Krylov Subspace Iterative Methods for Nonsymmetric Indefinite Linear Systems

A.T. Chronopoulos *

Abstract

The Arnoldi, the Odir and the GMRES methods have been proposed as iterative methods to solve general nonsymmetric linear systems. Since the number of direction vectors to be kept in storage can increase up to the dimension of the linear system the restarted methods or the truncated methods are used. Here we study these restarted methods for nonsymmetric indefinite problems and we prove convergence under certain conditions on the matrix of coefficients. We also derive the Minimal IOM(k) method which is based on the truncated Arnoldi method (IOM(k)) with an error minimization property.

*The author is an assistant professor at the Computer Science Dept., University of Minnesota, Minneapolis, Minnesota 55455. This work was supported by NSF (CCR-8722260) and by the AHPCRC at the University of Minnesota

1 Introduction

In this article we consider a linear system of equations

$$Ax = f \tag{1}$$

where A is a nonsingular nonsymmetric matrix of order n. If the symmetric part of A is positive definite then the matrix will be called *definite*. We define the minimal polynomial of a nonzero vector v with respect to matrix A to be the least degree monic polynomial $q_k(\lambda)$ so that $q_k(A)v = 0$.

Krylov subspace iterative methods for a nonsymmetric linear system obtain approximations to its solution by projecting onto subspaces generated by subsets of the Krylov vectors $\{r_0, Ar_0, \ldots, A^i r_0, \ldots\}$. Some of these iterative methods minimize (at each iteration) an error functional over the projected subspaces. This property will be called *error minimization* property. This property guarantees monotone convergence to the solution. Various formulations of Krylov subspace iterative methods have been derived by Arnoldi, Axelsson, Concus and Golub, S. Eisenstat, H. Elman, M. Schultz, Saad, Vinsome and Young (see [1], [3], [7], [8], [11], [12], [13], [14]). For a taxonomy of the various methods see [2]. Here we consider Arnoldi, Odir, and GMRES which can be applied to general nonsymmetric matrices (see [1], [14], [11], [12]). These methods have been proposed in the following forms. A *full orthogonalization* method in which the size of the projected Krylov subspace increases by one at each iteration. A *restarted* method in which m iterations are completed in a *cycle* and then the method restarts. A *truncated* method in which each new direction vector is orthogonalized against the $k-1$ preceding direction vectors. Let Arnoldi, Odir and GMRES denote the full orthogonalization methods Arnoldi(m), Odir(m) and GMRES(m) the restarted methods and IOM(k), Odir(k) and MIOM(k) (Minimal IOM(k) see [6]) denote the truncated methods.

In this article we prove that Odir(m) and GMRES(m) converge for some nonsymmetric indefinite problems and we give residual error bounds. We propose an implementation of GMRES(m) using the Householder QR in solving the linear least squares problems (of dimension m). We introduce the Minimal IOM(k) (MIOM(k)). MIOM(k) uses the truncated Arnoldi direction vectors and has an error minimization property. We study its convergence properties and give a recursive implementation similar to IOM(k) (see [11]). In section 2 we review the s-step Minimal Residual method and present

convergence theorems for nonsymmetric indefinite matrices. In section 3 we review the Odir method and prove results on the convergence of Odir(m). In section 4 we present the GMRES with a Householder QR approach to solve the associated linear least squares problems and prove convergence results. In section 5 we introduce MIOM(k) and study its convergence properties.

2 The s-step Minimal Residual Method

The s-step Minimal Residual method (s-MR) is equivalent to GMRES(s) and Odir(s) (see [5]). It has been proved that s-MR converges for definite matrices, all symmetric and skew-symmetric matrices and for a class of non-symmetric indefinite matrices (see [5]).

Let x_i and r_i be the solution and residual vectors at the i-th iteration of s-MR. The $i+1$-the iteration consists of

$$x_{i+1} = x_i + {a_i}^1 r_i + \ldots + {a_i}^s A^{s-1} r_i \tag{2}$$

$$r_{i+1} = r_i - {a_i}^1 r_i - \ldots - {a_i}^s A^{s-1} r_i \tag{3}$$

where the scalars ${a_i}^j, j = 1, \ldots, s$ are selected so that x_{i+1} minimizes $E(x)$ over the affine subspace $L_i = \{x_i + \sum_{j=0}^{s-1} a_j A^j r_i, \quad a_j \in \mathbf{R}\}$.

Remark 2.1: The solution x_i in s-MR equals (in exact arithmetic) x_{im} in Odir(m) and GMRES(m) with $m = s$ because these methods approximate the solution using the same Krylov subspace.

For the matrix A let $M = (A + A^T)/2$ and $N = (A - A^T)/2$ be its symmetric and skew symmetric parts. The matrix M^2 (N^2) is symmetric and positive (negative) definite (see [5]). Let the eigenvalues of a symmetric matrix B of dimension n be denoted by $\lambda_n(B) \leq \ldots \leq \lambda_1(B)$. The following convergence result is of interest for nonsymmetric indefinite matrices.

