

$$\|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\| = 1 \cdot 10^{-3}. \tag{39}$$

	Standard iterative methods			
	RRGMRES		LSQR	
j	$\lVert\hat{\boldsymbol{x}} - \boldsymbol{x}_j\rVert$	$\lVert\boldsymbol{b} - A\boldsymbol{x}_j\rVert$	$\lVert\hat{\boldsymbol{x}} - \boldsymbol{x}_j\rVert$	$\lVert\boldsymbol{b} - A\boldsymbol{x}_j\rVert$
1	$2.59 \cdot 10^{+0}$	$1.04 \cdot 10^{+0}$	$7.88 \cdot 10^{+0}$	$4.96 \cdot 10^{+0}$
2	$1.31 \cdot 10^{+0}$	$9.48 \cdot 10^{-2}$	$6.53 \cdot 10^{-1}$	$6.13 \cdot 10^{-2}$
3	$\underline{6.82 \cdot 10^{-2}}$	$\underline{4.18 \cdot 10^{-2}}$	$\underline{1.55 \cdot 10^{-1}}$	$\underline{4.18 \cdot 10^{-2}}$
4	$1.18 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$	$9.90 \cdot 10^{-1}$	$4.15 \cdot 10^{-2}$
5	$6.60 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$	$5.26 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$

	Decomposition methods with W given by (41)			
	RRGMRES		LSQR	
j	$\lVert\hat{\boldsymbol{x}} - \boldsymbol{x}_j\rVert$	$\lVert\boldsymbol{b} - A\boldsymbol{x}_j\rVert$	$\lVert\hat{\boldsymbol{x}} - \boldsymbol{x}_j\rVert$	$\lVert\boldsymbol{b} - A\boldsymbol{x}_j\rVert$
1	$5.89 \cdot 10^{-1}$	$5.71 \cdot 10^{-2}$	$5.65 \cdot 10^{-1}$	$5.65 \cdot 10^{-2}$
2	$\underline{4.99 \cdot 10^{-2}}$	$\underline{4.18 \cdot 10^{-2}}$	$\underline{1.43 \cdot 10^{-1}}$	$\underline{4.18 \cdot 10^{-2}}$
3	$1.19 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$	$9.79 \cdot 10^{-1}$	$4.15 \cdot 10^{-2}$
4	$6.95 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$	$5.32 \cdot 10^{+0}$	$4.15 \cdot 10^{-2}$
5	$1.11 \cdot 10^{+1}$	$4.15 \cdot 10^{-2}$	$3.16 \cdot 10^{+1}$	$4.15 \cdot 10^{-2}$

