INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS, VOL. 23, 177–190 (1996)

NEWTON–GMRES ALGORITHM APPLIED TO COMPRESSIBLE FLOWS

RÉMI CHOQUET AND JOCELYNE ERHEL

INRIA, Campus de Beaulieu, F-35042 Rennes, France

SUMMARY

This paper addresses the resolution of non-linear problems arising from an implicit time discretization in CFD problems. We study the convergence of the Newton–GMRES algorithm with a Jacobian approximated by a finite difference scheme and with restarting in GMRES. In our numerical experiments we observe, as predicted by the theory, the impact of the matrix-free approximations. A second-order scheme clearly improves the convergence in the Newton process.

KEY WORDS non-linear problem; approximate Jacobian; convergence; iterative linear solver; restarting

1. INTRODUCTION

Many scientific applications lead to a non-linear system of equations. We consider here the numerical simulation of steady state compressible flows. Implicit time discretizations allow us to use large time steps. On the other hand, at each time step a non-linear system of equations must be solved. Because of memory requirements, we want to use a so-called matrix-free algorithm. Several authors (see e.g. References 1 and 2) have considered inexact Newton methods where the Newton equations are solved approximately by an iterative solver. Moreover, since the Jacobian is required only through a matrix–vector product, it can be approximated by a finite difference scheme.[3,4] The resulting matrix-free algorithm, which we call Newton–MF-GMRES, has been studied there with no restarting in GMRES.

Here we extend these results to GMRES with restarting, denoted GMRES(m), as designed in Reference 5. Global convergence of Newton can be enhanced by a line search backtracking procedure provided that the approximate solution given by the iterative solver is a descent direction.[6] We give a sufficient condition on the stopping criterion of GMRES(m) to guarantee this result. The quadratic local convergence of the basic Newton iterations is no longer achieved with the Newton–MF-GMRES method. As in Reference 3, but in the context of restarting, we give here sufficient conditions on the stopping criterion and the approximation of the Jacobian to obtain a linear local convergence. We introduce a centred second-order difference quotient to approximate the Jacobian. This scheme is more expensive than the usual first- order difference quotient, but it is more accurate and leads to a better Newton convergence.

We apply the Newton–MF-GMRES(m) algorithm to the numerical solution of the compressible Navier–Stokes equations. We present results for two steady state problems. We study in detail the convergence of Newton and GMRES for one implicit time step and also for the stationary non-linear

CCC 0271–2091/96/020177–14
© 1996 by John Wiley & Sons, Ltd.

Received May 1995
Revised November 1995

problem with no time derivative. As expected, the Newton convergence is improved by using an accurate Jacobian approximation. The improvement comes mainly from the better approximation at each restart of GMRES.

The paper is organized as follows. In Section 2 we describe the method under consideration and introduce an equivalent GMRES scheme. In Section 3 we study the global and local convergence of the Newton iterations. In Section 4 we present and analyse the numerical results. We give conclusions in Section 5.

Throughout the paper, $\|.\|$ denotes the Euclidean norm $\|.\|_2$.

2. NUMERICAL METHOD

We describe in this section the inexact Newton algorithm, combined with a GMRES solver with restarting and an approximation of the Jacobian. The problem to solve is to find $u \in \mathbb{R}^N$ satisfying $F(u) = 0$. The Jacobian is denoted $J(u)$.

Algorithm. Inexact Newton–MF-GMRES

```
* $\eta$ is the relative tolerance for the residual norm;
convergence := false;
choose $u_0$;
$i := 0$;
until convergence do
   * approximately solve $J(u_i)\delta_i = -F(u_i)$;
   call MF-GMRES($m$, $J(u_i)$, $F(u_i)$, $\delta_i$);
   $u_{i+1} = u_i + \delta_i$;
   if $\|F(u_{i+1}) < \|\eta\|F(u_0)\|$ then
   convergence := true;
   else
   $i = i + 1$;
   endif
enddo
```

At each Newton iteration the linear system is approximately solved by GMRES with restarting after m steps[5] and a matrix-free Jacobian estimation. The algorithm is denoted MF-GMRES(m, J, F, δ).

Algorithm. MF-GMRES(m, J, F, δ)

```
* This algorithm solves approximately $J\delta = -F$;
$\xi$ is the relative tolerance for the residual norm;
convergence := false;
choose $\delta_0$;
until convergence do
   $q_1$ := appr($J*\delta_0$);
   $r_0 := -F - q_1$;
   $\beta := \|r_0\|$;
   $v_1 := r_0/\beta$;
   for $k = 1, \ldots, m$ do
      $q_{k+1}$ := appr($J * v_k$);
      $p := q_{k+1}$;
```

```
        for $l = 1, \dots, k$ do
          $h_{lk} := v_l^{\mathrm{T}} p$;
          $p := p - h_{lk} v_l$;
        endfor
        $h_{k+1,k} := \|p\|_2$;
          $v_{k+1} := p/h_{k+1,k}$;
      endfor;
      $y_m = \mathrm{argmin}_{y \in \mathbf{R}^m} \|\beta e_1 - \bar{H}_m y\|$;
      $\delta_m := \delta_0 + V_m y_m$;
      $\rho_m = \|\beta e_1 - \bar{H}_m y_m\|$;
        if $\rho_m < \xi \|F\|$ then
        convergence := true;
        $\delta := \delta_{\mathrm{m}}$;
      else
        $\delta_0 = \delta_{\mathrm{m}}$;
      endif
    enddo
```

