Weighted FOM and GMRES for solving nonsymmetric linear systems

Azeddine Essai

Laboratoire d'Analyse Numérique et d'Optimisation, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq Cedex, France
E-mail: essai@ano.univ-lille1.fr

Received 4 June 1998; revised 24 September 1998
Communicated by C. Brezinski

This paper presents two new methods called WFOM and WGMRES, which are variants of FOM and GMRES, for solving large and sparse nonsymmetric linear systems. To accelerate the convergence, these new methods use a different inner product instead of the Euclidean one. Furthermore, at each restart, a different inner product is chosen. The weighted Arnoldi process is introduced for implementing these methods. After describing the weighted methods, we give the relations that link them to FOM and GMRES. Experimental results are presented to show the good performances of the new methods compared to FOM(m) and GMRES(m).

Keywords: nonsymmetric linear systems, sparse matrices, iterative methods, Krylov subspaces, Arnoldi, FOM, GMRES

AMS subject classification: 65F10, 65F50

1. Introduction

Many scientific applications require the solution of linear systems of the form

$$Ax = b, \tag{1}$$

where A is a $n \times n$ nonsingular real matrix, $b \in \mathbb{R}^n$ the right-hand side, and the vector x is the solution of the linear system. In general, these systems are large, sparse and nonsymmetric.

To solve such systems, Saad proposed the FOM method [13], and later Saad and Schultz proposed the GMRES method [18]. Both methods use the Arnoldi process to construct an orthonormal basis $V_m = [v_1, \ldots, v_m]$ of the Krylov subspace

$$K_m(A, r_0) = \operatorname{span}\big(r_0, Ar_0, \ldots, A^{m-1}r_0\big),$$

where x_0 is an initial guess and $r_0 = b - Ax_0$ is the first residual vector.

© J.C. Baltzer AG, Science Publishers

The following algorithm describes the Arnoldi process for constructing an orthonormal basis of the Krylov subspace $K_m(A, v)$. This algorithm uses the modified Gram–Schmidt process and starts with the vector $v_1 = v/\|v\|_2$.

Algorithm 1. Arnoldi process.
For $j = 1, \ldots, m$
$w = Av_j$,
For $i = 1, \ldots, j$
$h_{i,j} = (w, v_i)_2$,
$w = w - h_{i,j}v_i$,
End
$h_{j+1,j} = \|w\|_2$, *if* $h_{j+1,j} = 0$ *Stop*,
$v_{j+1} = w/h_{j+1,j}$,
End.

Let the matrix $H_m \in \mathbb{R}^{m\times m}$ be the Hessenberg matrix whose nonzero entries are the scalars $h_{i,j}$ constructed by the Arnoldi process. Let us define the matrix $\overline{H}_m \in \mathbb{R}^{(m+1)\times m}$ by

$$\overline{H}_m = \begin{pmatrix} \boxed{\begin{matrix} & H_m & \end{matrix}} \\ 0 \quad \ldots \quad 0 \quad h_{m+1,m} \end{pmatrix} = \begin{pmatrix} H_m \\ h_{m+1,m}e_m^{\mathrm{T}} \end{pmatrix}.$$

It is known that the matrices built by the Arnoldi process satisfy the following relations:

$$V_m^{\mathrm{T}}V_m = I_m, \tag{2}$$
$$AV_m = V_{m+1}\overline{H}_m, \tag{3}$$
$$H_m = V_m^{\mathrm{T}}AV_m. \tag{4}$$

The FOM and GMRES methods generate iterates $x_m = x_0 + \omega_m$, where ω_m is a correction in the Krylov subspace $K_m(A, r_0)$.

In the FOM method, the iterate x_m^{F} is chosen such that its residual r_m^{F} is orthogonal to $K_m(A, r_0)$, i.e.,

$$V_m^{\mathrm{T}}r_m^{\mathrm{F}} = 0.$$

At each iteration of the GMRES method, x_m^{G} is chosen such that its residual r_m^{G} has a minimum norm, i.e.,

$$\left\|r_m^{\mathrm{G}}\right\|_2 = \min_{x\in x_0+K_m(A,r_0)} \|b - Ax\|_2.$$

The GMRES method can be associated to other methods in order to obtain information on eigenvalues. Thus, hybrid methods, wich are more powerful, are obtained [6,12,14]. In 1995, Morgan proposed a restarted GMRES method augmented with eigenvectors [11]. The GMRES method can also be preconditioned to be more efficient [16,17].

In this paper, we discuss two new iterative methods based on a modified Arnoldi process. The new process uses, instead of the Euclidean scalar product, another one, denoted by $(\cdot,\cdot)_D$, where D is a chosen diagonal matrix. The idea of changing the inner product is to accelerate the convergence of the components of the residual which are far away from zero. To achieve this, an appropriate weight is associated to each term of the inner product. A natural choice of these weights is the entries of the first residual. The new Arnoldi process is called the weighted Arnoldi process. It builds a D-orthonormal basis of the Krylov subspace $K_m(A, v)$.

The paper is organized as follows. In section 2, we introduce the weighted Arnoldi process and give some of its properties. We define, in section 3, the weighted FOM method and the weighted GMRES method. The algebraic links between the weighted methods and the usual ones are stated in section 4. We give, in section 5, some numerical examples to compare the weighted methods with the initial ones and we conclude.

