East Asian Math J 15(1999), No 2, pp 345-354

ORTHOGONALIZATION PROCESS USING SYSTEMS

SUCHEOL YI

ABSTRACT Orthogonalization can be done by the well known Gram-Schmidt process or by using Householder transformations. In this paper, we introduced an alternative process using linear systems

1. Introduction

A given set of linearly independent vectors can be replaced by an orthonormal basis that spans the same space. One of the well known algorithm for this purpose is the Gram-Schmidt process and it is widely used in algorithms for solving linear systems. We first consider iterative methods for solving large linear systems that use orthogonalization process to generate basis vectors for some spaces The GMRES method [8] given by Saad and Schultz is a Krylov subspace method for solving a linear system

$$Ax = b, \quad \text{where} \quad A \in R^{n\times n} \quad \text{is nonsingular.} \tag{1}$$

The kth iterate of GMRES is defined as $x_k = x_0 + z_k$ for a given initial guess $x_0 \in R^n$ and the correction z_k is chosen to minimize the norm of the residual vector $r(z) = r_0 - Az$, where $r_0 = b - Ax_0$, over the kth Krylov subspace $K_k(r_0, A) \equiv \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$ at the kth step, i e ,

$$\|r_0 - Az_k\|_2 = \min_{z\in K_k(r_0,A)} \|r_0 - Az\|_2. \tag{2}$$

Received April 15, 1999

Key words and phrases GMRES, QMR, symmetric QMR, and Krylov subspace methods

In the GMRES method, the kth correction z_k is determined by maintaining an orthonormal basis for the space $K_k(r_0, A)$ together with a $(k+1) \times k$ upper Hessenberg matrix H_k. An orthonormal basis for $K_k(r_0, A)$ can be obtained by the Arnoldi process and it can be summarized as follows:

ALGORITHM 1.1 Arnoldi Process

Initialize: Choose an initial vector v_1 with $\|v_1\|_2 = 1$.

Iterate: For $k = 1, 2, \ldots$, do:

$h_{i,k} = v_i^T A v_k, i = 1, 2, \ldots, k,$

$\tilde{v}_{k+1} = A v_k - \sum_{i=1}^{k} h_{i,k} v_i.$

Set $h_{k+1,k} = \|\tilde{v}_{k+1}\|_2$.

If $h_{k+1,k} = 0$, stop, otherwise,

$v_{k+1} = \tilde{v}_{k+1} / h_{k+1,k}$

For convenience, classical Gram-Schmidt orthogonalization is used in the algorithm above. In practice, it is usual to implement Arnoldi's method using the modified Gram-Schmidt process [6], which has superior computational properties. Even if the classical Gram-Schmidt process were replaced by the modified Gram-Schmidt process it is known that it may fail to perform well if the vectors on which it operates are not fairly independent as measured by the condition number. Björck [1] has shown that orthogonalization based on Householder transformations has great reliability and Walker [9] suggested an implementation using Householder transformations in the Arnoldi process.

The essential difficulty with the GMRES method is increasing storage and cost per iteration in applying the Arnoldi process due to the calculations of inner products and other computations in the long recursion formula when k gets larger. Restarting the Arnoldi process periodically every m iterations for some suitable value of m could be a possible remedy. The resulting restarted version of GMRES is denoted by GMRES(m). If A is symmetric in Arnoldi's method, then we have a short recursion formula In this symmetric case, Arnoldi's method with short recursion is often called the symmetric Lanczos process. MINRES [7] can be viewed as a specialization of the usual GMRES approach to the symmetric case, in which the short recurrence symmetric Lanczos process is used to generate an orthonormal basis for

$K_k(r_0, A)$.