The matrix $\bar{H}_m = (h_{lk})$ is an upper Hessenberg matrix of order $(m+1) \times m$. Usually a QR factorization of $\bar{H}_m$ using Givens rotations is employed to solve the least squares problem $\min_{y \in \mathbf{R}^m} \|\beta e_1 - \bar{H}_m y\|$. To simplify here, we have presented a version where each cycle goes until completion. Actually, the test of convergence is done also inside the cycle, but without computing δ_k, only by estimating ρ_k. Since the Jacobian is only approximated, the usual relations of GMRES are no longer satisfied, in particular the following.

Remark 1

In the algorithm MF-GMRES it is possible to get

$$JV_m \neq V_{m+1}\bar{H}_m, \qquad \rho_m \neq \| -F - J\delta_m \|.$$

2.1. Equivalence with a perturbed GMRES algorithm

In order to study the convergence of the inexact Newton method, we will first prove that each cycle of GMRES with an approximation of J is equivalent to a cycle of GMRES with an exact matrix–vector product, but where the right-hand side and the matrix are perturbed. The perturbations are directly related to the errors in the approximation of J. We follow the lines of Brown's proof,[3] but we introduce also a perturbation of the right-hand side to take into account the restarting procedure.

We define the matrix $\Sigma_m = (\sigma_1, \dots, \sigma_m)$ with $\sigma_k = q_{k+1} - Jv_k$ and the vector $\sigma_0 = q_1 - J\delta_0$.

Proposition 1

Each cycle of MF-GMRES(m, J, F, δ) with an initial guess δ_0 is equivalent to a cycle of GMRES(m, $\tilde{J}$, F, δ) with the same initial guess, where

$$\tilde{J} = J + \Sigma_m V_m^{\mathrm{T}}, \qquad \tilde{F} = F + \sigma_0 - \Sigma_m V_m^{\mathrm{T}} \delta_0,$$

assuming that the matrix $\tilde{J}$ is non-singular.

To prove this proposition, we will prove the following two lemmas where we denote with a tilde the variables occurring in GMRES(m, $\tilde{J}$, $\tilde{F}$, δ). The proposition follows readily from these lemmas.

Lemma 1

$\bar{H}_k = \tilde{\bar{H}}_k$ and $V_{k+1} = \tilde{V}_{k+1}$ for all $k = 1, \dots, m$.

Proof. First we show that the initial residuals are the same:

$$\tilde{r}_0 = -\tilde{F} - \tilde{J}\delta_0 = -(F + \sigma_0 + J\delta_0) = -F - q_1 = r_0.$$

Thus we have $\tilde{v}_1 = v_1$. Then we show the lemma by induction on k. We first note that

$$q_{k+1} = Jv_k + \sigma_k = (J + \Sigma_m V_m^{\mathrm{T}})v_k = \tilde{J}v_k.$$

Now, assuming that $\bar{H}_{k-1} = \tilde{\bar{H}}_{k-1}$ and $V_k = \tilde{V}_k$, we get, before the loop on $l = 1, \dots, k$, $\tilde{p} = \tilde{J}v_k = q_{k+1} = p$, so that $\tilde{h}_{lk} = h_{lk}$ for $l = 1, \dots, k+1$ and $\tilde{v}_{k+1} = v_{k+1}$. □

Lemma 2

Assuming that $\tilde{J}$ is non-singular, then $\rho_m = \tilde{\rho}_m$ and $\delta_m = \tilde{\delta}_m$.

Proof

$$\tilde{\rho}_m = \min_{y \in \mathbb{R}^m} \|\tilde{\beta}e_1 - \tilde{\bar{H}}_m y\|_2 = \min_{y \in R^m} \|\beta e_1 - \bar{H}_m y\|_2 = \rho_m.$$

Thus $y_m = \tilde{y}_m$ as $\tilde{J}$ is non-singular and $\delta_m = \tilde{\delta}_m$. □

This equivalence allows us to introduce an equivalent residual, the norm of which is used in MF-GMRES to test convergence.

Definition 1

At each cycle of MF-GMRES with an initial guess δ_0 the equivalent residual is given by

$$r_m = -\tilde{F} - \tilde{J}\delta_m \quad \text{with} \quad \delta_m = \delta_0 + V_m y_m \quad \text{and} \quad \|r_m\| = \rho_m.$$

2.2. Impact of perturbations on the residual

Since each cycle computes only an equivalent residual r_m, it is in general different from the initial residual r'_0 at the following cycle. Thus the sequence ρ_m in MF-GMRES is not necessarily non-increasing through the cycles. More precisely, we have the following result.