Theorem 2.1: Assume that the degree of the minimal polynomial r_0 is greater than $2 \leq s$. If (a) $d = \lambda_n(M^2) + \lambda_n(N^2)$ or (b) $d = -[\lambda_1(N^2) + \lambda_1(M^2)] > 0$, then the matrix A^2 is definite and s-MR converges to the solution. The residuals satisfy

$$||r_{i+1}||_2^2 \leq c||r_i||_2^2 \,, \tag{4}$$

where $c = [1 - d^2/\lambda_1(A^{2T}A^2)]$.
Proof: In [5].

We prove the following convergence result for s-MR for normal nonsymmetric indefinite matrices. This result is of special interest in classifying the class of matrices for which Odir and GMRES take many steps before any significant residual norm reduction is observed (see [10]).

Theorem 2.2: Assume that the degree of the minimal polynomial of r_0 is greater than or equal to 8 and the matrix A is normal indefinite, with eigenvalues $\lambda_j = \alpha_j \pm \mathbf{i}\beta_j$, where $\mathbf{i} = \sqrt{-1}$ and $j = 1, \ldots, n$. Define the following regions in the (α-β) complex plane: (i)the east and west open cone defined by $\beta = \pm\alpha$ (ii) the open cones in the complement of region (i), (iii) the two two-sided smallest open cones defined by $\beta = .198913\alpha$ and $\beta = .668179\alpha$ and $\beta = -.198913\alpha$ and $\beta = -.668179\alpha$, (iv) the open cones in the complement of region (iii) Then the following statements hold:
(a) If $2 \leq s$ and the eigenvalues of A are confined to either region (i) or (ii), then s-MR converges. Furthermore, the associated residuals satisfy

$$||r_{i+1}||_2^2 \leq c^{1/\nu}||r_i||_2^2 \tag{5}$$

where $c = [1 - d^\nu/\lambda_1(A^{\nu T}A^\nu)]$, $\nu = 2$ and d is given by

$$d = \mathrm{Min}_{1\leq j\leq n}\{|\alpha_j^2 - \beta_j^2|\} > 0$$

(b) If $8 \leq s$ and the eigenvalues of A are confined to either region (iii) or (iv), then s-MR converges. Furthermore, the associated residuals satisfy

$$||r_{i+1}||_2^2 \leq c^{1/\nu}||r_i||_2^2 \tag{6}$$

where $c = [1 - d^\nu/\lambda_1(A^{\nu T}A^\nu)]$, $\nu = 8$ and d is given by

$$d = \mathrm{Min}_{1\leq j\leq n}\{|\alpha_j^8 + \beta_j^8 + 70\alpha_j^4\beta_j^4 - 28(\alpha_j^6\beta_j^2 + \alpha_j^2\beta_j^6)|\} > 0$$

Proof: Since A is normal there is a unitary transformation U such that $A = U^H\mathrm{Diag}(\lambda_j)U$. Then

$$(A^2 + A^{2T})/2 = U^H\mathrm{Diag}((\alpha_j^2 - \beta_j^2))U \tag{7}$$

and

$$(A^8 + A^{8T})/2 = U^H\mathrm{Diag}(\alpha_j^8 + \beta_j^8 + 70\alpha_j^4\beta_j^4 - 28(\alpha_j^6\beta_j^2 + \alpha_j^2\beta_j^6))U \tag{8}$$

If we consider only the single steepest descent direction $p_i = A^{s-1}r_i$ for s =2, 4, 6, 8 then

$$d = \text{Min}_{v \neq 0} |(v, A^s v)/(v, v)|. \tag{9}$$

For the particular choice of s =2, 8 it is easy to identify the inclusion regions (i)-(iv) and verify that the inequality on the residual norm holds. We prove the case $s \geq 8$. For $0 \leq l,\ k$ let $\mu_{l,k}$ be the moments $(A^l r_l)^T A^l r_k$. Consider the approximate solution $\bar{x}_{i+1} = x_i + \frac{\mu_{0,8}}{\mu_{8,8}} A^7 r_i$ then $E(x_{i+1}) \leq E(\bar{x}_{i+1}) < E(x_i)$ provided that $\mu_{0,8} \neq 0$. Now, $|\mu_{0,8}| = |r_i^T A^8 r_i| = |r_i^T (A^8 + A^8 T)/2 r_i| \geq d$. The inclusion region defined by (iii) is determined by solving the inequality $\alpha^8 + \beta^8 + 70\alpha^4\beta^4 - 28(\alpha^6\beta^2 + \alpha^2\beta^6) > 0$. Equality holds for $\beta = \pm.198913\alpha, \pm.668179\alpha$ (where the slopes are given accurate in six digits) . $\square$.