For the nonsymmetric Lanczos method, one starts with two nonzero vectors v_1 and w_1 and then generates basis vectors $\{v_j\}$ for $K_k(v_1, A)$ and $\{w_j\}$ for $K_k(w_1, A^T)$ such that the bi-orthogonality condition

$$w_j^T v_k = \begin{cases} \delta_{j,k} \neq 0 \text{ if } k = j \\ 0 \quad \text{otherwise} \end{cases}$$

holds. The point is that the two bases can be built with just three-term recurrences, but we will lose the orthogonality of the basis elements. The BCG method given by Fletcher [2] is a typical example of applying the nonsymmetric Lanczos process. The QMR method introduced by Freund and Nachtigal [3] is another Krylov subspace method for solving linear systems, which is based on the nonsymmetric Lanczos process. The solutions of QMR can be obtained by solving a quasi-minimization problem. In the symmetric indefinite case without preconditioning, symmetric QMR [4] is obtained using the same approach as MINRES. However, in solving the systems of the preconditioned system

$$A'y = b', \; x = M_2^{-1}y, \text{ where } A' = M_1^{-1}AM_2^{-1} \text{ and } b' = M_1^{-1}b,$$

symmetric QMR is also implemented by solving a quasi-minimization problem and it is clear that symmetric QMR is equivalent to MINRES with no preconditioning.

The GMRES method is the most widely used for solving general linear systems and symmetric QMR is a very effective method for solving symmetric indefinite linear systems. In section 2, we establish a theoretical result about Ritz values. In section 3, we give a new approach of orthogonalization process by solving linear systems and we present numerical experiments in section 4.

2. On theoretical behavior

In the standard GMRES implementation, the Arnoldi process starting with $v_1 = r_0/\|r_0\|_2$ is applied and it can be readily seen that the vectors $v_1, \ldots, v_k$ that are generated by Arnoldi's method form an orthonormal basis for the space $K_k(r_0, A)$. The following relation

$$AV_k = V_{k+1}H_k \tag{3}$$

is also obtained from the Arnoldi process, where $V_k = (v_1, \ldots, v_k)$ and H_k is a $(k+1) \times k$ upper Hessenberg matrix constructed by the orthonormalizing scalars $h_{i,k}$. Any vector $z \in K_k(r_0, A)$ can be parametrized as $z = V_k y$ for some $y \in R^k$ and then the least-squares problem (2) is equivalent to

$$\min_{y \in R^k} \| \|r_0\|_2 e_1^{k+1} - H_k y \|_2 \tag{4}$$

by the equation (3), where e_1^{k+1} is the first column of the identity matrix I_{k+1}. This upper Hessenberg-least-squares problem can be solved by using Givens rotations because of the form of H_k.

DEFINITION 2.1. Suppose that $S \in R^{n \times n}$ is symmetric and $U \in R^{n \times k}$ is such that $U^T U = I$. If

$$Z^T (U^T S U) Z = \text{ diag } (\lambda_1, \ldots, \lambda_k)$$

is the Schur decomposition of $U^T S U$, then the λ_i are called Ritz values and y_i, the ith column of UZ, are called Ritz vectors.

We can see that $P_k = V_k V_k^T$ is the unique orthogonal projection onto $K_k(r_0, A)$ and it is clear that $V_k^T P_k A V_k = V_k^T P_k A P_k V_k$ is symmetric if A is symmetric. Then it follows that there exists an orthogonal matrix Q_k such that

$$Q_k^T (V_k^T P_k A V_k) Q_k = \text{ diag } (\theta_1, \ldots, \theta_k) = D_k. \tag{5}$$

On the other hand, we have

$$\begin{aligned} V_k^T P_k A V_k &= V_k^T P_k V_{k+1} H_k \quad \text{by the relation (3)} \\ &= V_k^T V_k V_k^T V_{k+1} H_k \\ &= \bar{H}_k, \end{aligned} \tag{6}$$

where $\bar{H}_k$ is the $k \times k$ upper Hessenberg matrix that is obtained by deleting the last row of H_k. In fact, it can be easily shown that $\bar{H}_k$ is a tridiagonal symmetric matrix if A is symmetric. Then the equation

$Q_k^T \bar{H}_k Q_k = D_k$ is satisfied by the equations (5) and (6), which implies that the θ_i are also eigenvalues of $\bar{H}_k$.

