

---

# A general algorithm for compressible and incompressible flow. Stability analysis and explicit time integration

Stability  
analysis

141

---

N. Nigro, M. Storti and S. Idelsohn

*Grupo de Tecnología Mecánica del INTEC, Universidad Nacional del Litoral and CONICET, Santa Fe, Argentina*

*<http://venusunl.edu.ar/gtm-enghtml>*

## Introduction

Solving flow problems in compressible or incompressible regimes has been, historically, a very different matter. There are mainly two difficulties associated with incompressible flows, namely a compatibility condition on the interpolation spaces and bad conditioning of the system matrices. Compressible codes generally use equal-order interpolations. If such a compressible code is used in a nearly incompressible flow pattern, i.e. with a Mach number  $M \ll 1$ , then oscillations in pressure ("checkerboard modes") are likely to appear. Moreover, these oscillations are likely to occur in locally incompressible regions, like stagnation points, embedded in a globally compressible flow[1]. The subject has been extensively studied by numerical analysts and the conclusion is that the interpolation spaces must satisfy a stability condition named the LBB condition[2,3]. On the other hand, compressible flows can be solved with explicit schemes. If the objective is to reach the steady state, then, with techniques like local time stepping and absorbing boundary conditions, an efficient, easy to code and highly parallelizable algorithm requiring a very low amount of core memory is obtained. However, in the limit of incompressible flows, the rate of convergence drops like  $1/M$  owing to bad conditioning, and the strategy becomes unfeasible.

In this work, these two difficulties are addressed. First, a stabilized algorithm is presented for the incompressible or nearly incompressible Navier-Stokes equations, allowing equal-order interpolations. The stabilization terms are obtained as a straightforward extension of the SUPG (Streamline Upwind Petrov-Galerkin) method to multidimensional advective-diffusive systems of equations. Second, it is shown how the explicit scheme can be used in the

---

The authors wish to express their gratitude to Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) for its financial support.

International Journal of Numerical  
Methods for Heat & Fluid Flow  
Vol. 7 No. 2/3, 1997, pp. 141-168.  
© MCB University Press, 0961-5539

incompressible regime through the use of a specially devised preconditioning. Using these techniques, rates of convergence independent of Mach number are achieved and the compressible explicit code can be used in the incompressible limit.

**Stabilized methods for incompressible flow**  
*Conservative form of the Navier-Stokes equations*

Navier-Stokes equations written in conservative form are:

$$\mathbf{F}_{a,i}^i = \mathbf{F}_{d,i}^i + \mathbf{b} \quad (1)$$

Implicitly we assume  $(\cdot)_{,i} = \partial(\cdot)/\partial x_i$ .

$\mathbf{U} \in \mathbb{R}^5$  is the fluid local state vector, where  $\mathbf{U} = [\rho, \rho \mathbf{u}^T, \rho e]^T$ , with  $\rho, \mathbf{u}, e$  as the density, velocity and total energy of the fluid.  $\mathbf{F}_a, \mathbf{F}_d \in \mathbb{R}^{5 \times 3}$  are the advective and diffusive fluxes respectively, which depend on the state vector and its gradient as:

$$\mathbf{F}_a(\mathbf{U}) = \begin{bmatrix} \rho \mathbf{u}^T \\ \rho \mathbf{u} \mathbf{u}^T + p \mathbf{I}_{3 \times 3} \\ (\rho e + p) \mathbf{u}^T \end{bmatrix} \quad \mathbf{F}_d(\mathbf{U}, \nabla \mathbf{U}) = \begin{bmatrix} 0 \\ \boldsymbol{\tau} \\ (\boldsymbol{\tau} \cdot \mathbf{u} - \mathbf{q})^T \end{bmatrix} \quad (2)$$

where  $p, \boldsymbol{\tau}, \mathbf{q}$  are the thermodynamic pressure, the stress deviatoric tensor and the heat flux vector respectively, with

$$\tau_{jk} = \mu(u_{j,k} + u_{k,j}) + \lambda u_{l,l} \delta_{jk} \quad (3.1)$$

$$q_j = -\kappa \theta_{,j} \quad (3.2)$$

where  $\mu$  is the dynamic viscosity,  $\lambda$  is the second viscosity coefficient,  $\kappa$  the thermal conductivity and  $\theta$  the temperature. The right-hand side term  $\mathbf{b}$  involves body forces and external heat-sources. The description of the mathematical model finishes with an introduction of the state equation of the fluid and the relation between the energy and two of the thermodynamic variables of the fluid

$$\theta = \theta(p, \rho), \quad e = e(p, \rho) \quad (4)$$

*Numerical spatial discretization – A Petrov-Galerkin formulation overview*

The numerical formulation is based on the Petrov-Galerkin weighted residual method which allows test functions that can be different from the interpolation ones and not necessarily continuous. This method introduces the numerical dissipation needed to stabilize the system in advection-dominated problems, keeping the consistency with the continuum problem, and is one of the most referenced in the CFD area by the finite element method[4,5].

To fix ideas suppose that:

$$\mathbf{F}_a^i(\mathbf{U}) = \mathbf{A}_i \mathbf{U}, \quad \mathbf{F}_d^i(\mathbf{U}, \nabla \mathbf{U}) = \mathbf{K}_{i,j} \mathbf{U}_{,j} \quad (5)$$

where  $\mathbf{A}_j$  and  $\mathbf{K}_{ij}$  are constant matrices. For each node  $\mathbf{a}$  there is an interpolation function  $N_{\mathbf{a}}$  (hat type in 1D, bilinear in 2D and multilinear in general) and a test function  $W_{\mathbf{a}} = N_{\mathbf{a}} + P_{\mathbf{a}}$ , where  $P_{\mathbf{a}}$  is called the perturbation function. The standard Galerkin method is recovered when we impose  $P_{\mathbf{a}} \equiv 0$ . The  $W_{\mathbf{a}}$  (and, of course,  $P_{\mathbf{a}}$ ) are not necessarily continuous through the inter-element boundaries. The variational formulation employed is:

$$\int_{\Omega} (\mathbf{N}_{\mathbf{a}}^T \mathbf{A}_i \mathbf{U}_{,i} + \mathbf{N}_{\mathbf{a},i}^T \mathbf{K}_{ij} \mathbf{U}_{,j}) d\Omega + \sum_{e=1}^{N_{el}} \int_{\Omega_e} \mathbf{P}_{\mathbf{a}}^{eT} (\mathbf{A}_i \mathbf{U}_{,i} + \mathbf{K}_{ij} \mathbf{U}_{,j} - \mathbf{b}) d\Omega = \int_{\Omega} \mathbf{N}_{\mathbf{a}}^T \mathbf{b} d\Omega + \int_{\Gamma} \mathbf{N}_{\mathbf{a}}^T \mathbf{h} d\Gamma \quad (6)$$

where  $\mathbf{h}$  is the diffusive flux imposed on the boundary  $\Gamma_h$ . The Euler-Lagrange form is obtained through classical integration by parts:

$$\sum_{e=1}^{N_{el}} \int_{\Omega_e} \mathbf{W}_{\mathbf{a}}^T (\mathbf{A}_i \mathbf{U}_{,i} - \mathbf{K}_{ij} \mathbf{U}_{,j} - \mathbf{b}) d\Omega + \int_{\Gamma_{int}} \mathbf{N}_{\mathbf{a}}^T [n_i \mathbf{K}_{ij} \mathbf{U}_{,j}] d\Gamma + \int_{\Gamma_h} \mathbf{N}_{\mathbf{a}}^T (n_i \mathbf{K}_{ij} \mathbf{U}_{,j} - \mathbf{h}) d\Gamma = 0 \quad (7)$$

where:

$$[n_i \mathbf{K}_{ij} \mathbf{U}_{,j}](\mathbf{x}) = n_i(\mathbf{x}) \{ (\mathbf{K}_{ij} \mathbf{U}_{,j})(\mathbf{x}^+) - (\mathbf{K}_{ij} \mathbf{U}_{,j})(\mathbf{x}^-) \} \quad (8)$$

is the jump in the diffusive flux throughout the inter-element boundary,  $\mathbf{x}$  is a point which lies there,  $\mathbf{x}^{\pm}$  are points belonging to each side of the boundary and  $\Gamma_{int}$  is the inter-element contour. Consistency is warranted because the continuum solution is also solution of this variational formulation.

In the following sections, the type of perturbation function used is described. The simple scalar one-dimensional problem is investigated and then these results are extended to multidimensional systems.

*One-dimensional scalar advection-diffusion equation.* In the scalar case:

$$\mathbf{a} u_{,x} = k u_{,xx} \quad (9)$$

is the equation to be solved where  $\mathbf{a}$  is the advection velocity of the fluid,  $k$  is the diffusivity and  $u$  the dependent variable. By a centred finite difference scheme or, similarly, by a finite element (Galerkin method) formulation, the following linear system on a homogeneous grid of spacing  $\Delta x$  is obtained:

$$\mathbf{a} \frac{u_{i+1} - u_{i-1}}{2\Delta x} = k \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} \quad (10)$$

The truncation error is  $O(\Delta x^2)$ , but it produces wiggles for advection-dominated problems characterized by a Péclet number  $Pe = u\Delta x/k \gg 1$ . In a finite difference context it is well known that the previous defect is solved by the introduction of artificial diffusion or by an upwinding technique. Exact nodal values can be obtained if the numerical diffusion is adjusted according to a

certain expression which is then called the “magic function”. This phenomenon is known as “super-convergence”. In an FEM context, this scheme can be cast as a Petrov-Galerkin formulation if the perturbation function for the node  $i$  is taken as:

$$P_i = \tau a N_{i,x}, \quad \tau = \frac{\Delta x}{2|a|} \psi(Pe) \tag{11}$$

where the magic function is defined as (see Figure 1):

$$Pe = \frac{u\Delta x}{2|a|}, \quad \psi(Pe) = \coth(Pe) - \frac{1}{Pe} \tag{12}$$

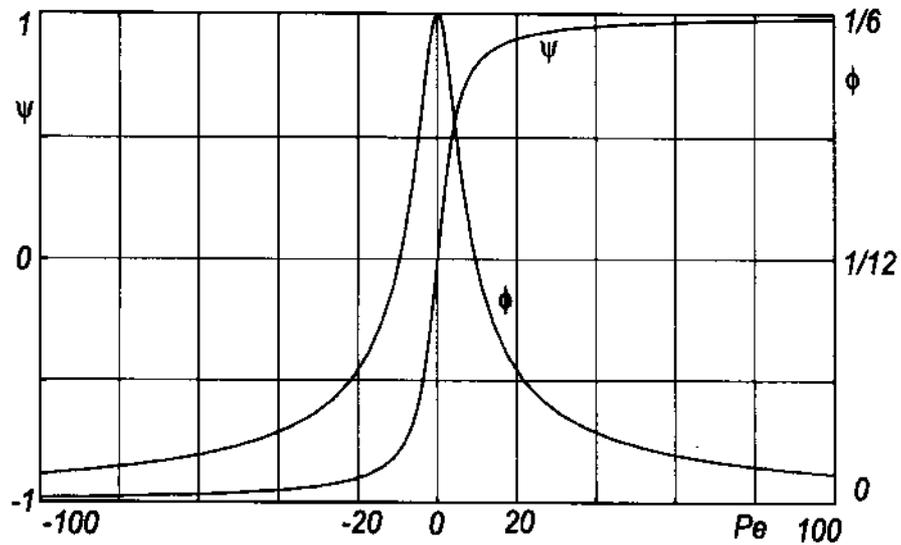


Figure 1.  
Magic functions  $\psi$  and  $\phi$  ( $Pe$ )

$\tau$  is called, in SUPG terminology, an intrinsic time scale. An interesting advantage of this kind of test functions is that it produces a conservative scheme.