2. Weighted Arnoldi process

Before giving a complete description of the weighted Arnoldi process, let us define the D-scalar product. If u and v are two vectors of $\mathbb{R}^n$, their D-scalar product is

$$(u, v)_D = v^{\mathrm{T}} D u = \sum_{i=1}^{n} d_i u_i v_i,$$

where $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is a diagonal matrix. Let us denote by d the real vector $d = (d_1, d_2, \ldots, d_n)^{\mathrm{T}}$.

This inner product is well defined if and only if the matrix D is positive definite, that is, $d_i > 0\ \forall i \in \{1, \ldots, n\}$.

In this case, we can define the D-norm $\|\cdot\|_D$ associated with this inner product by

$$\|u\|_D = \sqrt{(u, u)_D} = \sqrt{u^{\mathrm{T}} D u} = \sqrt{\sum_{i=1}^{n} d_i u_i^2} \quad \forall u \in \mathbb{R}^n.$$

We choose the vector d such that $\|d\|_2 = \sqrt{n}$. This choice enables us to recover the Euclidean norm if all the elements of d are equal.

The following algorithm describes the new Arnoldi process which uses the D-inner product $(\cdot,\cdot)_D$ to construct a D-orthonormal basis of $K_m(A, v)$ starting with the vector $\widetilde{v}_1 = v/\|v\|_D$.

Algorithm 2. Weighted Arnoldi process.

For $j = 1, \ldots, m$

$w = A\widetilde{v}_j$,

For $i = 1, \ldots, j$
$\widetilde{h}_{i,j} = (w, \widetilde{v}_i)_D$,
$w = w - \widetilde{h}_{i,j}\widetilde{v}_i$,
End
$\widetilde{h}_{j+1,j} = \|w\|_D$, *if* $\widetilde{h}_{j+1,j} = 0$ *Stop*,
$\widetilde{v}_{j+1} = w/\widetilde{h}_{j+1,j}$,
End.

The weighted process differs from a left preconditioning by the matrix D.

The basis $\widetilde{V}_m = [\widetilde{v}_1, \ldots, \widetilde{v}_m]$ constructed by this algorithm is D-orthonormal, thus it holds

$$\widetilde{V}_m^{\mathrm{T}} D \widetilde{V}_m = I_m. \tag{5}$$

The square Hessenberg matrix $\widetilde{H}_m$ whose nonzero entries are the scalars $\widetilde{h}_{i,j}$, constructed by the weighted Arnoldi process, can be expressed in the form

$$\widetilde{H}_m = \widetilde{V}_m^{\mathrm{T}} D A \widetilde{V}_m. \tag{6}$$

Let us define the matrix $\overline{\widetilde{H}}_m \in \mathbb{R}^{(m+1)\times m}$ as

$$\overline{\widetilde{H}}_m = \begin{pmatrix} \widetilde{H}_m \\ \widetilde{h}_{m+1,m} e_m^{\mathrm{T}} \end{pmatrix}.$$

As in the Arnoldi process, the matrices constructed by the weighted Arnoldi process satisfy the relationship

$$A\widetilde{V}_m = \widetilde{V}_{m+1}\overline{\widetilde{H}}_m. \tag{7}$$

Now, let us give some relations between the matrices generated by the two processes, namely Arnoldi and weighted Arnoldi.

Proposition 1. Assume that algorithms 1 and 2 do not break down before the mth step. Then, there exists an upper triangular matrix U_m such that

$$\widetilde{V}_m = V_m U_m, \tag{8}$$

$$U_m = V_m^{\mathrm{T}} \widetilde{V}_m, \tag{9}$$

$$U_m^{-1} = \widetilde{V}_m^{\mathrm{T}} D V_m, \tag{10}$$

and we can express $\overline{\widetilde{H}}_m$ in terms of $\overline{H}_m$ as

$$\overline{\widetilde{H}}_m = U_{m+1}^{-1} \overline{H}_m U_m. \tag{11}$$

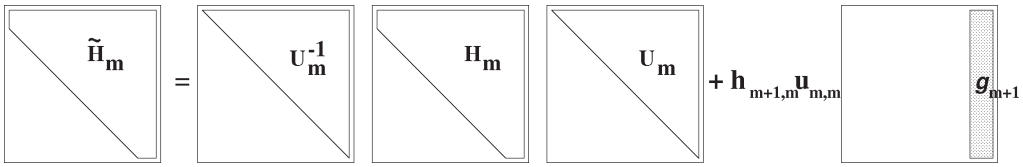

Figure 1. The matrix $\widetilde{H}_m$ in terms of H_m.

Proof. As V_j and $\widetilde{V}_j$ are two bases of the Krylov subspace $K_j(A, v)$ for all $j \in \{1, \ldots, m\}$, we can express $\widetilde{V}_m$ in terms of V_m as

$$\widetilde{V}_m = V_m U_m,$$

where U_m is an $m \times m$ upper triangular matrix.

If we multiply (8) on the left by V_m^{T}, we obtain (9), and if we multiply (8) on the left by $\widetilde{V}_m^{\mathrm{T}} D$, we get (10), thanks to the D-orthogonality of the matrix $\widetilde{V}_m$.