On the subspace $K_k(r_0, A)$ of R^n, we also have

$$P_k A = P_k A P_k = V_k \bar{H}_k V_k^T. \tag{7}$$

Any vector z in $K_k(r_0, A)$ can be written as $z = V_k y$ for some $y \in R^k$. Then we can see that the equation $P_k A z = \lambda z$ is equivalent to the equation $\bar{H}_k y = \lambda y$ and it also follows that Ritz values are eigenvalues of $P_k A$ if we consider eigenvectors only over the k-dimensional subspace $K_k(r_0, A)$ of R^n. We summarize this result in the following theorem.

THEOREM 2.2 *If A is symmetric and the columns of V_k are the Arnoldi basis vectors with $v_1 = r_0/\|r_0\|_2$ for $K_k(r_0, A)$, then the Ritz values (eigenvalues of $P_k A$ in that we consider eigenvectors of that $n \times n$ matrix only over the k-dimensional subspace $K_k(r_0, A)$) are the same as the eigenvalues of $\bar{H}_k$.*

3. On orthogonalization

Suppose that we have linearly independent vectors $v_1, \ldots, v_k$ which are all in R^n. For a given vector v to be orthogonalized against the vectors $v_1, \ldots, v_k$, we set $w = v - \sum_{i=1}^{k} \alpha_i v_i$ and find $\alpha = (\alpha_1, \ldots, \alpha_k)^T$ such that $v_i^T w = 0$ for $i = 1, \ldots, k$, i.e., we need to find vectors w and α such that (i) $v = w + V_k \alpha$, where $V_k = (v_1, \ldots, v_k)$ and (ii) $V_k^T w = \theta$, where θ is the zero vector in R^k. Then the orthogonalization process is equivalent to solving the following symmetric linear system

$$\begin{pmatrix} I_n & V_k \\ V_k^T & 0_k \end{pmatrix} \begin{pmatrix} w \\ \alpha \end{pmatrix} = \begin{pmatrix} v \\ \theta \end{pmatrix}, \tag{8}$$

where I_n is the identity matrix of dimension $n \times n$ and 0_k is the zero matrix of dimension $k \times k$.

Suppose that we have an $n \times n$ real matrix A and partition A as $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. If A_{11} is nonsingular, then it is known that the

partitioned matrix A can be factorized as

$$A = \begin{pmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ 0 & S \end{pmatrix},$$

where $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ is called the Schur complement of A_{11} in A and we assume that the sizes of all matrices in the factorization are appropriate for matrix multiplication. Then the coefficient matrix in (8) can be factorized as

$$\begin{pmatrix} I_n & V_k \\ V_k^T & 0_k \end{pmatrix} = \begin{pmatrix} I_n & \bar{0}_k \\ V_k^T & I_k \end{pmatrix} \begin{pmatrix} I_n & V_k \\ \bar{0}_k^T & -V_k^T V_k \end{pmatrix},$$

where the size of the zero matrix $\bar{0}_k$ is $n \times k$. Therefore, solving the symmetric linear system (8) is equivalent to solving the following non-symmetric linear system

$$\begin{pmatrix} I_n & V_k \\ \bar{0}_k^T & -V_k^T V_k \end{pmatrix} \begin{pmatrix} w \\ \alpha \end{pmatrix} = \begin{pmatrix} v \\ -V_k^T v \end{pmatrix}. \tag{9}$$

Then solving equations (8) and (9) may offer a new approach to orthogonalization.

4. Numerical experiments

We present numerical experiments that show the performance of orthogonalization by solving linear systems. We used MATLAB and used double precision on Sun Microsystems workstations in all experiments.

To generate a matrix V having 100 linearly independent columns, we set $V = \text{rand}(4000, 100)$ and then V is the matrix of dimension 4000×100 with random entries chosen from a uniform distribution on the interval (0,1). For a vector v to be orthogonalized against the columns of the matrix V, we set $v = u + V\alpha$, where $u = \text{eps}(\text{rand}(4000, 1) - 0\ 5\text{ones}(4000, 1)), \alpha = \text{rand}(100, 1)$, eps is the machine epsilon, and ones(4000,1) is the vector having 4000 components which are all ones. The vector v is then nearly in the space $\text{span}\{v_i | i = 1, \ldots, 100\}$, where v_i is the ith column of V. Then the resulting systems have dimensions