Note that in theorem 2.2 we only considered the s-MR with $2, 8 \leq s$ and identified the normal matrices for which the method converges. However, the class of normal matrices for which convergence is guaranteed is much larger. This can be shown by verifying that the inclusion regions of A^s for $s = 4$ or 6 are different from (i)-(iv) and nonempty.

Remark 2.2: Let us assume that in a Krylov subspace iterative method $x_{i+s} = x_i + a_i p_i + \ldots + a_{i+s-1} p_{i+s-1}$ and that the residual r_{i+s} is minimized over the subspace $\{p_i, \ldots, p_{i+s-1}\}$. If the residual can be proved to be orthogonal to $\{Ar_i, \ldots, A^s r_i\}$ then the residual norm bounds between $||r_{i+1}||_2^2$ and $||r_i||_2^2$ proved for s-MR (see (5) and (6) above) also apply between the residual norms of r_{i+s} and r_i (of the Krylov subspace method under consideration). In this case we state that the iterations $i, \ldots, i+s-1$ of the iterative method contain one iteration of the s-MR method.

Using remark 2.1 to one cycle of iterations of Odir(m) or GMRES(m) with $s \leq m$ shows that one cycle contains one iteration of s-MR. We next discuss the convergence of the restarted Odir method.

3 The Odir Method

Odir forms $A^T A$-orthogonal direction vectors and minimizes (at each iteration) the residual norm $||f - Ax_i||_2$ for definite matrices [14]. However Odir(k) may not converge even if A is definite. For nonsymmetric indefinite matrices Odir is guaranteed to converge if n iterations are taken. However

convergence is not warranted for Odir(m). In this section we find conditions under which Odir(m) converges for indefinite matrices. Let x_0 be an initial guess to the solution of (1) and let $r_0 = b - Ax_0$ be the initial residual. Let $j_i = 0$ for Odir and $j_i = \max(0, i-k+1)$ for Odir(k). The Odir and Odir(k) algorithms can be summarized as follows.

Algorithm 3.1 Odir and Odir(k)

Compute r_0 and set $p_0 = r_0$.

For $i = 0, 1, \ldots$ **until** convergence **do**

1. $a_i = \frac{r_i^T A p_i}{(Ap_i)^T Ap_i}$
2. $x_{i+1} = x_i + a_i p_i$
3. $r_{i+1} = r_i - a_i A p_i$
4. $b_j^i = -\frac{(A^2 p_i)^T Ap_j}{(Ap_j)^T Ap_j}$ for $j_i \leq j \leq i$
5. $p_{i+1} = Ap_i + \sum_{j=j_i}^{i} b_j^i p_j$
6. $Ap_{i+1} = A^2 p_i + \sum_{j=j_i}^{i} b_j^i Ap_j$

EndFor

In Odir the storage requirements for the vectors p_i and Ap_i increase up to the dimension n of the coefficient matrix. This is not practical for large problems so the truncated version Odir(k) has been proposed in [14]. If the matrix is symmetric or skew symmetric it can be easily proven [14] that Odir(2) is equivalent to Odir. There are examples of nonsymmetric definite matrices for which Odir(k) does not converge.

The following proposition which was proved in [14] for definite matrices relates the direction vectors of Odir and GCR (see [8]). The definiteness of A was used to prove that all *steplengths* a_i in Odir are nonzero. However for indefinite matrices some steplengths may be zero.

Proposition 3.1: Assume that in Odir(m) that the steplengths $a_0, \ldots, a_i$ for $i \leq m-1$ are not zero then for $1 \leq j \leq i+1$ the direction vectors (generated by Odir(m)) are:

$$p_j = \frac{-1}{a_{j-1}} \bar{p}_j \tag{10}$$

where $\bar{p}_0 = p_0 = r_0$ and

$$\bar{p}_j = r_j + \sum_{\mu=0}^{j-1} \bar{b}_\mu^j \bar{p}_\mu \tag{11}$$

and

$$\bar{b}_\mu^j = -\frac{(Ar_j, A\bar{p}_\mu)}{(A\bar{p}_\mu, A\bar{p}_\mu)}, \quad 0 \le \mu \le j \tag{12}$$

Proof: See [14]. □

The following remark explains why this result does not hold for Odir(k).