Proposition 2

We consider two consecutive cycles. For the first cycle, let us denote by δ_0 the initial solution, r_0 the initial residual, r_m the equivalent residual after the cycle, $\delta_m = \delta_0 + V_m y_m$ the solution obtained and σ_0 and $\Sigma_m = (\sigma_1, \dots, \sigma_m)$ the perturbations induced by the approximation of the Jacobian. For the second cycle, let us denote by $\delta'_0 = \delta_m$ the initial solution and r'_0 the initial residual with an approximation σ'_0. The difference between r'_0 and r_m is given by

$$\|r'_0 - r_m\| \leqslant \|\sigma_0\| + \|\sigma'_0\| + \|\Sigma_m\| \|y_m\|.$$

Proof. We use the definitions previously given to compute r_m and r'_0:

$$\begin{aligned} r_m &= -\tilde{F} - \tilde{J}\delta_m \\ &= -F - \sigma_0 + \Sigma_m V_m^{\mathrm{T}}\delta_0 - (J + \Sigma_m V_m^{\mathrm{T}})(\delta_0 + V_m y_m) \\ &= -F - \sigma_0 - J\delta_0 - JV_m y_m - \Sigma_m y_m, \\ r'_0 &= -F - J(\delta_0 + V_m y_m) - \sigma'_0, \\ r'_0 - r_m &= \sigma_0 - \sigma'_0 + \Sigma_m y_m. \end{aligned}$$

We simply get the result by taking the norms. □

The initial perturbations of the type σ_0 appear alone, whereas the subsequent perturbations, occurring during the basis construction, are multiplied by y_m. Since $\|y_m\|$ is a multiple of $\beta = \|r_0\|$, it becomes smaller and smaller at each restart. Therefore we can expect a larger impact of the approximation on the initial residual than of the approximation in the Krylov vectors. This conjecture will be confirmed numerically by the experiments of the next sections.

Another way of measuring the impact of these perturbations is through the condition number of the Krylov subspace, as defined in Reference 7. We have shown, by equivalence with an exact GMRES, that Newton–GMRES builds a Krylov subspace $\tilde{K}(m, \tilde{v}_1, \tilde{J})$, where $\tilde{v}_1 = r_0/\|r_0\|$. Let us denote by $\hat{K}(m, \hat{v}_1, J)$ the Krylov subspace generated by the exact GMRES cycle using the exact Jacobian J with the same initial solution δ_0. Here $\hat{v}_1 = \hat{r}_0/\|\hat{r}_0\|$, with $\hat{r}_0 = r_0 + \sigma_0$. Adapting the definitions of Reference 7, the distance between the two Krylov subspaces is bounded using the Krylov condition number μ as follows.

Proposition 3

For sufficiently small perturbations σ_k, $k = 0, \ldots, m$, the distance between the computed subspace $\tilde{K}$ and the true Krylov subspace $\hat{K}$ is bounded at the first order in $\max(\|\sigma_0\|/\|r_0\|, \|\Sigma_m\|/\|J\|)$ by

$$d(\tilde{K}, \hat{K}) \leqslant \mu \max(\alpha\|\sigma_0\|/\|r_0\|, \|\Sigma_m\|/\|J\|),$$

where μ is the Krylov condition number and α is a constant.

Proof. Let $\phi = \max(\|\tilde{v}_1 - \hat{v}_1\|, \|J - \tilde{J}\|/\|J\|)$. The definition of the condition number μ implies

$$d(\tilde{K}, \hat{K}) \leqslant [\mu + o(\phi)]\phi.$$

Clearly, $\|J - \tilde{J}| = \|\Sigma_m\|$ and $\|\tilde{v}_1 - \hat{v}_1\| \leqslant \alpha\|\sigma_0\|/\|r_0\|$, where α is some constant, which gives the result. □

Here too the perturbation in the initial residual has more impact because it is divided by $\|r_0\|$, which becomes smaller and smaller, whereas the impact of Σ_m is reduced by $\|J\|$.

3. CONVERGENCE ANALYSIS

Thus now, for each iteration i of the Newton method, we introduce the following notation: δ_0 is the initial guess in the last cycle of MF-GMRES, $J_i = J(u_i)$ and $F_i = F(u_i)$, $\tilde{J}_i$ and $\tilde{F}_i$ are the perturbed matrix and right-hand sides of the equivalent GMRES for this last cycle, δ_i is the computed solution and r_i is the equivalent residual. We then have the following.

Proposition 4

The inexact Newton algorithm builds the following sequence: u_0 is given,

$$u_{i+1} = u_i + \delta_i \quad \text{with} \quad \tilde{J}_i\delta_i = -(\tilde{F}_i + r_i).$$

We are now able to study the impact on the Newton convergence of the approximation of the Jacobian (given by the perturbations in the matrix and the right-hand sides) and the approximation of the solution (given by the estimated residual).