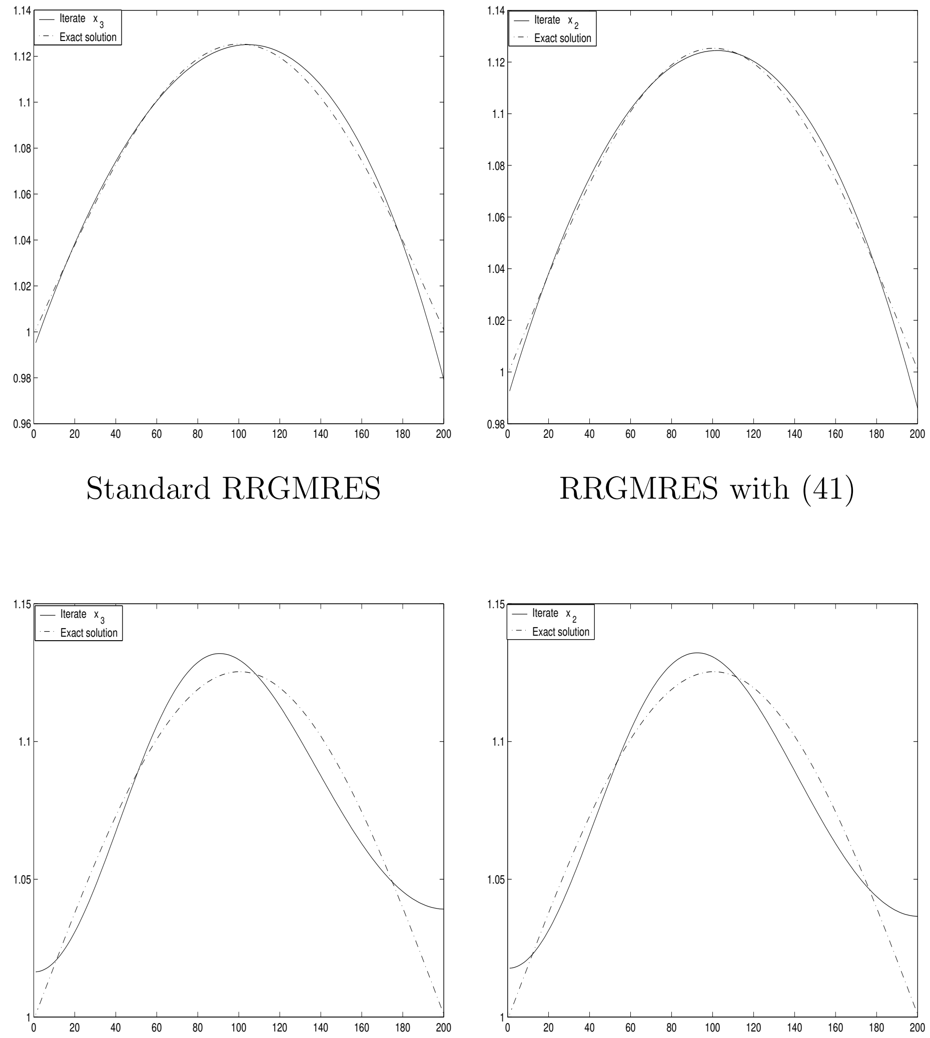
Table 1
Example 4.1. Error and residual error for computed iterates $\boldsymbol{x}_j$. Values for iterate determined by the discrepancy principle (20) underlined; $\|\boldsymbol{e}\| = 4.20 \cdot 10^{-2}$.

All computations were carried out in Matlab with machine epsilon $2 \cdot 10^{-16}$. The parameter η in (20) is set to unity in all examples.

Example 4.1. The Fredholm integral equation of the first kind

$$\int_0^\pi \exp(s\cos(t))x(t)dt = 2\frac{\sinh(s)}{s}, \qquad 0 \le s \le \frac{\pi}{2}, \tag{40}$$

discussed by Baart [1] is frequently used to illustrate the performance of numerical methods for the solution of ill-posed problems. We used the Matlab code baart from Regularization Tools [11] for the discretization of (40) by a Galerkin method with orthonormal box functions as test and trial functions, and determined the nonsymmetric matrix $A \in \mathbb{R}^{200\times 200}$ and the scaled discrete approximation $\breve{\boldsymbol{x}} \in \mathbb{R}^{200}$ of the solution $x(t) = \sin(t)$ of (40). The matrix A is of ill-determined rank; it has condition number $\kappa(A) = 5.2 \cdot 10^{18}$, where $\kappa(A) = \|A\|\|A^{-1}\|$. Let $\boldsymbol{c} = [1, 1, \ldots, 1]^T \in \mathbb{R}^{200}$, define $\hat{\boldsymbol{x}} = \breve{\boldsymbol{x}} + \boldsymbol{c}$, and let $\hat{\boldsymbol{b}}$ be given by (38). The error vector $\boldsymbol{e} \in \mathbb{R}^{200}$ was determined by the Matlab



Standard RRGMRES — RRGMRES with (41)

Standard LSQR — LSQR with (41)

Fig. 1. Example 4.1. Approximate solutions $\boldsymbol{x}_j$ determined by the discrepancy principle (20) using standard and augmented iterative methods (continuous curves) and the exact solution $\hat{\boldsymbol{x}}$ of the error-free system (3) (dash-dotted curves).