The multidimensional version of (1) without source term  $\mathbf{B}$  is:

$$\mathbf{A}\mathbf{U}_{,x} = \mathbf{K}\mathbf{U}_{,xx}, \quad \mathbf{U} \in \mathbb{R}^m, \quad \mathbf{A}, \mathbf{K} \in \mathbb{R}^{m \times m} \tag{13}$$

where  $m$  is the number of equations of the system. The generalization from the scalar to the system case consists in finding a similarity transformation to carry the system to a set of  $m$  uncoupled advective-diffusive scalar equations. A new variable vector  $\mathbf{U} = \mathbf{S}\mathbf{V}$  is defined, where  $\mathbf{S}$  is a non-singular change of basis matrix to be determined, and the whole system is multiplied by  $\mathbf{S}^{-1}\mathbf{K}^{-1}$ . Then the problem is reduced to find a matrix  $\mathbf{S}$  that solves the following eigenvalue problem:

$$\mathbf{S}^{-1}\mathbf{K}^{-1}\mathbf{A}\mathbf{S} = \mathbf{A} \tag{14}$$

with  $\Lambda$  diagonal matrix. For symmetrizable parabolic systems, it is always possible to find the above matrix. After some algebra[6], the perturbation function can be expressed as:

$$\mathbf{P}_i^T = \mathbf{N}_{i,x}^T \mathbf{A} \boldsymbol{\tau}_x, \quad \boldsymbol{\tau}_x = \frac{\Delta x^2}{2} \phi(\mathbf{K}^{-1} \mathbf{A} \Delta x) \mathbf{K}^{-1} \quad (15)$$

where the  $\phi$  function is associated to the  $\psi$  one and defined as:

$$\phi(x) = \psi(x)/x \quad (16)$$

(see Figure 1). Note that, as:

$$\phi(0), \text{ then } \phi \text{ is regular at } x=0 \quad (17)$$

Note the parallelism between equations (11) and (15).

An interesting property of this choice for the matrix of intrinsic time scales is that it extends the phenomenon of superconvergence to systems, i.e. exact nodal values are obtained for a restricted class of problems (constant grid size, no source term, constant Jacobians). As far as the authors are aware, this result has not been reported previously in the literature. With respect to the additional cost, the basic problem is the eigenvalue decomposition in (14). It can be shown that, for the Navier-Stokes equations, the eigenvalue problem can be reduced by symmetry and invariance considerations from a quintic to a cubic, which is then solved by Cardano's formula. Once the eigenvalues are obtained, the eigenvectors follow by standard algebra.

*Application to the Navier-Stokes equations* The method described so far is not applicable to the Navier-Stokes system since the corresponding  $\mathbf{K}_{ij}$  matrices are singular in this case. For this reason, the eigenvalue problem (14) is degenerated and the limit value of its eigenvalue is undetermined. To solve this problem, the system is regularized, adding a small positive quantity  $\varepsilon$  to the null diagonal element of the  $\mathbf{K}_{ij}$  matrices. We stress the fact that this parameter  $\varepsilon$  is only introduced in order to compute the weight functions for the SUPG method so that the discrete scheme continues to be a weighted residual formulation of the unmodified (i.e. without the  $\varepsilon$ ) set of equations. The influence of this parameter on the precision and stability of the scheme will be analysed in the next section.

With respect to the extension to the multidimensional case, the following composition rule for the matrix of intrinsic time scales is adopted:

$$\boldsymbol{\tau}^{-1} = \boldsymbol{\tau}_x^{-1} + \boldsymbol{\tau}_y^{-1} + \boldsymbol{\tau}_z^{-1} \quad (18)$$

where

$$\boldsymbol{\tau}_i = (\Delta x_i^2/2) \phi(\mathbf{K}_{ii}^{-1} \mathbf{A}_i \Delta x_i) \mathbf{K}_{ii}^{-1} \quad i = 1, 2, 3 \quad (19)$$

*Analysis of the stabilizing terms for the incompressibility condition in the Stokes equations*

In what follows, the kind of scheme obtained for the Stokes equations will be demonstrated, and it will be compared to other types of stabilization technique. The Stokes equations under consideration are:

$$\nabla \cdot \mathbf{u} = 0$$

146

$$\nabla p - \nu \nabla^2 \mathbf{u} = \mathbf{f}$$

(20)

This system can be put in the context of homogeneous linear advective-diffusive systems (13) through the following definition of the flux Jacobians:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix}, \quad \mathbf{A}_j k_j = \begin{bmatrix} \mathbf{0} & \mathbf{k} \\ \mathbf{k}^T & 0 \end{bmatrix}, \quad \mathbf{K}_{jl} = \delta_{jl} \begin{bmatrix} \nu \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix} \quad (21)$$

$\mathbf{k}$  is an arbitrary wave number vector. Now we will compute the matrix of intrinsic time scales and the weight functions from expressions (15,18,19). As previously mentioned, the  $\mathbf{K}_{jj}$  are singular and, therefore, we add a parameter  $\epsilon$  to the null diagonal term:

$$\mathbf{K}_{ij} = \delta_{ij} \begin{bmatrix} \nu \mathbf{I} & 0 \\ 0 & \epsilon \end{bmatrix} \quad (22)$$

Now, the matrices of intrinsic time scales can be computed with the aid of equation (19) and the result is:

$$\boldsymbol{\tau}_x = \text{diag} \left\{ \frac{h\xi}{2} \sqrt{\frac{\epsilon}{\nu}}, \frac{h^2}{12\nu}, \frac{h^2}{12\nu}, \frac{h\xi}{2} \sqrt{\frac{\nu}{\epsilon}} \right\} \quad (23)$$

where  $\xi = \psi(h/\sqrt{\epsilon\nu})$  and  $\text{diag}\{a, b, c, \dots\}$  is a diagonal matrix with diagonal entries  $a, b, c, \dots$ . Similar expressions are obtained for the other directions. Finally the expression for  $\boldsymbol{\tau}$  is obtained from (18):

$$\boldsymbol{\tau} = \text{diag}\{\alpha, \alpha, \alpha, \beta\}, \quad \alpha = \frac{\sqrt{\epsilon} h^2 \xi}{2h\sqrt{\nu} + 24\nu\sqrt{\epsilon}\xi}, \quad \beta = \frac{h\xi}{6} \sqrt{\frac{\nu}{\epsilon}} \quad (24)$$

Note that in the limit  $\epsilon \rightarrow 0$  we have:  $\xi \rightarrow 1$ ,  $\alpha \rightarrow (h/2) \sqrt{\epsilon/\nu}$  and  $\beta \rightarrow (h/6) \sqrt{\nu/\epsilon}$ .

The discrete system corresponds to a centred (Galerkin) discretization of the following modified system:

$$-\nu \nabla^2 \mathbf{u} - \beta \nabla(\nabla \cdot \mathbf{u}) + \nabla p = \mathbf{f}$$

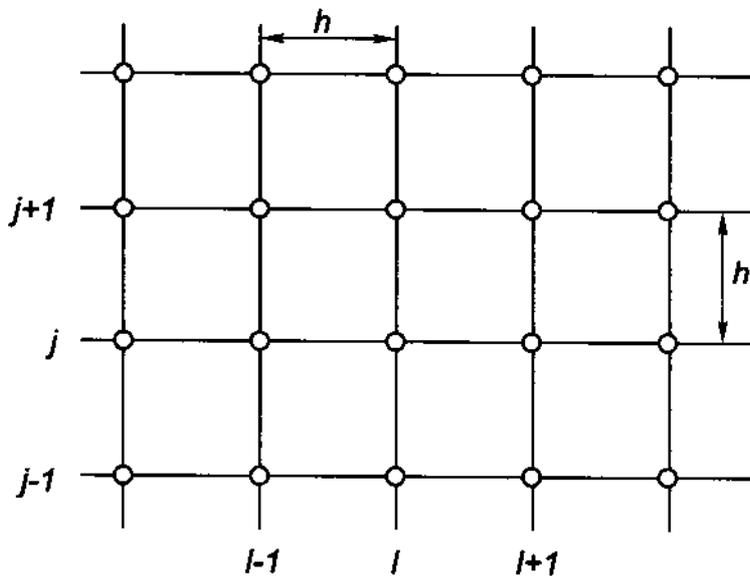
(25)

$$\nabla \cdot \mathbf{u} - \alpha \nabla^2 p = -\alpha(\nabla \cdot \mathbf{f})$$

and it can be obtained from the original Stokes system by adding  $-\beta$  times the gradient of the continuity equation to the momentum equation, and  $-\alpha$  times the divergence of the momentum equation to the continuity equation. Since the systems are equivalent, the resulting stabilized scheme has the same precision

(in the sense of order of truncation error) independently from  $\alpha$  and  $\beta$ . However, the term  $-\alpha \nabla^2 p$  changes completely the nature of the equation since, now, it is a second-order equation on  $p$  and an additional boundary condition has to be imposed on pressure.

With respect to the influence of  $\alpha$  and  $\beta$  on stability, an in-depth discrete Fourier analysis has been made, but only the relevant results will be presented here. For a complete analysis see [6]. Here, the analysis will be limited to a structured grid composed of  $Q1/Q1$  elements (quadrangles with bilinear interpolation for velocities and pressures) (see Figure 2).



**Figure 2.**  
Structured mesh  
composed of  $Q1/Q1$   
elements

Moreover, only one-dimensional variations are considered so that the index  $j$  can be dropped and the resulting discrete system of equations is:

$$\begin{aligned}
 \frac{p_{l+1} - p_{l-1}}{2h} - (\beta + \nu) \frac{u_{l+1} - 2u_l + u_{l-1}}{h^2} &= \frac{f_{x,l+1} + 4f_{x,l} + f_{x,l-1}}{6} \\
 -(\beta + \nu) \frac{v_{l+1} - 2v_l + v_{l-1}}{h^2} &= \frac{f_{y,l+1} + 4f_{y,l} + f_{y,l-1}}{6} \\
 -\alpha \frac{p_{l+1} - 2p_l + p_{l-1}}{h^2} + \frac{u_{l+1} - u_{l-1}}{2h} &= -\alpha \frac{f_{l+1} - f_{l-1}}{2h}
 \end{aligned} \tag{26}$$

First, note that if  $\alpha = 0$  (Galerkin formulation), the pressure is determined up to an arbitrary component of the form  $p_l = \bar{p}(-1)^l$ . This high-frequency mode is detected in the form of spurious oscillations in pressure and is termed a