Remark 3.1: For Odir(k) it is not possible to prove a similar result involving only k direction vectors. To see this, consider $p_{k+1} = Ap_k + \sum_{\mu=1}^{k} b_\mu^k p_\mu$. Using the assumptions of proposition 3.1 with $i = k+1$ following its proof we can obtain $p_{k+2} = -\frac{1}{a_{k+1}}[r_{k+2} + \sum_{\mu=0}^{k+1} {b_\mu}^{k+1} p_\mu]$. This expression involves the direction vector p_0. Therefore it is not possible (for general nonsymmetric matrices) to map the Odir(k) direction vectors into the Omin(k) direction vectors. Also, using this expression for p_{k+2} we can write the steplength $a_{k+2} = -\frac{(r_{k+2}, Ar_{k+2}) + {b_0}^{k+1}(r_{k+2}, Ap_0)}{a_{k+1}{||Ap_{k+2}||_2}^2}$, because r_{k+2} is orthogonal to $Ap_{k+1} \ldots Ap_1$. However in general $(r_{k+2}, Ap_0) \neq 0$. Thus a_{k+2} maybe zero even for A definite.

We next express the direction vectors p_i generated by Odir (or Odir(m)) in terms of preceding direction vectors and $A^l r_i$, where $0 \le l$ and r_i is the residual vectors. This enables us to embed an s-dimensional steepest descent iteration in these methods. We can then establish convergence and obtain error bounds.

Remark 3.2: The direction vectors p_j $(1 \le j)$ in Odir can expressed in the form:

$$p_j = A^j r_0 + \sum_{j=0}^{j-1} \bar{b}_\mu^{j-1} p_\mu \tag{13}$$

where

$$\bar{b}_\mu^{j-1} = -\frac{(A^{j+1} r_i, Ap_\mu)}{(Ap_\mu, Ap_\mu)}, \quad 0 \le \mu \le j-1 \tag{14}$$

To see this one uses algorithm 3.1 (4.-5.) and the $A^T A$-orthogonality of the direction vectors.

The following proposition expresses the direction vectors in terms of powers of A applied to a residual r_j $(1 \le j)$ and preceding direction vectors.

Proposition 3.2: Assume $1 \leq i$ and that the steplength a_{i-1} (in Odir) is not zero. Then for $i \leq j$ the direction vectors are given by:

$$p_j = -\frac{1}{a_{i-1}} \bar{p}_j, \tag{15}$$

where

$$\bar{p}_j = A^{j-i} r_i + \sum_{\mu=0}^{j-1} \bar{b}_\mu^{j-1} p_\mu, \tag{16}$$

and

$$\bar{b}_\mu^j = -\frac{(A^{j-i+1} r_i, Ap_\mu)}{(Ap_\mu, Ap_\mu)}, \quad 0 \leq \mu \leq j-1 \tag{17}$$

Proof: Recall that the direction vector p_i is defined by the recursion $p_i = Ap_{i-1} + \sum_{\mu=0}^{i-1} b_\mu^{i-1} p_\mu$. Since $a_{i-1} \neq 0$ we obtain that $Ap_{i-1} = \frac{[r_i - r_{i-1}]}{a_{i-1}}$ and $r_{i-1} \in \{p_0, \ldots, p_{i-1}\}$. Therefore, $p_i = \frac{1}{a_{i-1}} \bar{p}_i$ where $\bar{p}_i = r_i + \sum_{\mu=0}^{i-1} \bar{b}_\mu^{i-1} p_\mu$. The parameters $\bar{b}_\mu^{i-1} = -\frac{(Ar_i, Ap_\mu)}{(Ap_\mu, Ap_\mu)}$, $0 \leq \mu \leq j-1$ are determined by the $A^T A$-orthogonality of p_i against p_μ. Now inductively we can form p_j for $i < j$. □

We next prove that under the assumption that A^s is definite for some $s \leq m$ then Odir(m) reduces the residual error norm.

Theorem 3.1: Let s $(0 < s)$ be an integer and assume that $(v, A^s v) \neq 0$ for nonzero vectors v. For $j = 0, 1, \ldots$ assume that the degree of the minimal polynomial r_0 is greater than $(j+1)s$ then there exists some k $(1 \leq k \leq s)$ such that the steplength a_{js+k-1} in Odir is not zero and $||r_{js+k-2}|| < ||r_{js+k-1}||$.

Proof: We use induction on j. For $j = 0$ assume that $a_i = \frac{(r_i, Ap_i)}{(Ap_i, Ap_i)} = 0$ for all $0 \leq i \leq s-1$. Since r_i is orthogonal to Ap_l for $l < i$ using remark 3.2 we obtain $(r_0, A^s r_0) = 0$, which contradicts the hypothesis. Thus $a_{k-1} \neq 0$ for some $k < s$. For $1 \leq j$, we must prove that there exists index i such that $js + 1 \leq i \leq (j+1)s$ with $a_{i-1} \neq 0$. From the induction hypothesis there is a largest positive integer l such that $j(s-1) + 1 \leq l \leq js$ and $a_{l-1} \neq 0$. Assume that $a_i = 0$ for $l < i$ and $r_i = r_l$. This implies that $(r_i, Ap_i) = 0$. Now applying proposition 3.2 and using the orthogonality of r_i and Ap_j (for $j < i$) we conclude that $(r_l, A^{i-l} r_l) = 0$. Because of the assumptions this implies that $i \leq l + s - 1$. This proves that $a_{js+k-1} \neq 0$ for some k $(1 \leq k \leq s)$. □

We next use this theorem and the results of section 2 to obtain residual error bounds for Odir(m).

