ELSEVIER

Comput. Methods Appl. Mech. Engrg. 153 (1998) 1–44

Computer methods in applied mechanics and engineering

A comparative study of different sets of variables for solving compressible and incompressible flows

Guillermo Hauke[a,*,1], Thomas J.R. Hughes[b,2]

[a]*Area de Mecanica de Fluidos, Centro Politecnico Superior, C/Maria de Luna 3, University of Zaragoza, 50015 Zaragoza, Spain*
[b]*Division of Mechanics and Computation, Durand Building, Stanford University, Stanford, CA 94305-4040, USA*

Received 15 April 1994

Abstract

A globally conservative Galerkin/least-squares formulation which attains correct shock structure is developed for any choice of variables. Only the choice of entropy variables satisfies exactly the discrete Clausius–Duhem inequality without any dissipative mechanisms, whereas for the rest of the variables, artificial diffusion is required to guarantee entropy production. The limit of the formulation is well defined for entropy variables and the primitive variables $(p, \boldsymbol{u}, T)$, leading to conservative incompressible formulations. The approach is stable for any continuous interpolations, both for compressible and incompressible flows. A comparative study of different variables is performed, indicating that entropy variables and the primitive variables $(p, \boldsymbol{u}, T)$ possess the most attributes for practical problem solving.

1. Introduction

Fluid mechanics is subdivided into two main areas: incompressible flows and compressible flows. Liquids are typically modeled as incompressible. However, liquids, like all media, are compressible and occasionally this property is important such as in underwater acoustics or explosions. Gases, on the other hand, are typically viewed as compressible but at very low speeds they often behave almost incompressibly, such as in flows about automobiles or underhood flows. Consequently, there is no clear physical dividing line between compressible and incompressible behavior in fluids; this is similar to the situation in solids as well. However, in fluids it is rare that an individual is an expert in both incompressible *and* compressible flows. In fact, in computational fluids dynamics it is typical that very different concerns and numerical methods arise in incompressible and compressible flows. For example, people are concerned with element locking, pressure oscillations and the Babuška–Brezzi condition in the incompressible regime, whereas conservation, entropy production and shock capturing are prime concerns in the compressible regime. Different sets of variables are also used: primitive variables for the incompressible case and primarily conservation variables for the compressible case. Likewise, it is common to use different order interpolations in incompressible flows, whereas equal-order are the rule for compressible flows. Even the procedures to solve the system of equations differ. In incompressible flows segregated solvers are widely used while in compressible flows coupled solvers are preferred. This dichotomy is puzzling. It would seem that it should be possible to approach all of fluid mechanics from a more unified point of view. This thought is the motivation of the present work. Here, we build upon the work introduced in [1].

* Corresponding author.

[1] Formerly graduate Research Assistant, Stanford University. Supported by a grant of the Ministerio de Educación y Ciencia, Spain.

[2] Professor of Mechanical Engineering and Chairman of the Division of Mechanics and Computation.

0045-7825/98/$19.00 © 1998 Elsevier Science S.A. All rights reserved
PII S0045-7825(97)00043-1

In order to study the viability of a unified approach for compressible and incompressible flows we need to choose a starting formulation. Issues such as the changing character of the system of PDE's, upwinding, handling of the incompressibility constraint and using uniform interpolation functions for the whole range of flows need to be successfully treated by the algorithm. Classical methods to handle the incompressibility constraint, such as mixed methods, projection methods or penalty methods, are not optimal or simply do not work for compressible flows. At the same time, commonly used compressible solvers fail to handle adequately the incompressibility constraint. The Galerkin/least-squares method has been successfully applied to incompressible flows (see [2] and references therein) and compressible flows [3,4]. In particular, the Galerkin/least-squares term can stabilize the varied computational difficulties that are present in both types of flows. Thus, it arises as a natural formulation upon which to build a unified approach.

A necessary step to achieve our goal is to use the same variables for the whole spectrum of flows. Conservation variables, which have been thought for a long time to be the optimal variables for compressible flows, cannot be employed in the incompressible limit and therefore need to be abandoned. In principle, any set of variables with two independent thermodynamic properties (for a simple compressible substance) may be employed to compute compressible flows. We will show here, by introducing a transformation of variables in the variational formulation, how to develop a stable method for any choice of variables. However, each set of variables possesses unique properties; therefore the accuracy, convergence, computational cost and robustness of the algorithm depend upon this choice. Since the choice of variables has not been given enough attention previously, we will examine and compare in detail the performance of the most important sets of variables in the compressible regime: conservation variables, entropy variables and two types of primitive variables. It will be shown that of these, only two survive in the incompressible limit, namely, entropy variables and the primitive variables including pressure.

Generally, the methods available for calculating incompressible flows are nonconservative. In this case, the fluxes along the boundaries may be out of balance, creating numerical errors which are of importance in industrial applications. Therefore, developing a conservative incompressible formulation is very appealing from the practical standpoint. The method developed here will inherit in the incompressible limit this global conservation property.

We begin in Section 2 by presenting a method for the compressible Navier–Stokes equations using *any* set of variables. The formulation is based on the Galerkin/least-squares method for entropy variables and then extended to any set of variables. In Section 3 this approach is shown to embrace incompressible flows as well for *certain* variables. Section 4 describes solution procedures, namely, the iterative strategy and preconditioning. We demonstrate our approach in Section 5 through computation of various compressible and incompressible flows. For compressible flows, these include inviscid and viscous computations, most of which are supersonic. The comparisons performed for incompressible flows are based on the driven cavity problem and the flow around a circular cylinder. We summarize our findings and draw conclusions in Section 6.

2. A Galerkin/least-squares formulation for any set of variables

The Navier–Stokes equations can be expressed in conservative form as

$$\boldsymbol{U}_{,t} + \boldsymbol{F}^{\mathrm{adv}}_{i,i} = \boldsymbol{F}^{\mathrm{diff}}_{i,i} + \mathscr{F} \tag{1}$$

where $\boldsymbol{U}$ is the vector of conservation variables, $\boldsymbol{F}^{\mathrm{adv}}_i$ is the advective flux in the ith-direction, $\boldsymbol{F}^{\mathrm{diff}}_i$ is the diffusive flux in the ith-direction, and $\mathscr{F}$ is the source vector. An inferior comma represents partial differentiation and the summation convention on repeated indices is applied throughout.

Using any set of variables $\boldsymbol{Y}$, it is possible to rewrite (1) in quasi-linear form as

$$\boldsymbol{A}_0\boldsymbol{Y}_{,t} + \boldsymbol{A}_i\boldsymbol{Y}_{,i} = (\boldsymbol{K}_{ij}\boldsymbol{Y}_{,j})_{,i} + \mathscr{F} \tag{2}$$

where $\boldsymbol{A}_0 = \boldsymbol{U}_{,\boldsymbol{Y}}$, $\boldsymbol{A}_i = \boldsymbol{F}^{\mathrm{adv}}_{i,\boldsymbol{Y}}$ is the ith Euler Jacobian matrix, and $\boldsymbol{K} = [\boldsymbol{K}_{ij}]$ is the diffusivity matrix where $\boldsymbol{K}_{ij}\boldsymbol{Y}_{,j} = \boldsymbol{F}^{\mathrm{diff}}_i$.

In particular, consider the set of entropy variables $\boldsymbol{V}$ [5],

$$\boldsymbol{V}=\frac{1}{T}\begin{Bmatrix}\tilde{\mu}-|\boldsymbol{u}|^2/2\\ u_1\\ u_2\\ u_3\\ -1\end{Bmatrix} \tag{3}$$

where $\tilde{\mu}=e+p/\rho-Ts$ is the chemical potential per unit mass, e is the internal energy, p is the pressure, ρ is the density, s is the entropy, u_i are the cartesian velocity components and T is the absolute temperature. Eq. (2) then becomes a *symmetric advective–diffusive system* which can be written as

$$\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,t}+\tilde{\boldsymbol{A}}_i\boldsymbol{V}_{,i}=(\tilde{\boldsymbol{K}}_{ij}\boldsymbol{V}_{,j})_{,i}+\tilde{\boldsymbol{\mathcal{F}}} \tag{4}$$

Two basic properties characterize this form of the equations. The first one is the symmetry and positivity of the matrices: $\tilde{\boldsymbol{A}}_0$ is symmetric positive-definite; the $\tilde{\boldsymbol{A}}_i$'s are symmetric; and $\tilde{\boldsymbol{K}}=[\tilde{\boldsymbol{K}}_{ij}]$ is symmetric positive-semidefinite. Secondly, as it is shown in [5,6], appropriately defined discrete methods based on this form of the equations satisfy the Clausius–Duhem inequality.

These properties have been exploited by Hughes et al. [7,8] to extend the SUPG and the Galerkin/least-squares finite element methods to systems of equations, resulting in effective techniques for computing high speed compressible flows. Refs. [3,4,9] present examples of this approach.

Before presenting the Galerkin/least-squares method, we need to introduce some notation. Consider a space-time domain, where the time interval $I=]0,T[$ is subdivided into N intervals $I_n=]t_n,t_{n+1}[$, $n=0,1,\ldots,N-1$. We define for each time interval $Q_n=\Omega\times I_n$ and $P_n=\Gamma\times I_n$, where Ω is the spatial domain and Γ its boundary. Finally, the 'slab' Q_n is decomposed into elements Q_n^e, $e=1,2,\ldots,(n_{\mathrm{el}})_n$.

Given a trial solution space $\mathcal{S}_{\boldsymbol{V}}$ and weighting function space $\mathcal{V}_{\boldsymbol{V}}$, this method can be stated: Within each Q_n, $n=0,1,\ldots,N-1$, find $\boldsymbol{V}\in\mathcal{S}_{\boldsymbol{V}}$ such that $\forall\,\boldsymbol{W}\in\mathcal{V}_{\boldsymbol{V}}$:

$$\begin{aligned}
&\int_{Q_n}(-\boldsymbol{W}_{,t}\cdot\boldsymbol{U}(\boldsymbol{V})-\boldsymbol{W}_{,i}\cdot\boldsymbol{F}_i^{\mathrm{adv}}(\boldsymbol{V})+\boldsymbol{W}_{,i}\cdot\tilde{\boldsymbol{K}}_{ij}\boldsymbol{V}_{,j}-\boldsymbol{W}\cdot\tilde{\boldsymbol{\mathcal{F}}})\,\mathrm{d}Q\\
&\quad+\int_{\Omega}(\boldsymbol{W}(t_{n+1}^-)\cdot\boldsymbol{U}(\boldsymbol{V}(t_{n+1}^-))-\boldsymbol{W}(t_n^+)\cdot\boldsymbol{U}(\boldsymbol{V}(t_n^-)))\,\mathrm{d}\Omega\\
&\quad+\sum_{e=1}^{(n_{\mathrm{el}})_n}\int_{Q_n^e}(\tilde{\mathcal{L}}\boldsymbol{W})\cdot\tilde{\tau}(\tilde{\mathcal{L}}\boldsymbol{V}-\tilde{\boldsymbol{\mathcal{F}}})\,\mathrm{d}Q+\sum_{e=1}^{(n_{\mathrm{el}})_n}\int_{Q_n^e}\tilde{\nu}^h g^{ij}\boldsymbol{W}_{,i}\cdot\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,j}\,\mathrm{d}Q\\
&=\int_{P_n}\boldsymbol{W}\cdot(-\boldsymbol{F}_i^{\mathrm{adv}}(\boldsymbol{V})+\boldsymbol{F}_i^{\mathrm{diff}}(\boldsymbol{V}))n_i\,\mathrm{d}P
\end{aligned} \tag{5}$$

The first and last integrals are the Galerkin contributions in integrated by parts form. The second integral weakly enforces the continuity in time of the solution from one slab to the next. In the second integral we have used the following notation for the limits

$$W(t_n^{\pm})=\lim_{\epsilon\to 0^{\pm}}W(t_n+\epsilon) \tag{6}$$

The third integral is the *least-squares* term, where $\tilde{\mathcal{L}}$ denotes the advective–diffusive operator, i.e.

$$\tilde{\mathcal{L}}=\tilde{\boldsymbol{A}}_0\frac{\partial}{\partial t}+\tilde{\boldsymbol{A}}_i\frac{\partial}{\partial x_i}-\frac{\partial}{\partial x_i}\left(\tilde{\boldsymbol{K}}_{ij}\frac{\partial}{\partial x_j}\right) \tag{7}$$

The fourth integral is the so-called *discontinuity capturing operator*, which is written in terms of the contravariant metric tensor g^{ij}, defined by

$$g^{ij}=[\xi_{k,i}\quad\xi_{k,j}]^{-1} \tag{8}$$

where ξ_k, $k=1,2,3$, are the local element coordinates, and $\tilde{\nu}^h$ is a scalar function of the residual, namely, $\tilde{\mathcal{L}}\boldsymbol{V}-\tilde{\boldsymbol{\mathcal{F}}}$. For example, the Hughes–Mallet [10] version of the operator is

$$\tilde{\nu}^h_{\mathrm{HM}} = \max\left(0, \left[\frac{(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})\cdot \tilde{\boldsymbol{A}}_0^{-1}(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})}{g^{ij}\boldsymbol{V}_{,i}\cdot\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,j}}\right]^{1/2} - \frac{(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})\cdot\tilde{\boldsymbol{\tau}}(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})}{g^{ij}\boldsymbol{V}_{,i}\cdot\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,j}}\right) \tag{9}$$

Other versions of $\tilde{\nu}^h$ are

$$\tilde{\nu}^h_{\mathrm{quad}} = 2\frac{(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})\cdot\tilde{\boldsymbol{\tau}}(\tilde{\mathscr{L}}\boldsymbol{V} - \tilde{\boldsymbol{\mathscr{F}}})}{g^{ij}\boldsymbol{V}_{,i}\cdot\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,j}} \tag{10}$$

and

$$\tilde{\nu}^h_{\mathrm{min}} = \min(\tilde{\nu}^h_{\mathrm{HM}}, \tilde{\nu}^h_{\mathrm{quad}}) \tag{11}$$

Having established the method for entropy variables, we may transform it to any set of variables. Using the chain rule, it is easy to see that

$$\boldsymbol{V}_{,i} = \boldsymbol{V}_{,\boldsymbol{Y}}\boldsymbol{Y}_{,i} \tag{12}$$

and therefore the following relations among the coefficient matrices for entropy variables and the variables $\boldsymbol{Y}$ apply

$$\begin{aligned}
\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,\boldsymbol{Y}} &= \boldsymbol{A}_0\\
\tilde{\boldsymbol{A}}_i\boldsymbol{V}_{,\boldsymbol{Y}} &= \boldsymbol{A}_i\\
\tilde{\boldsymbol{K}}_{i,j}\boldsymbol{V}_{,\boldsymbol{Y}} &= \boldsymbol{K}_{ij}
\end{aligned} \tag{13}$$

Thus, after changing variables under the integral sign, (5) can be rewritten as follows: Within each Q_n, $n = 0, 1, \ldots, N-1$, find $\boldsymbol{Y} \in \mathscr{S}_{\boldsymbol{Y}}$ such that $\forall\, \boldsymbol{W} \in \mathscr{V}_{\boldsymbol{Y}}$:

$$\begin{aligned}
&\int_{Q_n} (-\boldsymbol{W}_{,t}\cdot\boldsymbol{U}(\boldsymbol{Y}) - \boldsymbol{W}_{,i}\cdot\boldsymbol{F}_i^{\mathrm{adv}}(\boldsymbol{Y}) + \boldsymbol{W}_{,i}\cdot\boldsymbol{K}_{ij}\boldsymbol{Y}_{,j} - \boldsymbol{W}\cdot\boldsymbol{\mathscr{F}})\,\mathrm{d}Q\\
&\quad + \int_{\Omega} (\boldsymbol{W}(t_{n+1}^-)\cdot\boldsymbol{U}(\boldsymbol{Y}(t_{n+1}^-)) - \boldsymbol{W}(t_n^-)\cdot\boldsymbol{U}(\boldsymbol{Y}(t_n^-)))\,\mathrm{d}\Omega\\
&\quad + \sum_{e=1}^{(n_{\mathrm{el}})_n}\int_{Q_n^e} (\mathscr{L}^{\mathrm{T}}\boldsymbol{W})\cdot\boldsymbol{\tau}(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})\,\mathrm{d}Q + \sum_{e=1}^{(n_{\mathrm{el}})_n}\int_{Q_n^e} \nu^h g^{ij}\boldsymbol{W}_{,i}\cdot\boldsymbol{A}_0\boldsymbol{Y}_{,j}\,\mathrm{d}Q\\
&= \int_{P_n} \boldsymbol{W}\cdot(-\boldsymbol{F}_i^{\mathrm{adv}}(\boldsymbol{Y}) + \boldsymbol{F}_i^{\mathrm{diff}}(\boldsymbol{Y}))n_i\,\mathrm{d}P
\end{aligned} \tag{14}$$

The first and last integrals constitute the Galerkin terms expressed as a function of the variables $\boldsymbol{Y}$. The jump term remains essentially unchanged. The least-squares contribution is written in terms of the differential operator $\mathscr{L}$ which is given by

$$\mathscr{L} = \boldsymbol{A}_0\frac{\partial}{\partial t} + \boldsymbol{A}_i\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}\frac{\partial}{\partial x_j}\right) \tag{15}$$

and $\mathscr{L}^{\mathrm{T}}$, which is defined by

$$\mathscr{L}^{\mathrm{T}} = \boldsymbol{A}_0^{\mathrm{T}}\frac{\partial}{\partial t} + \boldsymbol{A}_i^{\mathrm{T}}\frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}^{\mathrm{T}}\frac{\partial}{\partial x_j}\right) \tag{16}$$

Note that when entropy variables are used, $\tilde{\mathscr{L}} = \tilde{\mathscr{L}}^{\mathrm{T}}$ because of the symmetry of the coefficient matrices and (5) is recovered. We assume

$$\boldsymbol{\tau} = \boldsymbol{Y}_{,\boldsymbol{V}}\tilde{\boldsymbol{\tau}} \tag{17}$$

Note that this expression neglects the spatial derivative of $\boldsymbol{V}_{,\boldsymbol{Y}}$ emanating from the diffusion term. In SUPG, the diffusion term's action on the weighting function is omitted. Finally, the discontinuity capturing operator transforms in a similar fashion. The artificial diffusivity coefficients appear as

$$\nu^h_{\mathrm{HM}} = \max\left(0, \left[\frac{(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})\cdot \tilde{\boldsymbol{A}}_0^{-1}(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})}{g^{ij}\boldsymbol{Y}_{,i}\cdot \boldsymbol{A}_0^{\mathrm{DC}}\boldsymbol{Y}_{,j}}\right]^{1/2} - \frac{(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})\cdot \tilde{\boldsymbol{\tau}}(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})}{g^{ij}\boldsymbol{Y}_{,i}\cdot \boldsymbol{A}_0^{\mathrm{DC}}\boldsymbol{Y}_{,j}}\right) \tag{18}$$

$$\nu^h_{\mathrm{quad}} = 2\frac{(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})\cdot \tilde{\boldsymbol{\tau}}(\mathscr{L}\boldsymbol{Y} - \boldsymbol{\mathscr{F}})}{g^{ij}\boldsymbol{Y}_{,i}\cdot \boldsymbol{A}_0^{\mathrm{DC}}\boldsymbol{Y}_{,j}} \tag{19}$$

where

$$\boldsymbol{A}_0^{\mathrm{DC}} = \boldsymbol{V}_{,\boldsymbol{Y}}^{\mathrm{T}}\tilde{\boldsymbol{A}}_0\boldsymbol{V}_{,\boldsymbol{Y}} = \boldsymbol{V}_{,\boldsymbol{Y}}^{\mathrm{T}}\boldsymbol{A}_0 \tag{20}$$

REMARK 1. In general, the matrices $\boldsymbol{A}_0$, $\boldsymbol{A}_i$ and $\boldsymbol{K} = [\boldsymbol{K}_{ij}]$ are neither symmetric nor positive. This can have an impact on preconditioning and convergence of iterative solvers.