4100×4100. We also used the vector $x = \text{ones}(4100, 1)$ for the initial guess in applying the symmetric QMR and GMRES methods to solve equations (8) and (9) with no preconditioning. In all experiments, we used solid and dashdot curves for the true residual norms generated by symmetric QMR and GMRES when they are applied to linear systems (8) and (9) with $V_k = V$, respectively. Figure 1 shows that both symmetric QMR and GMRES solve the linear systems (8) and (9) successfully up to approximately 10^{-10} level of residual norm reduction

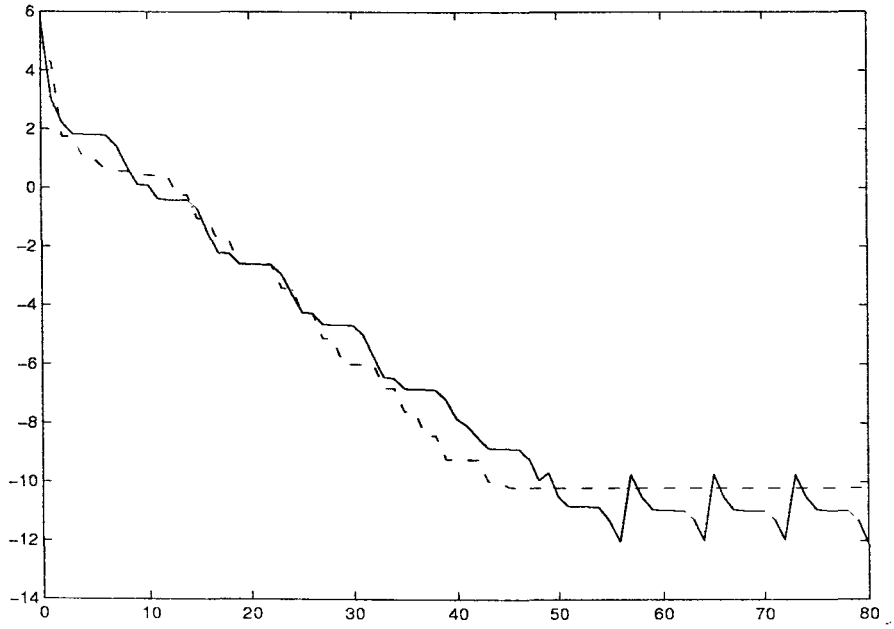

Figure 1: Log_{10} of the true residual norms vs. iterations. Solid curve: GMRES applied to (9); dashdot curve: symmetric QMR applied to (8)

In the following Figure 2, we plotted the residual norm reductions versus floating-point operation counts for symmetric QMR and GMRES. As shown in Figure 2, symmetric QMR and GMRES need about

the same number of operations to reach around 10^{-10} level of residual norm reduction, although GMRES needs fewer operations than the symmetric QMR method does.

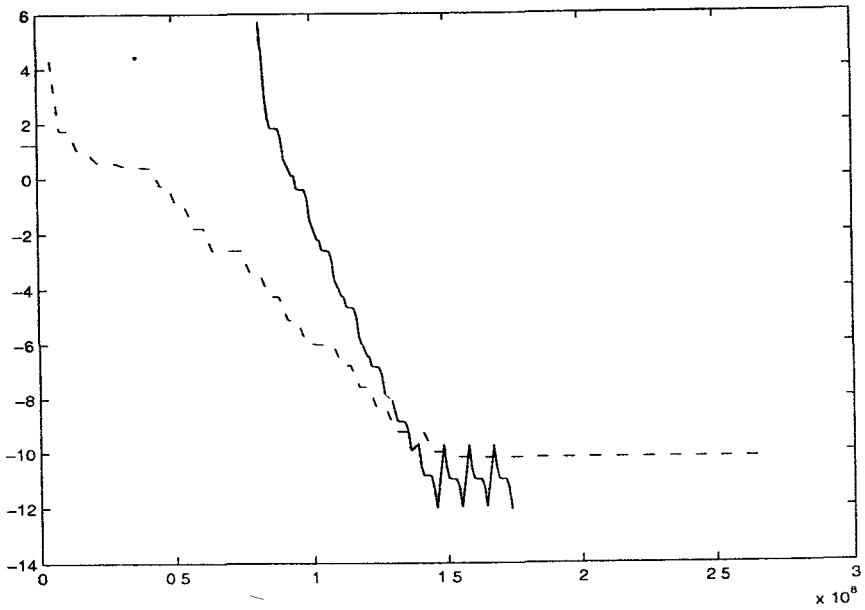

Figure 2. Log_{10} of the true residual norms vs. the number of floating-point operations. Solid curve: GMRES applied to (9); dashdot curve: symmetric QMR applied to (8).