Corollary 3.2: Under the assumptions of theorems 2.1-2.2 one cycle of Odir(m) with $m = ks$ contains k iterations of an s-MR method. Thus we obtain the following residual error bound

$$||r_m||_2{}^2 \leq c^k ||r_0||_2{}^2 \tag{18}$$

where c is the residual norm bounds constant in theorem 2.1-2.2.

Proof: Using remark 3.2, proposition 3.2 and theorem 3.1 for $0 \leq j \leq k-1$ we can show that $r_{(j+1)s}$ is orthogonal to $A^l r_{js}$, for $1 \leq k \leq s$. This implies that the iterations $js+l$, $0 \leq l \leq s-1$ contain one s-MR iteration. The residual error bound follows from theorems 2.1, 2.2 and 3.1 and remark 2.2 $\square$.

4 The GMRES method

In this section we study the convergence properties of the restarted GMRES(s) method. The Arnoldi method (see [1]) generates an orthonormal basis $V_m = [v_1, v_2, \ldots, v_m]$. $\bar{\bar{H}}_m = {V_m}^T A V_m$ is an (m x m) upper Hessenberg matrix generated by the Arnoldi method. The GMRES method [12] consists of an Arnoldi method and an error minimization step. We next present one cycle of the restarted GMRES(m) method. The norm of the residual is monitored for the convergence check.

Algorithm 4.1 GMRES(m)

Compute r_0 and set $v_1 = \frac{r_0}{\|r_0\|}$

For $j = 1, \ldots, m-1$

1. $h_{i,j} = v_i^T A v_j,\ 1 \leq i \leq j$
2. $\hat{v}_{j+1} = A v_j - \sum_{i=1}^{j} h_{i,j} v_i$
3. $h_{j+1,j} = \|\hat{v}_{j+1}\|_2$
4. $v_{j+1} = \hat{v}_{j+1} / \|\hat{v}_{j+1}\|_2$

EndFor

Form the approximate solution: $x_m = x_0 + V_m y_m$, where $y_m \in \mathbf{R}^m$ minimizes

$$\bar{J}(y) = \|\beta e_1 - \bar{H}_m y\| \quad e_1 = [1, \ldots, 0]^T. \tag{19}$$

The $(m+1)$ x m matrix $\bar{H}_m$ is the same as $\bar{\bar{H}}_m$ except for an additional row whose only nonzero element is $h_{m+1,m}$ in the $(m+1, m)$ position. It was shown in [12] that minimizing the error functional $\bar{J}(y)$ is equivalent to solving:

$$\mathrm{Min}_{x \in x_0 + \mathbf{K}_m} \|b - Ax\|_2 \tag{20}$$

where $\mathbf{K}_m = \mathrm{span}\{r_0, Ar_0, \ldots A^{m-1} r_0\}$ is the Krylov subspace of dimension m.

We next present a Householder QR method for minimizing $\bar{J}(y)$.

Proposition 4.1: The Householder based QR factorization can be efficiently utilized to solve the linear least squares problem of minimizing $\bar{J}(y)$ without causing fill-in to $\bar{H}_m$.

Proof: Let us consider as an example the matrix $\bar{H}_m$ for $m = 6$, with x representing the nonzero entries.

$$\begin{bmatrix} x & x & x & x & x & x \\ x & x & x & x & x & x \\ 0 & x & x & x & x & x \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & x \end{bmatrix} \tag{21}$$

Now, before beginning the QR reduction, we apply row permutations $P = P_{m,m+1} \ldots P_{1,2}$. This moves the first row of $\bar{H}_m$to the bottom of the matrix, and leaves us with a new matrix $H_m = P\bar{H}_m$ consisting of a upper triangular square matrix of size m augmented with a single dense row. For $m = 6$ it has the form:

$$\begin{bmatrix} x & x & x & x & x & x \\ 0 & x & x & x & x & x \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & x \\ x & x & x & x & x & x \end{bmatrix} \tag{22}$$

The diagonal entries of the matrix H_m are the entries $h_{j+1,j}$ of the matrix $\bar{H}_m$. These entries (norms of Arnoldi vectors) are nonzero if the degree of the minimal polynomial of r_0 is greater than m. Now the error functional to be minimized can be written as

$$J(y) = |||[\beta e_{m+1} - H_m y]||| = ||P[\beta e_1 - \bar{H}_m y]|||. \tag{23}$$

We note that $J(y) = \bar{J}(y)$. $\square$

In the past (see [11]) the restarted Arnoldi method for solving linear systems used LU factorization to compute y_m as a solution of $\bar{\bar{H}}_m y = \beta e_1$. A robust Arnoldi(m) can be defined by minimizing $\bar{\bar{J}}(y) = \|\beta e_1 - \bar{\bar{H}}_m y\|$ to compute y_m. However Arnoldi(m) does not minimize the residual norm (as GMRES(m) does).The Givens QR decomposition of the matrix $\bar{\bar{H}}_m$ has been used in [4] to compute y_m. The following result was also proved in [4] in a different way.

