ALGORITHMS FOR RANGE RESTRICTED ITERATIVE METHODS FOR LINEAR DISCRETE ILL-POSED PROBLEMS

ARTHUR NEUMAN*, LOTHAR REICHEL†, AND HASSANE SADOK‡

Abstract. Range restricted iterative methods based on the Arnoldi process are attractive for the solution of large nonsymmetric linear discrete ill-posed problems with error-contaminated data (right-hand side). Several derivations of this type of iterative methods are compared in *A. Neuman, L. Reichel, and H. Sadok: Implementations of range restricted iterative methods for linear discrete ill-posed problems, Linear Algebra Appl., in press.* We describe MATLAB codes for the best of these implementations. MATLAB codes for range restricted iterative methods for symmetric linear discrete ill-posed problems are also presented.

Key words. linear discrete ill-posed problem, iterative method, truncated iteration,

1. Introduction.

We consider minimal residual iterative methods for the computation of approximate solutions of linear systems of equations

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

with a large nonsymmetric or symmetric matrix A with many "tiny" singular values of different orders of magnitude. In particular, A is severely ill-conditioned and may be singular. Linear systems of equations (1.1) with such a matrix are commonly referred to as linear discrete ill-posed problems. They arise, for instance, from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel.

The right-hand side vector in many linear discrete ill-posed problems that arise in science and engineering represents measured data and is contaminated by an error $e \in \mathbb{R}^m$. Let $\hat{b} \in \mathbb{R}^m$ denote the unknown error-free right-hand side associated with the right-hand side b of (1.1), i.e.,

$$b = \hat{b} + e. \tag{1.2}$$

Let $A^\dagger$ denote the Moore-Penrose pseudoinverse of the matrix A. We would like to determine an approximation of $\hat{x} = A^\dagger \hat{b}$ by computing an approximate solution of (1.1). Note that the vector $A^\dagger b = \hat{x} + A^\dagger e$ generally is a meaningless approximation of $\hat{x}$ because typically $A^\dagger e$ is of much larger norm than $\hat{x}$.

A meaningful approximation of $\hat{x}$ often can be computed by applying a Krylov subspace iterative method with initial iterate $x_0 = 0$ to the solution of (1.1) and by terminating the computations after sufficiently few iterations. We will comment on stopping criteria below.

GMRES is a popular iterative method for the solution of large nonsymmetric linear systems of equations that arise from the discretization of well-posed problems; see, e.g., Saad [13]. It has been observed in [2, 8] that when solving linear discrete ill-posed problems (1.1) with an error-contaminated right-hand side b, a variation of

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: aneuman@kent.edu

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu

‡Laboratoire de Mathématiques Pures et Appliquées, Université du Littoral, Batiment H. Poincarré, 50 Rue F. Buisson, BP 699, 62228 Calais cedex, France. E-mail: sadok@lmpa.univ-littoral.fr

GMRES, referred to as Range Restricted GMRES (RRGMRES), often delivers more accurate approximations of $\hat{x}$ than GMRES. RRGMRES requires all iterates to live in the range of A. A new implementation of RRGMRES has recently been described in [10], where also a variant for symmetric linear systems is described. The latter requires only short recursion relations. Therefore the storage requirement is independent of the number of iterations. We present MATLAB codes for these implementations.

This paper is organized as follows. Section 2 reviews the RRGMRES implementation proposed in [10] for nonsymmetric linear discrete ill-posed problems. Simplifications of the computations made possible when the matrix A in (1.1) is symmetric are also commented on. Section 3 discusses criteria for terminating the iterations, and Section 4 is concerned with breakdown of the iterations. The software is briefly described in Section 5. Further details on the software can be found in the primer.

2. RRGMRES. We first consider linear discrete ill-posed problems (1.1) with a nonsymmetric matrix A. The kth iterate, x_k, determined by RRGMRES with initial iterate $x_0 = 0$ solves the minimization problem

$$\min_{x \in \mathbb{K}_k(A,Ab)} \|Ax - b\|, \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean vector norm and

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

is a Krylov subspace.

We outline how the minimization problem (2.1) can be solved with the aid of the Arnoldi process. Application of k steps of the Arnoldi process to A with initial vector $v_1 = b/\|b\|$ gives the Arnoldi decomposition

$$AV_k = V_{k+1} \bar{H}_k, \tag{2.2}$$

where $V_{k+1} = [v_1, v_2, \ldots, v_k, v_{k+1}] \in \mathbb{R}^{m \times (k+1)}$ has orthonormal columns, which span the Krylov subspace

$$\mathbb{K}_{k+1}(A, b) = \text{span}\{b, Ab, \ldots, A^k b\}.$$

We assume this subspace to be of dimension $k + 1$. In the very rare event that $\dim(\mathbb{K}_{k+1}(A, b)) \leq k$, the Arnoldi process breaks down. This situation is discussed in Section 4. A thorough discussion on the Arnoldi process can be found in, e.g., [13].