“checkerboard mode”. A deeper Fourier analysis can be performed by putting  $f_{x,l}$  and  $f_{y,l}$  in the form of a plane wave  $f_{x,l} = \hat{f}_x e^{ikx_l}$ , with the wave number  $|k| < \pi/h$  and  $x_l = lh$ . It turns out to be that  $u_l, v_l$  and  $p_l$  have, up to a constant, the same form  $p_l = \hat{p} e^{ikx_l}$ , etc. Replacing these expressions in (26) we arrive at:

$$\hat{p} = G_{p,f_x}(\mathbf{k}, \alpha, \beta, \dots) \hat{f}_x \tag{27}$$

with:

$$(1/h)G_{p,f_x} = \frac{-i \sin(kh) \{ \frac{1}{3}[2 + \cos(kh)] + (\gamma/4) \sin^2(kh/2) \}}{\gamma \sin^4(kh/2) + \sin^2(kh)} \tag{28}$$

$$\gamma = 16\alpha(\nu + \beta)/h^2 \tag{29}$$

$G_{p,f_x}$  is a response function and we call  $\gamma$  the stability parameter. In Figure 3, we can see the response curves as a function of  $kh$  for several values of  $\gamma$ . For  $\gamma = 0$ ,  $G_{p,f_x} \rightarrow \infty$  as  $k \rightarrow \pi/h$ . This corresponds to the unstable mode  $p_l \propto (-1)^l$  that we mentioned previously for the Galerkin formulation. If  $\gamma$  is very low, the response function is bounded and  $G_{p,f_x} = 0$  for  $kh = \pi$ , but it has a sharp peak at some point near  $kh \rightarrow \pi$ . This means that, notwithstanding the fact that the response function is bounded, spurious oscillations could occur. On the other hand, if  $\gamma$  is too high, the response curve drops below the exact response curve, giving rise to an over-diffusive algorithm. It can be seen that, for a stability parameter  $\gamma \approx 1$ , the response curve is closest to the exact one. This analysis is valid for all stabilized methods, provided that the stabilizing terms can be put in the form of (25). For instance, the case  $\alpha = \beta = 0$  corresponds to the Galerkin non-stabilized case. Obviously,  $\gamma = 0$  and the scheme is unstable. The case

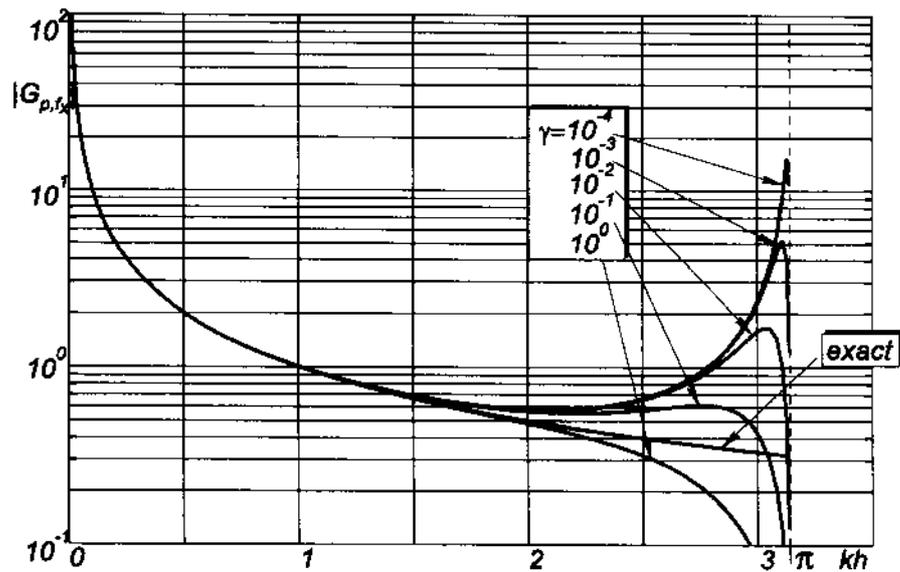


Figure 3.  
Response curves for  
several global stability  
parameters  $\gamma$

$\beta = 0$ ,  $\alpha = \alpha' h^2 / 2\nu$  corresponds to the stabilized method of Hughes *et al.*[2] ( $\alpha'$  here corresponds to  $\alpha$  in their work). For this method  $\gamma = 8\alpha'$ . In that work, it is reported that the scheme is highly oscillatory for  $\alpha' < 0.01$  ( $\gamma < 0.08$ ), whereas no oscillations were observed for  $\alpha' > 0.1$  ( $\gamma > 0.8$ ). These results confirm the stability analysis performed here. On the other hand, the scheme from Frey *et al.*[3] (in the Stokes regime and for linear elements) is obtained with  $\alpha = mh^2/8\nu$ ,  $\beta = \lambda m u^2 h^2 / 4\nu$ , where  $u$  is the local velocity and  $\lambda, m$  are  $\mathcal{O}(1)$  constants, defined in their work. On the other hand, in our method  $\gamma$  is linked to  $\varepsilon$  through equations (24) and (29). However, the predicted value for  $\gamma$  falls always in the interval  $0.4 < \gamma < 1.4$ , irrespective of the value chosen for  $\varepsilon$ . In conclusion, the method presented here is stable and is “parameter free” in the sense that, notwithstanding the fact that a “free” parameter  $\varepsilon$  is introduced, stability is assured independently of the value of  $\varepsilon$ .

#### Numerical examples

The first example is the well known test of the square, lid-driven cavity. Here, we present results for  $Re = 1, 40, 100$ . Figure 4 shows the problem description. In Figures 5-7, the velocity field is shown for each Reynolds number and included in it is the reference position of the principal vortex centre obtained from Schreiber and Keller[7]. A mesh composed of  $20 \times 20$  elements has been used in all cases. The results presented here are in close agreement with the referenced ones.

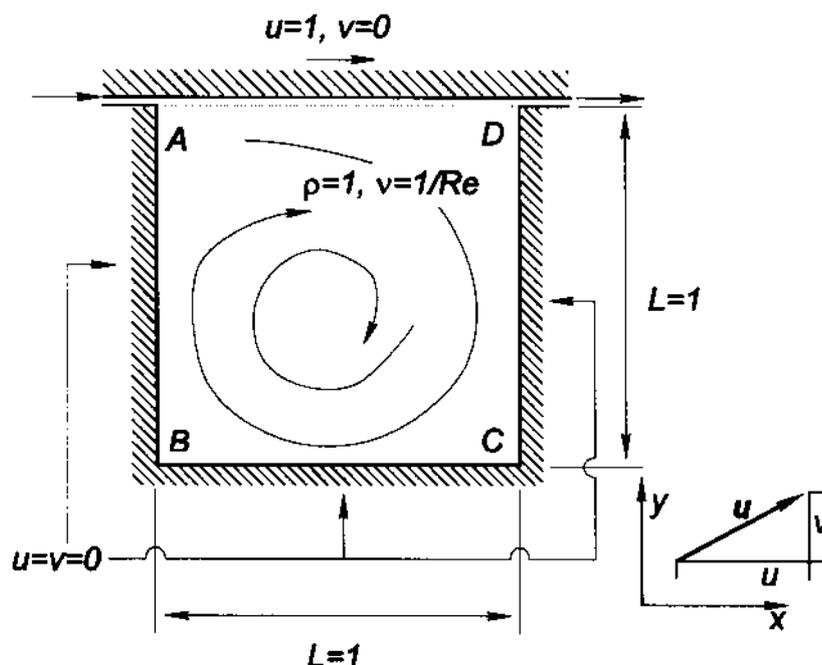


Figure 4.  
Lid-driven square-cavity  
– problem description

---

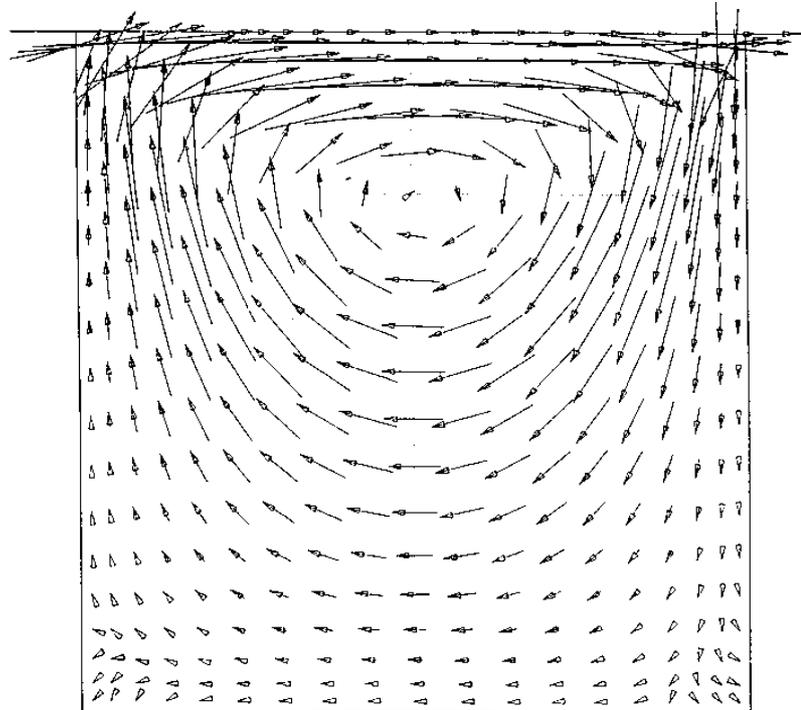
HFF  
7,2/3

150

---

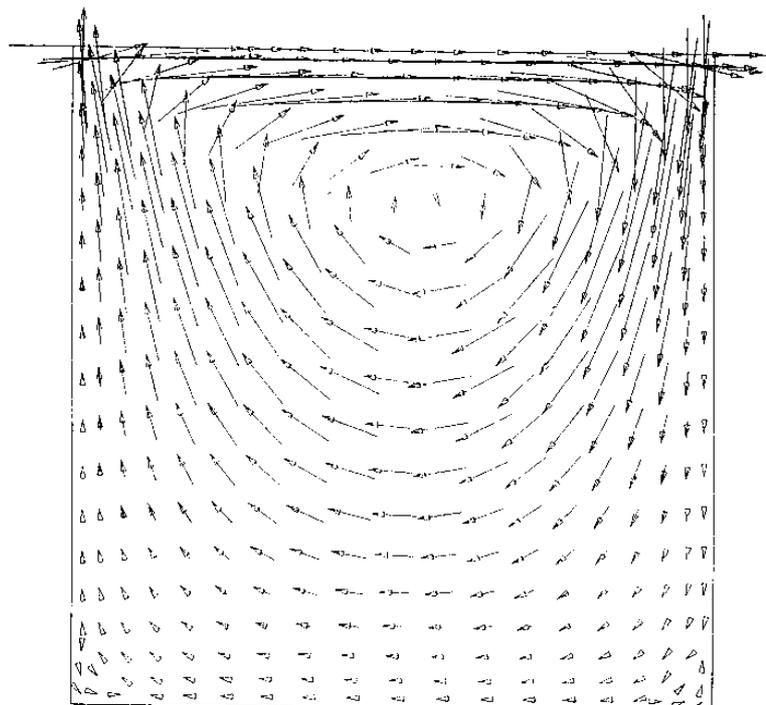
**Figure 5.**  
Lid-driven square-cavity  
– velocity field for  
 $Re = 1$