Proposition 4.2: GMRES(m) and Arnoldi(m) give the same solution for singular $\bar{\bar{H}}_m$.

Proof: We must prove that both methods give the same y_m. Suppose the matrix $\bar{\bar{H}}_m$ has rank $m-1$. The application of the first $(m-1)$ elementary Householder transformations to H_m make $(m+1)$-th row completely zero. So the m-th elementary Householder transformation equals the identity matrix and the m-th row remains unchanged. Therefore in the backsolution stage of the least squares problem the m-th row (its right-hand side entry being 0) will not contribute in forming y_m. Thus the functionals $J(y)$ and $\bar{J}(y)$ are minimized at the same point y_m. $\square$

proposition 4.3: GMRES(m) and Arnoldi(m) converge if and only if $(v_1, A^i v_1) \neq 0$ for some $i = 1, \ldots, m$. Assume that

Proof: Assume that $(v_1, A^i v_1) \neq 0$ for $i = 1, \ldots, m$. Consider the transformed linear least squares

$$\text{Min}_{y \in R^m} J(y)$$

The last row of H_m is zero. This implies that the first m rows form an upper triangular matrix and $Q_m = I_m$. Then $||r_m|| = ||r_0||$ for any $k \leq m$. Conversely, if $(v_1, A^i v_1) \neq 0$ for some $i = 1, \ldots, m$ then the s-MR method (with $s = m$) converges. So remark 2.1 gives the proof.

Corollary 4.1: Under the assumptions of theorems 2.1-2.2 one cycle of GMRES(m) method with $m = ks$ contains k iterations of an s-MR method. Thus

$$||r_m||_2{}^2 \leq c^k ||r_0||_2{}^2 \tag{24}$$

where c is the constant factor inequalities 2.2-2.6.
Proof: It follows from the equivalence of GMRES(m) and Odir(m) and corollary 3.2. □. In the next section IOM(k) is modified to an iterative method with an error minimization property.

5 The Minimal IOM(k) method

IOM(k) (see [11]) is based on the truncated Arnoldi to form a partially orthonormal sequence of vectors $\{v_1, v_2, \ldots\}$. Partially means that the every k consecutive vectors are orthogonal. Then a banded upper Hessenberg linear system is solved to form the approximate solution. The IOM(k) method does not have an error minimization property. We generalize the IOM(k) to the Minimal IOM(k) which has an error minimization property. It is proved that for symmetric, skew symmetric matrices or matrices of the form $I - N$ where N is skew symmetric MIOM(k) is identical to GMRES.

The defining equation for the direction vectors in IOM(k) is:

$$\hat{v}_{j+1} = Av_j - \sum_{i=j-k+1}^{j} h_{ij} v_i \tag{25}$$

with $h_{ij} = (Av_j, v_i)$, $i = j-k+1, 2, \ldots, j$ and $h_{j+1,j} = ||\hat{v}_{j+1}||$. The matrix $\bar{\bar{H}}_k$ is banded upper Hessenberg. Let us consider an approximation to the solution of (1) based on the truncated Arnoldi generated vectors $V_m = [v_1, \ldots, v_m]$. The approximate solution is formed: $x_m = x_0 + V_m y_m$, where y_m solves $\bar{\bar{H}}_m y_m = \beta e_1$. If the LU decomposition is used to solve this linear system of dimension m a recursion is derived in [11] that requires only k vectors be stored in updating x_m.

Let us consider the problem

$$\mathrm{Min}_{z \in L_m} ||f - A[x_0 + z]||, \tag{26}$$

where L_m is the space spanned by the vectors in V_m. This requires that the residual error $\hat{J}(y) = ||\beta v_1 - AV_m y|| = ||V_m[\beta e_1 - \bar{H}_m y]||$ be minimized. We use $\bar{H}_m$ to denote the $(m+1)$ x m matrix which is the same as $\bar{\bar{H}}_m$ except for an additional row whose only nonzero element is $h_{m+1,m}$ in the $(m+1, m)$

position. This matrix for $m = 6$ and $k = 3$ has the following form.

$$\begin{bmatrix} x & x & x & 0 & 0 & 0 \\ x & x & x & x & 0 & 0 \\ 0 & x & x & x & x & 0 \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & x \end{bmatrix} \tag{27}$$

Now

$$\text{Min}_{y \in R^m} \quad \hat{J}(y) \leq ||V_m^T V_m|| \quad \text{Min}_{y \in R^m} \bar{J}(y), \tag{28}$$

where $\bar{J}(y) = |||[\beta e_1 - \bar{H}_m y]|||$. Thus minimizing $\bar{J}$ gives only a suboptimal solution. This suboptimal *error minimization method* will be called Minimal IOM(k) (MIOM(k)). Equality holds when $||V_m^T V_m|| = 1$.