REMARK 2. Consistency in the variational sense is preserved for any set of variables, that is, the exact solution satisfies the weak form. However, stability depends on the variable choice.

REMARK 3. The finite element method (14) is globally conservative (see [8]) and attains the correct shock structure for *any* choice of variables. An important aspect of the formulation is that it is based on the correct weak form of the Navier–Stokes equations, reproducing the right jump conditions at shocks for any variable set.

REMARK 4. The choice $\boldsymbol{Y} = \boldsymbol{V}$ (entropy variables) results in satisfaction of the entropy production inequality for the discrete solution without the additional dissipative mechanisms, namely, the least-squares and discontinuity capturing terms. For $\boldsymbol{Y} \neq \boldsymbol{V}$, entropy production is contingent upon the presence of these terms. In particular, let $\boldsymbol{Y} \in \mathscr{S}_{\boldsymbol{Y}}$ be the solution of the discrete equations and $\tilde{\boldsymbol{V}}(\boldsymbol{Y}) \in \mathscr{S}_{\boldsymbol{Y}}$ be the interpolant of $\boldsymbol{V}(\boldsymbol{Y})$. The discrete Clausius–Duhem inequality for the finite element method is derived by setting $\boldsymbol{W} = \tilde{\boldsymbol{V}}(\boldsymbol{Y}) \in \mathscr{S}_{\boldsymbol{Y}}$ in (14). Note that

$$\tilde{\boldsymbol{V}}(\boldsymbol{Y})\cdot \mathscr{L}\boldsymbol{Y} = \boldsymbol{V}(\boldsymbol{Y})\cdot \mathscr{L}\boldsymbol{Y} + (\tilde{\boldsymbol{V}}(\boldsymbol{Y}) - \boldsymbol{V}(\boldsymbol{Y}))\cdot \mathscr{L}\boldsymbol{Y}$$

The first term on the right-hand side results in the exact Clausius–Duhem inequality [6] whereas the second term is the error introduced by the variable choice. This term, which is zero for entropy variables but nonzero for other variables, must be compensated for with the additional dissipative mechanisms.

REMARK 5. Satisfaction of the entropy production inequality guarantees convergence to the entropy satisfying solution.

3. A conservative formulation for compressible and incompressible flows

Consider a *general divariant fluid*, in which all thermodynamic quantities are defined in terms of two independent variables, e.g. ρ, T or ρ, e. An incompressible fluid can be regarded as a divariant fluid which is characterized by constant density and the equation of state for the internal energy per unit mass

$$\mathrm{d}e = c_v\,\mathrm{d}T \tag{21}$$

where the specific heat at constant volume is a given function of the temperature, i.e. $c_v = c_v(T)$. These relations are sufficient to determine all the thermodynamic properties, such as enthalpy h, entropy s and chemical potential $\tilde{\mu}$, which can be expressed as

$$h = e + \frac{p}{\rho} \tag{22}$$

$$\mathrm{d}s = \frac{1}{T}\mathrm{d}e \tag{23}$$

$$\tilde{\mu} = h - Ts \tag{24}$$

Note that in this case the specific heats are equal,

$$c_p = c_v \tag{25}$$

and the compressibility coefficients are zero, i.e.

$$\alpha_p = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial T}\right)_p = 0 \tag{26}$$

$$\beta_T = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial p}\right)_T = 0 \tag{27}$$

If c_v is constant, these equations can be simplified to

$$e = c_v T \tag{28}$$

$$h = c_v T + \frac{p}{\rho} \tag{29}$$

$$s = c_v \ln T \tag{30}$$

$$\tilde{\mu} = c_v T + \frac{p}{\rho} - c_v T \ln T \tag{31}$$

In obtaining these relations we have assumed the following reference values:

$$e_0 = c_v T_0 \tag{32}$$

$$s_0 = c_v \ln T_0 \tag{33}$$

The incompressible limit for a general fluid can be obtained by taking α_p and β_T to zero. For example, consider conservation and primitive variables employing density, viz.

$$\boldsymbol{U} = \rho \begin{Bmatrix} 1 \\ u_1 \\ u_2 \\ u_3 \\ e + \dfrac{|\boldsymbol{u}|^2}{2} \end{Bmatrix} \tag{34}$$

$$\boldsymbol{Y} = \begin{Bmatrix} \rho \\ u_1 \\ u_2 \\ u_3 \\ T \end{Bmatrix} \tag{35}$$

For these variables, the Euler Jacobians are a function of a, the speed of sound, and $\bar{\gamma}$, which have the following limiting behavior

$$a^2 = \frac{c_p}{\rho c_v \beta_T} \to \infty \tag{36}$$

$$\bar{\gamma} = \frac{\alpha_p}{\rho c_v \beta_T} \to \frac{0}{0} \tag{37}$$

so the first Jacobians become

$$\boldsymbol{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \infty & \mp\frac{0}{0} & \mp\frac{0}{0} & \mp\frac{0}{0} & \frac{0}{0} \\ -u_{12} & u_2 & u_1 & 0 & 0 \\ -u_{31} & y_3 & 0 & u_1 & 0 \\ \pm\infty & -\frac{0}{0} & \mp\frac{0}{0} & \mp\frac{0}{0} & \pm\frac{0}{0} \end{bmatrix} \quad \text{(conservation variables)} \tag{38}$$

$$\boldsymbol{A}_1 = \begin{bmatrix} u_1 & \rho & 0 & 0 & 0 \\ \infty & 2\rho u_1 & 0 & 0 & \frac{0}{0} \\ u_{12} & \rho u_2 & \rho u_1 & 0 & 0 \\ u_{31} & \rho u_3 & 0 & \rho u_1 & 0 \\ \pm\infty & -e_3^p + \rho u_1^2 & \rho u_{12} & \rho u_{31} & \pm\frac{0}{0} \end{bmatrix} \quad \text{(primitive variables employing density)} \tag{39}$$

and likewise for the rest of the Euler Jacobians and coefficient matrices, which can be found in Appendix A. Therefore, the incompressible limit is not well-behaved. Note that density does not make sense as a variable in the incompressible limit; it becomes constant and therefore some of its coefficients must go to infinity in order to maintain a finite contribution to changes of pressure.

However, if the quasi-linear form of the Navier–Stokes equations (2) is expressed in terms of entropy variables or primitive variables employing pressure, viz.

$$\boldsymbol{V} = \frac{1}{T}\begin{Bmatrix} \tilde{\mu} - \frac{|\boldsymbol{u}|^2}{2} \\ u_1 \\ u_2 \\ u_3 \\ -1 \end{Bmatrix} \tag{40}$$

$$\boldsymbol{Y} = \begin{Bmatrix} p \\ u_1 \\ u_2 \\ u_3 \\ T \end{Bmatrix} \tag{41}$$

the assumed equations of state can be directly substituted in the coefficient matrices and the incompressible limit can be shown to exist. For instance, the first Euler Jacobian for primitive variables employing pressure is

$$\boldsymbol{A}_1 = \begin{bmatrix} \rho\beta_T u_1 & \rho & 0 & 0 & -\rho\alpha_p u_1 \\ \rho\beta_T u_1^2 + 1 & 2\rho u_1 & 0 & 0 & -\rho\alpha_p u_1^2 \\ \rho\beta_T u_1 u_2 & \rho u_2 & \rho u_1 & 0 & -\rho\alpha_p u_1 u_2 \\ \rho\beta_T u_3 u_1 & \rho u_3 & 0 & \rho u_1 & -\rho\alpha_p u_3 u_1 \\ u_1 e_2^p & e_3^p + \rho u_1^2 & \rho u_1 u_2 & \rho u_3 u_1 & u_1 e_4^p \end{bmatrix} \tag{42}$$

where

$$k = \frac{|\boldsymbol{u}^2|}{2}, \qquad e_1 = h + k$$

$$e_1^p = \rho\beta_T e_1 - \alpha_p T, \qquad e_2^p = e_1^p - 1$$

$$e_3^p = \rho e_1, \qquad e_4^p = -\rho\alpha_p e_1 + \rho c_p$$

The limit α_p, $\beta_T \to 0$ is seen to be well defined. In the case of entropy variables, the first Euler Jacobian is [6]

$$\tilde{\boldsymbol{A}}_1 = \frac{\beta_T T}{v^2}\begin{bmatrix} u_1 & c_1 & u_1u_2 & u_3u_1 & u_1e_3 \\ & u_1\left(u_1^2 + 3\dfrac{v}{\beta_T}\right) & u_2c_1 & u_3c_1 & e_1\dfrac{v}{\beta_T} + u_1^2e_4 \\ & & u_1c_2 & u_1u_2u_3 & u_1u_2e_4 \\ & \text{symm} & & u_1c_3 & u_3u_1e_4 \\ & & & & u_1\left(e_5 + 2e_1\dfrac{v}{\beta_T}\right) \end{bmatrix} \tag{43}$$

where

$$k = \frac{|\boldsymbol{u}^2|}{2}, \qquad d = \frac{v\alpha_p T}{\beta_T}, \qquad \bar{\gamma} = \frac{v\alpha_p}{\beta_T c_v}$$

$$c_i = u_i^2 + \frac{v}{\beta_T}, \qquad e_1 = h + k, \qquad e_2 = e_1 - d$$

$$e_3 = e_2 + \frac{v}{\beta_T}, \qquad e_4 = e_2 + 2\frac{v}{\beta_T}$$

$$e_5 = e_1^2 - 2e_1 d + \frac{v(2k + c_p T)}{\beta_T}$$

Multiplying through by the common factor β_T, it can be seen that again the incompressible limit is well defined.

In this way, method (14) becomes a new technique for computing *both* incompressible and compressible flows.

REMARK 1. Note that the incompressible limit is well defined for entropy variables and the primitive variables $(p, \boldsymbol{u}, T)$, but not for conservation variables and the primitive variables $(\rho, \boldsymbol{u}, T)$. In general, the limit will not be well behaved for sets where ρ is a variable. Of particular interest among all possible sets of variables are the primitive variables $(p, \boldsymbol{u}, T)$. These have the greatest potential for efficient implementation due to the simplicity of coefficient matrices. (For analogous considerations in classical elasticity see Hughes [11, Chapter 4].)

REMARK 2. Regarding the practical issue of boundary condition specification, the most cumbersome variables are conservation variables and entropy variables, whereas the most convenient ones are primitive variables.

REMARK 3. In the incompressible case, the matrices $\boldsymbol{A}_0$ and $\tilde{\boldsymbol{A}}_0$ have rank 4, and $\tilde{\boldsymbol{A}}_0$ is positive-semidefinite.

REMARK 4. Global conservation is retained in the incompressible limit. Again, the choice $\boldsymbol{Y} = \boldsymbol{V}$ (entropy variables) results in satisfaction of the entropy production inequality for the discrete solution without the additional dissipative mechanisms.

REMARK 5. This method accommodates thermally coupled and isothermal incompressible flows. Thermally coupled incompressible flows are obtained by allowing the viscosity to depend on the temperature. Isothermal incompressible flows can be obtained by specifying a constant viscosity.

REMARK 6. The Boussinesq approximation may be attained by allowing the density to be temperature-dependent in the body force term.

REMARK 7. The following simple, diagonal form of the matrix $\boldsymbol{\tau}$ is a natural extension of previous research on incompressible flows for primitive variables (see, for instance, [2]),

$$\boldsymbol{\tau}_{\text{inc}} = \text{diag}(\tau_c, \tau_m, \tau_m, \tau_m, \tau_e) \tag{44}$$

where

$$\tau_c = \frac{|\boldsymbol{u}|h_1^e}{2}\min(1, \text{Re}^h) \tag{45}$$

$$\tau_m = \min\left(\frac{\Delta t}{\rho}, \frac{h_2^e}{2\rho|\boldsymbol{u}|}, \frac{m^e (h_3^e)^2}{4\mu}\right) \tag{46}$$

$$\tau_e = \min\left(\frac{\Delta t}{\rho c_v}, \frac{h_2^e}{2\rho c_v|\boldsymbol{u}|}, \frac{m^e (h_3^e)^2}{4\kappa}\right) \tag{47}$$

The above expressions have been written in terms of h_i^e, $i = 1, 2, 3$ measures of the element size; Δt, the time step; μ and κ, the viscosity and the thermal conductivity; the element Reynolds number,

$$\mathrm{Re}^h = \frac{\rho|\boldsymbol{u}|h_1^e}{2\mu} \tag{48}$$

and $m^e = \min(1/3, 2C^e)$, where C^e is a constant arising from an inverse estimate of the second derivatives of the shape functions (see [2] for details). The h_i^e are defined in terms of the metric tensors g^{ij} and $g_{ij} = [g^{ij}]^{-1}$,

$$h_1^e|\boldsymbol{u}| = 2(u_i g^{ij} u_j)^{1/2} \tag{49}$$

$$\frac{|\boldsymbol{u}|}{h_2^e} = \frac{1}{2}(u_i g_{ij} u_j)^{1/2} \tag{50}$$

$$h_3^e = h_2^e \tag{51}$$

For steady state computations the Δt term is omitted, making the long term solution independent of the time step.

It is possible to obtain a diagonal matrix for entropy variables based on the advection and diffusion of the variables, such as $\tilde{\boldsymbol{\tau}}_{\mathrm{inc}} = \mathrm{diag}(\boldsymbol{V}_{,\boldsymbol{Y}}\boldsymbol{\tau}_{\mathrm{inc}})$, which is symmetric, positive-definite. But this simple choice, as will be shown in the section on numerical examples, is not good. It is therefore necessary to find an improved matrix $\boldsymbol{\tau}$. If we wish to retain the properties $\tilde{\boldsymbol{\tau}}_{\mathrm{inc}}$ is symmetric and positive-definite and $\tilde{\boldsymbol{\tau}}_{\mathrm{inc}} = \boldsymbol{V}_{,\boldsymbol{Y}}\boldsymbol{\tau}_{\mathrm{inc}}$, it can be shown that the matrices for primitive and entropy variables have the form,

$$\boldsymbol{\tau} = \tilde{\boldsymbol{T}}_{\mathrm{U}}\tilde{\boldsymbol{D}}\tilde{\boldsymbol{T}}_{U}^{\mathrm{T}}\boldsymbol{V}_{\boldsymbol{Y}}^{\mathrm{T}} \tag{52}$$

$$\tilde{\boldsymbol{\tau}} = (\boldsymbol{V}_{,\boldsymbol{Y}}\tilde{\boldsymbol{T}}_{\mathrm{U}})\tilde{\boldsymbol{D}}(\boldsymbol{V}_{,\boldsymbol{Y}}\tilde{\boldsymbol{T}}_{\mathrm{U}})^{\mathrm{T}} \tag{53}$$

where $\tilde{\boldsymbol{D}}$ is a diagonal, positive-definite matrix and $\boldsymbol{T}_{\mathrm{U}}$ is an upper triangular matrix with unit diagonals. Furthermore, there are 10 constants in two dimensions and 15 in three dimensions to be chosen. Since the diagonal stabilizing matrix has proved to work well in the context of pressure primitive variables, it is plausible that the new design should depart as little as possible from it. Therefore, the diagonal entries of the new matrix are chosen to equal the $\boldsymbol{\tau}_{\mathrm{inc}}$ for the Navier–Stokes equations in non-conservative form, namely (44)–(47). The last column is chosen to be zero so that the residual of the energy equation does not influence the velocity field when primitive variables are employed. Numerical experiments also suggest that nonzero terms in the last column degrade accuracy. The off diagonal parameters of the second, third and fourth rows can be set to zero. These choices lead to the following matrix

$$\boldsymbol{\tau}_{\mathrm{inc}} = \begin{bmatrix} \tau_c & \rho u_1 \tau_m & \rho u_2 \tau_m & \rho u_3 \tau_m & 0 \\ 0 & \tau_m & 0 & 0 & 0 \\ 0 & 0 & \tau_m & 0 & 0 \\ 0 & 0 & 0 & \tau_m & 0 \\ -\bar{e}_1 \tau_e & -u_1 \tau_e & -u_2 \tau_e & -u_3 \tau_e & \tau_e \end{bmatrix} \tag{54}$$

and

$$\tilde{\boldsymbol{\tau}}_{\mathrm{inc}} = \boldsymbol{V}_{,Y}\boldsymbol{\tau}_{\mathrm{inc}} = \begin{bmatrix} \tilde{\tau}_{11} & u_1\dfrac{\bar{e}_1}{T^2}\tau_e & u_2\dfrac{\bar{e}_1}{T^2}\tau_e & u_3\dfrac{\bar{e}_1}{T^2}\tau_e & \dfrac{\bar{e}_1}{T^2}\tau_e \\ & \dfrac{1}{T}\tau_m + \dfrac{u_1^2}{T^2}\tau_e & \dfrac{u_1u_2}{T^2}\tau_e & \dfrac{u_3u_1}{T^2}\tau_e & -\dfrac{u_1}{T^2}\tau_e \\ & & \dfrac{1}{T}\tau_m + \dfrac{u_2^2}{T^2}\tau_e & \dfrac{u_2u_3}{T^2}\tau_e & -\dfrac{u_2}{T^2}\tau_e \\ & \text{symm.} & & \dfrac{1}{T}\tau_m + \dfrac{u_3^2}{T^2}\tau_e & -\dfrac{u_3}{T^2}\tau_e \\ & & & & \dfrac{1}{T^2}\tau_e \end{bmatrix} \tag{55}$$

where

$$\tilde{\tau}_{11} = \frac{1}{\rho T}\tau_c + \left(\frac{\bar{e}_1}{T}\right)^2\tau_e \tag{56}$$

$$\bar{e}_1 = h - \frac{|\boldsymbol{u}|^2}{2} \tag{57}$$

According to (52), this corresponds to

$$\tilde{\boldsymbol{T}}_{\mathrm{U}} = \begin{bmatrix} 1 & \rho u_1 & \rho u_2 & \rho u_3 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{58}$$

$$\tilde{\boldsymbol{D}} = \mathrm{diag}(\rho T\tau_c, T\tau_m, T\tau_m, T\tau_m, T^2\tau_e) \tag{59}$$

Note that these matrices are not unique nor necessarily optimal. Further research is necessary to define the optimal choices.