3.1. Global convergence

Quite often a line search backtracking technique is combined with the Newton iterations to improve the global convergence. This is possible in our framework if the solution given by GMRES is a descent direction. Indeed, we know that[6,8] if δ_i evaluated by GMRES is a descent direction at u_i, i.e.

$$F_i^{\mathrm{T}} J_i \delta_i < 0,$$

there exists $\mu > 0$ satisfying

$$\|F(u_i + \mu\delta_i)\| < \|F_i\|.$$

In Reference 4 it is shown that with no restarting and with an exact Jacobian the solution given by GMRES is a descent direction. In Reference 3 a condition is given for the solution of MF-GMRES with an approximation of the Jacobian but no restarting to be a descent direction. Here we extend this result to GMRES with restarting and we express the condition using the estimated norm of the residual ρ_m.

Proposition 5

We consider here a cycle of MF-GMRES(m, J, F, δ) with an initial guess δ_0. Assuming that $\tilde{J}$ is a non-singular, then $\delta_m = \delta_0 + V_m y_m$ is a descent direction for J if

$$\rho_m < \|F\|_2 - \|\sigma_0\|_2 - \|\Sigma_m\|_2\|y_m\|_2.$$

Proof

$$\begin{aligned} J\delta_m &= J\delta_0 + JV_m y_m \\ &= -(F + \sigma_0 + r_0) + JV_m y_m \\ &= -F - \sigma_0 - r_m - \Sigma_m y_m, \\ F^{\mathrm{T}} J\delta_m &\leqslant \|F\|_2(\|r_m\|_2 + \|\sigma_0\|_2 + \|\Sigma_m\|_2\|y_m\|_2 - \|F\|_2). \end{aligned}$$

The proposition follows. □

Provided that the residual in GMRES is small enough, we can use the following backtracking algorithm at each Newton iteration.

Algorithm. Backtracking

```
* Let $f(u) = F(u)^{\mathrm{T}}F(u)$ and $\delta$ be a descent direction at $u$,
this algorithm computes a scalar $\mu$ satisfying $f(u + \mu\delta) < f(u)$.
choose $\alpha \in (0, \frac{1}{2})$
$\mu = 1$
while $f(u + \mu\delta) > f(u) + \mu\alpha F^{\mathrm{T}}J\delta$ do
```

```
    $\mu = \rho\mu$ with $\rho \in (0, 1)$
end do
```

3.2. Local convergence

The basic Newton method is known to converge quadratically near the solution, whereas inexact Newton methods where the linear system is solved iteratively are known to converge only linearly or at most superlinearly. Here we prove also a local linear convergence when using the MF-GMRES linear solver.

Theorem 1

Let F: $\mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in an open convex set $D \subset \mathbb{R}^n$. Assume that there exists u_* and $1 > \alpha > 0$ such that

$$\begin{cases} N(u_*, \alpha) \subset D, \\ F(u_*) = 0, \\ J(u_*)^{-1} \text{ exists with } \|J(u_*)^{-1}\| = M_1, \\ J \in \mathrm{Lip}_\gamma(N(u_*, \alpha)), \\ M = \sup_{u \in N(u_*, \alpha)} \|J(u)\|, \\ M_1(\gamma + M)\alpha \leqslant \frac{1}{2}, \\ M_1(\gamma + 6M)\alpha < 1. \end{cases}$$

Let a sequence $u_1, u_2, \ldots$ be generated by

$$u_{i+1} = u_i - \tilde{J}_i^{-1}(\tilde{F}_i + r_i),$$

with

(A1) $\|\tilde{J}_i - J_i\| \leqslant \|F_i\|$,

(A2) $\|\tilde{F}_i - F_i\| \leqslant \alpha\|F_i\|$,

(A3) $\|r_i\| \leqslant \alpha\|F_i\|$.

Then for all $u_0 \in N(u_*, \alpha)$ the sequence $u_1, u_2, \ldots$ is well defined and converges linearly to u_*. More precisely, we have

$$\|u_{i+1} - u_*\| \leqslant M_1(\gamma + 6M)\alpha\|u_i - u_*\|.$$

Proof. We prove the theorem by recurrence, assuming $u_i \in N(u_*, \alpha)$, which is true for $i = 0$.

1. First we show that $\tilde{J}_i$ is non-singular. We have $\|F_i\| \leqslant M\|u_i - u_*\| \leqslant M\alpha$ and

$$\begin{aligned} \|J(u_*)^{-1}[\tilde{J}_i - J(u_*)]\| &\leqslant \|J(u_*)^{-1}\| \|[\tilde{J}_i - J_i)] + [J_i - J(u_*)]\| \\ &\leqslant M_1(\|F_i\| + \gamma\alpha) \\ &\leqslant M_1(M + \gamma)\alpha \\ &\leqslant \tfrac{1}{2}, \end{aligned}$$

so that $\tilde{J}_i$ is non-singular and, using Reference 6, $\|\tilde{J}_i^{-1}\| \leqslant 2M_1$.