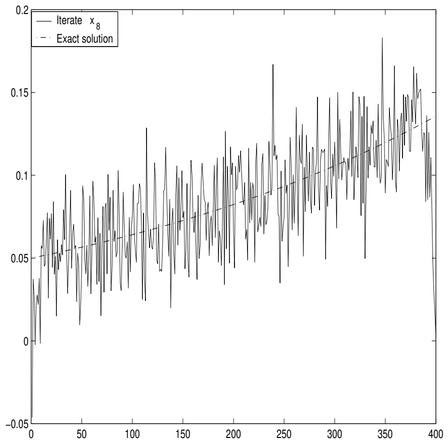
random number generator `randn` with seed 99 and scaled to satisfy (39). Then $\|\boldsymbol{e}\| = 4.20 \cdot 10^{-2}$. The right-hand side vector $\boldsymbol{b}$ in (1) is determined by (2).

Table 1 reports results for (standard) RRGMRES and LSQR, as well as for decomposition methods based on these iterative methods with

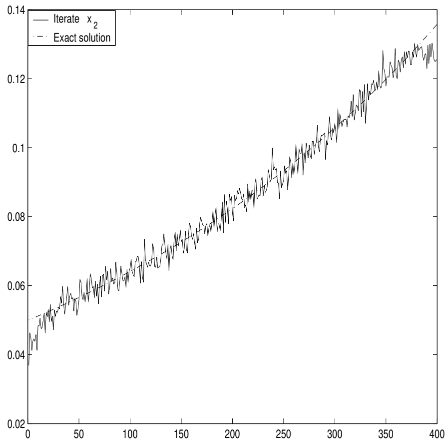
$$W = \frac{1}{\sqrt{200}}[1, 1, \ldots, 1]^T \in \mathbb{R}^{200}. \tag{41}$$

The table displays the residual error (21) and the error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j\|$ for $1 \leq j \leq 5$. The underlined values in the table mark the iterates determined by the discrepancy principle (20). These iterates approximate $\hat{\boldsymbol{x}}$ the best. They are displayed in Figure 1, which also shows the exact solution $\hat{\boldsymbol{x}}$ of the error-free problem (3). The table and figure show the decomposition methods to yield better approximations of $\hat{\boldsymbol{x}}$ than the standard iterative methods.

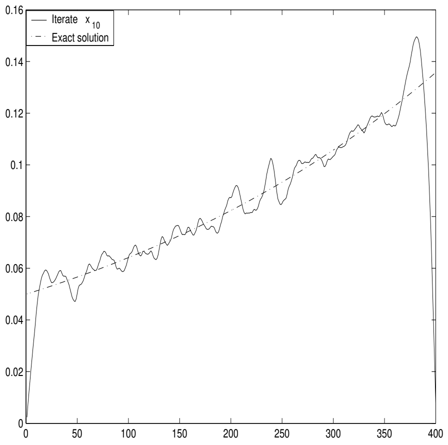
We remark that the success of decomposition with a given matrix W depends on the form of the solution. For instance, when the vector $\boldsymbol{c}$ is set to zero, the decomposition methods yield only minor improvements compared with the standard iterative methods. GMRES does not perform as well as RRGMRES for the present example, and we therefore do not report results for the former method. □



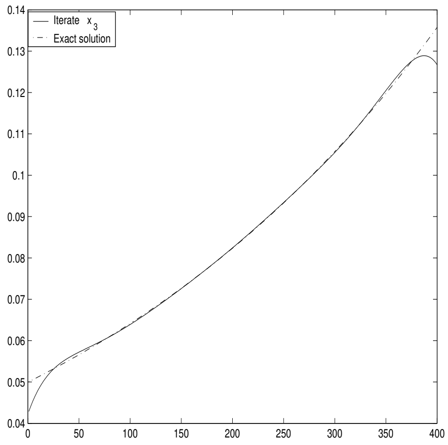
GMRES



GMRES with linear space W_2



RRGMRES



RRGMRES with linear space W_2

Fig. 2. Example 4.2. Approximate solutions $\boldsymbol{x}_j$ determined by the discrepancy principle (20) using standard GMRES and RRGMRES as well as GMRES- and RRGMRES-based decomposition methods (continuous curves), and the exact solution $\hat{\boldsymbol{x}}$ of the error-free system (3) (dash-dotted curves).