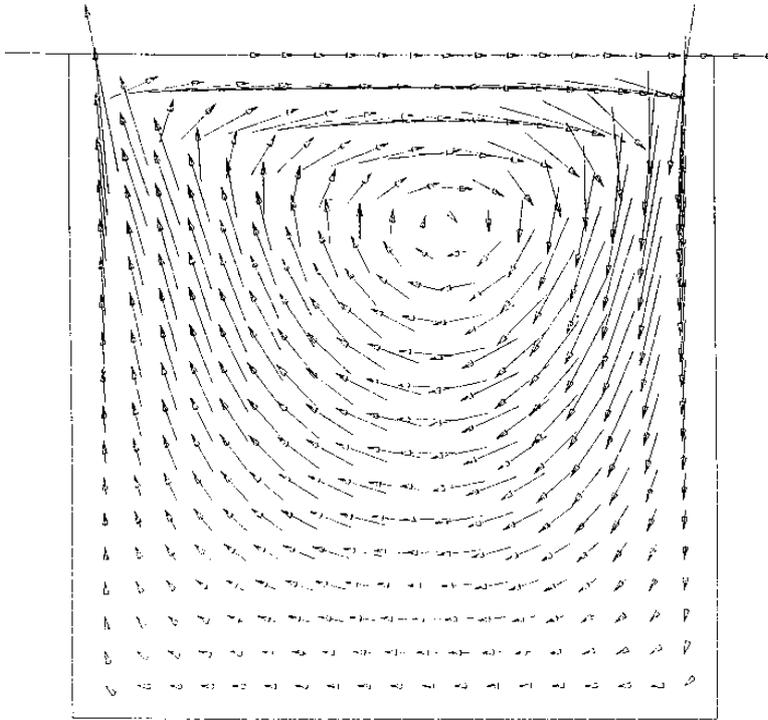
---



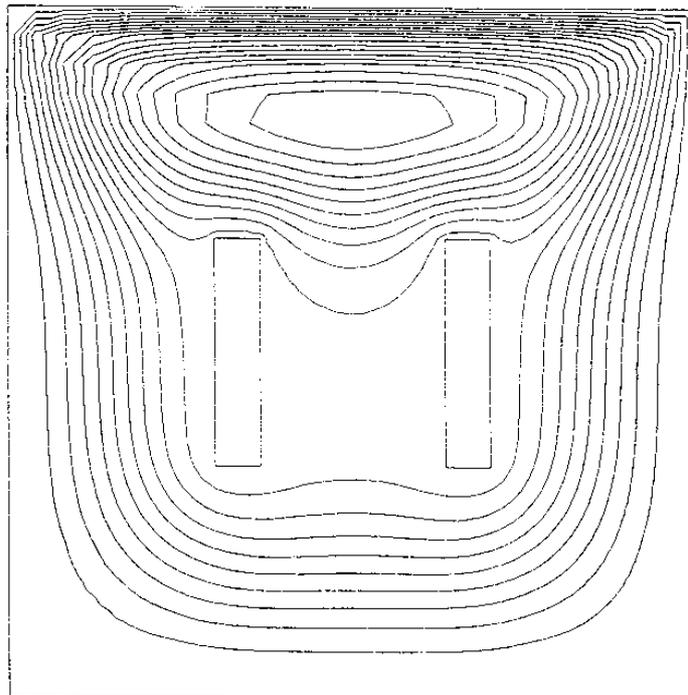
**Figure 6.**  
Lid-driven square-cavity  
– velocity field for  
 $Re = 40$

---





**Figure 7.**  
Lid-driven square-cavity  
– velocity field for  
 $Re = 100$



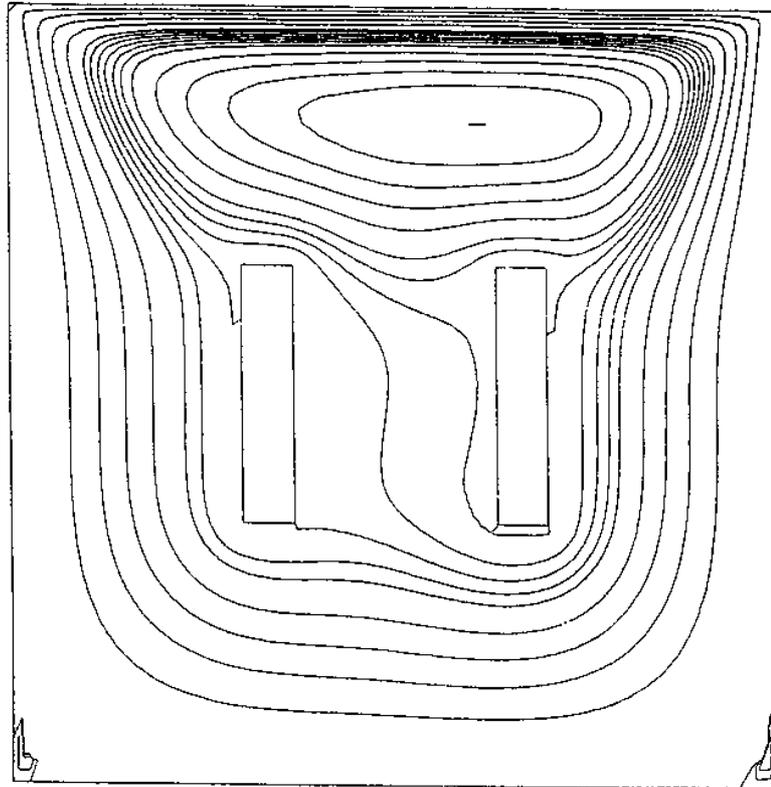
**Figure 8.**  
Multiply-connected  
lid-driven square-cavity  
– streamlines for  
 $Re = 0$

---

HFF  
7,2/3

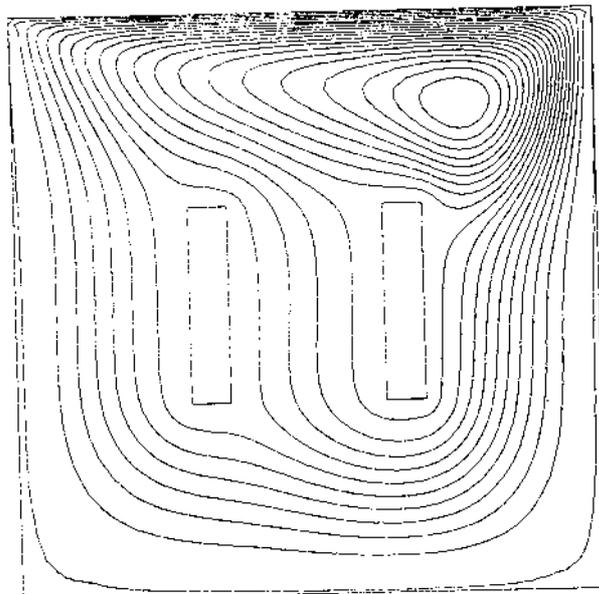
152

---



**Figure 9.**  
Multiply-connected  
lid-driven square-cavity  
– streamlines for  
 $Re = 50$

---



**Figure 10.**  
Multiply-connected  
lid-driven square-cavity  
– streamlines for  
 $Re = 250$

---

The second example consists of a lid-driven flow in a square cavity but, now, with two internal obstacles. This problem is referenced as the multiply-connected lid-driven square cavity[8]. The rectangular obstacles are centred at  $(\mathbf{x}, \mathbf{y}) = (\frac{1}{3}, \frac{1}{2})$  and  $(\frac{2}{3}, \frac{1}{2})$ , and they have a thickness of  $\frac{1}{5}$  and a height of  $\frac{1}{3}$ . We solved for  $Re = 0, 50, 250$  and in Figures 8-10 the streamlines derived from the numerical results are shown. They were compared with published results by Lipke and Wagner[8], using a similar quantity of nodes, ( $\approx 3,600$ ) and appear to be in close agreement with them.

In all cases, Mach numbers below 0.05 were used, and  $\varepsilon$  was selected according to  $\varepsilon = \tilde{\varepsilon} h^2/\nu$ , with  $\tilde{\varepsilon} = 1$ .

### Iterative solution of large systems with explicit pseudo-temporal schemes

A very common procedure in order to find steady states from non-linear equations arising in computational fluid dynamics, say  $\mathbf{F}(\mathbf{x}) = 0$ , is to iterate an explicit or implicit temporal scheme until convergence[9,10]:  $\mathbf{M}\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), t \rightarrow \infty$ . Here  $\mathbf{x} \in \mathbb{R}^N$  is the state vector,  $\mathbf{F}$  is a map from  $\mathbb{R}^N$  onto itself that represents the discrete system of equations,  $\mathbf{M}$  is the mass matrix of the system and the dot represents the time derivative. If an explicit scheme is used, comparatively low amounts of core memory and large CPU times are required. Moreover, the CPU time highly depends on the conditioning of the system. Bad conditioning of the system is caused by several multiplicative factors like: large variations in element size through the mesh, large variations in edge sizes for a given element and local incompressible ( $M \rightarrow 0$ ) or transonic ( $M \rightarrow 1$ ) behaviour, where  $M$  is the Mach number. On the other hand, if an implicit temporal scheme coupled to a direct solver is preferred, very large convergence rates are achieved, but it requires a large amount of memory to factorize the associated matrix. This problem can be partially overcome by solving the linear subproblems by an iterative “black-box” solver like GMRES (generalized minimal residual), BiCG (bi-conjugate gradient) etc. The amount of core memory is drastically reduced at the expense of an increased CPU time which, again, highly depends on the conditioning of the system. The success of an iterative scheme (either explicit or implicit) is, then, related to improving the conditioning of the system. For instance, bad conditioning arising from large variations in mesh element size can be removed through the use of a “local time stepping” strategy. Local time stepping can be seen as modifying the mass matrix of the original system to  $\tilde{\mathbf{M}}\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$  where  $\tilde{\mathbf{M}}$  is a diagonal matrix that includes a factor proportional to the maximum admissible time step based on a local stability analysis. As is well known, this modifies the temporal evolution of the state vector in such a way that it has no more physical sense. Owing to this fact, it is only applicable when looking for a steady state. Much in the same way, we look for mass matrices which correct the bad conditioning associated with different characteristic speeds at incompressible ( $M \rightarrow 0$ ) and transonic ( $M \rightarrow 1$ ) regimes. Of course, some restrictions in the degree of connectivity of the proposed mass matrix

have to be imposed. Otherwise, arbitrarily complex mass matrices can be proposed, which reduce substantially the condition of the system, but whose cost of computation is so high that the overall processing time is increased. Here the focus is on the search for mass matrices, i.e. block-diagonal matrices with each block connecting only the unknowns of each node. For these matrices, the cost of the inversion (and then the cost per preconditioning cycle) is  $\mathcal{O}(N_{\text{nod}})$ , where  $N_{\text{nod}}$  is the total number of nodes in the mesh.