2. Therefore u_{i+1} is well defined and

$$\begin{aligned} u_{i+1} - u_* &= u_i - u_* - \tilde{J}_i^{-1}\tilde{F}_i - \tilde{J}_i^{-1}\tilde{r}_i \\ &= \tilde{J}_i^{-1}[F(u_*) - F_i - J_i(u_* - u_i) + (J_i - \tilde{J}_i)(u_* - u_i) + (F_i - \tilde{F}_i) - r_i]. \end{aligned}$$

Since $J \in \mathrm{Lip}_\gamma(N(u_*, \alpha))$, we have

$$\|F(u_*) - F_i - J_i(u_* - u_i)\| \leqslant \frac{\gamma}{2}\|u_i - u_*\|^2.$$

It should be noted that the inequality above is the quadratic term in the exact Newton procedure. Now we deal with the errors due to the inexact Jacobian and to the approximate solution by GMRES. By (A1)–(A3) we get

$$\|(J_i - \tilde{J}_i)(u_* - u_i) + F_i - \tilde{F}_i - r_i\| \leqslant 3\|F_i\|\alpha.$$

Thus

$$\|u_{i+1} - u_*\| \leqslant M_1(\gamma + 6M)\alpha\|u_i - u_*\|.$$

From this theorem we readily get a convergence result for our framework.

Corollary 1

Under the same assumptions as above on J, the inexact Newton–MF-GMRES algorithm builds a sequence of iterates u_i which is well defined and convergent to the solution u_* provided that u_0 is close enough to u_*, that the approximation of J is accurate enough and that the tolerance ξ in MF-GMRES is small enough.

Proof. We simply need to check the assumptions (A1)–(A3) for the last cycle of MF-GMRES. We have

$$\begin{aligned} &\|\tilde{J} - J\| = \|\Sigma_m\| \leqslant \sqrt{m} \max_{1 \leqslant k \leqslant m} \|\sigma_k\|, \\ &\|\tilde{F} - F\| \leqslant \sigma_0\| + \|\Sigma_m\|\|V_m^{\mathrm{T}}\delta_0\|, \\ &\|r\| = \rho \leqslant \xi\|F\|. \end{aligned}$$

Thus, if $\max_{0 \leqslant k \leqslant m} \|\sigma_k\|$ and ξ are small enough, we get the result wanted. □

The assumptions on J contain the term M_1M which actually reflects the condition number of the Jacobian. It indicates that a good preconditioner would improve the convergence of the Newton iterations.

The errors σ_0 and σ_k which appear when approximating the Jacobian must be small not only to ensure a local convergence but also to get a descent direction. We study now these errors when using a finite difference scheme to approximate the Jacobian.

4. APPROXIMATION OF THE JACOBIAN BY A FINITE DIFFERENCE SCHEME

A finite difference scheme introduces a step τ to approximate $J(u)v$. This parameter should be small enough to get an accurate approximation. However, it is well known that, because of rounding errors, this step must be large enough compared with the machine precision ε. We consider in the following a first-order and a centred second-order scheme.

4.1. *First-order finite difference scheme*

We use the first-order approximation

$$J(u)v \approx \frac{F(u+\tau v)-F(u)}{\tau} \tag{1}$$

In the absence of rounding errors, if $J(u)$ is $\mathrm{Lip}_{\gamma}(N(u_{*}, \tau))$, then[6]

$$\|\sigma_k\| = \left\| J(u)v_k - \frac{F(u+\tau v_k)-F(u)}{\tau} \right\| \leqslant \frac{\tau\gamma}{2}.$$

However, rounding errors are of the order of ε/τ, so that, as in Reference 6, we advocate choosing

$$\tau = \frac{\sqrt{\varepsilon}\,(u^{\mathrm{T}}v)}{\|v\|_2}$$

to get a global error $\|\sigma_k\| = O(\sqrt{\varepsilon})$.

This error may be too large and may slow down the convergence. Therefore we study a more accurate second-order scheme.

4.2. *Second-order finite difference scheme*

Now we introduce a centred second-order scheme defined by

$$J(u)v \approx \frac{F(u+\tau v)-F(u-\tau v)}{2\tau}. \tag{2}$$

If $J(u)$ is sufficiently regular, the approximation error is of the order of τ^2 whereas rounding errors are of the order of ε/τ. Thus now we choose

$$\tau = \sqrt[3]{\varepsilon}\,\frac{(u^{\mathrm{T}}v)}{\|v\|_2}$$

to get a global error $\|\sigma_k\| = O(\varepsilon^{2/3})$.

The smaller error is obtained at a larger CPU time, since this scheme involves twice as much work as the first-order scheme above. We now study the effects of both schemes with numerical examples.