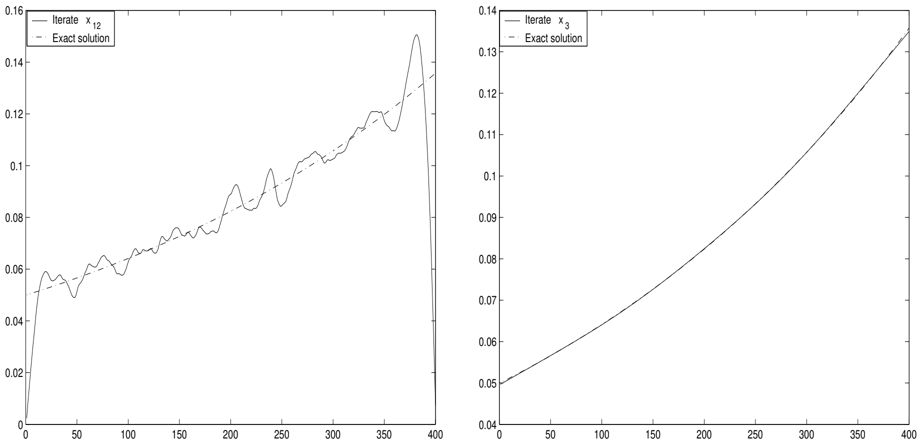
	Standard iterative methods					
	GMRES		RRGMRES		LSQR	
j	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$
1	$8.60\cdot10^{-1}$	$9.24\cdot10^{-3}$	$9.27\cdot10^{-1}$	$1.10\cdot10^{-2}$	$9.27\cdot10^{-1}$	$1.10\cdot10^{-2}$
2	$6.38\cdot10^{-1}$	$3.26\cdot10^{-3}$	$7.71\cdot10^{-1}$	$5.07\cdot10^{-3}$	$7.81\cdot10^{-1}$	$5.22\cdot10^{-3}$
3	$5.05\cdot10^{-1}$	$9.45\cdot10^{-4}$	$6.00\cdot10^{-1}$	$1.46\cdot10^{-3}$	$6.12\cdot10^{-1}$	$1.54\cdot10^{-3}$
4	$4.07\cdot10^{-1}$	$3.78\cdot10^{-4}$	$4.95\cdot10^{-1}$	$6.94\cdot10^{-4}$	$5.17\cdot10^{-1}$	$8.24\cdot10^{-4}$
5	$3.70\cdot10^{-1}$	$2.40\cdot10^{-4}$	$4.43\cdot10^{-1}$	$3.98\cdot10^{-4}$	$4.68\cdot10^{-1}$	$4.67\cdot10^{-4}$
6	$3.61\cdot10^{-1}$	$1.75\cdot10^{-4}$	$3.83\cdot10^{-1}$	$2.37\cdot10^{-4}$	$4.08\cdot10^{-1}$	$2.81\cdot10^{-4}$
7	$4.14\cdot10^{-1}$	$1.57\cdot10^{-4}$	$3.46\cdot10^{-1}$	$1.89\cdot10^{-4}$	$3.80\cdot10^{-1}$	$2.25\cdot10^{-4}$
8	**$5.22\cdot10^{-1}$**	**$1.51\cdot10^{-4}$**	$3.19\cdot10^{-1}$	$1.67\cdot10^{-4}$	$3.47\cdot10^{-1}$	$1.82\cdot10^{-4}$
9	$7.74\cdot10^{-1}$	$1.47\cdot10^{-4}$	$2.91\cdot10^{-1}$	$1.57\cdot10^{-4}$	$3.22\cdot10^{-1}$	$1.68\cdot10^{-4}$
10	$1.37\cdot10^{+0}$	$1.43\cdot10^{-4}$	**$2.78\cdot10^{-1}$**	**$1.53\cdot10^{-4}$**	$3.04\cdot10^{-1}$	$1.60\cdot10^{-4}$
11	$2.22\cdot10^{+0}$	$1.40\cdot10^{-4}$	$2.74\cdot10^{-1}$	$1.51\cdot10^{-4}$	$2.87\cdot10^{-1}$	$1.55\cdot10^{-4}$
12	$3.80\cdot10^{+0}$	$1.35\cdot10^{-4}$	$2.93\cdot10^{-1}$	$1.49\cdot10^{-4}$	**$2.79\cdot10^{-1}$**	**$1.53\cdot10^{-4}$**