Using relations (7) and (8), we obtain

$$AV_m U_m = V_{m+1} U_{m+1} \overline{\widetilde{H}}_m,$$

then using (3) we have

$$V_{m+1} \overline{H}_m U_m = V_{m+1} U_{m+1} \overline{\widetilde{H}}_m.$$

As V_{m+1} is orthonormal, we get

$$\overline{H}_m U_m = U_{m+1} \overline{\widetilde{H}}_m.$$

Multiplying the last equality on the left by the matrix U_{m+1}^{-1}, we obtain formula (11). □

Relation (11) leads us to wonder if the matrices H_m and $\widetilde{H}_m$ are similar. Unfortunately, this is false, as will be shown in the following proposition.

Proposition 2. Under the same assumptions as in proposition 1, we can express $\widetilde{H}_m$ in terms of H_m by

$$\widetilde{H}_m = U_m^{-1} H_m U_m + h_{m+1,m} u_{m,m} g_{m+1} e_m^{\mathrm{T}}, \tag{12}$$

where $g_{m+1} \in \mathbb{R}^m$ is obtained from column $m+1$ of the matrix U_{m+1}^{-1} by deleting its last component.

Therefore, the matrices H_m and $\widetilde{H}_m$ are not similar in general.

Proof. In order to avoid any confusion, we will denote by $g_{i,j}$ $(1 \leqslant i, j \leqslant m)$ the entries of the matrix U_m^{-1}.

We split the matrix U_{m+1}^{-1} as follows:

$$U_{m+1}^{-1} = \begin{bmatrix} & U_m^{-1} & & g_{m+1} \\ 0 & \ldots & 0 & g_{m+1,m+1} \end{bmatrix} = \begin{bmatrix} & \widehat{U_{m+1}^{-1}} & & \\ 0 & \ldots & 0 & g_{m+1,m+1} \end{bmatrix},$$

where $g_{m+1} = (g_{1,m+1}, g_{2,m+1}, \ldots, g_{m,m+1})^{\mathrm{T}}$. So, the matrix $\widehat{U_{m+1}^{-1}}$ is obtained from the matrix U_{m+1}^{-1} by removing its last row.

From relation (11), we can write

$$\widetilde{H}_m = \widehat{U_{m+1}^{-1}}\overline{H}_m U_m,$$

then we have the relations

$$\widetilde{H}_m = \begin{bmatrix} U_m^{-1} & g_{m+1} \end{bmatrix} \begin{bmatrix} H_m \\ h_{m+1,m} e_m^{\mathrm{T}} \end{bmatrix} U_m = U_m^{-1} H_m U_m + h_{m+1,m} g_{m+1} e_m^{\mathrm{T}} U_m.$$

As the matrix U_m is upper triangular, the matrix product $e_m^{\mathrm{T}} U_m$ is equal to $u_{m,m} e_m^{\mathrm{T}}$. Therefore, we obtain formula (12). □

Remark. It is easy to see from (12) that H_m and $\widetilde{H}_m$ are similar if and only if $u_{m,m} = 0$ or $g_{m+1} = 0$. The former means that we have reached the degree of the minimal polynomial of A for v, the latter is satisfied if v_{m+1} and $\widetilde{v}_{m+1}$ are linearly dependent.

Corollary 1. Under the same assumptions as in proposition 1, we can express H_m in terms of $\widetilde{H}_m$ by

$$H_m = U_m^{-1} \widetilde{H}_m U_m + \frac{h_{m+1,m}}{u_{m+1,m+1}} u_{m+1} e_m^{\mathrm{T}}, \tag{13}$$

where $u_{m+1} \in \mathbb{R}^m$ is obtained from column $m+1$ of the matrix U_{m+1} by deleting its last component.

Proof. This corollary follows immediately from relation (12) by multiplying it on the left by U_m and on the right by U_m^{-1}, and using the relation

$$U_m g_{m+1} = -\frac{1}{u_{m+1,m+1}} u_{m+1}.$$

This relation can be easily shown by writing that $U_{m+1} U_{m+1}^{-1} = I_{m+1}$, and splitting the matrices U_{m+1} and U_{m+1}^{-1}. □

Complexity

Let us denote by N_{nz} the number of nonzero elements of the matrix A. Both the Arnoldi and the weighted Arnoldi processes require m steps and, at each step, a matrix–vector product is computed, then the cost is $2mN_{\mathrm{nz}}$ operations.

The Euclidean inner product costs $2n$ operations, and the D-inner product costs $3n$ operations. Then, step j requires approximately $2jn$ operations for Arnoldi and $3jn$ operations for weighted Arnoldi. We add $2jn$ operations to form the vector $j+1$ in the two processes. Therefore, the total number of operations required to perform m steps for these processes is approximately the following:

Process	Number of operations
Arnoldi	$2mN_{\text{nz}} + 2m^2n$
Weighted Arnoldi	$2mN_{\text{nz}} + (5/2)m^2n$

3. The weighted FOM and weighted GMRES methods

Like all Krylov methods, the mth ($m \geqslant 1$) iterate x_m of the weighted FOM and weighted GMRES methods belongs to the affine Krylov subspace $x_0 + K_m(A, r_0)$.