5. NUMERICAL EXPERIMENTS

5.1. *Equations and numerical schemes*

We consider the implicit solution of a compressible, Newtonian and viscous fluid without source terms. The Navier–Stokes equations governing the flow are written in the conservative form

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho U) = 0\\
&\frac{\partial(\rho U)}{\partial t} + \operatorname{div}(\rho U \otimes U) + \nabla p = \operatorname{div}(\mu S),\\
&\frac{\partial e}{\partial t} + \operatorname{div}[(e+p)U] = \operatorname{div}(\mu S U) + \operatorname{div}(\kappa \nabla T).
\end{aligned} \tag{3}$$

Here ρ is the density, U is the velocity, T is the temperature, $e = \rho(T + \|U\|^2)/2)$ is the total energy density, $p = (\gamma = 1)\rho T$ is the pressure, $S = (\nabla U + \nabla U^{\mathrm{T}}) - \frac{2}{3}\operatorname{div}(U)I$ is the deformation tensor and

$\kappa=\gamma\mu/Pr$ is defined via the parameters $\gamma=1{\cdot}4$, $Pr=0{\cdot}72$ for air and $\mu=1/Re$ given by the Sutherland law

$$\mu=\mu_\infty\left(\frac{T}{T_\infty}\right)^{1\cdot5}\frac{T_\infty+110}{T+110},$$

where the subscript ∞ denotes reference quantities.

To solve (3), we use the conservative variables $(\rho,\ \rho U,\ e)$.[9] The convective terms are upwinded thanks to a finite volume formulation for the Euler part of the equations. The Riemann problem at each interior interface is solved by approximating the flux with the Osher scheme.[10] This scheme is differentiable if and only if $U\cdot v\neq 0$ (where U and v are respectively the value of the velocity and the normal at each interface). We also use the Steger–Warming flux splitting[11] to approximate the flux at the freestream boundary. For the diffusive term we use a standard P_1 finite element formulation. Roughly speaking, after a mixed finite volume/finite element P_1 formulation we have to solve

$$u_{,t}+G(u)=0, \tag{4}$$

where $u\in\mathbb{R}^N$ is composed of blocks of the variables $(\rho,\ \rho U,\ e)$ and G is a non- linear function in $\mathbb{R}^N$. Of course, G depends on the space discretization used.

Here we consider only steady state solutions of (4). We have used the software developed at INRIA[9] where time is discretized by an explicit Euler scheme given by

$$u(n+1)=u(n)+\Delta t_nG(u). \tag{5}$$

Then we have implemented an implicit backward Euler scheme given by

$$u(n+1)+\Delta t_nG(u(n+1))-u(n)=0, \tag{6}$$

where each time step gives rise to the non-linear problem $F(u)=0$, where $F(u)=u+\Delta t_nG(u)-u(n)$. Assuming that $\|\partial G(u)/\partial u\|$ is bounded, we can always find a time step Δt_n such that the Jacobian $J(u)=I+\Delta t_n\partial G(u)/\partial u$ of (6) is non-singular.

We have also implemented the resolution of the stationary non-linear problem with no time derivative (corresponding to an infinite time step):

$$G(u)=0. \tag{7}$$

Here we cannot prove that the Jacobian $\partial G/\partial u$ is non-singular.

In both cases, since the Jacobian is not explicitly known, we solve (6) and (7) by the inexact matrix-free algorithms given in Section 2, including a backtracking strategy.

For numerical purposes we consider the following two steady state problems in two-dimensional space: *problem 1*—a viscous flow at Mach 0·8 and $Re=5000$ around an NACA0012 aerofoil with an incidence of 3°; *problem 2*—an inviscid flow at Mach 1·2 around an NACA0012 aerofoil with an incidence of 7°. We discretize the domain by a finite element mesh with 801 nodes and 1516 triangles; thus $N=3204$. All the computations are done on a SPARC-IPX workstation.

5.2. Numerical studies of the backward Euler integration

The explicit Euler scheme (5) is used during the first 100 steps where the solution varies greatly, with a local time step denoted $\Delta t_n(\mathrm{exp})$. Reasonably large time steps $\Delta t_n(\mathrm{imp})=CFL\Delta t_n(\mathrm{exp})$ can then be used in the implicit scheme (6) while guaranteeing the convergence of Newton iterations. We analyse in detail the first implicit non-linear problem (6) where $u(n)$ is given by the last explicit iteration. We use the inexact Newton–MF-GMRES algorithm where GMRES is restarted up to four times and every m iterations. The convergence threshold in GMRES is chosen to be $\xi=10^{-4}$. We want to study the impact of the approximation in the Jacobian on both the GMRES and the Newton convergence. We first approximate the Jacobian by the first-order scheme (1). Then we replace the approximation by the second-order scheme (2), first only for the initial residual, then also in the Arnoldi process. In order to measure the impact of the approximation, we compare at the end of each restart the equivalent residual $\rho_m = \|r_m\|$ given by the least squares problem with the true residual (also approximated) $\| -F - \text{appr}(J\delta_m)\|$. These quantities should be equal if the Jacobian was computed exactly (see Remark 1). We also compute the gradient $F^{\mathrm{T}}J\delta/\|\delta\|$ to check whether backtracking can be applied.

The first results are given in Tables I–VII, with the following contents in each column:

1. the number of the Newton iteration i; a letter 'b' indicates a backtracking to decrease $\|F_i^{\mathrm{T}}F_i\|$
2. the relative residual $\|F_i/F_0\|$ before the iterate i
3. the equivalent residual $\rho_m/\|F_i\|$ after each cycle of MF-GMRES
4. the true residual approximated by $\| -F_i - \text{appr}(J_i\delta_m)\|/\|F_i\|$
5. the gradient $F_i^{\mathrm{T}}J_i\delta_m/\|\delta_m\|$.