REMARK 8. If entropy variables or primitive variables are employed, the same formulation can be used to compute compressible and incompressible flows. The difficulties to overcome, though, are to design a $\boldsymbol{\tau}$ matrix and a discontinuity capturing operator which are valid for both types of flows. In particular, the compressible $\boldsymbol{\tau}$ matrix and the discontinuity capturing operators are not well defined in the incompressible limit. Conversely, the incompressible $\boldsymbol{\tau}$ is not effective for transonic and supersonic flows. A first attempt to define a $\boldsymbol{\tau}$ matrix suitable for both compressible and incompressible flows has led to the following expression:

$$\boldsymbol{\tau} = \frac{\boldsymbol{\tau}_{\mathrm{inc}} + \left(\dfrac{M}{M_{\mathrm{ref}}}\right)^k \boldsymbol{\tau}_{\mathrm{comp}}}{1 + \left(\dfrac{M}{M_{\mathrm{ref}}}\right)^k} \tag{60}$$

where $\boldsymbol{\tau}_{\mathrm{inc}}$ is the incompressible stabilization matrix, $\boldsymbol{\tau}_{\mathrm{comp}}$ is the compressible one, M is the Mach number, and k and M_{ref} are constants. The simple matching works well in practice for a wide range of values of the constants, namely $1 \leq k \leq 4$ and $\frac{1}{3} \leq M_{\mathrm{ref}} \leq 1$.

4. The solution of the system of equations

A nonlinear algebraic system of equations must be solved in each time slab. For more details in the derivation of the algorithm implemented here, the reader is referred to the work of Shakib et al. [3], which is closely followed in this section. The algorithm requires solving a linear algebraic system of equations at each iteration. Direction solvers based on Gaussian elimination require enormous storage and computational cost. Ideally, a

better approach would be to employ iterative solvers, such as preconditioned GMRES. When the steady solution is sought, it is not necessary to accurately solve the equation system at each time step. Therefore, performing only one Newton iteration per time step and approximately solving the linear system of equations accelerates the convergence towards the steady state. In this respect, iterative techniques have an advantage over direct solvers because the accuracy in the solver can be easily controlled. However, iterative techniques are very sensitive to the condition number of the system and therefore some form of preconditioning is usually crucial. We have employed the GMRES algorithm with two levels of preconditioning.

The first level is based on nodal block-diagonal scaling. Let the linear system of equations be denoted by

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{61}$$

where $\boldsymbol{x}$ is the vector of unknowns, $\boldsymbol{A}$ is the left-hand side matrix and $\boldsymbol{b}$ is the right-hand side vector. The nodal block-diagonal preconditioner $\boldsymbol{C}$ is formed by extracting the $n_{\text{dof}} \times n_{\text{dof}}$ nodal diagonal blocks from $\boldsymbol{A}$ where n_{dof} is the number of degrees of freedom per node

$$\boldsymbol{C} = \text{block-diag}(\boldsymbol{A})$$

Since the blocks are not symmetric unless entropy variables are employed, $\boldsymbol{C}$ is factored with an LU decomposition

$$\boldsymbol{C} = \boldsymbol{L}\boldsymbol{U} \tag{62}$$

and the preconditioned system of equations becomes

$$(\boldsymbol{L}^{-1}\boldsymbol{A}\boldsymbol{U}^{-1})(\boldsymbol{U}\boldsymbol{x}) = (\boldsymbol{L}^{-1}\boldsymbol{b}) \tag{63}$$

The second level of preconditioning employs element-by-element approximate factorization strategies [12], namely,

(1) Gauss–Seidel element-by-element preconditioning (GS-EBE).
(2) Nonsymmetric Cholesky element-by-element preconditioning (NC-EBE).

REMARK 1. In the incompressible case the nodal blocks are not in general positive-definite. However, experience indicates the LU factorization can usually be performed without row exchanges, but it may fail in situations where the velocity vector is zero everywhere, such as an initial condition where the fluid is at rest. In this case, the zero condition may be perturbed to a small nonzero value.

REMARK 2. The system of equations is poorly conditioned in the incompressible case. Therefore more iterations are required for convergence.

Fig. 1 compares for an incompressible flow, the efficiency of the direct solver versus the preconditioned GMRES solver. Here, the problem under consideration is the driven cavity flow, at a Reynolds number of 400, with a discretization of 1616 elements. The computations include only the block-diagonal preconditioner.

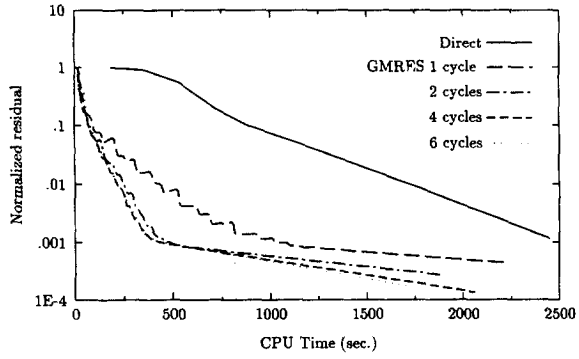


Fig. 1. Comparison of solvers for incompressible flows. Driven cavity flow, Re = 400.

The number of GMRES iterations increases significantly with the tolerance, for instance, for a tolerance as large as 0.5, the number of iterations can already increase to 40. Rather than the tolerance, the total number of iterations seems to be the limiting factor for the iterations, and so this is what is varied in the computations. In the comparisons presented in Fig. 1 the dimension of the Krylov space was set to 10, so for a 2-cycle GMRES computation, the number of iterations per time step will usually be 20.

5. Numerical examples

In this section we compare results obtained with entropy variables, conservation variables, density primitive variables $(\rho, \boldsymbol{u}, T)$ and pressure primitive variables $(p, \boldsymbol{u}, T)$. Since this section will focus on comparing the sets of variables, fairly coarse meshes are employed in all the examples. Here, we will be mainly concerned with steady state solutions which were reached by marching an implicit algorithm in time, and solving the resulting systems of equations via the preconditioned GMRES algorithm. In the GMRES algorithm, the tolerance was set to 0.1, the dimension of the Krylov space to 10 and the maximum number of cycles equalled 2. Unless otherwise specified, the solutions were started with free stream values, using a local time-stepping strategy. The finite element spaces were formed by linear triangles or bilinear quadrilaterals in space and constants in time, in which 3-point or 2×2 Gaussian quadrature, respectively, was employed for integration in space. The computations presented were performed on a CONVEX-C1 in double precision.

The solutions of the compressible Navier–Stokes equations assumed the perfect gas model and a constant Prandtl number of 0.72, while the Prandtl number was 1.0 for incompressible flows. The Hughes–Mallet discontinuity capturing operator (9) was employed for inviscid calculations while the 'minimum' operator (11) was used for Navier–Stokes computations. No discontinuity capturing operator was employed for incompressible flows. Equal order interpolations were used throughout.

The following definitions for the non-dimensional parameters will be used throughout: Pressure coefficient $C_p = (p - p_\infty)/\frac{1}{2}\rho_\infty u_\infty^2$, skin friction coefficient $C_f = \tau_{\text{wall}}/\frac{1}{2}\rho_\infty u_\infty^2$, and heat-flux coefficient $C_h = q_{\text{wall}}/\frac{1}{2}\rho_\infty u_\infty^3$. All these quantities are computed using the consistent boundary flux method [3]. τ_{ij} will denote the viscous stress tensor components and q_i the heat flux vector components. Finally, by 'residual' we mean here the residual of the nonlinear *discrete* system of equations.

5.1. Poiseuille flow

The fully developed flow inside a rectangular channel, between two infinite parallel plates (see Fig. 2), is characterized by a quadratic velocity profile and a decreasing linear pressure field. This simple example illustrates some errors that arise in coarse meshes at high Reynolds numbers as a consequence of the nonlinear change of variables. In the context of entropy variables (assuming momentarily that the flow is isothermal), if

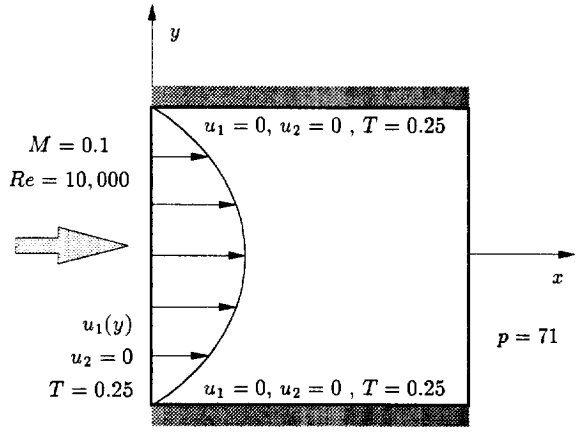


Fig. 2. Poiseuille flow problem.

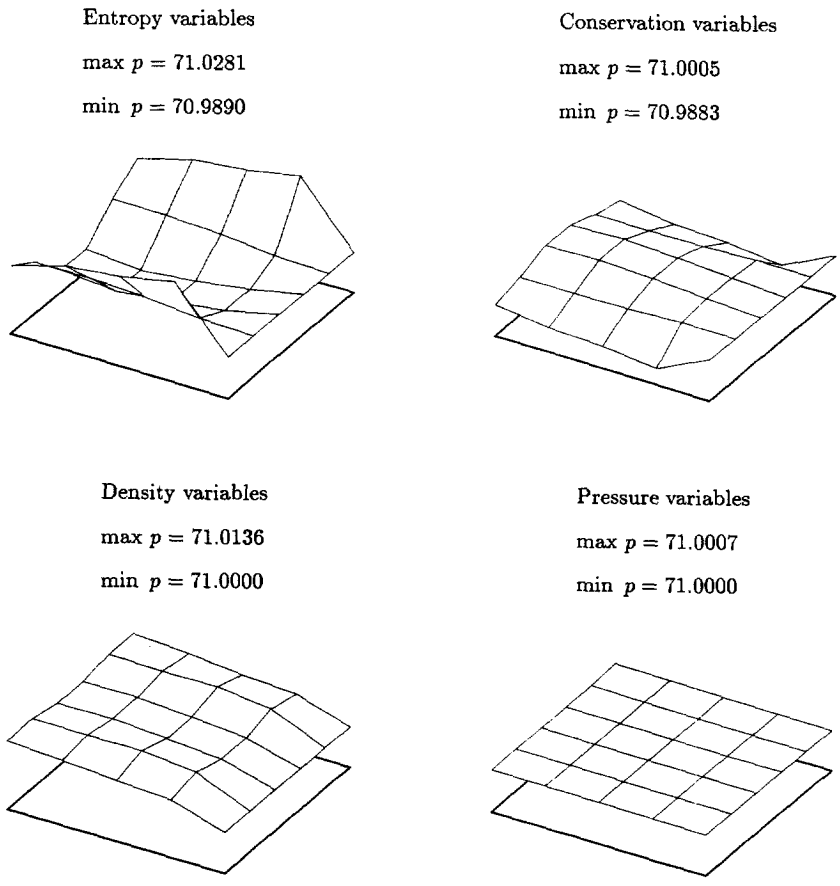


Fig. 3. Pressure elevations for Poiseuille flow, Re = 10 000.

the velocity components in the second, third and fourth slots of V are, say, linearly interpolated, then, under equal order interpolation for all the variables, the square of the velocity in the first slot of V will also be linearly interpolated, which should be, however, quadratic. This inconsistency in interpolating the variables causes the advective fluxes not to cancel in situations where they should, as in Poiseuille flow.

To illustrate the previous difficulty, a unit square was discretized into a mesh of 4×5 elements. The Reynolds number based on the channel width and the centerline velocity was set to 10 000 and the Mach number to 0.1. The boundary conditions, depicted in Fig. 2, are zero velocity and prescribed temperature along the walls, a parabolic velocity profile with prescribed temperature at the inlet boundary and constant pressure at the exit boundary.

Fig. 3 shows, at the same scale, the pressure surfaces generated by the different variables. The maximum theoretical value of the pressure field is 71.0008 and the minimum is 71.0000. The most accurate pressure gradient is given by pressure variables, while entropy variables give the least accurate pressure gradient. Conservation and density variables produce intermediate results. All these discrepancies vanish as the Reynolds number decreases or as the mesh is refined. No problems are detected in the velocity field.

5.2. One dimensional steady shock

The methods presented above are conservative and thus when a steady shock is given as the initial condition its position remains fixed with time. As an example, a Mach 2 shock is advanced in time on a mesh of 21×1 square elements which extends over $-2.1 \leq x \leq 2.1$, $-0.1 \leq y \leq 0.1$. The shock is placed at $x = 0$ and the initial conditions are given by

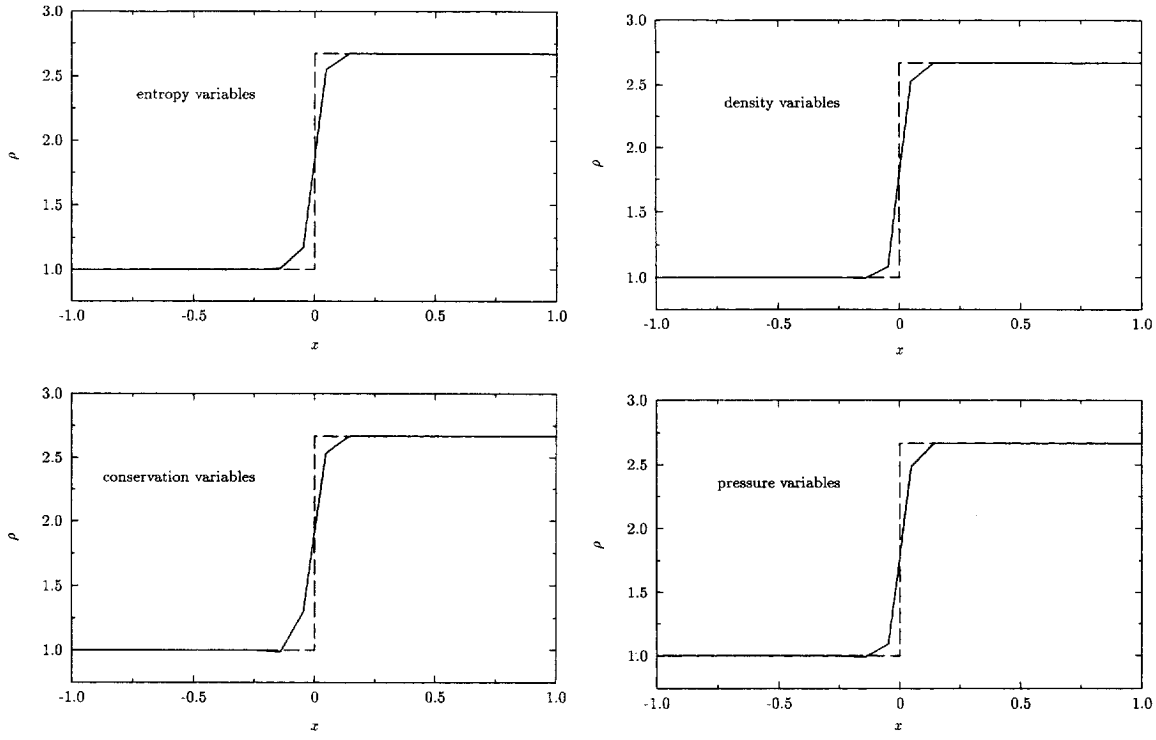


Fig. 4. One-dimensional shock, $M = 2$.

$$x<0\begin{cases}M=2\\ \rho=1\\ u_1=1\\ T=0.61941\ 10^{-3}\end{cases}\qquad x>0\begin{cases}M=0.57735\\ \rho=2.66667\\ u_1=0.37500\\ T=0.10453\ 10^{-2}\end{cases}$$

The y-velocity component, u_2, was set to zero in the entire domain. Density, temperature and velocity were specified as boundary conditions at the inflow boundary and temperature at the outflow.

Fig. 4 shows the density calculated using different variables after 100 time steps at a CFL number of 1. All the methods maintain the correct position and amplitude of the shock. For a nonconservative method the shock will move.

5.3. Oblique shock

This problem consists of an inviscid, Mach 2, uniform flow, which is sharply turned at an angle of 10° by a wall (see Fig. 5). An oblique shock, which forms an angle of 29.3° with the wall, is generated by the corner. The

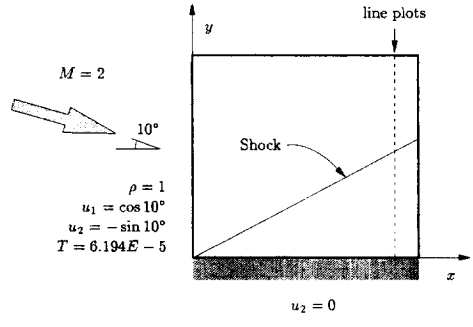


Fig. 5. Oblique shock. Problem schematic.