The iterate x_m^{WF} of the weighted FOM method is chosen such that its residual is D-orthogonal to the subspace $K_m(A, r_0)$, that is,

$$r_m^{\text{WF}} \perp_D K_m(A, r_0). \tag{14}$$

The iterate x_m^{WG} of the weighted GMRES method is chosen to minimize the residual D-norm in $x_0 + K_m(A, r_0)$. Hence, it is the solution of the least squares problem

$$\underset{x \in x_0 + K_m(A, r_0)}{\text{minimize}} \ \|b - Ax\|_D. \tag{15}$$

In these methods, we use the D-inner product and the D-norm, so, in order to compute the solution in the affine subspace $x_0 + K_m(A, r_0)$, we construct a D-orthonormal basis of the Krylov subspace $K_m(A, r_0)$ by the weighted Arnoldi process described in the previous section.

An iterate x_m of these two methods can be written as

$$x_m = x_0 + \widetilde{V}_m y_m,$$

where $y_m \in \mathbb{R}^m$.

Therefore, the corresponding residual $r_m = b - Ax_m$ satisfies

$$r_m = b - A\big(x_0 + \widetilde{V}_m y_m\big) = r_0 - A\widetilde{V}_m y_m = \widetilde{V}_{m+1}\big(\widetilde{\beta} e_1 - \overline{\widetilde{H}}_m y_m\big),$$

where $\widetilde{\beta} = \|r_0\|_D$ and e_1 is the first vector of the canonical basis.

Then, the weighted FOM method consists in finding the solution vector y_m^{WF} of the problem

$$\widetilde{V}_m^{\text{T}} D \widetilde{V}_{m+1}\big(\widetilde{\beta} e_1 - \overline{\widetilde{H}}_m y_m^{\text{WF}}\big) = 0,$$

which is equivalent to solving

$$\widetilde{H}_m y_m^{\text{WF}} = \widetilde{\beta} e_1. \tag{16}$$

As far as the weighted GMRES method is concerned, since the matrix $\widetilde{V}_{m+1}$ is D-orthonormal, we have

$$\|r_m\|_D^2 = \big\|\widetilde{V}_{m+1}\big(\widetilde{\beta} e_1 - \overline{\widetilde{H}}_m y_m\big)\big\|_D^2 = \big\|\widetilde{\beta} e_1 - \overline{\widetilde{H}}_m y_m\big\|_2^2,$$

and thus, problem (15) is reduced to finding the solution vector y_m^{WG} of the minimization problem

$$\underset{y\in\mathbb{R}^m}{\text{minimize}} \ \left\|\widetilde{\beta}e_1 - \overline{\widetilde{H}}_m y\right\|_2. \tag{17}$$

The solution of (16) and (17) can be obtained by making use of the QR decomposition of the matrices $\widetilde{H}_m$ and $\overline{\widetilde{H}}_m$, respectively, as for the FOM and GMRES algorithms [9,10].

In the case where m is equal to the degree of the minimal polynomial of A for r_0, the Krylov subspace $K_m(A, r_0)$ is invariant. Therefore, the iterate x_m obtained by both methods is the exact solution of the linear system (1).

The weighted FOM and weighted GMRES algorithms require the storage of $\widetilde{V}_m$, that is to say, m n-dimensional vectors at step m. Therefore, the value of m is limited by the storage constraint and by avoiding rounding errors. To address such problems, encountered also in FOM and GMRES, these algorithms can be restarted after m iterations. The corresponding algorithms are named the restarted weighted FOM and the restarted weighted GMRES, and they are denoted WFOM(m) and WGMRES(m). They are as follows:

Algorithm 3. WFOM(m) (WGMRES(m)).

1. Start: choose x_0, m and a tolerance ε, compute $r_0 = b - Ax_0$.
2. Choose the vector d such that $\|d\|_2 = \sqrt{n}$, compute $\widetilde{\beta} = \|r_0\|_D$ and $\widetilde{v}_1 = r_0/\widetilde{\beta}$.
3. Construct the D-orthonormal basis $\widetilde{V}_m$ by the weighted Arnoldi process, starting with the vector $\widetilde{v}_1$.
4. Form the approximation:
   WFOM: Solve the system $\widetilde{H}_m y_m = \widetilde{\beta}e_1$ by the QR factorization of $\widetilde{H}_m$, set $x_m = x_0 + \widetilde{V}_m y_m$, $r_m = b - Ax_m$.
   WGMRES: Compute $y_m = \arg\min_{y\in\mathbb{R}^m} \|\widetilde{\beta}e_1 - \overline{\widetilde{H}}_m y\|_2$ by the QR factorization of $\overline{\widetilde{H}}_m$, set $x_m = x_0 + \widetilde{V}_m y_m$, $r_m = b - Ax_m$.
5. Restart: if $\|r_m\|_2 \leqslant \varepsilon$ Stop else set $x_0 = x_m$, $r_0 = r_m$, and go to step 2.

These weighted processes differ from a left preconditioning of the system by the matrix D.

When a breakdown occurs at iteration j in the weighted Arnoldi process, in other words, when $\widetilde{h}_{j+1,j} = 0$, the process is stopped and the approximation of the solution is given by $x_j = x_0 + \widetilde{V}_j \widetilde{H}_j^{-1}(\widetilde{\beta}e_1)$ for both methods.