Tables I and II give the results for both problems with a first-order scheme. We observe that the equivalent residual is completely different from the true residual. In that case the convergence of Newton is rather slow. Results when using a second-order scheme for the initial residual at each restarting are given in Tables III and IV. For both problems the Newton convergence is improved with no requirement for backtracking. In problem 1 the equivalent residual of MF-GMRES now estimates accurately the true residual. In problem 2 there is still a large difference. Finally, Tables V and VI give the results when using a second-order scheme in all approximations of the Jacobian. The effect is not so impressive as in Tables III and IV.

Finally, we increase the local time steps for problem 2 so that the assumptions for the local convergence of Newton are not satisfied. We see through Table VII that backtracking is necessary in the first Newton iterations.

5.3. *Numerical studies of the non-discretized problem*

Then we solve the problem $G(u) = 0$ with no time discretization using an initial guess given by the first 100 steps of the explicit Euler scheme for problem 1 and the first 1000 steps of the explicit Euler scheme for problem 2. We compute 10 steps of the inexact Newton–MF-GMRES algorithm where GMRES is restarted up to four times and every $m = 20$ iterations steps, with the same convergence threshold. The results for the full test cases are given in Tables VIII and IX. The tables show the influence of the order of the finite difference scheme (given in the first column) on the relative residual

Table I. Observations for problem 1 with a first-order scheme, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\| -F_i - \text{appr}(J,\delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}}J_i\delta_m/\|\delta_m\|$
0	1·0	1·681793E-04	1·681824E-04	
–		7·998507E-05	7·999422E-05	−0·496986
1	1·380937E-03	2·058242E-02	2·058544E-02	
–		2·297247E-04	3·788322E-04	
–		9·358031E-05	3·062475E-04	−1·140616E-03
2	4·228783E-07	8·097340E-03	0·477054	
–		9·451777E-04	0·470251	
–		1·126173E-03	0·499369	
–		1·089406E-03	0·515916	−6·242091E-07
3	2·182093E-07			

Table II. Observations for problem 2 with a first-order scheme, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \mathrm{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	3·018444E-04	3·018057-04	
–		9·337174E-05	9·339440E-05	−0·202629
1	1·339627E-02	8·692537E-03	8·945976E-03	
–		7·533453E-05	2·251117E-03	−4·336211E-03
2	3·013312E-05	3·898022E-03	0·519220	
–		1·863874E-03	0·808648	
–		3·398747E-03	1·288501	
–		4·955945E-03	0·529647	−2·813971E-05
3	1·595999E-05	3·795039E-03	3·567834	
–		1·407613E-02	3·724042	
–		1·492283E-02	1·007644	
–		3·563932E-03	0·767542	−1·339620E-05
4	1·224939E-05			

Table III. Observations for problem 1 with a second-order scheme at each restarting, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \mathrm{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	1·681793E-04	1·681890-04	
–		7·999014E-05	7·999013E-05	−0·496986
1	1·380913E-03	2·058248E-02	2·058245E-04	
–		2·294992E-04	2·294790E-04	
–		8·559659E-05	8·559996E-05	−1·140582E-03
2	1·176692E-07	1·346541E-02	1·452551E-02	
–		1·116557E-04	7·522575E-03	
–		9·111143E-05	7·323161E-03	−5·954292E-08
3	6·2240364E-10			

Table IV. Observations for problem 2 with a second-order scheme at each restarting, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \mathrm{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	3·018444E-04	3·018499E-04	
–		9·338585E-05	9·338578E-05	−0·202629
1	1·339628E-02	8·677930E-03	8·677844E-03	
–		9·624604E-05	9·675667E-05	−4·336226E-03
2	2·069862E-06	8·260932E-03	8·696502E-03	
–		7·655946E-05	3·732129E-03	−7·566524E-07
3	5·747445E-09			

Table V. Observations for problem 1 with a second-order scheme, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \mathrm{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	1·681898E-04	1·681898E-04	
–		7·998903E-05	7·998903E-05	−0·496986
1	1·380904E-03	2·058190E-02	2·058190E-02	
–		2·295504E-04	2·295210E-04	
–		8·560229E-05	8·560299E-05	−1·140587E-03
2	1·176182E-07	1·349783E-02	1·445929E-02	
–		1·131854E-04	7·518289E-03	
–	8·494013E-05	7·531063E-03	−5·949319E-08	
3	6·262193E-10			

Table VI. Observations for problem 2 with a second-order scheme, $m = 15$, $CFL = 5{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \text{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	3·018737E-04	3·018737E-04	
–		9·340862E-05	9·340853E-05	−0·202629
1	1·339629E-02	8·590151E-03	8·590011E-03	
–		9·478208E-05	9·505118E-05	−4·336222E-03
2	2·055026E-06	8·093605E-03	8·707604E-03	
–		8·024169E-05	3·960275E-03	−7·527701E-07
3	6·038253E-09			