However, solving equation (8) can be preferred, since the condition number of the coefficient matrix in equation (9) might be larger than that of the coefficient matrix in equation (8) because of the matrix $V_k^T V_k$ and roundoff errors can be arisen in the calculation of that matrix. Here, the condition number is defined as the ratio of the largest and smallest singular values. As we can see, the linear system (9) can be solved blockwise, i.e., we solve the following equation for α

$$V_k^T V_k \alpha = V_k^T v \tag{10}$$

and set $w = v - V_k\alpha$. Since the coefficient matrix $V_k^T V_k$ in equation (10) is symmetric positive definite, we can use the Cholesky decomposition to solve equation (10), which is a direct method, because the dimension of the coefficient matrix in (10) is small in our test example We investigated condition numbers of the matrices $A = (V, v_s)$ and $B = (V, v_c)$, where v_s and v_c are vectors obtained by solving (8) and (10) using the symmetric QMR method with stopping tolerance 10^{-10} for residual norm reductions and Cholesky decomposition, respectively. We found that the condition numbers of matrices A and B are approximately 6.7941e+12 and 5.3300e+14, respectively. We also found that Euclidean norms of the vectors $V_k^T v_s$ and $V_k^T v_c$ are approximately 4.1369e-11 and 8 4901e-10, respectively. The above observation implies that we may have a better accuracy by solving equation (8) using symmetric QMR, even though the symmetric QMR method needs about 40% more operations to reach around 10^{-10} level of true residual norm reduction than solving equation (10) with Cholesky decomposition In our experiments, we could find that solving equation (9) with GMRES gives only almost the same amount of accuracy that solving the normal equation (10) with Cholesky decompositon produces. We think this phenomenon can be explained by condition numbers, the condition number of the coefficient matrix in (8) is about 1 8654e+4, whereas that of the coefficient matrix in (9) is approximately 1 0484e+5, i.e., the linear system (8) is better conditioned than the linear system (9) is.

5. Conclusion

In this paper, we have considered Krylov subspace methods for solving large linear systems and have introduced a new approach for orthogonalization that use linear systems. Our orthogonalization may be useful when we orthogonalize a vector against a given set of linearly independent vectors for a better accuracy, even though solving linear systems (8) by iterative linear solvers require more floating-point operations than solving the normal equation (10) by direct methods.

REFERENCES

[1] Å Bjorck, *Solving linear least-squares problems by Gram-Schmidt orthogonalization*, BIT **7** (1967), 1–21.

[2] R Fletcher, *Lecture Note in Mathematics*, vol 506, Springer-Verlag, Berlin, 1976, pp 73–89.

[3] R W Freund and N M Nachtigal, *QMR: A quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math. **60** (1991), 315–339

[4] R W. Freund and N M. Nachtigal, *A new Krylov subspace method for symmetric indefinite linear systems*, ORNL/TM-12754 (1994)

[5] R. W Freund and H. Zha, *Simplifications of the nonsymmetric Lanczos process and a new algorithm for Hermitian indefinite linear systems*, Murray Hill Bell Labs (1994)

[6] G H Golub and C. F Van Loan, *Matrix Computations*, The Johns Hopkins University press, 1989

[7] C C Paige and M. A Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J Numer. Anal. **12** (1975), 617–629.

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

[9] H. F. Walker, *Implementation of the GMRES method using Householder transformations*, SIAM J Sci Stat. Comput **9** (1988), 152–163

Department of Applied Mathematics
Changwon National University
Changwon 641-773, Korea
E-mail: scyi@sarim.changwon.ac.kr