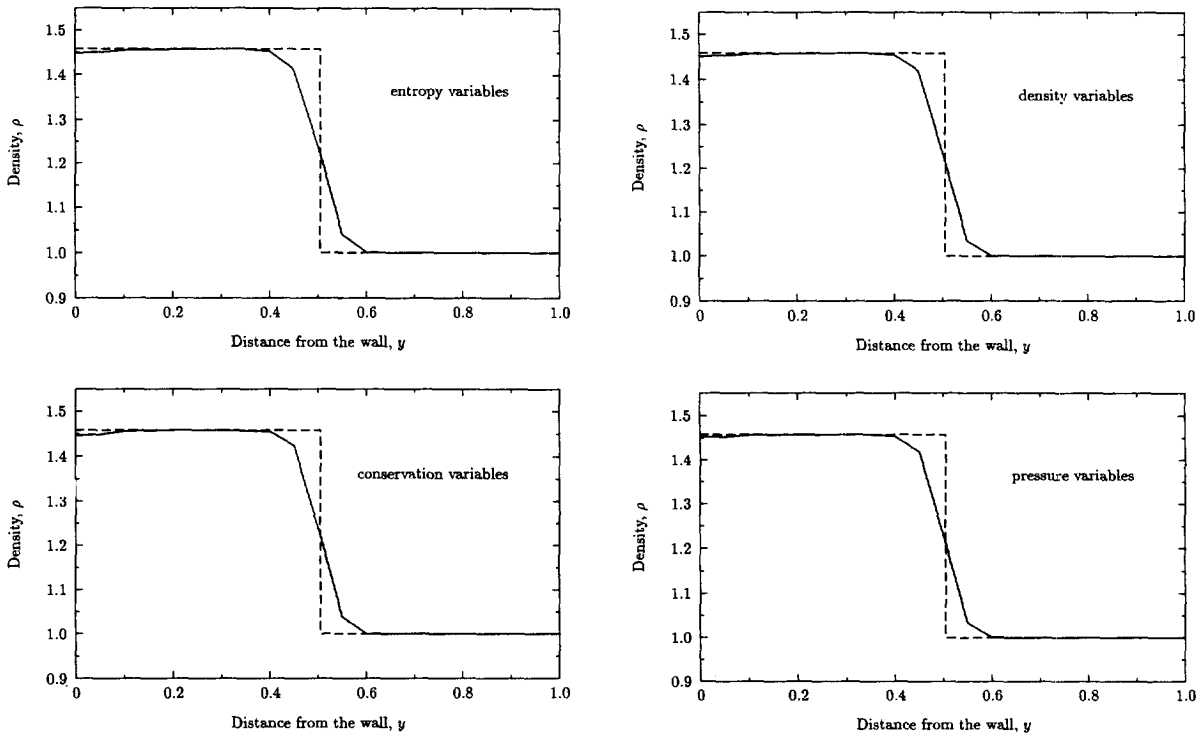


Fig. 6. Mach 2 oblique shock. Line cut at $x = 0.9$.

objective of this example is to show again that, independent of variable choice, correct shock structure is attained.

The computational domain $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ is discretized by a mesh of 20×20 square elements. The variables ρ, u_1, u_2 and T were prescribed at the inflow and top boundaries; none were prescribed at the exit; and zero normal velocity component was set at the wall, i.e. $u_2 = 0$.

Fig. 6 shows the density along the line $x = 0.9$. Note that the solutions are almost identical and resolve the shock within four elements, despite the shock being skew to the mesh.

Residual convergence is plotted in Fig. 7 for a CFL number of 10. All variables behave similarly, although for conservation variables the residual oscillates somewhat.

5.4. Flat plate

A Mach 3 viscous flow over a flat plate is chosen to compare the accuracy of wall quantities. From the leading edge of the plate, which is a singular point, a boundary layer and shock develop. The Reynolds number

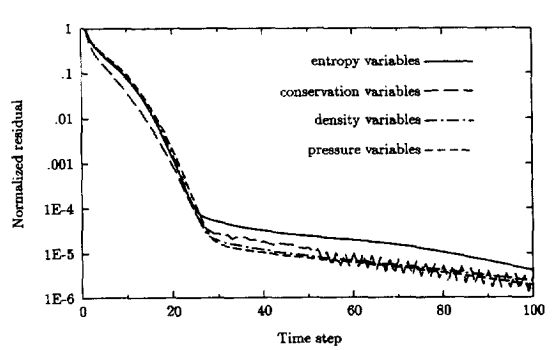


Fig. 7. Mach 2 oblique shock. Residual convergence.

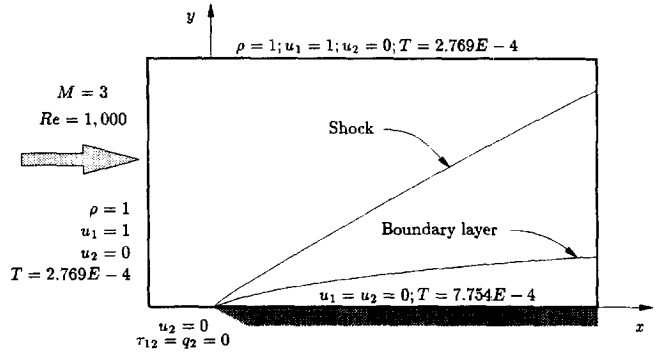


Fig. 8. Flat plate problem.

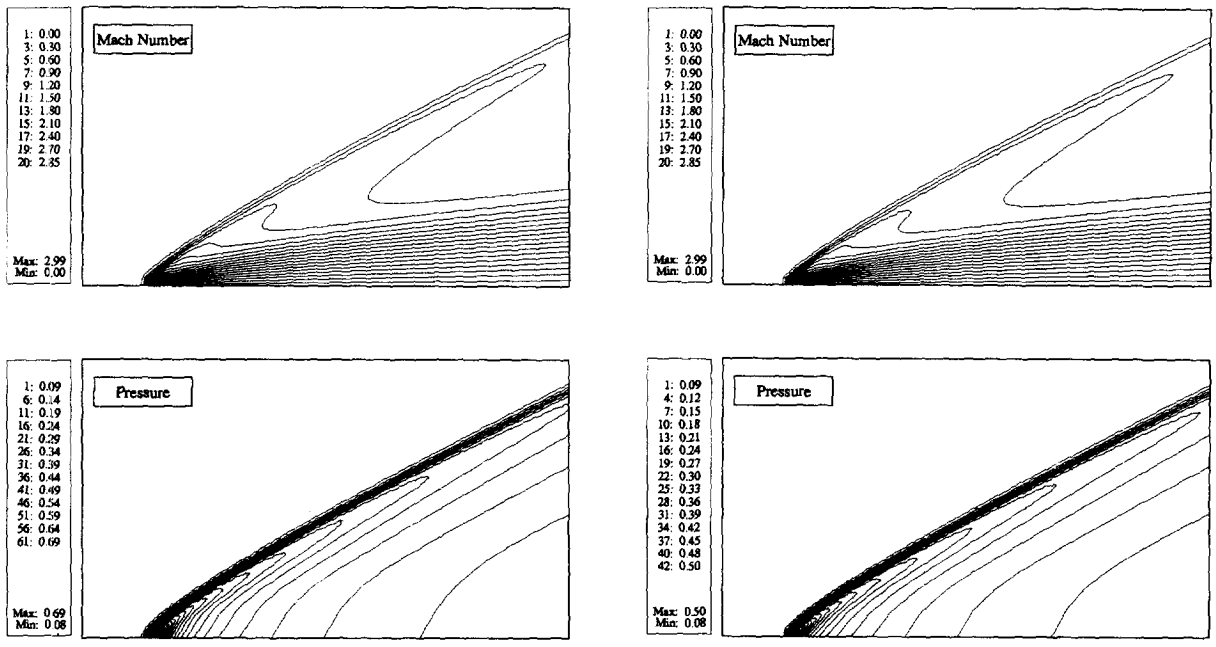


Fig. 9. Mach 3 flat plate. Entropy variables (7168 elements).

Fig. 10. Mach 3 flat plate. Conservation variables (7168 elements).

based on free stream values and unit length is 1000, and the following Sutherland viscosity law is used: $\mu = 0.0906T^{1.5}/(T + 0.0001406)$. The computational domain $-0.2 \leqslant x \leqslant 1.2$, $0 \leqslant y \leqslant 0.8$ was discretized into 112×64 (7168) square elements with edge size $h = 0.0125$. The leading edge of the plate was placed at $x = 0$, $y = 0$.

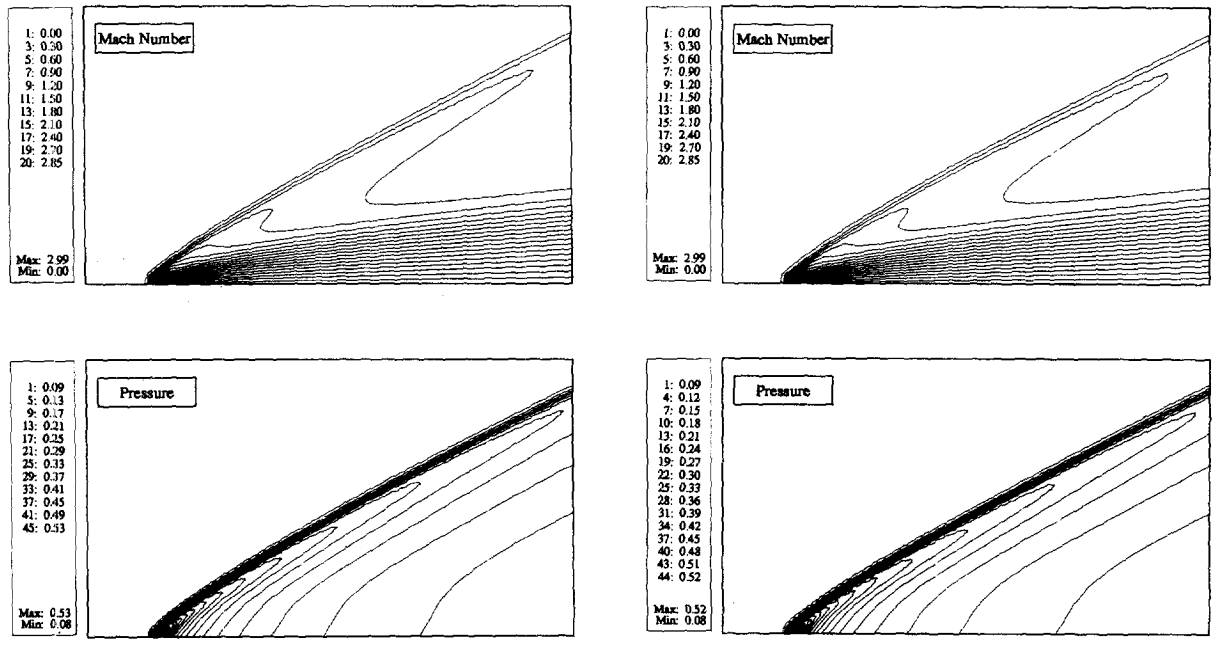


Fig. 11. Mach 3 flat plate. Density variables (7168 elements).

Fig. 12. Mach 3 flat plate. Pressure variables (7168 elements).

As illustrated in Fig. 8, the variables ρ, u_1, u_2 and T were prescribed at the inflow and top boundaries; zero vertical velocity, zero tangential viscous traction and zero heat flux were set on the symmetry line; and $u_1 = u_2 = 0$ and the stagnation temperature,

$$T_{\text{stag}} = T_\infty\left(1 + \frac{\gamma - 1}{2} M_\infty^2\right)$$

were imposed at the wall. Although it is theoretically necessary to prescribe one thermodynamic property at the subsonic portion of the exit boundary, no boundary condition was specified there. Conservation variables required special treatment, i.e., zero normal temperature gradient needed to be specified at the subsonic part of the exit boundary to stabilize the solution.

Figs. 9–12 show Mach number and pressure contours. The contours are practically identical for all variables. However, upon inspection of wall quantities in Fig. 13, we find that the results indeed depend on variable choice. Entropy variables are slightly less accurate predicting C_p and C_f than the rest of the variables, while conservation variables are slightly worse predicting C_h.

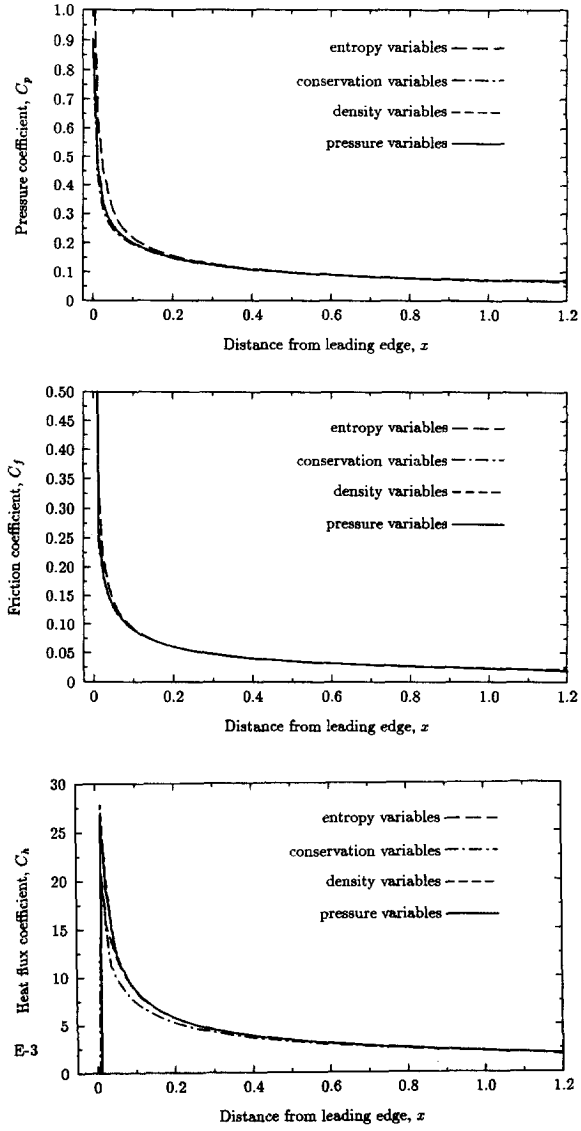


Fig. 13. Mach 3 flat plate. Wall quantities.

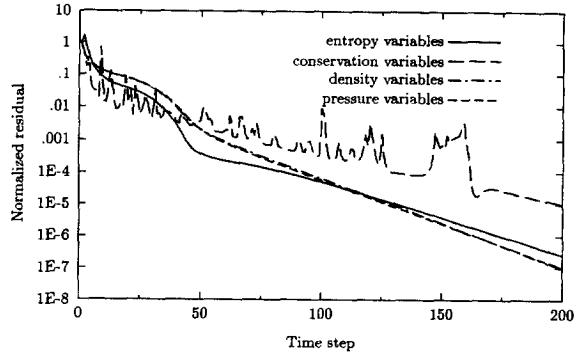


Fig. 14. Mach 3 flat plate. Residual convergence.

A plot of the normalized residual versus CPU time is shown in Fig. 14. The solution was advanced at a CFL number of 10 except for conservation variables, where for stability reasons, the CFL number was decreased to 5. All the methods behaved similarly, although the convergence with conservation variables is very oscillatory.

5.5. Compression corner

In the compression corner problem, a supersonic uniform flow over a plate is compressed at an angle of 10° degrees by a wall (see Fig. 15). A boundary layer and a shock are developed from the leading edge of the plate. The boundary layer separates around the corner of the plate, reattaching some distance downstream. The free stream Mach number is 3 and the Reynolds number based on free stream values and unit length is 16 800. The viscosity is given by the Sutherland law, $\mu = 0.00539T^{1.5}/(T + 0.0001406)$. Numerical accuracy of the variables is examined under more complex physics than that of the flat plate boundary layer. The computational domain, determined by

$$-0.1 \leqslant x \leqslant 1.0\,; \qquad 0 \leqslant y \leqslant 0.575$$

$$1.0 \leqslant x \leqslant 1.8\,; \qquad \tan 10°(x - 1.0) \leqslant y \leqslant 0.575 + \tan 10°(x - 1.0)$$

is discretized into a mesh of 104×39 (4056) elements (see Fig. 16), with a $\Delta y_{\min} = 6.47 \times 10^{-4}$ and a maximum aspect ratio of 26. The leading edge is placed at the origin.

Figs. 17–20 show contour plots of Mach number, density, pressure and temperature. The results are virtually the same for each variable set. However, differences are noticeable at the leading edge of the plate, which is a point of singularity. The results show that entropy variables effectively control the singularity, reducing overshoots and undershoots. Conservation variables and primitive variables, however, do not control the singularity as well as entropy variables. They are especially weak at controlling the undershoot in temperature, causing the spike in the Mach number. Note that these differences were not apparent in the flat plate problem, where the elements were squares. Entropy variables seem to handle singularities better with large aspect ratio elements than do the other variables.

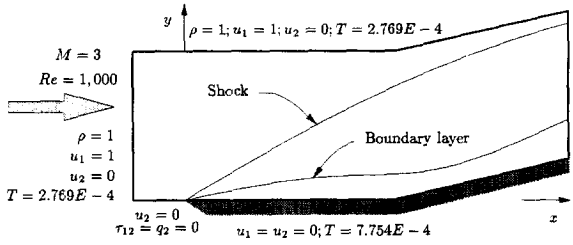


Fig. 15. Mach 3 compression corner. Problem schematic.

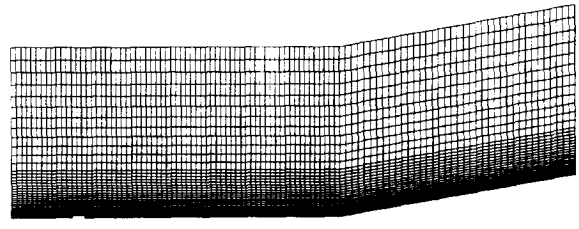


Fig. 16. Finite element mesh for the compression corner problem (4056 elements).

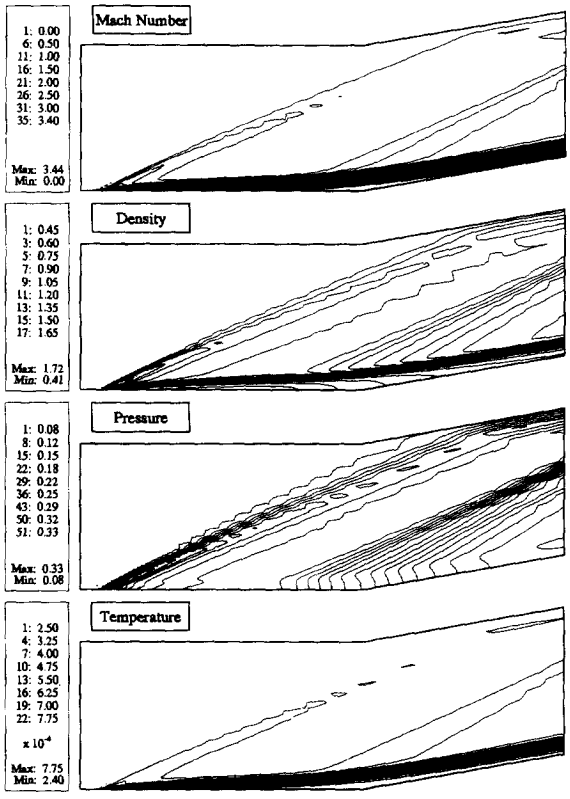


Fig. 17. Mach 3 compression corner. Entropy variables (4056 elements).