The restarted weighted GMRES algorithm minimizes the D-norm of the residual which changes at each restart. For this reason the Euclidean norm of the residual is not decreasing monotonously as is the case for GMRES.

4. Links between FOM, GMRES and their weighted versions

First, we will establish a link between FOM and WFOM. To achieve this, we will express the corrections $x_m^{\mathrm{WF}} - x_0$ belonging to $K_m(A, r_0)$ in the D-orthonormal basis $\widetilde{V}_m$ and in the orthonormal basis V_m as

$$x_m^{\mathrm{WF}} - x_0 = \widetilde{V}_m y_m^{\mathrm{WF}} = V_m \check{y}_m^{\mathrm{WF}},$$

where $y_m^{\mathrm{WF}} \in \mathbb{R}^m$ and $\check{y}_m^{\mathrm{WF}} \in \mathbb{R}^m$ satisfy the relation $y_m^{\mathrm{WF}} = U_m^{-1} \check{y}_m^{\mathrm{WF}}$.

The vector y_m^{WF} is the solution of the linear system (16):

$$\widetilde{H}_m y_m^{\mathrm{WF}} = \widetilde{\beta} e_1.$$

In this linear system, replacing y_m^{WF} by its expression and multiplying by U_m, we obtain

$$U_m \widetilde{H}_m U_m^{-1} \check{y}_m^{\mathrm{WF}} = U_m \widetilde{\beta} e_1 = \beta e_1,$$

where $\beta = \|r_0\|_2$.

Using relation (13) in corollary 1, we write

$$\left(H_m - \frac{h_{m+1,m}}{u_{m+1,m+1}} u_{m+1} e_m^{\mathrm{T}} \right) \check{y}_m^{\mathrm{WF}} = \beta e_1,$$

then $\check{y}_m^{\mathrm{WF}}$ is the solution of the linear system

$$\widehat{H}_m \check{y}_m^{\mathrm{WF}} = \beta e_1, \tag{18}$$

where the matrix $\widehat{H}_m$, whose first $m-1$ columns are equal to those of the matrix H_m (see figure 2), is

$$\widehat{H}_m = H_m - \frac{h_{m+1,m}}{u_{m+1,m+1}} u_{m+1} e_m^{\mathrm{T}}.$$

We deduce from this result that the knowledge of the components of $\widetilde{v}_{m+1}$ in the orthonormal basis allows us to construct the iterate x_m^{WF} from the Arnoldi process.

Now, we will establish a link between GMRES and WGMRES. As for FOM and WFOM, we write the correction $x_m^{\mathrm{WG}} - x_0$ in the form

$$x_m^{\mathrm{WG}} - x_0 = \widetilde{V}_m y_m^{\mathrm{WG}} = V_m \check{y}_m^{\mathrm{WG}}.$$

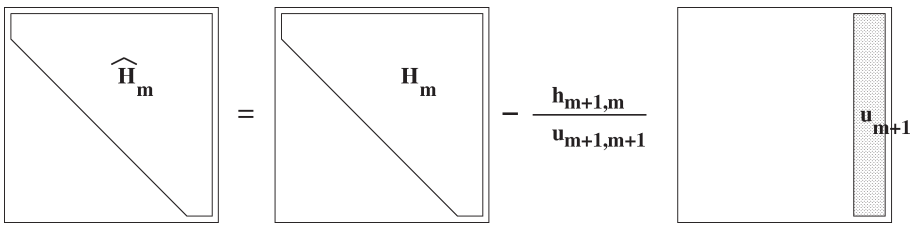

Figure 2. The matrix $\widehat{H}_m$ in terms of H_m.

As seen above, y_m^{WG} is the solution of the minimization problem (17), so

$$y_m^{\mathrm{WG}} = \arg\min_{y\in\mathbb{R}^m} \left\| \widetilde{\beta} e_1 - \overline{\widetilde{H}}_m y \right\|_2.$$

Since $\check{y}_m^{\mathrm{WG}}$ gives the components of $x_m^{\mathrm{WG}} - x_0$ in the basis of the orthonormal vectors, we obtain

$$\check{y}_m^{\mathrm{WG}} = \arg\min_{\check{y}\in\mathbb{R}^m} \left\| \widetilde{\beta} e_1 - \overline{\widetilde{H}}_m U_m^{-1} \check{y} \right\|_2,$$

then, using formula (11), we have

$$\check{y}_m^{\mathrm{WG}} = \arg\min_{\check{y}\in\mathbb{R}^m} \left\| U_{m+1}^{-1} \left(\beta e_1 - \overline{H}_m \check{y} \right) \right\|_2.$$

Finally, in order to get the relation between GMRES and WGMRES, we notice that $\check{y}_m^{\mathrm{WG}}$ is the solution of the same optimization problem as for GMRES but with the norm corresponding to the symmetric matrix $U_{m+1}^{-\mathrm{T}} U_{m+1}^{-1}$ instead of the Euclidean one. Thus, we obtain

$$\min_{\check{y}\in\mathbb{R}^m} \left\| \beta e_1 - \overline{H}_m \check{y} \right\|_{U_{m+1}^{-\mathrm{T}} U_{m+1}^{-1}}.$$

Thus, contrary to WFOM, it is necessary to know the components of all the vectors of $\widetilde{V}_{m+1}$ in the basis V_{m+1} in order to compute the iterate x_m^{WG} from the Arnoldi process.