*Convergence rate in explicit pseudo-temporal schemes*

*General considerations about rate of convergence in iterative systems.* Let us consider the one-step explicit pseudo-temporal scheme applied to a linear system of ODEs like:

$$\mathbf{M} \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} = \mathbf{K} \mathbf{U}^n - \mathbf{b} = \mathbf{R}^n \quad (30)$$

where  $\mathbf{K} = (\partial \mathbf{F} / \partial \mathbf{U})$  is the Jacobian of  $\mathbf{F}$  and  $\mathbf{b}$  is a constant right-hand side. The ultimate rate of convergence (r.o.c.) of the scheme is defined so as  $(\text{r.o.c.})^{-1}$  is the number of iterations needed to reduce the residual by a factor of ten, and can be computed as:

$$\text{r.o.c.} = \log_{10} \left\{ \lim_{n \rightarrow \infty} \frac{|\mathbf{R}^n|}{|\mathbf{R}^{n+1}|} \right\} \quad (31)$$

By standard eigenvalue decomposition, it can be shown that:

$$\lim_{n \rightarrow \infty} \frac{|\mathbf{R}^n|}{|\mathbf{R}^{n+1}|} = \min_{\mu=1}^N |1 + \Delta t \lambda_{\mu}| \quad (32)$$

where  $\{\lambda_{\mu}\}$  is the set of eigenvalues of  $\mathbf{M}^{-1} \mathbf{K}$ . To obtain the highest rate of convergence, we should choose a  $\Delta t$  as high as possible but, as is well known, such an explicit scheme has a maximum admissible time step  $\Delta t_{\text{crit}}$  owing to stability restrictions. On the other hand, the minimum in (32) will be attained when  $|\lambda_{\mu}|$  is small. For the smaller eigenvalues of systems coming from the discretization by standard methods like FEM or FDM, the product  $|\Delta t_{\text{crit}} \lambda_{\mu}|$  is much smaller than unity. The following approximation can then be made:

$$|1 + \Delta t_{\text{crit}} \lambda_{\mu}| \approx 1 + \Delta t_{\text{crit}} \text{Re} \{ \lambda_{\mu} \} \quad (33)$$

so that:

$$\text{r.o.c.} = 0.4343 \Delta t_{\text{crit}} \min_{\mu} \text{Re} \{ \lambda_{\mu} \} \quad (34)$$

The critical time step is fixed by the high frequency, small wavelength components, so that it is highly dependent on mesh size and discretization scheme. On the other hand, the minimum eigenvalue is usually associated with the smooth, large wavelength ones and can be computed directly from the

continuum system. To fix ideas, the set of eigenvalues  $\{\lambda_\mu\}$  for a simplified continuum system will be computed.

Consider a linear, homogeneous, one-dimensional system like:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = 0, \quad 0 < x < L, \quad 0 < t < \infty \quad (35)$$

where  $\mathbf{U} \in \mathbb{R}^2$  is the state vector and  $\mathbf{A} \in \mathbb{R}^{2 \times 2}$  is the advective flux Jacobian. We suppose that  $\mathbf{A}$  has two real eigenvalues  $\{\alpha^+, -\alpha^-\}$ , with  $\alpha^\pm > 0$ , corresponding to a right- and left-moving component respectively. The boundary and initial conditions are:

$$\mathbf{B}_0 \mathbf{U}(0, t) = 0$$

$$\mathbf{B}_L \mathbf{U}(L, t) = 0 \quad (36)$$

$$\mathbf{U}(x, 0) = \mathbf{U}_0(x)$$

The  $\mathbf{B}$ s are  $1 \times 2$  matrices since we have one ingoing and one outgoing wave at each boundary.

The set of eigenvalues  $\lambda_\mu$  can be determined by Laplace transform which, in practice, is equivalent to search for solutions of the form:  $\mathbf{U}(x, t) = e^{-\lambda t} \hat{\mathbf{U}}(x)$ . The rate of convergence is given then by the  $\lambda$  with the lowest real part. Replacing this particular form in (35):

$$-\lambda \hat{\mathbf{U}} + \mathbf{A} \frac{\partial \hat{\mathbf{U}}}{\partial x} = 0 \quad (37)$$

whose solution is:

$$\hat{\mathbf{U}}(x) = \exp\{\lambda \mathbf{A}^{-1} x\} \hat{\mathbf{U}}(0) \quad (38)$$

On the other hand, the boundary conditions are transformed to:

$$\mathbf{B}_0 \hat{\mathbf{U}}(0) = 0 \quad (39)$$

$$\mathbf{B}_L \hat{\mathbf{U}}(L) = \mathbf{B}_L \exp\{\lambda \mathbf{A}^{-1} L\} \hat{\mathbf{U}}(0) = 0$$

so that the equation we are looking for is:

$$\det \begin{bmatrix} \mathbf{B}_0 \\ \mathbf{B}_L e^{\lambda \mathbf{A}^{-1} L} \end{bmatrix} = 0 \quad (40)$$

Since the system is hyperbolic, a simpler expression can be obtained if we switch to a basis of  $\mathbb{R}^2$ , where  $\mathbf{A}$  is diagonal. Let  $\mathbf{S}$  be the change of basis matrix such that  $\mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \text{diag}\{\alpha^+, -\alpha^-\}$ , then:

$$\exp\{\lambda \mathbf{A}^{-1} L\} = \mathbf{S} \begin{bmatrix} e^{\lambda L/a^+} & 0 \\ 0 & e^{-\lambda L/a^-} \end{bmatrix} \mathbf{S}^{-1} \quad (41)$$

Transforming the boundary conditions into the eigencomponents:

$$\mathbf{B}'_0 = \mathbf{B}_0 \mathbf{S} \quad (42)$$

$$\mathbf{B}'_L = \mathbf{B}_L \mathbf{S}$$

and (40) simplifies to:

$$\det \begin{bmatrix} B'_{01} & B'_{02} \\ B'_{L1} e^{\lambda L/a^+} & B'_{L2} e^{-\lambda L/a^-} \end{bmatrix} = 0 \quad (43)$$

It can be shown that the  $\mathbf{B}'_{ij}$ s are related to the reflection coefficients  $R_{0,L}$  at the extremes of the domain through:

$$R_0 = \frac{B'_{02}}{B'_{01}}, \quad R_L = \frac{B'_{L1}}{B'_{L2}} \quad (44)$$

and, then, the equation (42) becomes:

$$e^{-\lambda L/a^-} - R_0 R_L e^{\lambda L/a^+} = 0 \quad (45)$$

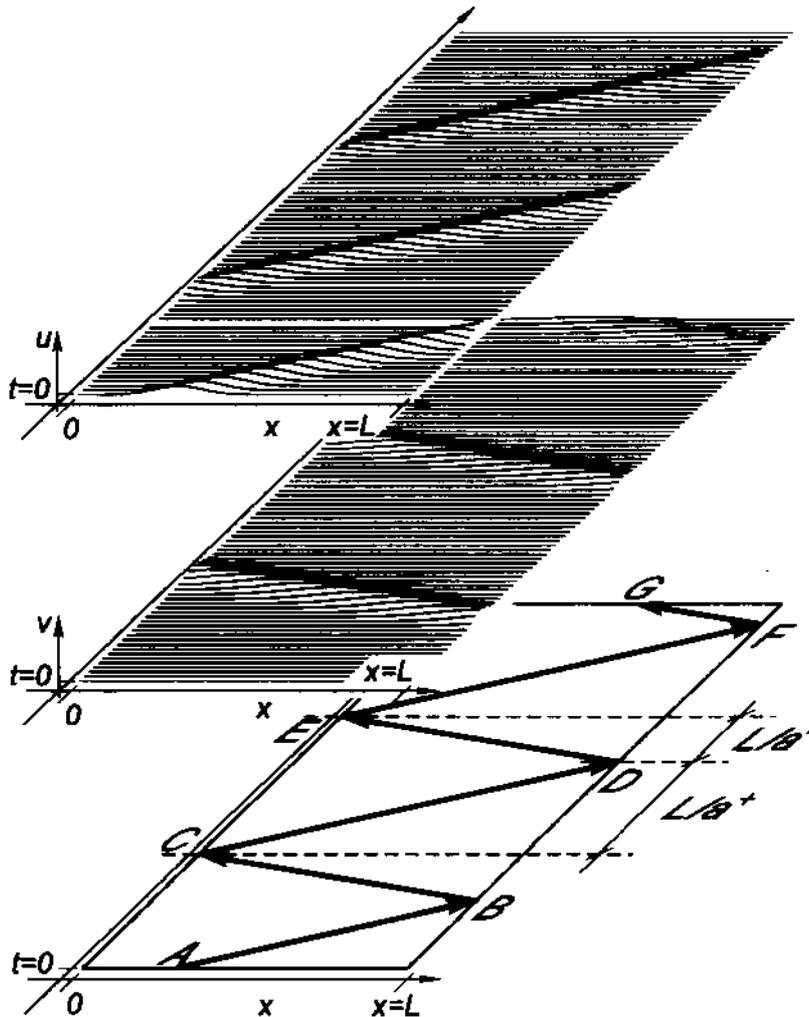
from which the following expression for the set of eigenvalues  $\{\lambda_\mu\}$  can be obtained:

$$\lambda_\mu \left( \frac{L}{a^+} + \frac{L}{a^-} \right) = \log(1/|R_0 R_L|) + i(2\mu + \sigma)\pi, \quad \mu = 0, \pm 1, \pm 2, \dots \quad (46)$$

where  $\sigma = 0(1)$  if  $R_0 R_L$  is positive (negative). The real part is, then:

$$\text{Re}\{\lambda_\mu\} = \frac{\log(1/|R_0 R_L|)}{\frac{L}{a^+} + \frac{L}{a^-}} \quad (47)$$

Equation (47) has a very clear physical meaning (see Figure 11). Consider a smooth wave-packet that is being reflected back and forth between the two boundaries. At each reflection, some part of the amplitude is absorbed and, since there is no dissipation in the interior, this is the unique mechanism of amortiguation. To compute the rate of amortiguation, a complete cycle has to be considered; for instance we begin to follow the packet just when it leaves the left boundary at  $C$ , until it reaches the same position at  $E$ , after being reflected at  $D$  and  $E$ .  $\text{Re}\{\lambda\}$  is, then, the logarithm of the amortiguation in the cycle divided by the total duration of the cycle. The expression in the



**Figure 11.** Convergence by absorption at the boundaries.  $L = 1$ ,  $a^+ = 1$ ,  $a^- = 2$ . The coefficients of reflection are  $R_{0,1} = +0.8$  at each boundary. The perturbation at  $t = 0$  is a right-going gaussian of width = 0.1 and unit amplitude in  $u$  centered at  $x = 0.3$ . It propagates to the right and hits the right boundary at  $t = 0.7$ . It is reflected to a left-going  $v$ -wave of amplitude 0.8 until it hits the left-boundary at  $t = 1.2$ . It is reflected there to a right-going  $u$ -wave of amplitude 0.64. At  $t = 1.5$  it is at the same place where the process started and the cycle restarts again with a total loss of amplitude in the cycle of  $|R_0 R_L| = 0.64$

denominator  $L[(1/a^+) + (1/a^-)]$  is the total duration of the cycle, and the expression  $R_0 R_L$  appearing in the numerator is the attenuation owing to the absorption at the boundaries. Usually, one has absorbing boundary conditions at the boundaries. For 1D systems, completely absorbing boundary conditions can be devised and  $R_0, R_L = 0$ . This implies an infinite rate of convergence, i.e. the error is eliminated in a finite time: the time that it takes to arrive at the completely absorbing boundary. However, in practical 2D or 3D situations, local non-reflecting boundary conditions are completely absorbing only for normal incident waves and, then, a globally non-infinite rate of convergence is observed.