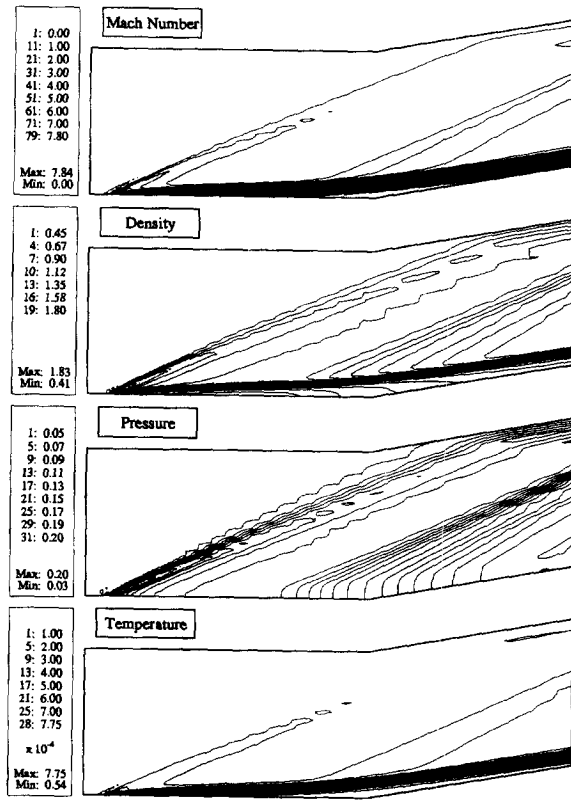


Fig. 18. Mach 3 compression corner. Conservation variables (4056 elements).

Fig. 21 analyzes the accuracy of wall quantities. C_p and C_f obtained with entropy variables are slightly less accurate than those obtained with other variables. Conservation variables are slightly worse for C_h than the other variables. Note that if no special outflow boundary condition treatment is used, conservation variables give rise to errors in the wall quantities at the outflow, especially in the heat transfer coefficient, C_h. Conservation variables require more complex treatment of boundary conditions than other variables.

Fig. 22 compares the residual convergence of solutions which were advanced at a CFL number of 10. Entropy and conservation variables present a higher initial drop in the residual but later, primitive variables possess a steeper slope. Conservation variables again converge in oscillatory fashion. It is assumed that this behavior is related to the implementation of the temperature boundary condition, since oscillations are absent for the adiabatic plate.

5.6. Inviscid double ellipse

The problem description is given in Fig. 23. The flow at the nose is suddenly brought to a stop, causing a strong bow shock around the body. A weak shock is also formed over the cockpit, called the canopy shock. The strong shock and the complex geometry challenge the robustness and stability of the methods. According to supersonic Euler calculations, the following boundary conditions were set for this problem. All the variables were prescribed at the inlet boundary, none at the exit and zero normal velocity was imposed at the wall. In order to reveal any differences among the various variables sets, the flow is computed on a very coarse mesh, consisting of 4448 elements and 2350 nodes (Fig. 24).

Fig. 25 shows the Mach number contours for entropy and density variables. Again, entropy variables are better able to control overshoots, producing a smoother solution than density variables near the shock. This is

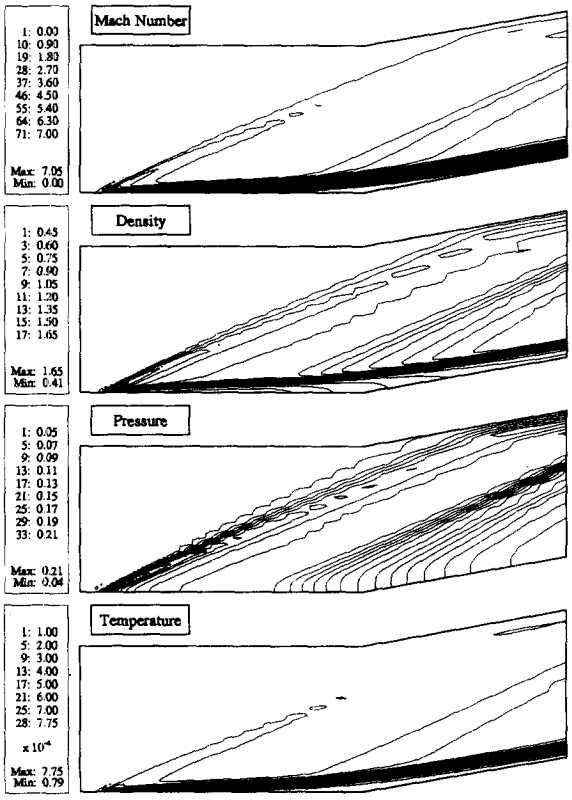


Fig. 19. Mach 3 compression corner. Density variables (4056 elements).

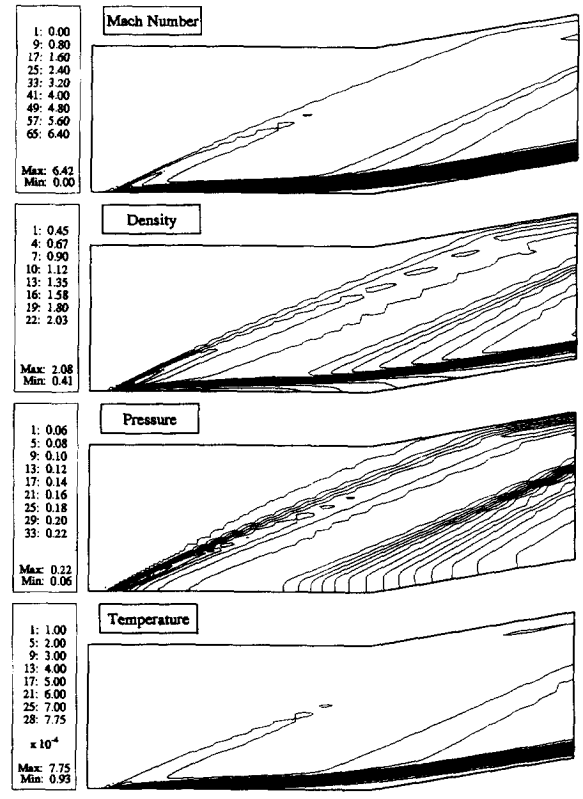


Fig. 20. Mach 3 compression corner. Pressure variables (4056 elements).

important because a large undershoot in temperature may terminate execution or cause divergence of the residual, which occurred for conservation and pressure variables.

REMARK. Overshoots and undershoots can be controlled by increasing artificial diffusion. Results obtained by doubling the value of the discontinuity capturing operator are shown in Fig. 26. Note that increasing the artificial diffusion was not necessary for weaker shocks, such as the Mach 2 oblique shock. The stronger the gradients, the larger the errors in the discrete Clausius–Duhem inequality, making necessary a larger correction with dissipative mechanisms. The optimal amount of artificial viscosity depends on the variables used and the strength of the shock. In conclusion, entropy variables, followed in second place by density variables, are the most robust because they need the least amount of artificial viscosity. Better results may be obtained by tuning τ as well as the discontinuity capturing operator for each particular variable set independently. This observation is consistent with the shock capturing operator and the least-squares term employed by Tezduyar's group for conservation variables (see [13,14]). Indeed, their diffusive mechanisms and, in particular, discontinuity capturing operator are more diffusive than the ones employed for entropy variables.

Note that in spite of the coarseness of the mesh, the shock extends over only four elements. Away from the shock the contours are practically the same for all variables.

Fig. 27 shows that wall pressure is the same for all variables. However, in Fig. 28 it can be seen that methods with increased artificial viscosity converge more slowly. In particular, conservation variables can not reduce the residual more than two orders of magnitude, whereas entropy variables present the best slope for convergence.

5.7. Driven cavity flow

The driven cavity flow is a classical problem to assess the performance of numerical methods for incompressible flows. Fig. 29 depicts the main flow features. Basically, the top boundary slides to the right,

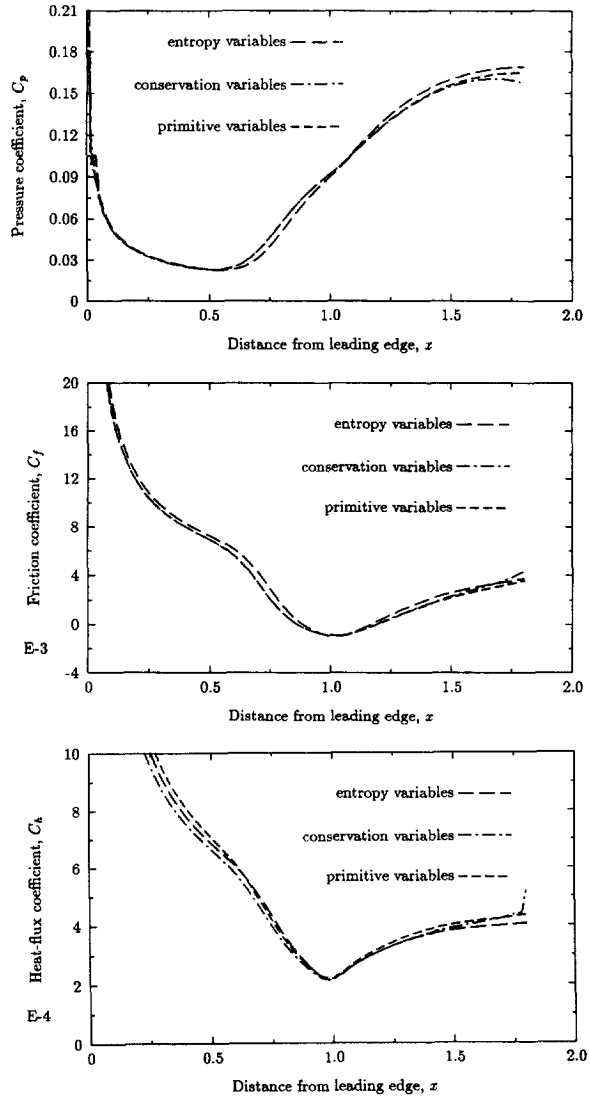


Fig. 21. Mach 3 compression corner. Wall quantities.

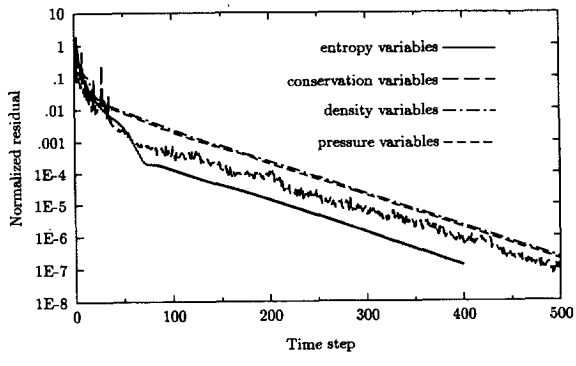


Fig. 22. Mach 3 compression corner. Residual convergence.

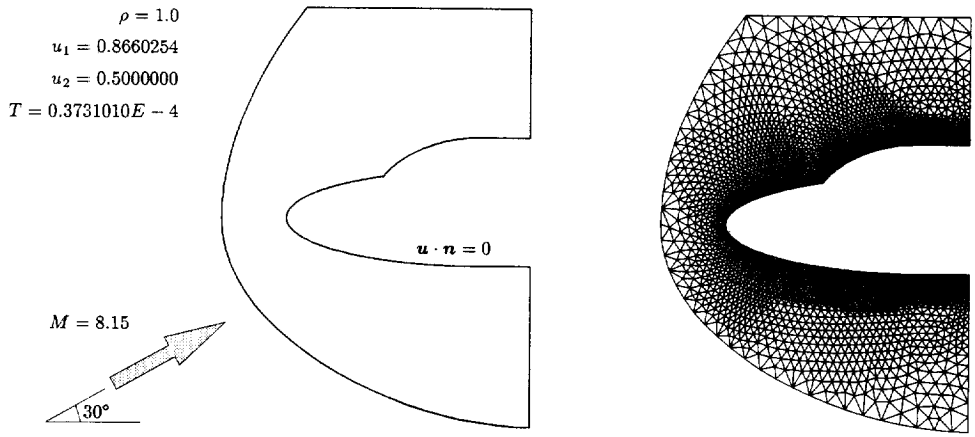


Fig. 23. Double ellipse problem.

Fig. 24. Mesh around the double ellipse (4448 elements).

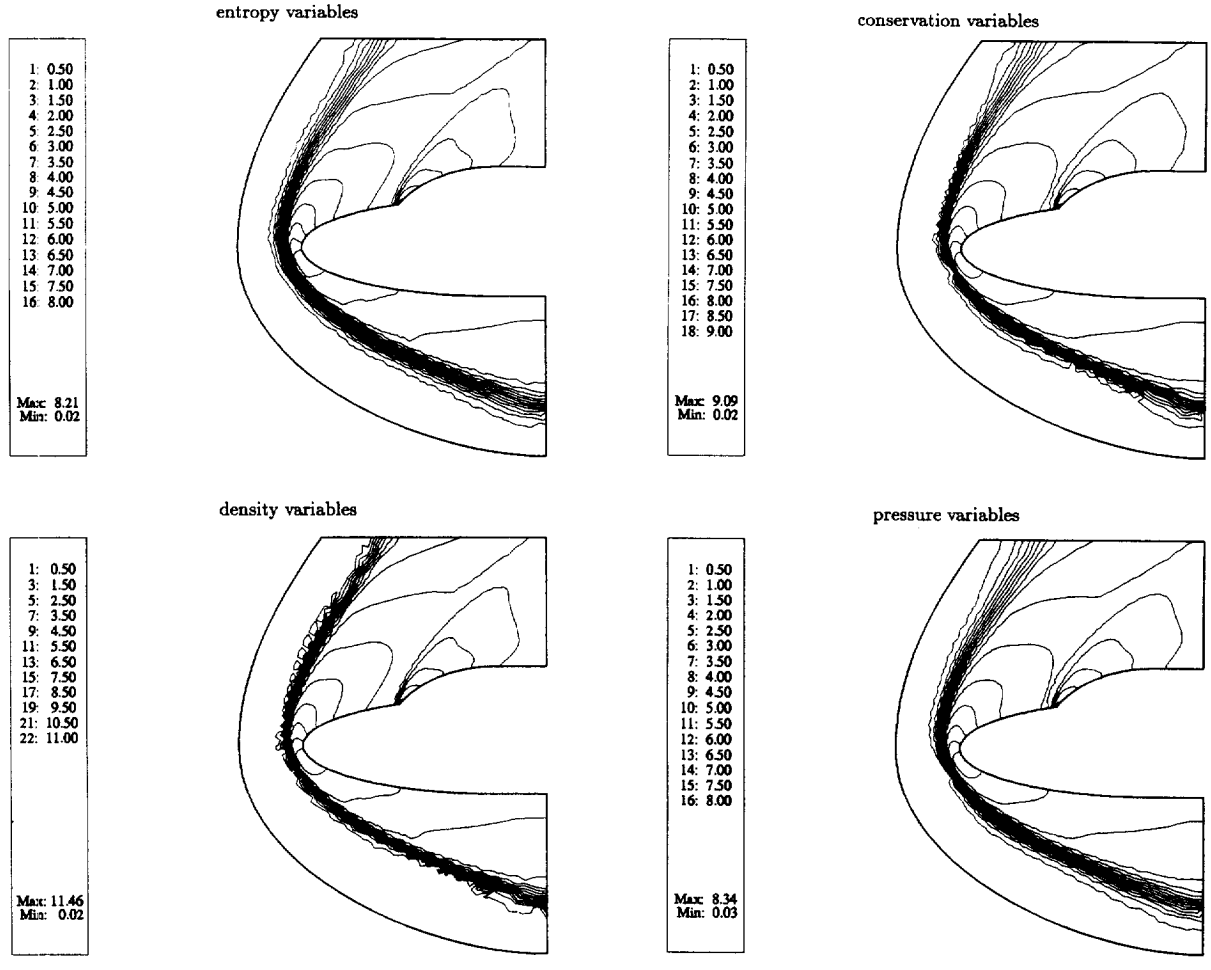
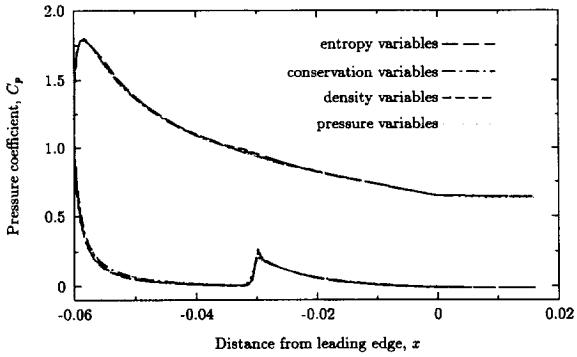


Fig. 25. Mach 8.15 double ellipse. Mach number contours.

Fig. 26. Mach 8.15 double ellipse. Mach number contours obtained with twice the discontinuity capturing operator.



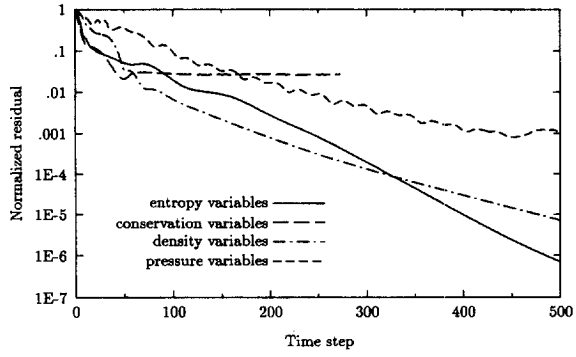


Fig. 27. Mach 8.15 double ellipse. Pressure coefficient.

Fig. 28. Mach 8.15 double ellipse. Residual convergence.