5. Numerical experiments

In this section, we will show some numerical examples in order to compare FOM and GMRES with their weighted versions. The experiments were performed in double-precision Fortran on an Alpha processor.

The vector d will be chosen as $d_i = \sqrt{n}|(r_0)_i|/\|r_0\|_2$. Such a choice favors the components of the residual which are far away from zero. In the following numerical examples, this choice will always be made.

The sparse and unsymmetric matrices tested here are from the Matrix Market Web server [2].[1] In all experiments, the right-hand side is a random vector with entries uniformly distributed in the interval $[0, 1]$. The initial guess is $x_0 = (0, \ldots, 0)$.

The convergence is monitored by means of the test

$$\|r\|_2/\|b\|_2 < \textit{epsilon},$$

the value of *epsilon* depending on the examples. The time is given in seconds for all examples.

The symbol * in the tables of results means that the convergence is not reached.

[1] http://math.nist.gov/MatrixMarket/.

Example 1: the matrix `add20`. This is a 2395×2395 matrix from the HAMM group, from the independent sets and generators of the Matrix Market, with 17319 nonzero entries. The estimated condition number is 1.76E+4.

This example was tested with various values of m (5, 10, 15, . . . , 80). WFOM(m) was always better than FOM(m), and WGMRES(m) was always better than GMRES(m).

In order to illustrate this example, figure 3 and table 1 give the results obtained for $m = 10$ with *epsilon* $= 10^{-12}$.

Example 2: the matrix `orsirr_1`. This is a 1030×1030 matrix arising from a 3D oil reservoir simulation, from the package OILGEN of the Harwell Boeing collection [5], with 6858 nonzero entries. The estimated condition number is 10E+2.

In table 2, we report the results given for different values of m. We take *epsilon* $= 10^{-11}$ and the algorithm is stopped if convergence is not reached after 2000 iterations.

We remark that, for all considered Krylov subspace sizes, each weighted method converges in less iterations and less time than its corresponding standard method. For the particular case $m = 10$, the WGMRES converges, while the residual norm of GMRES stagnates at 1.65E+01. The convergence of FOM and WFOM is slow, but,

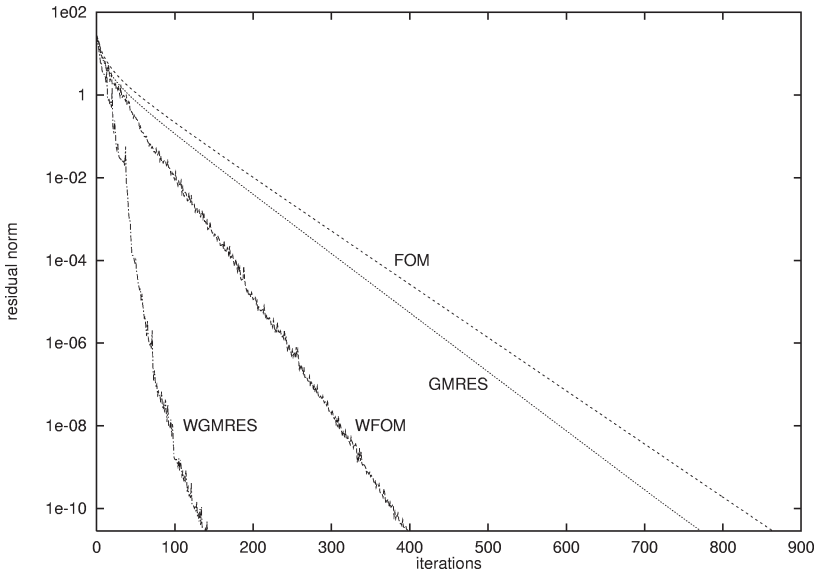

Figure 3. The matrix `add20` with $m = 10$.

Table 1

Method	FOM	WFOM	GMRES	WGMRES
Iterations	865	400	772	136
Time	79.44	37.34	75.25	13.17

Table 2
Results obtained with the matrix `orsirr_1`.

m	FOM		WFOM		GMRES		WGMRES	
	iter.	time	iter.	time	iter.	time	iter.	time
10	2000	75.35	2000	75.96	2000	73.66	1734	68.46
	$\|r\|_2 = 5.24\text{E}{-}04$		$\|r\|_2 = 2.26\text{E}{-}06$		$\|r\|_2 = 1.65\text{E}{+}01$			
20	800	77.62	630	66.89	815	82.75	198	**21.08**
30	325	63.40	167	35.67	300	57.56	120	25.59
40	142	44.35	82	**27.09**	156	47.88	64	21.93
50	64	**30.12**	53	27.90	77	35.85	49	24.48
60	51	31.78	39	27.31	54	**34.09**	39	27.32
70	42	36.37	32	31.58	42	35.94	28	27.05
80	36	39.01	26	31.21	34	37.13	27	33.01

at the maximal iteration 2000, the residual norm is reduced to 5.24E−04 for FOM and to 2.26E−06 for WFOM.

In table 2, we gave, in bold type, the relative minimal computational time of each method. The size of the Krylov subspace corresponding to the optimum for WGMRES (respectively, WFOM) is less than for GMRES (respectively, FOM); so the storage needed for the weighted methods to achieve optimal convergence is not higher.