	Decomposition methods with W given (43)					
	GMRES		RRGMRES		LSQR	
j	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$	$\|\hat{\boldsymbol{x}}-\boldsymbol{x}_j\|$	$\|\boldsymbol{b}-A\boldsymbol{x}_j\|$
1	$4.70\cdot10^{-2}$	$1.77\cdot10^{-4}$	$4.43\cdot10^{-2}$	$1.77\cdot10^{-4}$	$1.14\cdot10^{-2}$	$1.58\cdot10^{-4}$
2	**$6.49\cdot10^{-2}$**	**$1.54\cdot10^{-4}$**	$3.30\cdot10^{-2}$	$1.55\cdot10^{-4}$	$8.82\cdot10^{-3}$	$1.55\cdot10^{-4}$
3	$1.29\cdot10^{-1}$	$1.52\cdot10^{-4}$	**$2.86\cdot10^{-2}$**	**$1.54\cdot10^{-4}$**	**$3.08\cdot10^{-3}$**	**$1.54\cdot10^{-4}$**
4	$2.30\cdot10^{-1}$	$1.51\cdot10^{-4}$	$2.62\cdot10^{-2}$	$1.54\cdot10^{-4}$	$3.80\cdot10^{-3}$	$1.54\cdot10^{-4}$
5	$7.40\cdot10^{-1}$	$1.47\cdot10^{-4}$	$2.21\cdot10^{-2}$	$1.54\cdot10^{-4}$	$5.61\cdot10^{-3}$	$1.54\cdot10^{-4}$

Table 2
Example 4.2. Error and residual error for computed iterates $\boldsymbol{x}_j$. Values for iterate determined by the discrepancy principle (20) underlined; $\|\boldsymbol{e}\| = 1.54\cdot10^{-4}$.

Example 4.2. Consider the Fredholm integral equation of the first kind

$$\int_0^1 k(s,t)x(t)dt = \exp(s) + (1-e)s - 1, \qquad 0 \le s \le 1, \tag{42}$$



LSQR   LSQR with linear space W_2

Fig. 3. Example 4.2. Approximate solutions $\boldsymbol{x}_j$ determined by the discrepancy principle (20) using the standard LSQR method and an LSQR-based decomposition method (continuous curves), and the exact solution $\hat{\boldsymbol{x}}$ of the error-free system (3) (dash-dotted curves).

where

$$k(s,t) = \begin{cases} s(t-1), & s < t, \\ t(s-1), & s \geq t. \end{cases}$$

We discretized the integral equation by a Galerkin method with orthonormal box functions as test and trial functions using the Matlab program deriv2 from Regularization Tools [11] and obtained the symmetric indefinite matrix $A \in \mathbb{R}^{400\times400}$ and the scaled discrete approximation $\hat{\boldsymbol{x}} \in \mathbb{R}^{400}$ of the solution $x(t) = \exp(t)$ of (42). The error-free right-hand side vector is given by (38). The error vector $\boldsymbol{e} \in \mathbb{R}^{400}$ was determined by the Matlab random number generator randn with seed 111 and scaled to satisfy (39). Then $\|\boldsymbol{e}\| = 1.54 \cdot 10^{-4}$. The right-hand side vector $\boldsymbol{b}$ in (1) is given by (2).