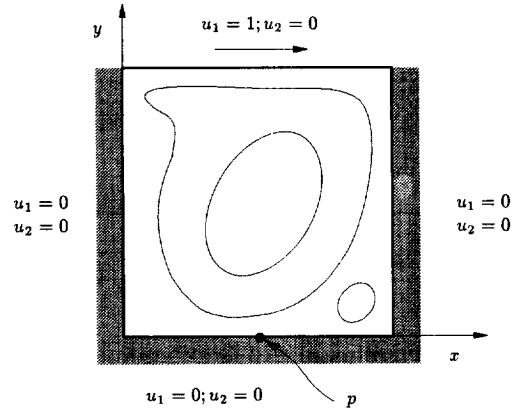


Fig. 29. Non-leaking cavity. Problem set up.

shearing the fluid within the domain and causing recirculation inside the cavity. A main vortex develops at the center of the cavity and, depending on the Reynolds number, secondary eddies may appear at the corners. The velocity field is discontinuous at both top corners, which are, therefore, singular points. The challenge of this problem is to control these singularities as well as accurately represent the smooth features of the flow. Fig. 29 also sketches the boundary conditions, which consist of zero velocity everywhere except along the top, where $u_1 = 1$ and $u_2 = 0$. These are supplemented with a reference temperature for the energy equation. Finally, the pressure space is constrained by fixing its value at one node.

In order to compare entropy and primitive variables, we considered first a Re = 1 incompressible isothermal flow, computed using a mesh of square elements with sides $h = 1/20$ and $h = 1/40$. The τ matrices employed in the calculations are (54) for primitive variables and (55) for entropy variables and their diagonal counterparts, namely, $\mathrm{diag}(\tau_{\mathrm{inc}})$ and $\mathrm{diag}(V_{,Y}\,\mathrm{diag}(\tau_{\mathrm{inc}}))$, respectively. Table 1 lists the values of the stream function and the vorticity at the center of the vortex for each case. Figs. 30 and 31 plot for the finest mesh the contours of streamlines, pressure and vorticity, the horizontal velocity component along the vertical centerline, and the vertical velocity component along the horizontal centerline. For this low Reynolds number case, both variable sets and both τ matrices perform similarly.

Consider now the Reynolds 400 incompressible isothermal flow. As shown in Fig. 32, the 40×40 mesh has been slightly refined at the top right and left corners to capture the singularity of the pressure field. Table 2 compares results. Observe that the diagonal τ matrix for entropy variables leads to very poor results. Clearly, the off diagonal components play a fundamental role in the accuracy. A good design of the τ matrix can significantly improve the accuracy of the entropy variables calculations. On the other hand, pressure variables are practically insensitive to the changes introduced in the off diagonals components. This indicates that there is flexibility with regard to off diagonal terms. In general, the new versions of the τ matrix, i.e. (54) and (55), give

Table 1
Non-leaking cavity flow, Re = 1. Value of the streamfunction, Ψ_c, and vorticity, ω_c, at the center of the main vortex

Variables	$h = 1/20$		$h = 1/40$	
	Ψ_c	ω_c	Ψ_c	ω_c
Primitive	−0.100388	−3.03788	−0.100092	−3.34077
Entropy	−0.100380	−3.03748	−0.100091	−3.34075
Diagonal τ matrix				
Primitive	−0.100388	−3.03789	−0.100092	−3.34078
Entropy	−0.100388	−3.03790	−0.100091	−3.34077
Benchmark[a]	−0.100	−3.232		

[a] Schreiber and Keller [17] $h = 1/120$.

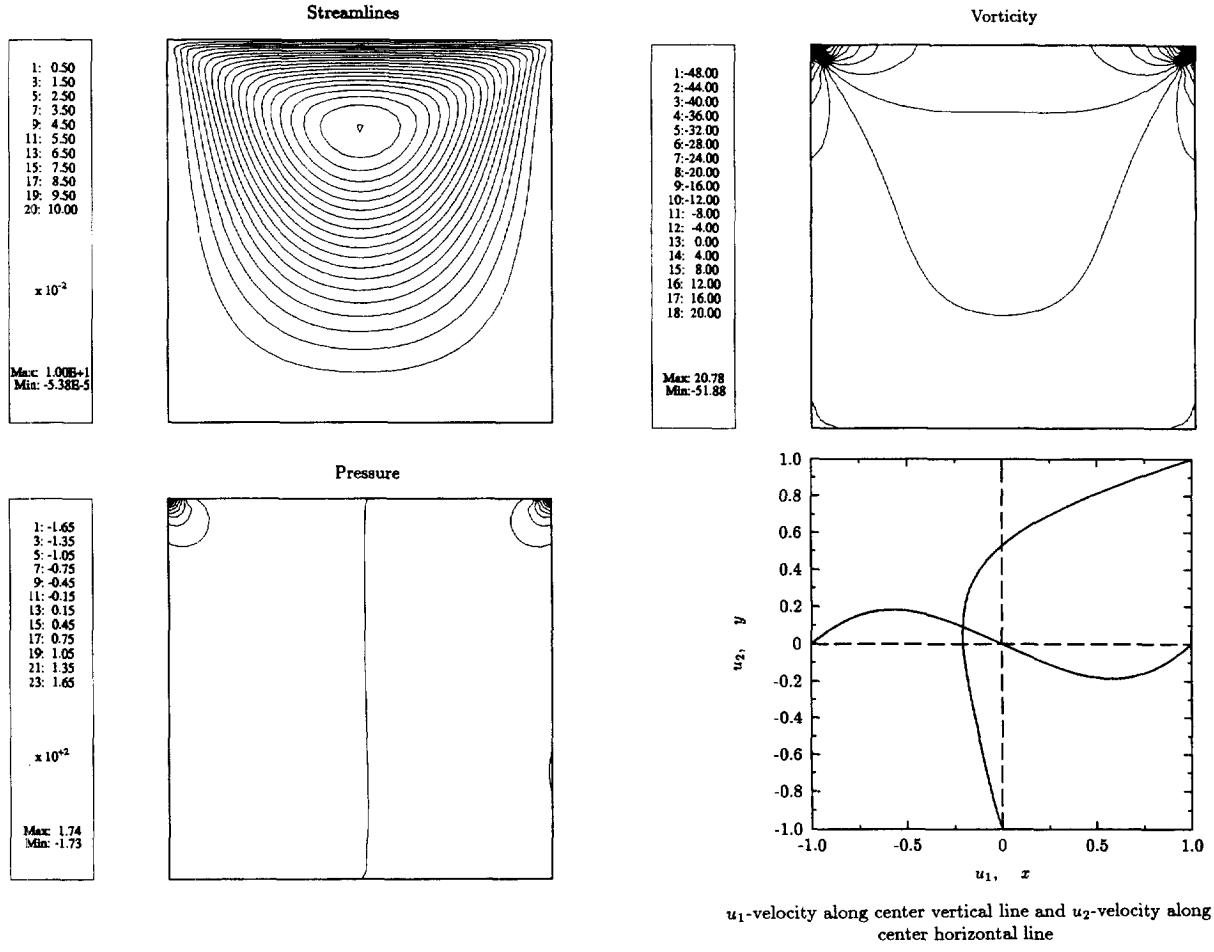
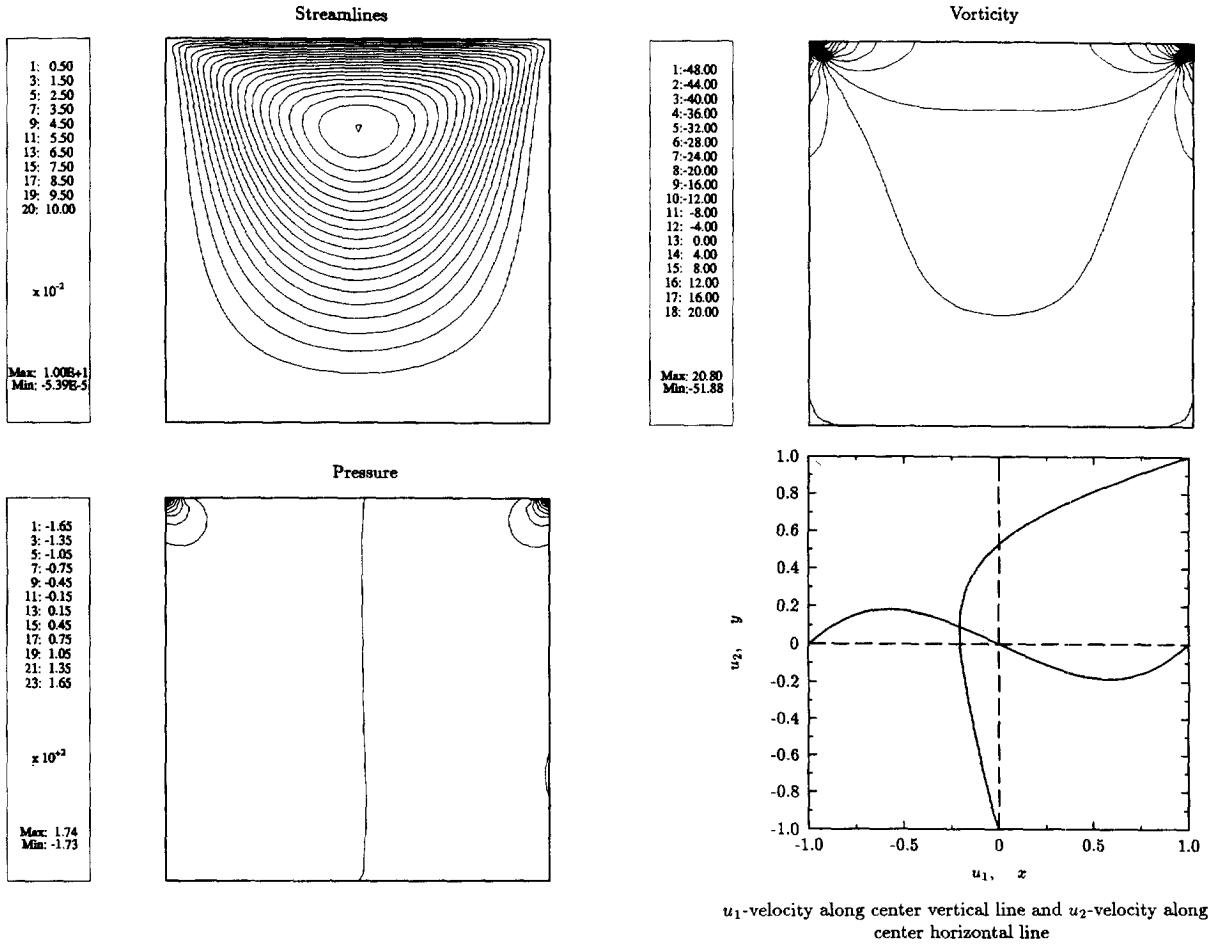


Fig. 30. Non-leaking cavity, Re = 1. Pressure variables.

good results. Streamlines, pressure, vorticity and velocity components along the vertical and horizontal center lines are displayed for pressure variables in Fig. 33 and for entropy variables in Fig. 34. Both solutions were obtained with the new versions of the τ matrix.

5.8. Circular cylinder, Re = 40

The flow around a stationary cylinder presents many varied physical characteristics depending on Reynolds number. For Reynolds numbers less than or equal to 40, the flow is steady and separates, causing a wake behind



u_1-velocity along center vertical line and u_2-velocity along center horizontal line

Fig. 31. Non-leaking cavity, Re = 1. Entropy variables.

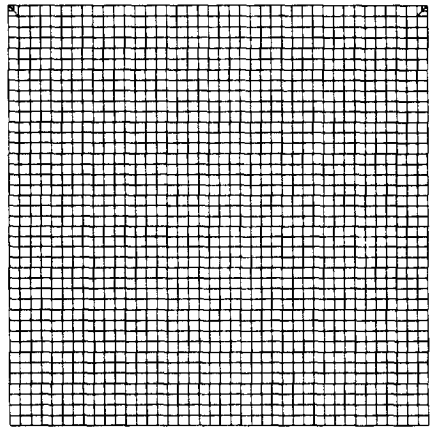


Fig. 32. Mesh for the Re = 400 non-leaking cavity, $h = 1/40$.

the cylinder to develop. The length of the wake and the point of separation, which depend on the Reynolds number, can be used to validate the method.

Geometry and boundary conditions are depicted in Fig. 35. The domain extends over $-4.5 \leqslant x \leqslant 15.5$, $-4.5 \leqslant y \leqslant 4.5$. The cylinder, of unit diameter, is placed at the origin. The boundary conditions consist of unit velocity and reference temperature along the inlet boundary; zero normal velocity component, zero tangential

Table 2
Non-leaking cavity flow, Re = 400. Value of the streamfunction, Ψ_c, and vorticity, ω_c, at the center of the main vortex

Variables	$h = 1/20$		$h = 1/40$	
	Ψ_c	ω_c	Ψ_c	ω_c
Primitive	−0.110467	−2.32711	−0.113802	−2.30023
Entropy	−0.115499	−2.53100	−0.115600	−2.33325
Diagonal τ matrix				
Primitive	−0.110893	−2.31357	−0.114054	−2.29780
Entropy	–	–	−0.059078	−1.55160
Benchmark[a]	−0.113909	−2.29469		

[a] Ghia et al. [15] $h = 1/256$.

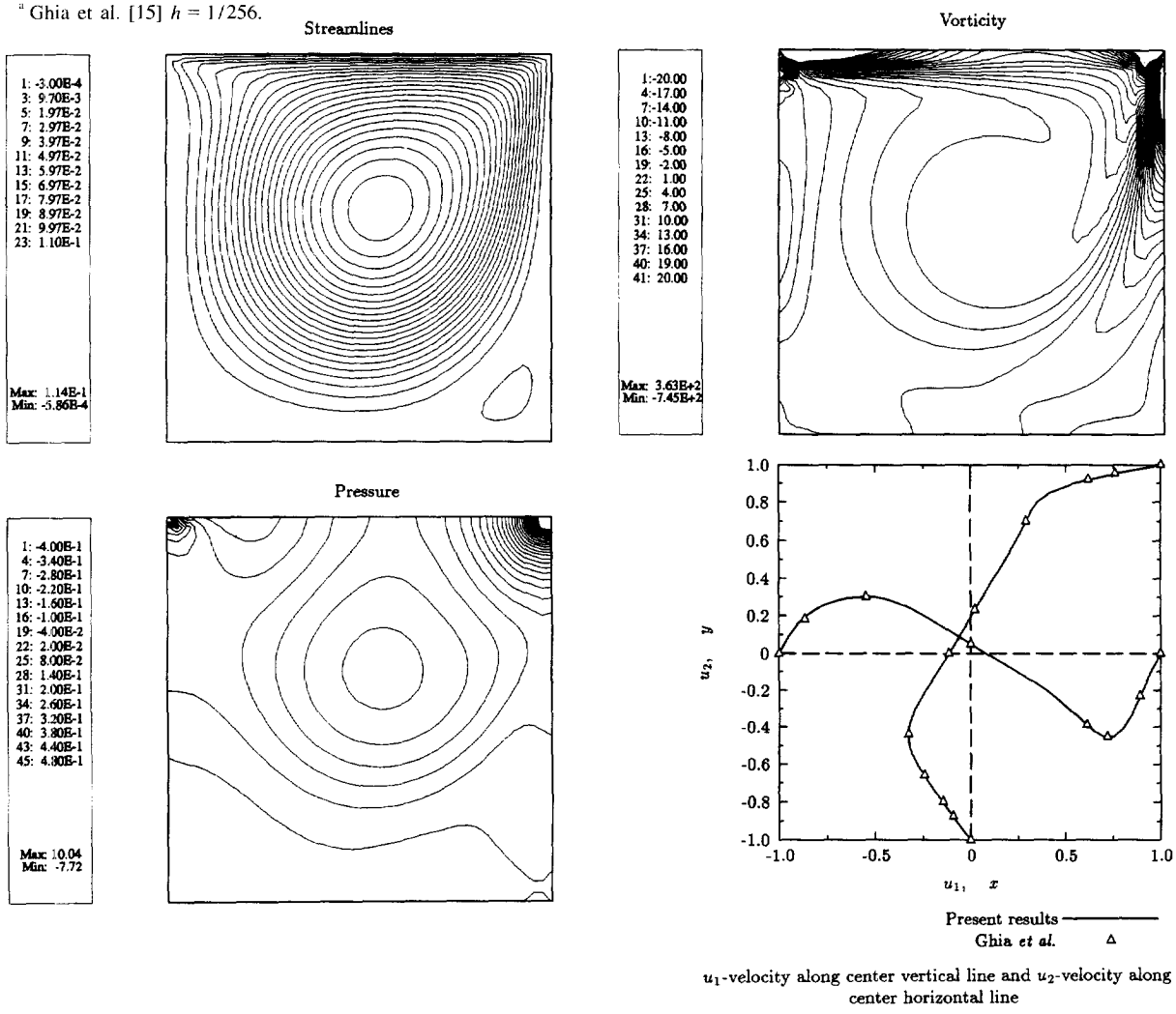
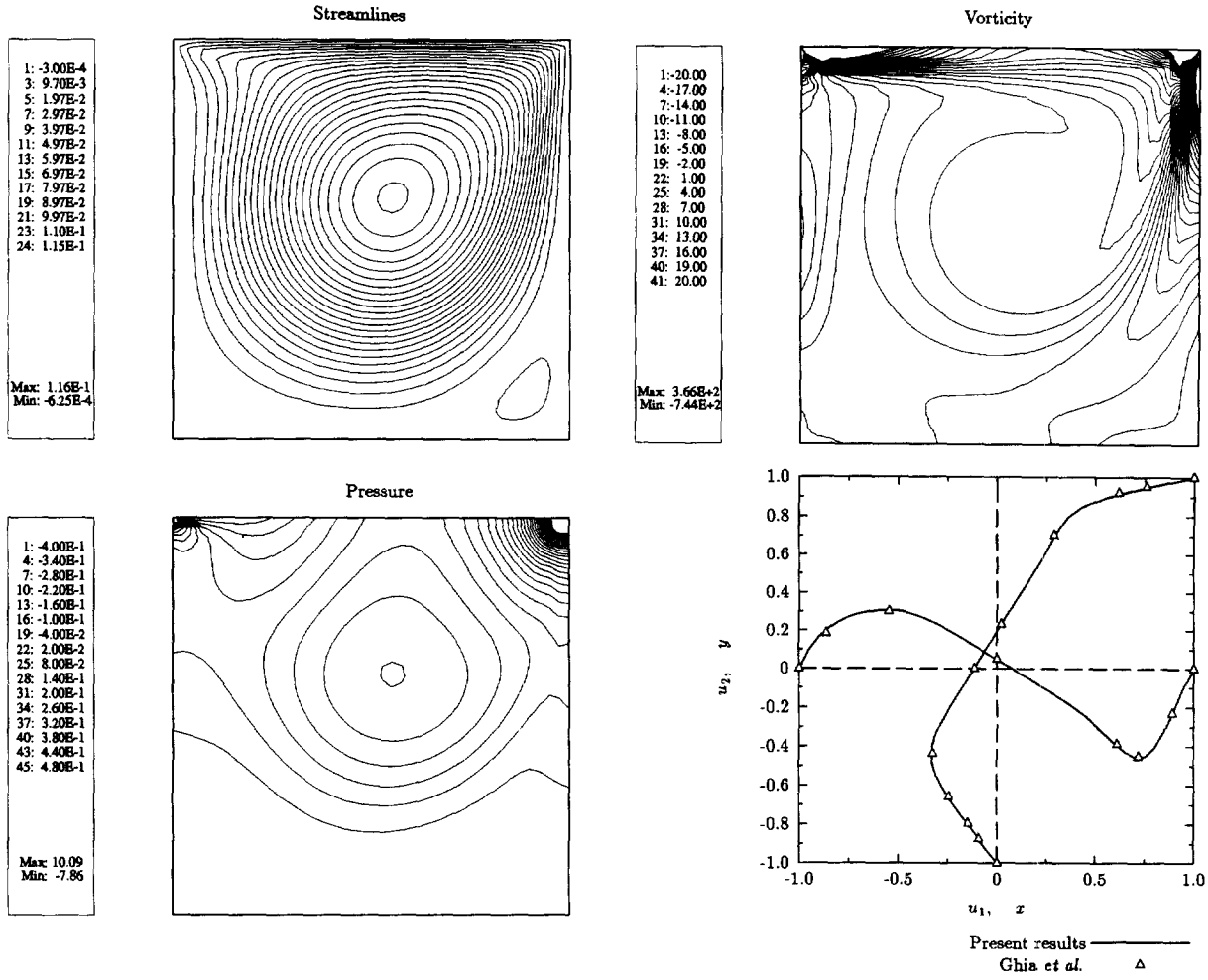


Fig. 33. Non-leaking cavity, Re = 400. Pressure variables.

viscous traction and zero heat flux along the horizontal boundaries; pressure together with zero traction and zero heat flux at the exit plane; and zero velocity components with zero heat flux at the wall of the cylinder.