Example 3: the matrix `fs_541_2`. This is a 541×541 matrix coming from the set SMTAPE of the Harwell Boeing collection and the package FACSIMILE for an atmospheric pollution problem, with 4285 nonzero entries. The estimated condition number is 7.7E+11.

For this example, we take $m = 40$ and *epsilon* $= 10^{-10}$. The results given in figure 4 and in table 3 show that, exceptionally, FOM has a better convergence than GMRES which stagnates at the value 1.16E+01. The WGMRES method converges, even if, until iteration 107, the WGMRES curve oscillates.

Example 4: the matrix `bfw782a`. This is a 782×782 matrix from the set BFWAVE, from the NEP collection, with 7514 nonzero entries. The estimated condition number is 4.6E+03. The results are given in figure 5 and in table 4, with $m = 20$ and *epsilon* $= 10^{-12}$.

This example is given to show that the improvement brought about by the weighted methods is simultaneously on the number of iterations and on the computational time.

Example 5: the matrix `memplus`. This is a 17758×17758 matrix from the set HAMM, from the independent sets and generators of the Matrix Market with 126150 nonzero entries. The condition number is not available. The results are given in figure 6 and in table 5, with $m = 30$ and *epsilon* $= 10^{-12}$.

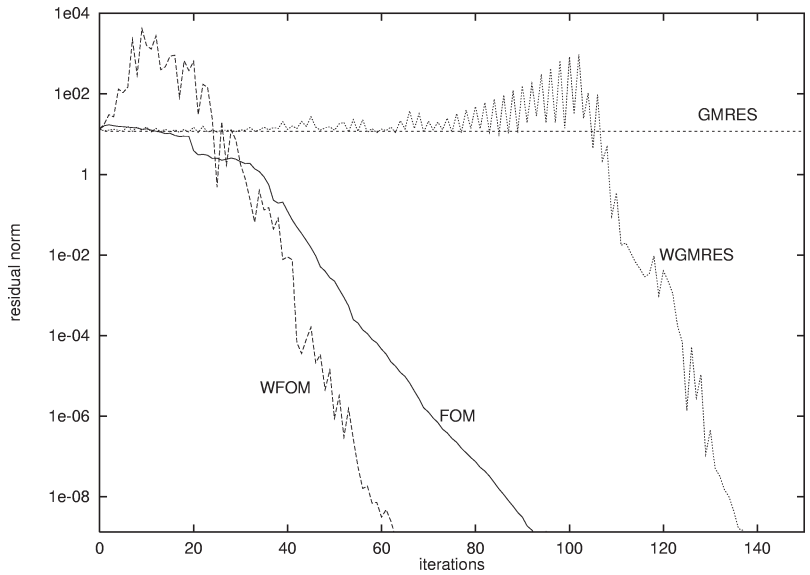

Figure 4. The matrix `fs_541_2` with $m = 40$.

Table 3

Method	FOM	WFOM	GMRES	WGMRES
Iterations	93	63	*	138
Time	14.67	12.75	*	27.50

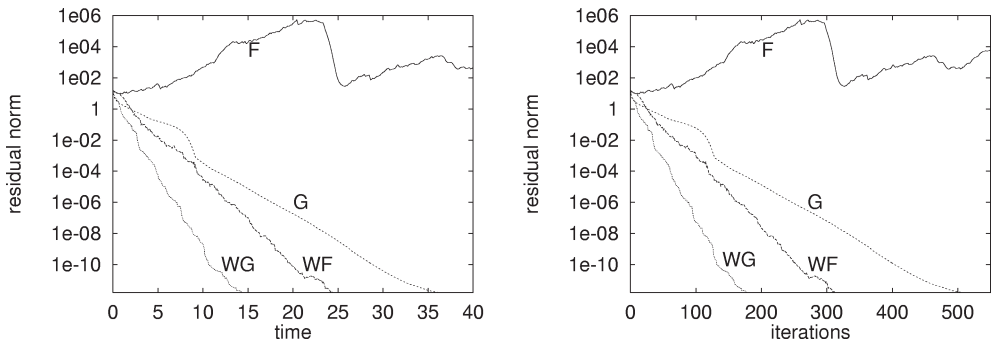

Figure 5. The matrix `bfw782a` with $m = 20$.

Table 4

Method	FOM	WFOM	GMRES	WGMRES
Iterations	*	314	506	178
Time	*	24.52	36.15	14.37

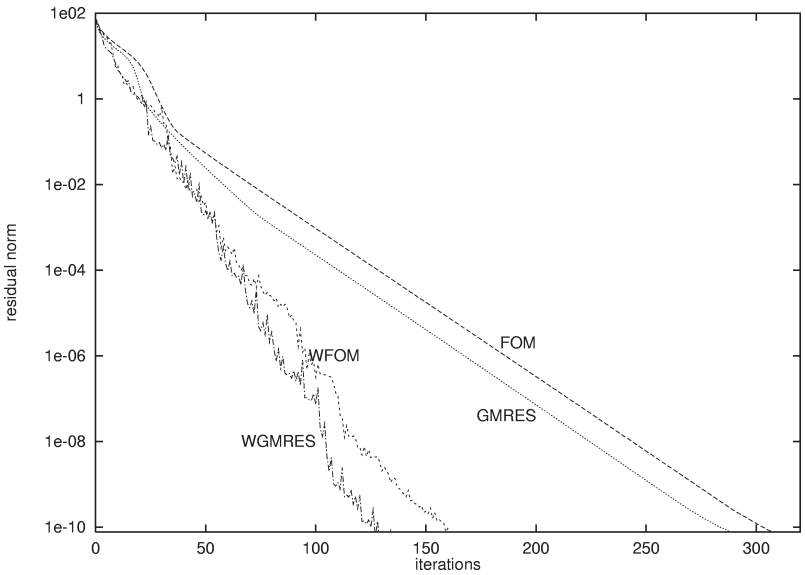

Figure 6. The matrix `memplus` with $m = 30$.