The matrix $\bar{H}_k \in \mathbb{R}^{(k+1) \times k}$ in (2.2) is of upper Hessenberg form with nonvanishing subdiagonal entries. The decomposition (2.2) is the basis for the most common implementation of standard GMRES; see Saad [13].

Introduce the QR factorization

$$\bar{H}_k = Q_{k+1} \bar{R}_k, \tag{2.3}$$

where the matrix $Q_{k+1} \in \mathbb{R}^{(k+1) \times (k+1)}$ is orthogonal and upper Hessenberg, and the matrix $\bar{R}_k \in \mathbb{R}^{(k+1) \times k}$ has a leading $k \times k$ upper triangular submatrix, R_k, and a vanishing last row. Let the matrix $W_k \in \mathbb{R}^{m \times k}$ consist of the first k columns of $V_{k+1} Q_{k+1}$. Then it follows from (2.2) and (2.3) that

$$AV_k = W_k R_k, \tag{2.4}$$

which shows that the range of W_k is $\mathbb{K}_k(A, Ab)$. The minimization problem (2.1) therefore can be written as

$$\begin{aligned}\min_{y\in\mathbb{R}^k} \|AW_k y - b\| &= \min_{y\in\mathbb{R}^k} \|A(AV_k)R_k^{-1}y - b\| \\ &= \min_{y\in\mathbb{R}^k} \|AV_{k+1}\bar{H}_k R_k^{-1} y - b\| \\ &= \min_{y\in\mathbb{R}^k} \|V_{k+2}\bar{H}_{k+1}Q_{k+1}\bar{I}_k y - b\| \\ &= \min_{y\in\mathbb{R}^k} \|\bar{H}_{k+1}Q_{k+1}\bar{I}_k y - e_1\|b\|\,\|,\end{aligned}$$

where $\bar{I}_k \in \mathbb{R}^{(k+1)\times k}$ consists of the first k columns of the identity matrix of order $k+1$, and e_1 denotes the first axis vector. The last equality follows from $V_{k+2}e_1 = b/\|b\|$.

Since the matrices $\bar{H}_{k+1}$ and Q_{k+1} are upper Hessenberg, the matrix

$$\widetilde{H}_k = \bar{H}_{k+1}Q_{k+1}\bar{I}_k \in \mathbb{R}^{(k+2)\times k}$$

vanishes below the sub-subdiagonal. The QR factorization

$$\widetilde{H}_k = Q'_{k+2}\widetilde{R}'_k \tag{2.5}$$

therefore can be computed in only $\mathcal{O}(k^2)$ arithmetic floating point operations. Here $Q'_{k+2} \in \mathbb{R}^{(k+2)\times(k+2)}$ is orthogonal and $\widetilde{R}'_k \in \mathbb{R}^{(k+2)\times k}$ has a leading $k \times k$ upper triangular submatrix and two vanishing last rows. We obtain

$$\min_{y\in\mathbb{R}^k} \|AW_k y - b\| = \min_{y\in\mathbb{R}^k} \|\widetilde{R}'_k y - (Q'_{k+2})^T e_1\|b\|\,\|. \tag{2.6}$$

An algorithm with MATLAB-like notation is described in [10], where also numerical examples are presented. The latter illustrate the superior performance of this implementation compared with other ones. The RRGMRES method described above stores the matrix V_{k+1} when computing the iterate x_k. Thus, the storage requirement grows with the number of iterations, and so does the computational effort per iteration. This can make it attractive to restart the algorithm periodically.

When A is symmetric, the matrix $\widetilde{H}_k$ is pentadiagonal. This makes it possible to implement the iterative method above with short recursion formulas. The number of terms of the recursion relation is bounded independently of the number of iterations. Therefore, the computational effort and storage requirement for the method can be bounded independently of the number of iterations k.

3. Stopping criteria. Assume that a bound δ for the norm of the error e in b is known. The discrepancy principle prescribes that we terminate the iterations with RRGMRES as soon as an iterate x_k that satisfies

$$\|Ax_k - b\| \le \eta\delta \tag{3.1}$$

has been found, where $\eta > 1$ is a user-specified constant independent of δ. This iterate is our computed approximation of $\hat{x}$. Properties of GMRES when used with this stopping criterion are discussed in [3]. Similar results hold for RRGMRES. A general discussion on the discrepancy principle can be found, e.g., in [5]. We provide MATLAB codes that use the stopping criterion (3.1). Iterations are carried out until the reduced minimization problem in the right-hand side of (2.6) for the first time has a residual error of norm bounded by $\eta\delta$.