The mesh, displayed in Fig. 36 along with a close-up of the cylinder region, consists of 4936 elements and 5063 nodes.

Figs. 37 and 39 show the pressure contours, stationary streamlines and temperature contours and Figs. 38 and 40 display a close up at the cylinder of the streamlines and the vorticity. The length of the wake and the



u_1-velocity along center vertical line and u_2-velocity along center horizontal line

Fig. 34. Non-leaking cavity, Re = 400. Entropy variables.

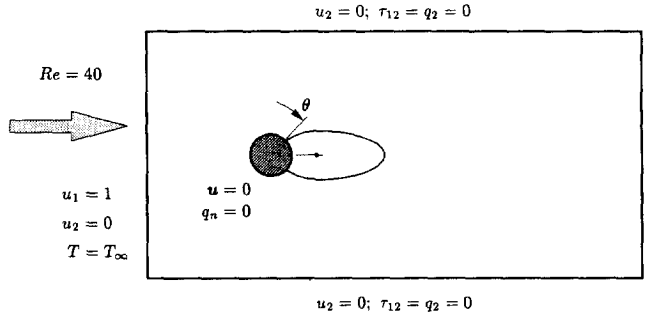


Fig. 35. Circular cylinder. Boundary conditions.

separation point obtained, presented in Table 3, correlate well with the experimental data of Coutanceau and Bouard [16]. Fig. 41 displays the friction coefficient obtained with the tangential component of the wall stress along the surface of the cylinder. Both sets of variables coincide exactly for low Reynolds number flows.

Again, there are difficulties in the convergence and stability of the solution for high Prandtl numbers with the iterative solvers.

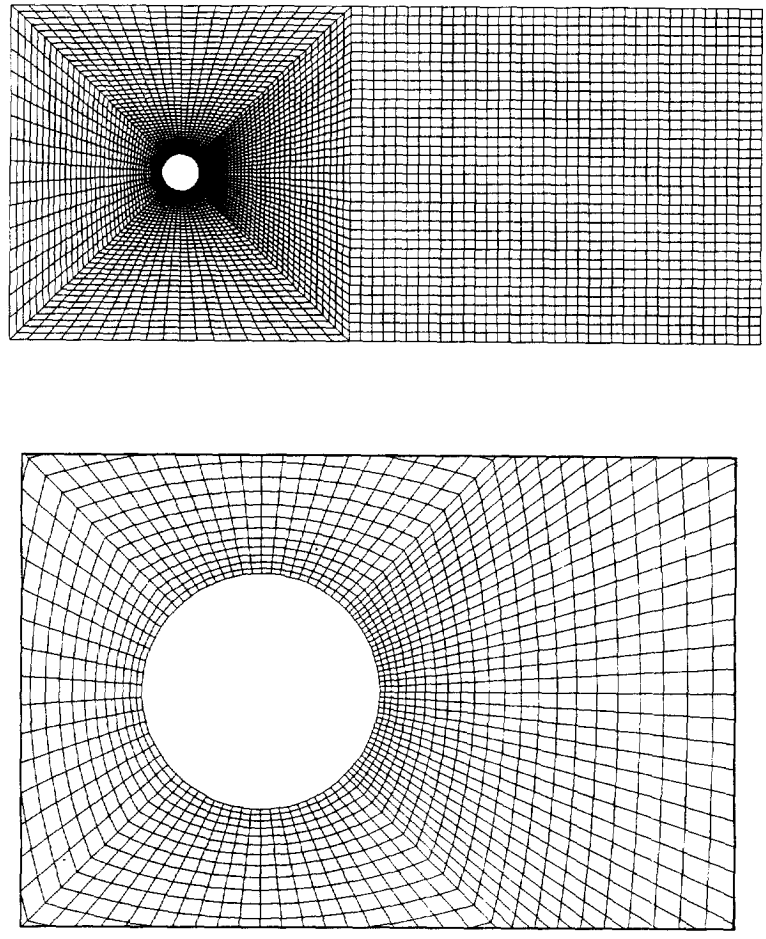


Fig. 36. Mesh for the circular cylinder (4936 elements, 5063 nodes).

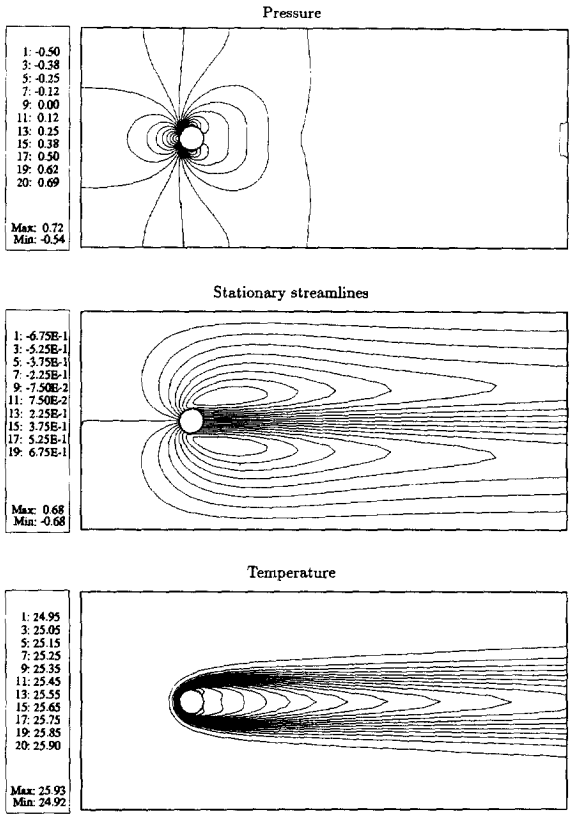
5.9. Circular cylinder, $Re = 100$

At Reynolds number 100, the solution around a stationary cylinder is periodic. Both temporal and spatial accuracy are of importance in capturing the features of this flow.

The same boundary conditions as the Reynolds 40 case are implemented. The flow is started from a uniform velocity field and constant pressure. The constant element in time, being first-order in time, is neither the most accurate nor efficient element to compute transients. Its damping ratio and phase error are very similar to those of the backward Euler scheme. In order to speed up the computations, 1-point quadrature was employed for the bilinear quadrilateral element. Reduced integration did not alter the period of the solution. For the same reason, no corrector passes were performed, that is, only one iteration was performed in each time step. For this flow, the Krylov space was set to 20 vectors and the maximum number of GMRES cycles was 5. These were only needed until $t = 4.0$. Afterwards, only 15 iterations per time step were performed.

To compute transients, it is necessary to solve the linear system of equations accurately. Since we are using an iterative solver, the tolerance of the GMRES algorithm has an important impact on the transient phase and the final period of the solution. This effect can be analyzed in Table 4 for pressure variables, where it is shown that the tolerance must be at least 0.001 to attain reasonable resolution.

Also, a study of convergence with respect to the time step, Δt, was conducted for pressure variables. For the constant in time element and the considered spatial resolution, $\Delta t = 0.1$ is not a small enough time step. Furthermore, obtaining a converged period requires smaller time steps than 0.025. For this time step, the



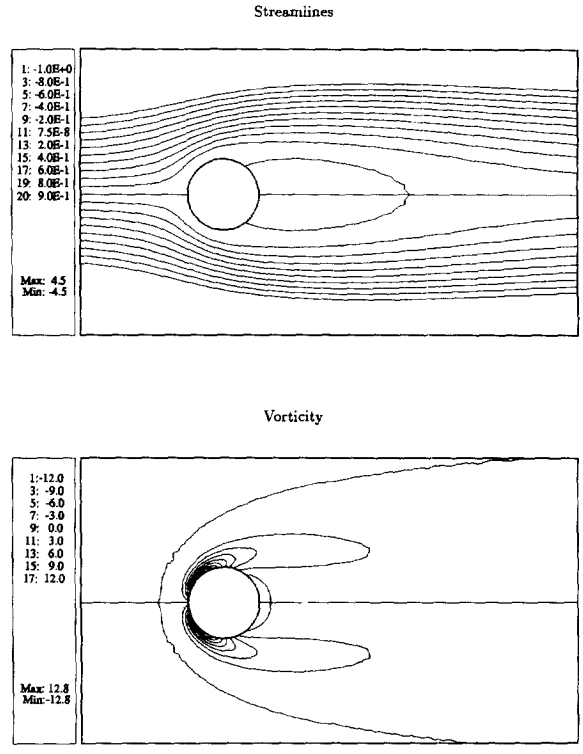


Fig. 37. Circular cylinder, Re = 40. Pressure variables.

Fig. 38. Circular cylinder, Re = 40. Pressure variables.

Table 3
Circular cylinder, Re = 40. Wake length and separation point

	Wake length	Separation point, θ
Experiment[a]	2.13	53.5°
Pressure	2.10	54.0°
Entropy	2.10	54.0

[a] Coutanceau and Bouard [16].

Table 4
Circular cylinder, Re = 100. Pressure variables. Tolerance (ϵ) of the GMRES solver; lift period; and Strouhal number

Δt	ϵ	T (period)	St
0.025	0.001	5.825	0.172
	0.01	6.025	0.166
	0.1	6.55	0.153
0.05	0.001	6.0	0.167
	0.01	6.4	0.156
	0.1	7.6	0.132
0.1	0.0001	6.4	0.156
	0.001	6.5	0.154
	0.01	7.1	0.141
	0.1	9.3	0.108

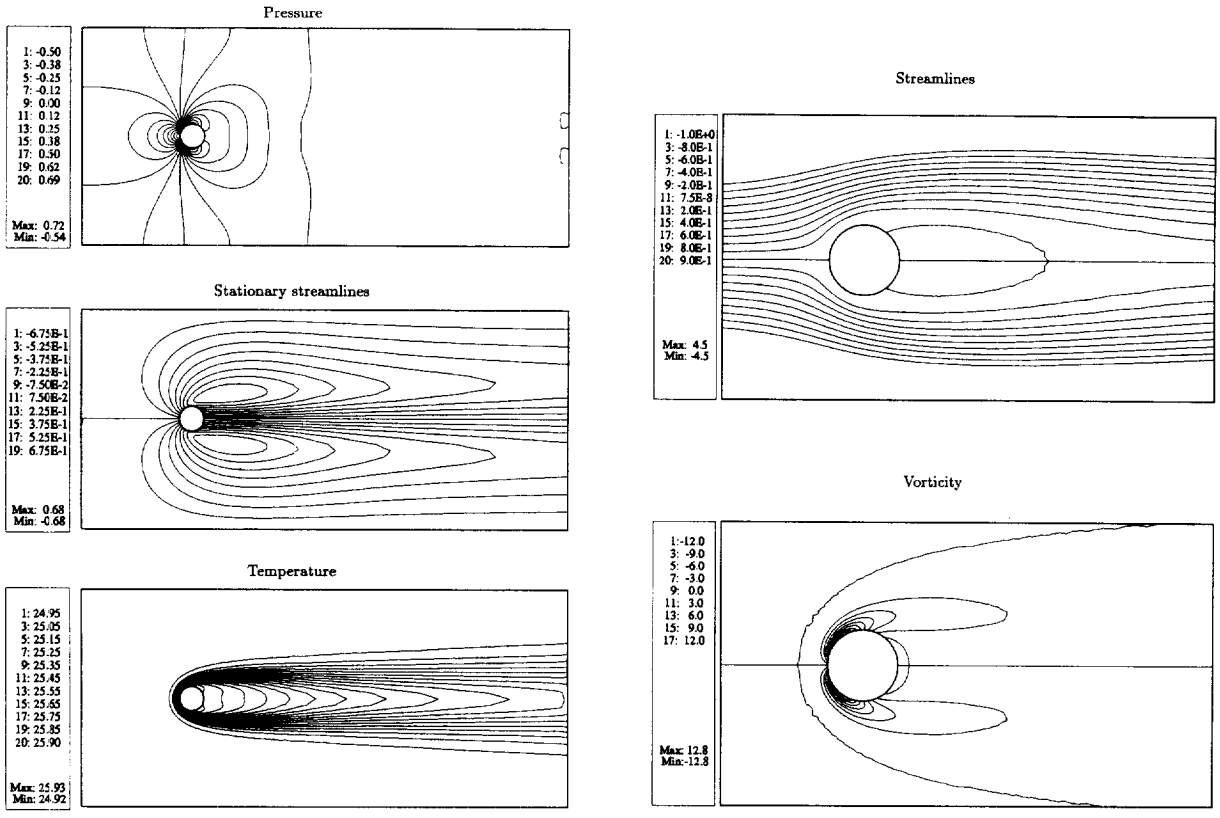


Fig. 39. Circular cylinder, Re = 40. Entropy variables.

Fig. 40. Circular cylinder, Re = 40. Entropy variables.

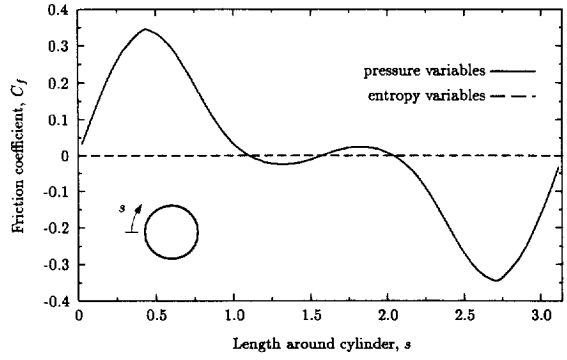


Fig. 41. Circular cylinder, Re = 40. Friction coefficient.

computed period was 5.825, equivalent to a Strouhal number of 0.172. Since the error is linear with Δt, the converged period can be extrapolated to 5.6. No further temporal refinement was attempted. Entropy variables give a much larger period. For the time step of 0.025 the computed period was 6.025, corresponding to a Strouhal number of 0.166.

The temporal evolution of the stationary streamlines can be followed in Figs. 42 and 45 for pressure variables and entropy variables, respectively. Once the flow becomes unsymmetric, the cylinder starts shedding very quickly. Figs. 43 and 46 display the lift coefficient, $C_L = F_y/(\frac{1}{2}\rho_\infty u_\infty^2 D)$, and the drag coefficient, $C_D = F_x/(\frac{1}{2}\rho_\infty u_\infty^2 D)$. The dimensionless coefficients have been expressed in terms of D, the cylinder diameter, and the

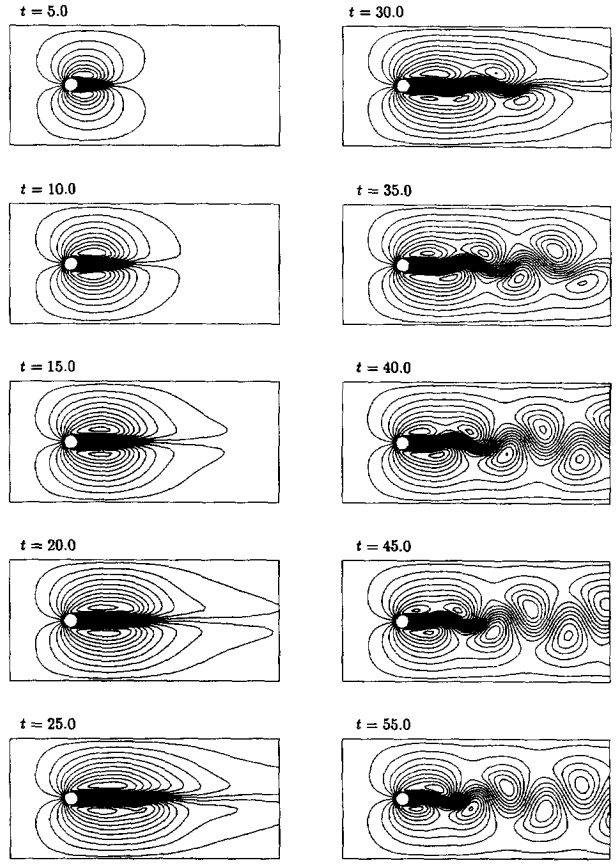


Fig. 42. Circular cylinder, Re = 100. Pressure variables. Stationary streamlines.