Table VII. Observations for problem 2 with a second-order scheme at each restarting, $m = 15$, $CFL = 500{\cdot}0$

i	$\|F_i\|/\|F_0\|$	$\rho_m/\|F_i\|$	$\|-F_i - \text{appr}(J, \delta_m)\|/\|F_i\|$	$F_i^{\mathrm{T}} J_i \delta_m/\|\delta_m\|$
0	1·0	0·330659	0·330659	
–		0·213414	0·213414	
–		0·142027	0·142028	
–		8·501748E-02	8·501752E-02	−111·572
0b	2·981211			
1	0·777835	0·555933	0·555934	
–		0·347357	0·347357	
–		0·212065	0·212065	
–		0·155792	0·155792	−147·262
1b	0·917935			
2	0·465581	0·525731	0·535731	
–		0·382918	0·382918	
–		0·254039	0·254039	
–		0·188227	0·188227	−100·297
3	0·254054	0·332741	0·332741	
–		0·212985	0·212985	
–		0·141447	0·141447	
–		9·725137E-02	9·725147E-02	−113·915
4	3·632507E-02	0·468147	0·468147	
–		0·347019	0·347019	
–		0·220597	0·220597	
–		0·153951	0·153951	−7·786406
5	5·431169E-03			

$\|G_{10}\|/\|G_0\|$ (given in the second column) and also show the cost (number of function evaluations of $G(u)$ and CPU time). The second-order scheme is only applied at each restarting for the case 2 and applied also during the Arnoldi process for the case 2f.

The use of a second-order scheme is of no help for problem 1, as observed in Table VIII. On the other hand, Table IX for problem 2 shows clearly the drawback of the first-order finite difference scheme. As in the previous test cases, the true residual in GMRES does not converge rapidly, slowing down the convergence in the Newton process. Here too the main improvement comes from applying the second-order scheme at each restarting. This adds a small overhead in time which can benefit the global convergence.

Table VIII. Direct solution with backtracking for problem 1

Order of FD scheme	$\|G_{10}\|/\|G_0\|$	# (evaluations)	CPU time (s)
1	5·49E-03	861	343
2	5·53E-03	891	361
2f	5·32E-03	1691	613

Table IX. Direct solution with backtracking for problem 2

Order of FD scheme	$\|G_{10}\|/\|G_0\|$	# (evaluations)	CPU time (s)
1	7·60E-04	861	324
2	1·14E-04	891	344
2f	1·11E-04	1691	620

6. CONCLUSIONS

In this paper we studied the convergence of a Newton–Krylov method where the linear system involving the Jacobian at each Newton iteration is solved by a restarted GMRES linear solver. We showed that the solution given by GMRES is a descent direction if the tolerance for convergence in GMRES is sufficiently small. A linear local convergence of Newton is guaranteed if this tolerance is small and also if the approximation of the Jacobian is accurate enough. These results can be applied to any matrix-free iterative linear solver.

We applied this algorithm to the numerical simulation of compressible flows, using an implicit time discretization. Numerical results on two problems confirm the theoretical study. In the light of our experiments we advocate the use of a second-order finite difference scheme to approximate the initial residual at each restart of GMRES. It clearly improves the convergence of Newton at a low CPU cost in each iteration.

REFERENCES

1. S. C. Eisenstat and H. F. Walker, 'Globally convergent inexact Newton methods', *SIAM J. Optim.*, **4**, 393–422 (1994).
2. P. N. Brown and Y. Saad, 'Convergence theory of nonlinear Newton– Krylov algorithms', *SIAM J. Optim.*, **4**, 297–330 (1994).
3. P. N. Brown, 'A local convergence theory for combined inexact- Newton/finite difference projection methods', *SIAM J. Numer. Anal.*, **24**, 407–434 (1987).
4. P. N. Brown and Y. Saad, 'Hybrid Krylov methods for nonlinear systems of equations', *SIAM J. Sci. Stat. Comput.*, **11**, 450–481 (1990).
5. Y. Saad and H. Schultz, 'GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems', *SIAM J. Sci. Stat. Comput.*, **7**, 856–869 (1986).
6. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
7. J.-F. Carpraux, S. K. Godunov and S. V. Kuznetsov, 'Condition number of the Krylov bases and subspaces', *Linear Algebra Appl.*, in press.
8. J. M. Ortega and W. C. Rheinholdt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic, New York, 1970.
9. B. Mohammadi, 'Fluid dynamics computation with NSC2KE. An user-guide, release 1.0', *Tech. Rep. RT-0164*, INRIA, 1994.
10. S. Osher and F. Solomon, 'Upwind difference schemes for the hyperbolic systems of conservation laws', *Math. Comput*, **38**, 339–374 (1982).
11. J. L. Steger and R. F. Warming, 'Flux vector splitting of the inviscid gasdynamic equations with application to finite-difference methods', *J. Comput. Phys.*, **40**, 263–293 (1981).