The least squares problem is solved as follows. Define $J(y) = ||\beta e_{m+1} - H_m y||$, where $H_m = P\bar{H}_m$ and P is a permutation matrix. Then it clear that $J(y) = \bar{J}(y)$. The matrix H_m has the form:

$$\begin{bmatrix} x & x & x & x & 0 & 0 \\ 0 & x & x & x & x & 0 \\ 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & x \\ x & x & x & 0 & 0 & 0 \end{bmatrix} \tag{29}$$

We use the row oriented Givens QR reduction to solve this least squares problem. So $H_m = Q_m U_m$ and U_m is a banded upper triangular matrix with bandwidth equal to k. However if the Householder were used then U_m can be a full upper triangular matrix.
We next summarize the properties of MIOM(k).

Proposition 5.1 : The following statements hold true for MIOM(k).

(i) Let the matrix A be symmetric, skew symmetric, or $I - N$ where N is skew symmetric. Then for $2 \leq k$ MIOM(k) is equivalent to GMRES.

(ii) The Givens QR method of the matrix H_m gives an orthogonal matrix Q_m and an upper triangular banded matrix U_m of width k such that $H_m = Q_m U_m$.

(iii) The approximate solution is $x_m = x_0 + V_m U_m^{-1} g_m$, where g_m equals the first m components of $\beta Q_m^T e_{m+1}$ and the residual norm equals $|\beta e_{m+1}^T Q_m^T e_{m+1}|$.

Proof: To prove (i) we use the special form of A and (25) to show that $V_m^T V_m = I$. So equality holds in (28). To prove (ii) we note that the Givens QR method will produce the Givens rotations $Q_{1,m+1}, \ldots, Q_{m,m+1}$ which are required to annihilate the entries of the last row of H_m. The application of rotation $Q_{j,m+1}$ may introduce one nonzero at the entry $(m+1, k+j)$ of the last row of H_m. No nonzero will be added to the $j-$th row. Thus the banded structure of the first m rows of H_m will be preserved in $Q_{j,m+1} \ldots Q_{1,m+1} H_m$. Finally, the $m+1$ row of the matrix H_m will be annihilated and U_m will equal the first m rows of $Q_m^T H_m$. The matrix Q_m^T is given by $Q_m^T = Q_{1,m+1} \ldots Q_{m,m+1}$ and $H_m = P \bar{H}_m = Q_m U_m$. To prove (iii) we note that the right hand side of the transformed linear least squares problem (minimizing $J(y)$) equals the first m components of $\beta Q_m^T P e_1 = \beta Q_m^T e_{m+1}$ and the residual error is $|e_{m+1}^T \beta Q_m^T P e_1]|$ (see [9]). $\square$

Remark 5.1: (i) MIOM(k) and IOM(k) do not improve the initial solution if $(v_1, A^j v_1) = 0$, for all $j = 1, \ldots, k$. To see this note that if $h_{1,j} = 0$, for all $j = 1, \ldots, k$ then the last row of H_m is zero. The first m rows form an upper triangular matrix and $Q_m = I_m$. The matrix $\bar{\bar{H}}_m$ is singular. Then $||r_m|| = ||r_0||$ for any $k \leq m$.

The IOM(k) method (see [11]) based on Givens QR determines y_m by minimizing the $\bar{\bar{J}}(y) = ||\bar{\bar{H}}_m y - \beta e_1||$

Proposition 5.2: The MIOM(k) and IOM(k) methods give the same approximate solution x_m if the matrix $\bar{\bar{H}}_m$ is singular.

Proof: Similar to Proposition 4.2. $\square$.

We next discuss how to implement MIOM(k) in a recursive form. This allows savings in the storage requirements for the direction vectors v_i. Only k of these vectors must be kept in store. This recursive implementation is similar to IOM(k) (see [11]). However MIOM(k) applies the Givens QR to H_m whereas IOM(k) applies the LU factorization to $\bar{H}_m$.

The matrix H_m is derived from the matrix H_{m-1} by adding the m-th column which has nonzeros $h_{j,m}$, for $j = m-k+1, \ldots, m+1$. This is done after permuting the m-th row of H_{m-1} to become the $(m+1)$-th row of (see matrix (29)) H_m. The entry $h_{m+1,m}$ is in the location (m, m). Because of the banded Hessenberg structure applying the first $(m-k-1)$ Givens rotations do not change its $m-$th column. In order to complete the QR reduction of

H_m the last $k+1$ Givens rotations applied to H_{m-1} must be applied between the entry $(m+1,m)$ of the last row and the $k+1$ the entries of the last column of the preceding $k+1$ rows.