*Rate of convergence in advective systems* Replacing (47) in (34):

$$\text{r.o.c.} = 0.4343 \Delta t_{\text{crit}} \frac{\log_{10}(1/|R_0 R_L|)}{L \left( \frac{1}{a^+} + \frac{1}{a^-} \right)} \quad (48)$$

The time step is made non-dimensional:  $\Delta t_{\text{crit}} C h / a_{\text{max}}$ , where  $C$  is the Courant number, which is restricted by the CFL condition to be smaller than unity,  $h$  is the mesh size and  $a_{\text{max}} = \max \{a^+, a^-\}$ . Replacing the above expression for  $\Delta t_{\text{crit}}$  in (48) we obtain:

$$\text{r.o.c.} = 0.4343 \frac{C \log_{10}(1/|R_0 R_L|)}{N(1 + \kappa)} \quad (49)$$

with:

$$\kappa = \frac{|a|_{\text{max}}}{|a|_{\text{min}}} = \frac{\max\{a^+, a^-\}}{\min\{a^+, a^-\}} \quad (50)$$

and  $N = L/h$  is the number of elements. This expression for the rate of convergence can be extended to multidimensional problems, replacing  $N$  by the number of elements in a characteristic direction

$$N \sim N_{\text{nod}}^{1/n_d}, \quad (51)$$

and  $\kappa$  by an equivalent condition number based on group velocities (this will be explained later). Here  $n_d$  is the number of spatial dimensions.

In the rest of the work, we will focus on how to improve the r.o.c.s based on the design of a good local preconditioning. This is done in two steps: the first one is to improve the condition number  $\kappa$  based only on an analysis of the continuum system. Once a good candidate is found, it must be verified by a standard stability analysis how the corresponding critical Courant number is modified, since it is included also in the expression for the r.o.c. (see equation (48)). This is highly dependent on the discretization scheme, both spatial and temporal. There exist cases where a given preconditioning improves the conditioning of the system but, after a stability analysis, it is shown that this "gain" is counteracted by a deterioration in the critical Courant number so that no overall gain is produced.

*Computation of condition number for multidimensional advective systems*

Now, the definition of the condition number is extended to the general multidimensional case, laying stress on the Euler equations. As Fourier analysis will be used, the analysis is restricted to the Cauchy problem (infinite domain), linear, homogeneous (constant coefficients) multidimensional advective systems of equations:

$$\mathbf{M} \frac{\partial \mathbf{U}}{\partial t} + \mathbf{A}_j \frac{\partial \mathbf{U}}{\partial x_j} = 0, \quad \forall \mathbf{x} \in \mathbb{R}^{n_d}, t > 0 \quad (52)$$

where  $\mathbf{U}(\mathbf{x}, t) \in \mathbb{R}^m$  is the state vector,  $\{\mathbf{A}_j\}_{j=1}^{n_d}$  are the Jacobians of the advective fluxes,  $\mathbf{A}_j \in \mathbb{R}^{m \times m}$ , and  $\mathbf{M} \in \mathbb{R}^{m \times m}$  is the mass matrix. Here, and in what follows, the Einstein convention will be adopted. It is assumed that the system is hyperbolic, i.e. for all  $\mathbf{k} \in \mathbb{R}^{n_d}$  the matrix  $\mathbf{k} \cdot \mathbf{A} = k_j \mathbf{A}_j$  is diagonalizable with real eigenvalues. Moreover, it is assumed that the system is symmetrizable, i.e. there exists a non-singular matrix  $\mathbf{S}$ , independent of  $\mathbf{k}$ , for which  $\mathbf{S}(\mathbf{k} \cdot \mathbf{A})\mathbf{S}^{-1}$  is symmetric for all  $\mathbf{k}$ . The Euler equations fall into both categories. We look for eigenfunctions in the form of plane waves:

$$\mathbf{U}(\mathbf{x}, t) = \hat{\mathbf{U}} \exp \{i(\mathbf{k} \cdot \mathbf{x} - \omega t)\} \quad (53)$$

Replacing (53) in (52), the following determinantal equation in  $\omega$  is obtained:

$$\det(-\omega \mathbf{M} + k_j \mathbf{A}_j) = 0 \quad (54)$$

and it results that  $\hat{\mathbf{U}}$  has to be an eigenvector corresponding to the eigenvalue  $\omega$ . For each  $\mathbf{k} \in \mathbb{R}^{n_d}$  a set of eigenvalues  $\{\omega_\mu(\mathbf{k})\}_{\mu=1}^m$  is obtained, which are called the ‘‘branches of eigenvalues’’. For the non-preconditioned system  $\mathbf{M} = \mathbf{I}$ , the eigenvalues are real since the system is hyperbolic. This feature must be kept by the preconditioning. For symmetrizable systems, a sufficient condition is that  $\mathbf{M}$  must be positive definite and symmetric in the basis where the Jacobians are symmetric. Coming back to the expression for the plane wave (53), it can be seen that constant amplitude planes  $\text{Re} \{i(\mathbf{k} \cdot \mathbf{x} - \omega t)\} = \text{constant}$ , have a characteristic phase velocity:

$$\mathbf{v}_{\phi\mu}(\mathbf{k}) = \frac{\omega_\mu}{k^2} \mathbf{k} \quad (55)$$

However, it can be shown that energy and information propagate at the group velocity:

$$\mathbf{v}_{g\mu} = \frac{\partial \omega_\mu}{\partial \mathbf{k}} \quad (56)$$

For a complete description on Fourier analysis of discrete systems the reader is referred to[11]. It is easy to see that the  $\omega_\mu$ s are homogeneous functions of degree one in  $\mathbf{k}$ , i.e.  $\omega_\mu(\alpha \mathbf{k}) = \alpha \omega_\mu(\mathbf{k})$  for  $\alpha > 0$ . As a consequence,  $\mathbf{v}_{g\mu}$  is homogeneous of degree 0 and, then, it depends only on the direction of  $\mathbf{k}$ :

$$\mathbf{v}_{g\mu}(\mathbf{k}) = \mathbf{v}_{g\mu}\left(\frac{\mathbf{k}}{k}\right) \quad (57)$$

The condition number  $\kappa$  is extended to the multidimensional case as:

$$\kappa = \frac{|\mathbf{v}_{g\mu}|_{\max}}{|\mathbf{v}_{g\mu}|_{\min}} \quad (58)$$

where the maximum and minimum are taken over all and  $|\mathbf{k}| = 1$  and  $\mu = 1, \dots, m$ . Some tips and details of the calculation of group velocities can be found in[12]

The compressible Euler equations in conservative form are:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}_i}{\partial x_i} = 0 \quad (59)$$

where:

$$\mathbf{U} = \begin{bmatrix} \rho \\ \rho \mathbf{u} \\ \rho e \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} \rho \mathbf{u}^T \\ \rho \mathbf{u} \mathbf{u}^T + p \mathbf{I}_{n_d \times n_d} \\ (\rho e + p) \mathbf{u}^T \end{bmatrix} \quad (60)$$

where  $\mathbf{U} \in \mathbb{R}^m$  is the state vector of the fluid,  $\mathbf{F} \in \mathbb{R}^{m \times n_d}$  are the advective fluxes,  $\rho$  the density,  $\mathbf{u}$  the velocity vector,  $p$  the pressure,  $e = p[\rho(\gamma - 1)] + u^2/2$  the total energy,  $\gamma = 1.4$  and  $m = 5$  in 3D ( $m = n_d + 2$  in general). The Jacobians of the fluxes are defined as:

$$\mathbf{A}_i = \frac{\partial \mathbf{F}_i}{\partial \mathbf{U}} \quad (61)$$

and can be written in compact form as:

$$k_i \mathbf{A}_i = \begin{bmatrix} 0 & \mathbf{k}^T & 0 \\ -(\mathbf{u} \cdot \mathbf{k}) \mathbf{u} + (\gamma - 1) \frac{u^2}{2} \mathbf{k} & \mathbf{u} \mathbf{k}^T + (\mathbf{u} \cdot \mathbf{k}) \mathbf{I} - (\gamma - 1) \mathbf{k} \mathbf{u}^T & (\gamma - 1) \mathbf{k} \\ (\mathbf{u} \cdot \mathbf{k}) [(\gamma - 1) u^2 - \gamma e] & (\frac{p}{\rho} + e) \mathbf{k}^T - (\gamma - 1) (\mathbf{u} \cdot \mathbf{k}) \mathbf{u}^T & \gamma (\mathbf{u} \cdot \mathbf{k}) \end{bmatrix} \quad (62)$$

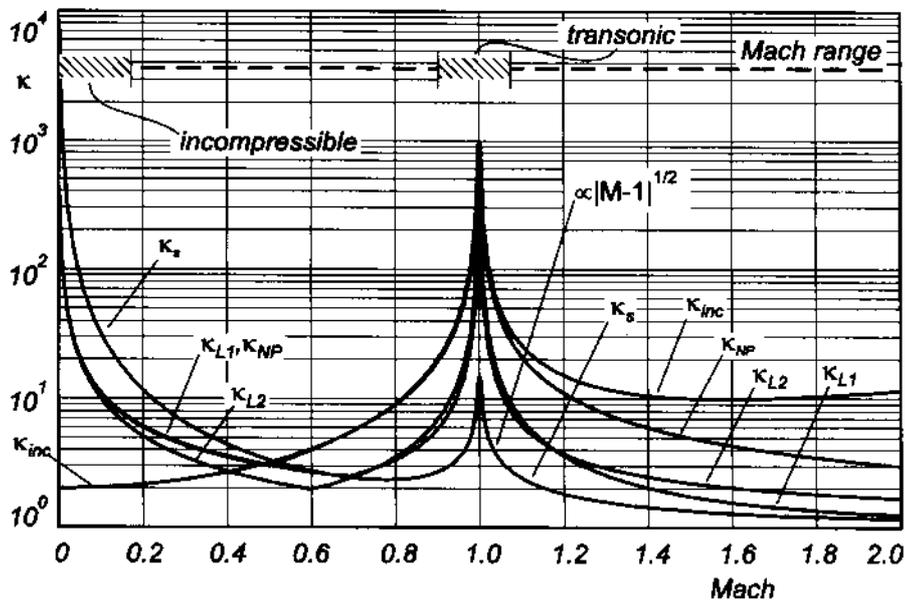
The system is symmetrizable and, then, hyperbolic, but it is not diagonalizable in the multidimensional case. However, it is (trivially) diagonalizable in the one-dimensional case. It is shown in [12] that an optimal preconditioning exists in such cases, defining the preconditioning matrix as the absolute value (in the matricial sense) of the one-dimensional Jacobian. This suggests several possibilities in the multidimensional case:

$$\mathbf{M}_{L1} = |\mathbf{A}_x| + |\mathbf{A}_y| + |\mathbf{A}_z| \quad (63)$$

$$\mathbf{M}_{L2} = \sqrt{\mathbf{A}_x^2 + \mathbf{A}_y^2 + \mathbf{A}_z^2} \quad (64)$$

$$\mathbf{M}_s = |s_i \mathbf{A}_i| \quad (65)$$

The first two correspond to different versions of the absolute value of the Jacobian vector and are optimal for diagonalizable systems. However, their performance is very poor for the Euler equations which are not diagonalizable, as can be seen in Figure 12. The condition number for the preconditioned systems, as well as for the non-preconditioned one, is shown. The last option (65)