Table 5

Method	FOM	WFOM	GMRES	WGMRES
Iterations	308	161	289	127
Time	1341.01	819.76	1245.49	681.79

This example is given to show the performance of WFOM and WGMRES on large sparse linear systems. The improvement factor relative to time is 1.63 for WFOM and 1.82 for WGMRES.

6. Conclusion and future work

In this paper, we proposed two new iterative methods, WFOM and WGMRES. The various examples studied illustrate the interest of these methods, even if their residual norms oscillate.

All relationships established between the methods FOM and GMRES in [3] and [4] are valid between WFOM and WGMRES. Particularly, the iterates of WGMRES can be obtained from those of WFOM by a minimal residual smoothing technique [19].

As the amount of operations of the weighted Arnoldi process is not much larger than that of the Arnoldi process, and as the weighted methods need fewer iterations to converge, the parallel versions of the weighted methods seem very interesting.

In section 4, we saw that the iterate of the WFOM method can be obtained from the Arnoldi process if the components of the vector $\widetilde{v}_{m+1}$ are known. Since this vector is the direction of the residual, the residual direction can be chosen at the restart. This idea can lead to a powerful method if the problem of the choice of the residual direction can be solved.

Future work will focus on solving the above open problem, trying to find the optimal vector d of weights, applying the weighting technique to the methods QMR [8] and TFQMR [7] based on the Lanczos process. Preconditioning the weighted methods by preconditioners such as SSOR and those based on the ILU factorization [17] and, finally, implementing all these methods on parallel computers is also under consideration.

References

[1] Z. Bai, D. Day, J. Demmel and J. Dongarra, A test matrix collection for non-Hermitian eigenvalue problems, Release 1.0 (1996).

[2] R.F. Boisvert, R. Pozo, K. Remington, R. Barrett and J.J. Dongarra, The Matrix Market: A web resource for test matrix collections, in: *The Quality of Numerical Software: Assessment and Enhancement*, ed. R.F. Boisvert (Chapman & Hall, London, 1997) pp. 125–137.

[3] P.N. Brown, A theoretical comparison of the Arnoldi and GMRES algorithms, SIAM J. Sci. Statist. Comput. 12 (1991) 58–78.

[4] J. Cullum and A. Greenbaum, Relations between Galerkin and norm-minimizing iterative methods for solving linear systems, SIAM J. Matrix Anal. Appl. 17 (1996) 223–247.

[5] I.S. Duff, R.G. Grimes and J.G. Lewis, User's guide for the Harwell–Boeing sparse matrix collection (Release I), Technical Report TR/PA/92/86, CERFACS, Toulouse, France (1992).

[6] H.C. Elman, Y. Saad and P. Saylor, A hybrid Chebyshev Krylov subspace algorithm for solving nonsymmetric systems for linear equations, SIAM J. Sci. Statist. Comput. 7 (1986) 840–855.

[7] R.W. Freund, A transpose-free quasi-minimal residual algorithm for non-Hermitian linear system, SIAM J. Sci. Comput. 14 (1993) 470–482.

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

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

[10] R.S. Martin, G. Peters and J.H. Wilkinson, The QR algorithm for real Hessenberg matrices, Numer. Math. 14 (1970) 219–231.

[11] R.B. Morgan, A restarted GMRES method augmented with eigenvectors, SIAM J. Matrix Anal. Appl. 16 (1995) 1154–1171.

[12] N.M. Nachtigal, L. Reichel and L.N. Trefethen, A hybrid GMRES algorithm for nonsymetric linear systems, SIAM J. Matrix Anal. Appl. 13 (1992) 796–825.

[13] Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, Math. Comp. 37 (1981) 105–126.

[14] Y. Saad, Least squares polynomials in the complex plane and their use for solving nonsymmetric linear systems, SIAM J. Sci. Statist. Comput. 7 (1987) 155–169.

[15] Y. Saad, SPARSKIT: a basic tool kit for sparse matrix computations, RIACS, NASA Ames Research Center (1991).

[16] Y. Saad, A flexible inner–outer preconditioned GMRES algorithm, SIAM J. Sci. Comput. 14 (1993) 461–469.

[17] Y. Saad, *Iterative Methods for Sparse Linear Systems* (PWS Publishing, New York, 1996).

[18] Y. Saad and M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Statist. Comput. 7 (1986) 856–869.
[19] L. Zhou and H.F. Walker, Residual smoothing techniques for iterative methods, SIAM J. Sci. Statist. Comput. 15 (1994) 297–312.