The right-hand side vector can be derived from the $(m-1)-$th step:

$$\bar{g}_m = Q_m \beta e_{m+1} = Q_{m,m+1} \bar{g}_{m-1} \tag{30}$$

The rotation $Q_{m,m+1}$ applied to $\bar{g}_{m-1}$ only affects the last two entries. The $(m+1)-$th entry of $\bar{g}_m$ is not included in g_m. Thus $g_m = [g_{m-1}, \gamma_m]^T$ by computing the last entry. Therefore

$$x_m = x_{m-1} + \gamma_m w_m \tag{31}$$

where w_i are the columns of the Nxm matrix:

$$W_m = V_m U_m^{-1} \tag{32}$$

The vectors w_m can be easily computed by the formula:

$$w_m = \frac{1}{u_{mm}} [v_m - \sum_{i=m-k+1}^{m-1} u_{im} w_i] \tag{33}$$

We now give a concise form of the MIOM(k) algorithm.

Algorithm 5.2 MIOM(k)

Compute r_0 and set $v_1 = \frac{r_0}{||r_0||}$

For $j = 1, 2, \ldots,$ **Until** Convergence **Do:**

$\hat{v}_{j+1} = Av_j$ - $\sum_{i=m-k+1}^{j} h_{ij} v_i$ with $h_{ij} = (Av_j, v_i)$, $i = m-k+1, 2, \ldots, j$

$h_{j+1,j} = ||\hat{v}_{j+1}||$

$v_{j+1} = \hat{v}_{j+1} / h_{j+1,j}$

Update the Givens QR factorization of H_m and compute γ_m.

Compute $w_m = \frac{1}{u_{mm}} [v_m - \sum_{i=m-k+1}^{m-1} u_{im} w_i]$

Compute $x_m = x_{m-1} + \gamma_m w_m$

EndFor

6 Conclusions

We proved some convergence results for the restarted Odir and GMRES methods for nonsymmetric indefinite linear systems. We studied an implementation of GMRES(m) using the Householder QR in solving the least squares problems involving the reduced upper Hessenberg matrices. We introduced MIOM(k) which is based on IOM(k) with an error minimization property. We studied its convergence properties and gave a recursive implementation which requires only k direction vectors to be kept in storage. For linear systems with symmetric or skew-symmetric coefficient matrices it is proved that MIOM(2) is equivalent to GMRES.

References

[1] W.E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9, (1951) 17-29.

[2] S. F. Ashby, T. A. Manteuffel, P. E. Saylor, *A Taxonomy for conjugate gradient methods* SIAM J. Numer. Anal., 27, (1990) 1542-1568.

[3] O. Axelsson, *A Generalized Conjugate Gradient, Least Squares Method*, Numer. Math. 51, 209-227 (1987).

[4] Brown P.N. (1991), *A Theoretical Comparison of the Arnoldi and GMRES Algorithms* SIAM J. Sci. Stat. Comp., **12**, 58-78.

[5] A. T. Chronopoulos, *s-step Iterative Methods for (Non)symmetric (In)definite Linear Systems* SIAM J. on Num. Analysis Vol. 28, No. 6, 1991.

[6] A. T. Chronopoulos, *Krylov Subspace Iterative Methods for Nonsymmetric Indefinite Linear Systems*, Dept. of C.Sci, U. of M. TR 90-21, April 1990.

[7] P. Concus and G. H. Golub *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, Lect. Notes in Econ. and Math. systems, 134, Springer-Verlag, Berlin.

[8] S.C. Eisenstat, H. C. Elman and M. H. Schultz *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal. 20, (1983) 345-357.

[9] Golub, G.H. and C.F. van Loan (1989), *Matrix Computations*, 2^{nd} edition, The Johns Hopkins University Press, Baltimore, Md.

[10] A. Greenbaum, *Matrices that generate the same Krylov varieties*, Proceed. IMA workshop on iterative methods, IMA, Univ. of Minn., Feb. 1992.

[11] Y. Saad, *Practical Use of Krylov Subspace Methods for Solving Indefinite and Nonsymmetric Linear Systems* SIAM J. Sci. Stat. Comp., 4, No 1,(1984).

[12] Y. Saad and M. Schultz, *GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems,* SIAM J. Sci. Stat. Comp., Vol. 7, (1986).

[13] P. Vinsome, *An iterative method for solving sparse sets of simultaneous equations*, Society of Petroleum Engineers of AIME, SPE 5729, 1976.

[14] D. M. Young and K. C. Jea, *Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods* Linear Algebra Appl., 34 (1980), 159-194.