**Figure 12.** Condition number of the continuum system for several preconditionings

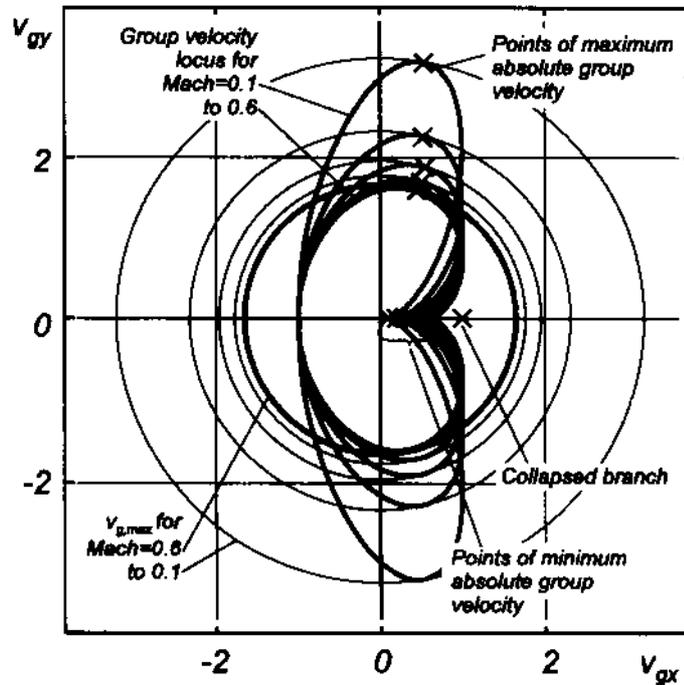
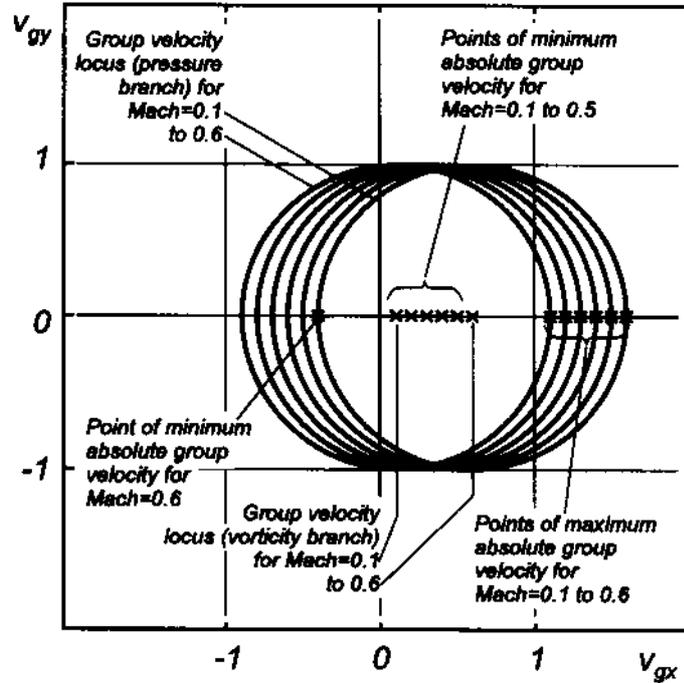
corresponds to the absolute value of the streamline projection of the Jacobians ( $\mathbf{s}$  stands for the streamline unit vector) and signifies a significant improvement since it can be shown that the conditioning behaves like  $\kappa_s \approx 2|1 - M|^{-1/2}$ .

*PMM for incompressible flows*

For the incompressible regime, none of the above-mentioned possibilities succeeded in giving a significant improvement in the conditioning of the system. In fact, the PMMs from equations (63) or (64) have a condition number  $\mathcal{O}(1)$  as  $M \rightarrow 0$ , so that it is clear that they can not correct the singularity. (A badly conditioned matrix cannot be corrected by a well-conditioned one). Regarding the streamline based  $\mathbf{M}_s$  from equation (65), the conditioning is even worse than the non-preconditioned system. In Figures 13 and 14 the locus of the group velocity vector can be seen in the  $(v_{gx}, v_{gy})$  plane for  $M = 0.1$  to  $0.6$  with  $\Delta M = 0.1$  and  $\mathbf{M} = \mathbf{I}, \mathbf{M}_s$ . As is well known, for the non-preconditioned system the locus for the branches corresponding to pressure waves are circles of unit radius centred at  $(M, 0)$ , whereas the entire branch for advection of vorticity is collapsed in a point at  $(M, 0)$ . The maximum absolute group velocity is attained for a pressure wave (i.e. a mode located on the pressure branch) propagating in the same direction of the flow. The minimum absolute group velocity corresponds to vorticity waves for  $0 \leq M \leq 0.5$  and to the pressure wave that propagates in the direction opposite to the flow for  $M \geq 0.5$ .

*Preconditioning by diagonal scaling* The preconditioning proposed for the incompressible case here is based on three different scaling parameters for the

**Figure 13.** Locus of group velocities for the non-preconditioned system  $M = I$ . The locus has been plotted for  $M = 0.1$  to  $0.6$  with a step of  $0.1$ . The branches of group velocities are indicated by points. For each Mach number two circles centred at the origin and with radii  $|v_g^j|_{\max, \min}$  have been drawn



**Figure 14.** Same as Figure 13 but for  $M = M_s$

continuity, momentum and energy equations, and a detailed analysis has been performed[12]. The preconditioned matrix is:

$$\mathbf{M}_{inc} = \text{diag}\{2M, 2M, 2M, 1/M\} \quad (66)$$

The locus of group velocities for  $M = 0.1$  to  $0.6$  is shown in Figure 15. Note that both circles at the maximum and minimum group velocities approach a defined value as  $M \rightarrow 0$ , so that the condition number remains bounded for  $\kappa_{inc} \rightarrow 2$  for  $M \rightarrow 0$ .

As we mentioned earlier, the stability of the discretized preconditioned system was checked. The integration scheme is the standard forward-Euler, and the spatial discretization is based on SUPG. For this combination, the stability analysis shows that the minimum admissible Courant number is always above 0.64, so that stability does not reduce the gain in the condition number. However, for a spatial discretization based on Taylor-Galerkin, the Courant number was so low that no gain in the overall processing time was obtained. This was interpreted as a consequence of the over-diffusive character of the Taylor-Galerkin method.

Regarding the cost of implementation of this preconditioning, it is negligible when compared with the cost of the evaluation of the residual. It amounts mainly to the multiplication of the residual of each equation by the inverse of

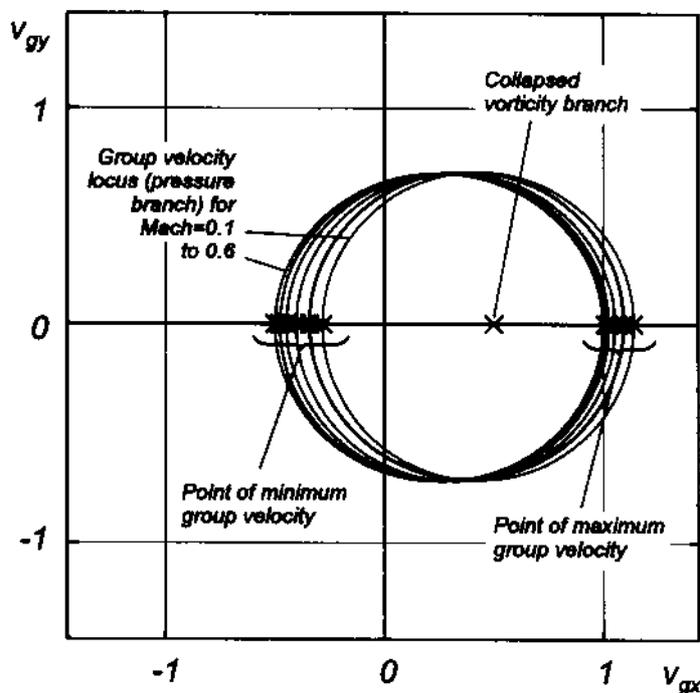


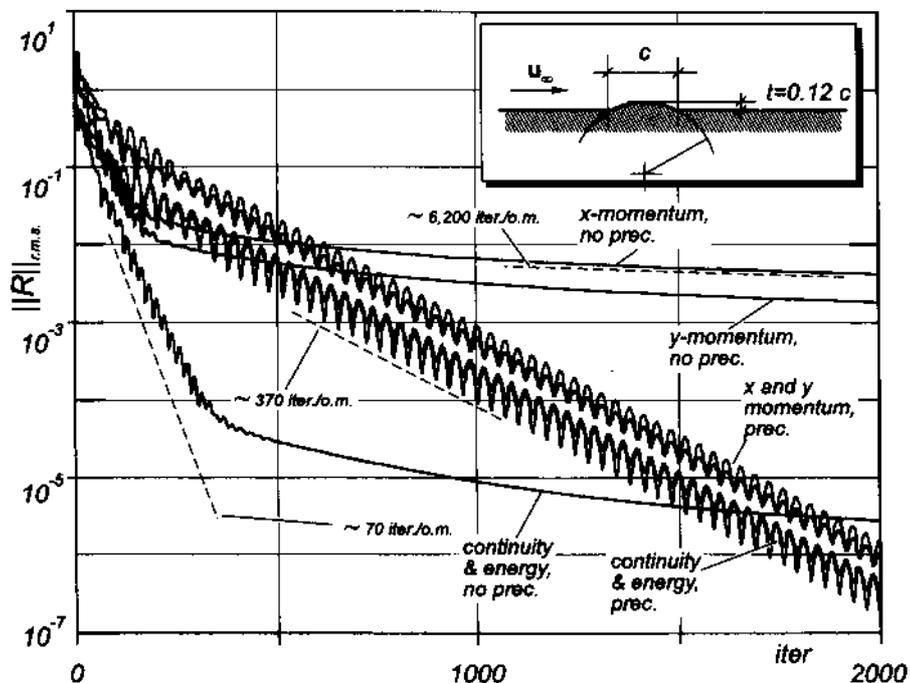
Figure 15.  
Same as Figure 13 but  
for  $M = M_{inc}$

the corresponding diagonal element in the preconditioning matrix. This multiplication must be performed for each node, at update time.

*Accuracy improvement.* Another topic to be discussed is how the preconditioning affects the accuracy of the solution. First of all, it should be pointed out that preconditioning and discretization do not commute, i.e. the discrete solution for the steady state of the preconditioned system is not the same as for the non-preconditioned one. Of course, both converge to the continuum steady solutions as the mesh is refined. The non-equivalence of the steady state discrete solutions is because the stabilizing “upwind” terms (in our case, taken from SUPG) are not the same. The question that arises is whether the numerical solution is improved or deteriorated by the preconditioning. It is well known that accuracy problems can arise in nearly incompressible flows. The subject has been extensively studied in the context of Navier-Stokes equations. For the Euler equations, the problem has received much less attention because it is not commonly used for solving incompressible inviscid flows with this system. However “checkerboard”-type instabilities have been reported in stagnation points[1]. Surprisingly enough, experiments have shown that the numerical solutions are much improved when this preconditioning is used, as will be reported in the numerical results. We do not have a full explanation of this phenomenon yet, but at first sight it seems reasonable that the stabilization scheme will work better for a well-conditioned problem.