When no bound δ for $\|e\|$ is available, another stopping criterion has to be used. Stopping rules based on the L-curve or Generalized Cross Validation (GCV) are popular options; see, e.g., [6] for discussions of these rules. We provide MATLAB codes that carry out a user-specified number of iterations. These codes are convenient to use with L-curve- or GCV-based stopping criteria, as well as when RRGMRES is applied in a restarted fashion.

4. Breakdown. The rare situation when the computations with the Arnoldi process (2.2) cannot be continued is referred to as breakdown. We first discuss this situation when A is nonsingular.

Assume that the Arnoldi process breaks down when seeking to generate the column v_{k+2} of V_{k+2}. Then the decomposition (2.2) with k replaced by $k+1$ reads

$$AV_{k+1} = V_{k+1}H_{k+1}, \tag{4.1}$$

where $V_{k+1} = [v_1, v_2, \ldots, v_{k+1}]^T \in \mathbb{R}^{m\times(k+1)}$ has orthonormal columns and $H_{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$ is of upper Hessenberg form. It follows from (4.1) that the spectrum of H_{k+1} is a subset of the spectrum of A. In particular, H_{k+1} is nonsingular, and therefore every columns of V_{k+1} lives in the range of A. It follows that

$$\min_{x\in\mathbb{K}_{k+1}(A,Ab)} \|Ax-b\| = \min_{y\in\mathbb{R}^{(k+1)}} \|AV_{k+1}y-b\| = \min_{y\in\mathbb{R}^{(k+1)}} \|H_{k+1}y - e_1\|b\|\,\| = 0,$$

where we have used (4.1) and the fact that H_{k+1} is nonsingular. Thus, in case of breakdown the minimization problem (2.1) can be solved exactly.

The handling of breakdown when A is singular is more complicated. An approach to determine a least-squares solution in this situation is described in [12] for standard GMRES. This approach can be adapted to the minimization problem of the present paper. We omit the details since breakdown is very rare.

5. Software. We provide MATLAB codes for the RRGMRES implementations described in [10] for the iterative solution of symmetric and nonsymmetric linear discrete ill-posed problems (1.1). The computations are terminated after a specified number of iterations or according to the discrepancy principle. Demos with a graphical user interface are included. The software is available from Netlib at

`http://www.netlbib.org/numeralgo/`

as the na33 package and is stored as a compressed archive in the file `rrgmrestbx.zip`. Installation details can be found in the `README.txt` file and in the primer of the package.

It is convenient to test the iterative methods on linear discrete ill-posed problems generated with MATLAB functions of Regularization Tools [7], such as the functions `baart.m`, `deriv2.m`, `phillips.m`, and `shaw.m`. The discrete ill-posed problems determined with these function are obtained by discretizing Fredholm integral equations of the first kind discussed in [1, 4, 9, 11, 14]. We provide MATLAB functions that discretize these integral equations in a different manner in order to have more test problems.

Acknowledgement. We would like to thank Giuseppe Rodriguez for valuable comments.

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), pp. 241–247.
[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, B. Lewis, and L. Reichel, On the regularizing properties of the GMRES method, Numer. Math., 91 (2002), pp. 605–625.
[4] L. M. Delves and J. L. Mohamed, Computational Methods for Integral Equations, Cambridge University Press, 1985; p. 310.
[5] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[6] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[7] P. C. Hansen, Regularization Tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007), pp. 189–194.
[8] P. C. Hansen and T. K. Jensen, Noise propagation in regularizing iterations for image deblurring, Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220.
[9] A. K. Louis and P. Maass, A mollifier method for linear operator equations of the first kind, Inverse Problems, 6 (1990), pp. 427-440.
[10] A. Neuman, L. Reichel, and H. Sadok, Implementations of range restricted iterative methods for linear discrete ill-posed problems, Linear Algebra Appl., in press.
[11] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM, 9 (1962), pp. 84–97.
[12] L. Reichel and Q. Ye, Breakdown-free GMRES for singular systems, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1001–1021.
[13] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, Philadelphia, 2003.
[14] C. B. Shaw, Jr., Improvements of the resolution of an instrument by numerical solution of an integral equation, J. Math. Anal. Appl., 37 (1972), pp. 83–112.