Define $W \in \mathbb{R}^{400\times2}$ by QR-factorization of

$$\hat{W} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & 400 \end{bmatrix} \in \mathbb{R}^{400\times2}, \qquad \hat{W} = WR, \tag{43}$$

i.e., W has orthonormal columns and $R \in \mathbb{R}^{2\times2}$ is upper triangular.

Table 2 displays results obtained with (standard) GMRES, RRGMRES, and LSQR, as well as with decomposition methods defined by W and these iterative methods. The table shows the residual errors $\|\boldsymbol{b} - A\boldsymbol{x}_j\|$ and the errors $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j\|$ for several values of j. The values for the iterates determined by the discrepancy principle are underlined. Figures 2 and 3 show the approximate solutions defined by the underlined iterates as well as $\hat{\boldsymbol{x}}$. The decomposition methods are seen to determine better approximations of $\hat{\boldsymbol{x}}$ with fewer matrix-vector product evaluations with the matrices A and A^T than the standard iterative methods. The LSQR-based decomposition method is seen to furnish the best approximation of $\hat{\boldsymbol{x}}$. We have found that when $\hat{\boldsymbol{x}}$ is the discretization of a smooth function, RRGMRES- and LSQR-based decomposition methods often yield better approximations of $\hat{\boldsymbol{x}}$ than the GMRES-based decomposition method using the same matrix W. This is the case in the present example. Note that the discrepancy principle does not always determine the iterate $\boldsymbol{x}_{j_{\rm opt}}$, cf. (4), that best approximates $\hat{\boldsymbol{x}}$. $\Box$

In both examples above, the columns of the matrix W represent smooth functions. An example where W is used to represent a discontinuity and the GMRES-based iterative method performs well is reported in [2]. Numerical experiments suggest that the matrix W should be chosen so that its range contains a fairly good approximation of $\hat{\boldsymbol{x}}$.

5 Conclusion and future work

Decomposition provides a unified approach to augmentation methods for the solution of linear discrete ill-posed problems. The present paper discusses applications of decomposition to several iterative methods. Applications to direct solution methods are also possible and will be discussed elsewhere, and so will the choice of W.

References

[1] M. L. Baart, The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems, IMA J. Numer. Anal. 2 (1982) 241–247.

[2] J. Baglama, L. Reichel, Augmented GMRES-type methods, submitted for publication.

[3] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

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

[5] D. Calvetti, B. Lewis, L. Reichel, On the regularizing properties of the GMRES method, Numer. Math. 91 (2002) 605–625.

[6] D. Calvetti, B. Lewis, L. Reichel, GMRES, L-curves and discrete ill-posed problems, BIT 42 (2002) 44–65.

[7] D. Calvetti, L. Reichel, A. Shuibi, Enriched Krylov subspace methods for ill-posed problems, Linear Algebra Appl. 362 (2003) 257–273.

[8] L. Eldén, A weighted pseudoinverse, generalized singular values, and constraint least squares problems, BIT 22 (1982) 487–502.

[9] G. H. Golub, C. F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.

[10] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman Scientific and Technical, Essex, 1995.

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

[12] P. C. Hansen, Rank Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.

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

[14] C. C. Paige, M. A. Saunders, Algorithm 583 LSQR: Sparse linear equations and least squares problems, ACM Trans. Math. Software 8 (1982) 195–209.

[15] L. Reichel, Q. Ye, Breakdown-free GMRES for singular systems, SIAM J. Matrix Anal. Appl. 26 (2005) 1001–1021.

[16] Y. Saad, M. H. Schultz, GMRES: A generalized minimal residual method for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. 7 (1986) 856–869.