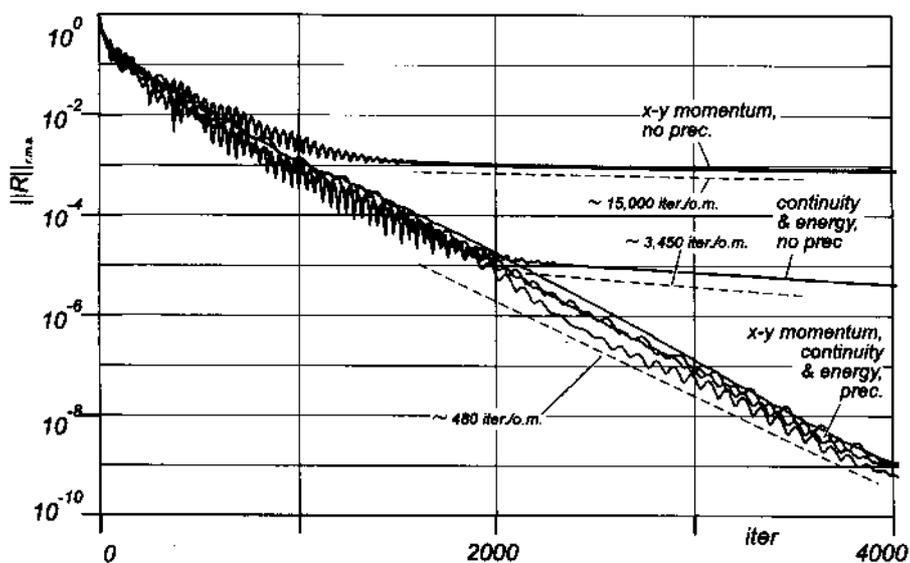
#### *Numerical results*

*Rate of convergence* In Figure 16 the convergence history can be seen for the circular bump (thickness = 12 per cent) at  $M = 10^{-3}$  with the non-preconditioned scheme and the preconditioning mass matrix presented in this paper. It can be seen that a significant improvement in the rate of convergence is achieved. Initially, for the non-preconditioned case, the residual has a relatively high r.o.c. of roughly  $(\text{r.o.c.})^{-1} = 70$  iterations/order with a highly oscillatory component, but it switches later to a smooth curve with a very low (and with a tendency to decrease) r.o.c. of  $(\text{r.o.c.})^{-1} = 6,200$  iterations/order. This behaviour is explained as follows: initially the error is mainly in the continuity equation which generates pressure waves which have a high group velocity and then a high rate of convergence. The oscillations are caused by reflections at the lateral slip boundaries (absorbing boundary conditions are used at the inlet and outlet boundaries). At a certain moment, the component of the error in the form of pressure waves has been dissipated, and that component in the form of vorticity remains almost with the same amplitude, since it has a much lower group velocity. From this point onwards, the r.o.c. is dominated by that of the vorticity waves, since they are the main component of the error. Since vorticity is propagated downstream, there is no possibility of reflections at the slip boundaries. This explains the smooth behaviour of the second part of the curve. With regard to the preconditioned case, all equations have almost the same rate of convergence of  $(\text{r.o.c.})^{-1} \approx 370$  iterations/order.



**Figure 16.** Convergence history for the bump at  $M = 10^{-3}$ , without (left) and with (right) preconditioning. Each curve corresponds to the r.m.s.-norm of the vector of nodal residuals per equation

The second example is a flow around a Joukowski profile (12 per cent thickness, 4.6 per cent camber) with an angle of attack  $\alpha = -0.89^\circ$  and a Mach number  $M = 10^{-3}$ . As in the previous case, Figure 17 shows the convergence histories for the non-preconditioned system and for the proposed preconditioning. The non-



**Figure 17.** Convergence history for the Joukowski profile (12 per cent thickness, 4.6 per cent camber) at an angle of attack  $\alpha = -0.8872^\circ$  and  $M = 10^{-3}$ , without (left) and with (right) preconditioning

preconditioned system has an r.o.c. ranging from  $(\text{r.o.c.})^{-1} = 3,450$  iterations/order for the energy and continuity equations, to 15,000 iterations/order for the momentum equations. Of course, the global r.o.c. of the system is given by the worst one. In contrast, for the preconditioned system  $(\text{r.o.c.})^{-1} \approx 480$  iterations/order for all the equations.

*Accuracy.* For the case of the circular bump considered in the previous paragraph, we take as reference a BEM (boundary element method) computation for incompressible potential flow, since the compressible effects are negligible, owing to the very low Mach number. We start with a relatively coarse mesh (11 elements on the chord) and in Figure 18 we show the  $C_p$  distribution curve on the  $x$ -axis and the profile with and without preconditioning. It can be seen that the checkerboard mode that is present for the standard (non-preconditioned) Euler code near the stagnation points is eliminated with the proposed preconditioning. In Figure 19 the numerical results obtained on a finer mesh can be seen, compared with the exact

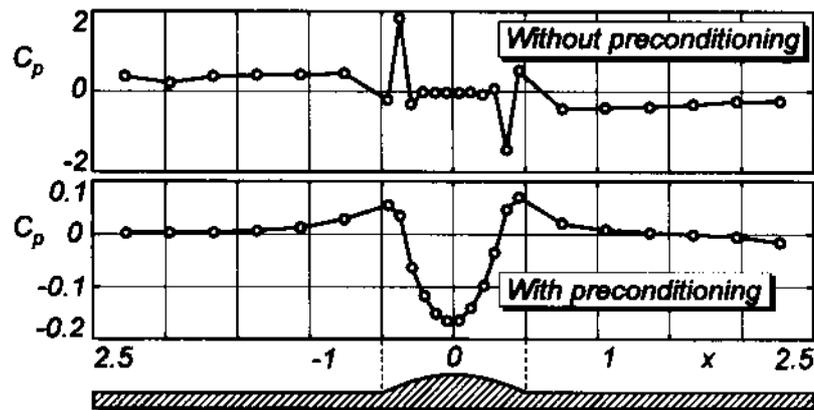


Figure 18.  
 $C_p$  distribution on the bump for the coarse mesh, with and without preconditioning

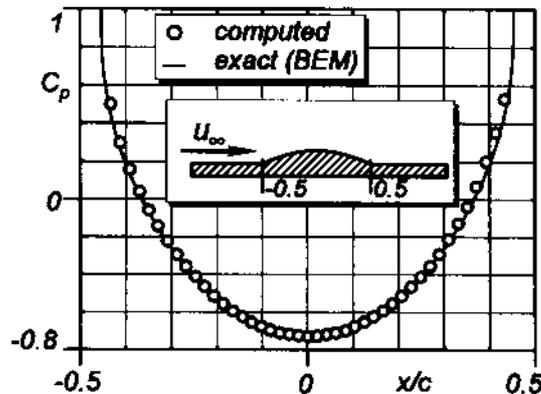
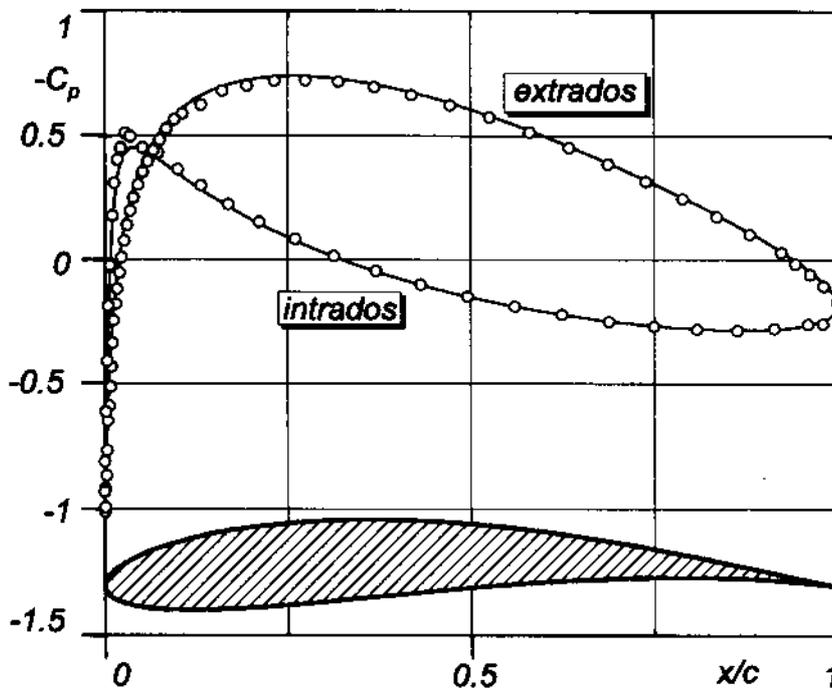


Figure 19.  
 $C_p$  distribution on the bump for the fine mesh, compared with a BEM (incompressible potential flow) computation



**Figure 20.**  
 $C_p$  distribution for the  
Joukowski profile,  
compared with the  
analytical solution by  
conformal mapping  
techniques  
(incompressible  
potential flow)

distribution. Finally, in Figure 20 the  $C_p$  distribution for the second example is shown (Joukowski profile at  $M = 10^{-3}$ , with an angle of attack  $\alpha = -0.89^\circ$ ) compared with the analytical potential flow solution obtained from conformal mapping techniques.

#### References

1. Bristeau, M., Glowinski, R., Dutto, L., Periaux, J. and Roge, G., "Compressible viscous flow calculations using compatible finite elements approximations", *International Journal for Numerical Methods in Fluids*, Vol. 11, 1990, pp. 719-49.
2. Hughes, T.J.R., Franca, L. and Balestra, M., "A new finite element formulation for CFD: V. Circumventing the Babuska-Brezzi condition: a stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations", *Computer Methods in Applied Mechanics and Engineering*, Vol. 59, 1986, pp. 85-99.
3. Frey, S.L., Franca, L.P. and Sampaio, R., "Stabilized methods for the incompressible Navier-Stokes flow", in Alder, H. *et al.* (Eds), *Numerical Methods in Engineering and Applied Sciences*, Proceedings of the International Congress on Numerical Methods in Engineering and Applied Sciences, held at Concepcion, Chile, 16-20 November 1992.
4. Hughes, T.J.R. and Tezduyar, T.E., "Finite element methods for first-order hyperbolic systems with particular emphasis on the compressible Euler equations", *Computer Methods in Applied Mechanics and Engineering*, Vol. 45, 1984, pp. 217-84.
5. Hughes, T.J.R. and Mallet, M., "A new finite element method for CFD: IV. A discontinuity-capturing operator for multidimensional advective-diffusive systems", *Computer Methods in Applied Mechanics and Engineering*, Vol. 58, 1986, pp. 329-36.

6. Storti, M., Nigro, N. and Idelsohn, S., "A Petrov-Galerkin technique for the solution of transonic and supersonic flows", *International Journal for Numerical Methods in Fluids*, Vol. 95, 1992, pp. 49-70.
7. Schreiber, R. and Keller, H.B., "Driven cavity flows by efficient numerical techniques", *J Comp. Phys*, Vol. 49, 1983, pp. 310-33.
8. Lipke, A. and Wagner, H., "Numerical solution of the Navier-Stokes equations in multiply connected domains", *Computers and Fluids*, Vol. 20 No. 1, 1991, pp. 19-27
9. Farmer, J., Martinelli, L. and Jameson, A., "Fast multigrid method for solving incompressible hydrodynamic problems with free surfaces", *AIAA Journal*, Vol. 32, 1994, pp. 1175-118.
10. Shakib, F. and Hughes, T.J.R., "A new finite element formulation for computational fluid dynamics: X. The compressible Euler and Navier-Stokes equations", *Computer Methods in Applied Mechanics and Engineering*, Vol. 89, 1991, pp. 141-219.
11. Vichnevetsky, R. and Bowles, J.B., "Analysis of numerical approximations of hyperbolic equations", *SIAM Studies in Applied Mathematics*, 1982.
12. Storti, M., Nigro, N. and Idelsohn, S., "Steady state incompressible flows using explicit schemes with an optimal local preconditioning", *Computer Methods in Applied Mechanics and Engineering*, Vol. 124, 1995, pp. 231-353.