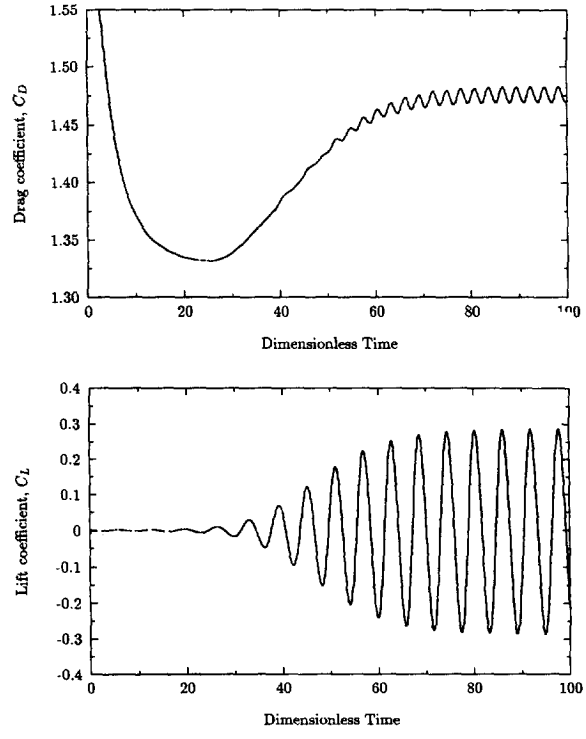


Fig. 43. Circular cylinder, Re = 100. Pressure variables. Evolution of the drag and lift coefficients.

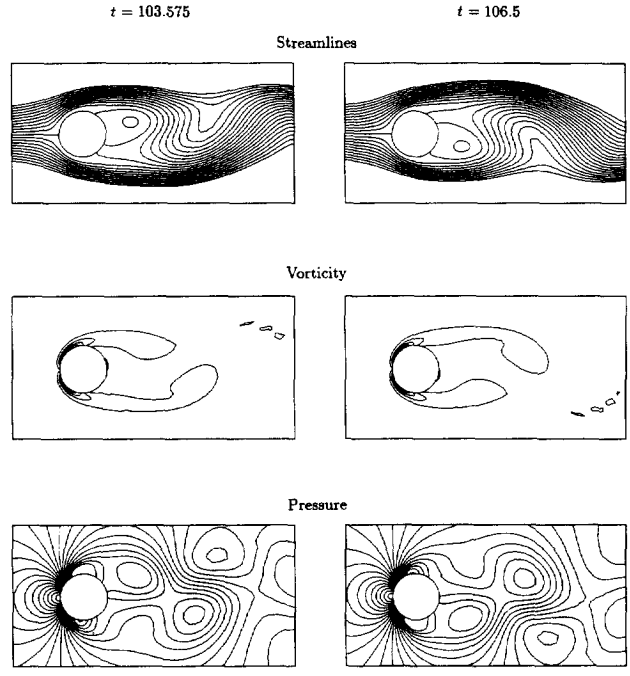


Fig. 44. Circular cylinder, Re = 100. Pressure variables. Solutions half a period apart.

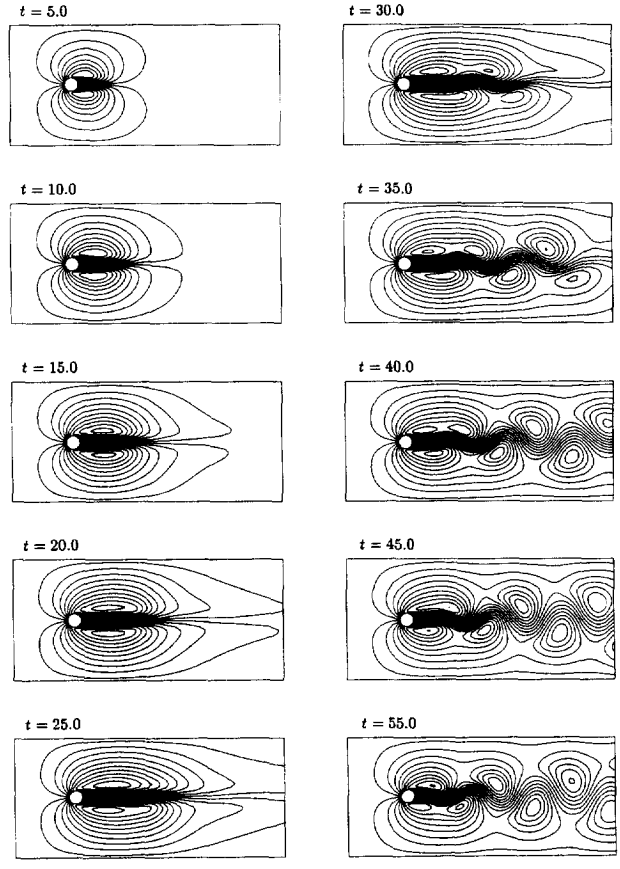


Fig. 45. Circular cylinder, Re = 100. Entropy variables. Stationary streamlines.

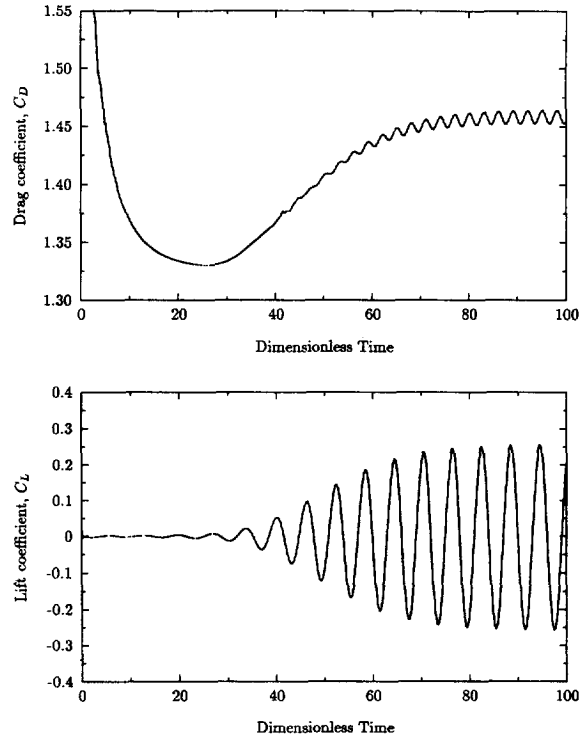


Fig. 46. Circular cylinder, Re = 100. Entropy variables. Evolution of the drag and lift coefficients.

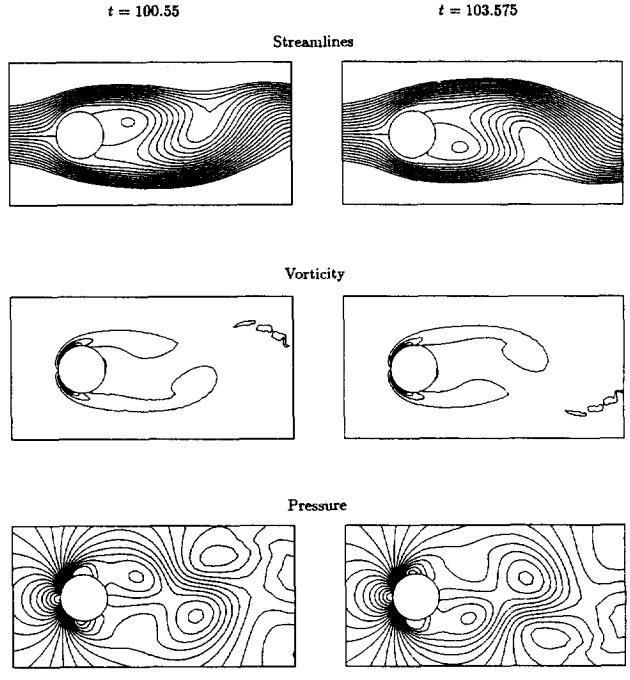


Fig. 47. Circular cylinder, Re = 100. Entropy variables, Solutions half a period apart.

total force acting on the cylinder, F_x, F_y. A close up at the cylinder is shown in Figs. 44 and 47 at instants of maximum and minimum lift. The effect of the reduced integration can be observed in the vorticity field, which has some minor oscillations. This effect is larger for entropy variables.

6. Conclusion

We have presented a Galerkin/lest squares method to solve the compressible Navier–Stokes equations which attains global conservation and correct shock structure for any choice of variables. Only the choice of entropy variables, however, satisfies exactly the discrete Clausius–Duhem inequality, whether or not additional dissipative mechanisms are present. Other variable choices require the presence of dissipative mechanisms, namely, the least squares term and the discontinuity capturing operator, to enforce entropy production. The incompressible limit is well behaved for entropy variables and the primitive variables (p, $\boldsymbol{u}$, T). In these cases, the limit of the formulation leads to *conservative* incompressible methods. This approach is stable for any combination of continuous interpolations, in particular, equal-order interpolations.

A comparison among the different variable sets led to the following conclusions.

(1) Entropy variables are the most robust in the presence of singularities. They are the least accurate for low speed channel flows, wall pressure and skin friction, while they are the most accurate for wall heat flux.
(2) Pressure variables are the most accurate on average for wall quantities and channel flows, but are not as robust as entropy variables in the presence of singularities and strong shocks. They are also the most convenient for boundary condition specification.
(3) Density variables are very similar to pressure variables, although they are not as accurate for low speed channel flows. However, they are more robust in the presence of singularities but always less robust than entropy variables.
(4) Conservation variables attain about the same level of accuracy as primitive variables for wall pressure and skin friction, although they are the least accurate for wall heat flux. They are also inaccurate for low speed channel flows. The specification of temperature boundary conditions is much more complex for conservation variables than for primitive or entropy variables. Errors in regions of outflow boundary layers require special treatment.
(5) The differences in accuracy are noticeable for coarse grids only. With sufficient refinement all the variables achieve similar accuracy.
(6) Entropy variables and pressure variables seems to possess the most attributes for practical problem solving.

Regarding incompressible flows, pressure variables are more accurate than entropy variables for high Reynolds number flows. However, a better design of the $\boldsymbol{\tau}$ matrix could further improve the accuracy of entropy variables. Pressure variables are also less sensitive than entropy variables to under integration.

Iterative techniques for incompressible flows need improvement. Until more robust solvers are developed, the unified approach will be at a disadvantage. Further research needs to be performed to define a $\boldsymbol{\tau}$ matrix and discontinuity capturing operator suitable for both compressible and incompressible flows.

Acknowledgments

We want to thank Ken Jansen and Jim Stewart for helpful comments, and especially to Roger Pierre for suggesting the decomposition of the incompressible $\boldsymbol{\tau}$ matrix.

Appendix A. Coefficient matrices for the Navier–Stokes equations expressed in terms of primitive variables and conservation variables

It is well known that the Navier–Stokes equations can be written as

$$\boldsymbol{U}_{,t} + \boldsymbol{F}^{\mathrm{adv}}_{i,i} = \boldsymbol{F}^{\mathrm{diff}}_{i,i} + \boldsymbol{\mathcal{F}} \tag{A.1}$$

where in three dimensions,

$$\boldsymbol{U} = \begin{Bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \end{Bmatrix} = \rho \begin{Bmatrix} 1 \\ u_1 \\ u_2 \\ u_3 \\ e^{\text{tot}} \end{Bmatrix} \quad \text{(conservation variables)} \tag{A.2}$$

$$\boldsymbol{F}_i^{\text{adv}} = \rho u_i \begin{Bmatrix} 1 \\ u_1 \\ u_2 \\ u_3 \\ e^{\text{tot}} \end{Bmatrix} + p \begin{Bmatrix} 0 \\ \delta_{1i} \\ \delta_{2i} \\ \delta_{3i} \\ u_i \end{Bmatrix} \quad \text{(Euler flux)} \tag{A.3}$$

$$\boldsymbol{F}_i^{\text{diff}} = \begin{Bmatrix} 0 \\ \tau_{1i} \\ \tau_{2i} \\ \tau_{3i} \\ \tau_{ij} u_j \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -q_i \end{Bmatrix} \quad \text{(diffusive flux)} \tag{A.4}$$

$$\boldsymbol{\mathcal{F}} = \rho \begin{Bmatrix} 0 \\ b_1 \\ b_2 \\ b_3 \\ b_i u_i + r \end{Bmatrix} \quad \text{(source vector)} \tag{A.5}$$

The above vectors are written in terms of density ρ, Cartesian velocity components $\boldsymbol{u} = \{u_1, u_2, u_3\}^{\mathrm{T}}$, total energy per unit mass $e^{\text{tot}} = e + |\boldsymbol{u}|^2/2$, where e is the internal energy per unit mass, thermodynamic pressure p, the viscous-stress tensor $\boldsymbol{\tau} = [\tau_{ij}]$, the heat-flux vector $\boldsymbol{q} = \{q_1, q_2, q_3\}^{\mathrm{T}}$, the body force vector per unit mass $\boldsymbol{b} = \{b_1, b_2, b_3\}^{\mathrm{T}}$, and the heat source per unit mass r. δ_{ij} is the Kronecker delta (i.e. $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$), a comma represents partial differentiation and the summation convention is used throughout.

Also, we are going to consider as constitutive relations a linear deformation law for the stress tensor and the Fourier's law for heat conduction, that is,

$$\tau_{ij} = \mu(u_{i,j} + u_{j,i}) + \lambda u_{k,k} \delta_{ij} \tag{A.6}$$

$$q_i = -\kappa T_{,i} \tag{A.7}$$

For convenience, we define the constant

$$\chi = \lambda + 2\mu \tag{A.8}$$

In these appendices, we present the Jacobians and matrices involved in the above computations for several sets of variables. For this purpose, given the equation of state $\tilde{\mu}(p, T)$, all the necessary parameters can be derived from the following thermodynamic relations,

$$s = -\frac{\partial \tilde{\mu}}{\partial T}, \qquad v = \frac{\partial \tilde{\mu}}{\partial p} \tag{A.9}$$

$$h = \tilde{\mu} + Ts, \qquad e = h - pv \tag{A.10}$$

$$\alpha_p = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_p = \frac{1}{v}\frac{\partial^2 \tilde{\mu}}{\partial p\, \partial T}, \qquad \beta_T = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T = -\frac{1}{v}\frac{\partial^2 \tilde{\mu}}{\partial p^2} \tag{A.11}$$

$$c_p = \left(\frac{\partial h}{\partial T}\right)_p = -T\frac{\partial^2 \tilde{\mu}}{\partial T^2}, \qquad c_v = c_p - \frac{\alpha_p^2 vT}{\beta_T} \tag{A.12}$$

Then, it can be shown that

$$\left(\frac{\partial p}{\partial \rho}\right)_T = \frac{1}{\rho \beta_T}, \qquad \left(\frac{\partial p}{\partial T}\right)_\rho = \frac{\alpha_p}{\beta_T} \tag{A.13}$$

$$\left(\frac{\partial e}{\partial \rho}\right)_T = -\frac{1}{\rho^2}\left(T\frac{\alpha_p}{\beta_T} - p\right), \qquad \left(\frac{\partial e}{\partial T}\right)_\rho = c_v \tag{A.14}$$

$$\left(\frac{\partial \rho}{\partial p}\right)_T = \rho \beta_T, \qquad \left(\frac{\partial \rho}{\partial T}\right)_p = -\rho \alpha_p \tag{A.15}$$

$$\left(\frac{\partial e}{\partial p}\right)_T = \frac{\beta_T p - \alpha_p T}{\rho} \qquad \left(\frac{\partial e}{\partial T}\right)_p = c_p - \frac{p\alpha_p}{\rho} \tag{A.16}$$

The following constants have been used to write the matrices

$$e_1^\rho = \rho\left(\frac{\partial e}{\partial \rho}\right)_T + e^{\text{tot}}, \qquad e_2^\rho = e_1^\rho + \left(\frac{\partial p}{\partial \rho}\right)_T \tag{A.17}$$

$$e_3^\rho = \rho e^{\text{tot}} + p, \qquad e_4^\rho = \rho\left(\frac{\partial e}{\partial T}\right)_\rho + \left(\frac{\partial p}{\partial T}\right)_\rho \tag{A.18}$$

$$e_1^p = \left(\frac{\partial \rho}{\partial p}\right)_T e^{\text{tot}} + \rho\left(\frac{\partial e}{\partial p}\right)_T, \qquad e_2^p = e_1^p + 1 \tag{A.19}$$

$$e_3^p = \rho e^{\text{tot}} + p, \qquad e_4^p = \left(\frac{\partial \rho}{\partial T}\right)_p e^{\text{tot}} + \rho\left(\frac{\partial e}{\partial T}\right)_p \tag{A.20}$$

$$s_1 = \frac{1}{\rho T}\left(\frac{\partial p}{\partial \rho}\right)_T, \qquad s_2 = \frac{\frac{T}{\rho}\left(\frac{\partial p}{\partial T}\right)_\rho - h}{T^2} + \frac{|\boldsymbol{u}|^2}{2T^2} \tag{A.21}$$

which after substituting the partial derivatives simplify to

$$e_1^\rho = e_1 - c_v T\bar{\gamma}, \qquad e_2^\rho = e_1^\rho + \frac{1}{\rho \beta_T} \tag{A.22}$$

$$e_3^\rho = \rho e_1, \qquad e_4^\rho = \rho c_v + \frac{\alpha_p}{\beta_T} \tag{A.23}$$

$$e_1^p = \rho \beta_T e_1 - \alpha_p T, \qquad e_2^p = e_1^p + 1 \tag{A.24}$$

$$e_3^p = \rho e_1, \qquad e_4^p = -\rho \alpha_p e_1 + \rho c_p \tag{A.25}$$

$$s_1 = \frac{1}{\rho^2 \beta_T T}, \qquad s_2 = \frac{1}{T^2}(d - \bar{e}_1) \tag{A.26}$$

where

$$v = \frac{1}{\rho}, \qquad d = \frac{v\alpha_p T}{\beta_T}, \qquad \bar{\gamma} = \frac{v\alpha_p}{\beta_T c_v},$$
$$k = \frac{|\boldsymbol{u}^2|}{2}, \qquad e_1 = h + k, \qquad \bar{e}_1 = h - k \tag{A.27}$$

A.1. Coefficient matrices for $(\rho, \boldsymbol{u}, T)$

Let the vector of primitive variables be

$$\boldsymbol{Y} = \begin{Bmatrix} \rho \\ u_1 \\ u_2 \\ u_3 \\ T \end{Bmatrix} \tag{A.28}$$

The matrix $\boldsymbol{A}_0 = \boldsymbol{U}_{,\boldsymbol{Y}}$ and its inverse $\boldsymbol{A}_0^{-1} = \boldsymbol{Y}_{,\boldsymbol{U}}$ can be written as

$$\boldsymbol{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ u_1 & \rho & 0 & 0 & 0 \\ u_2 & 0 & \rho & 0 & 0 \\ u_3 & 0 & 0 & \rho & 0 \\ e_1^\rho & \rho u_1 & \rho u_2 & \rho u_3 & \rho c_v \end{bmatrix} \tag{A.29}$$

and

$$\boldsymbol{A}_0^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\dfrac{u_1}{\rho} & \dfrac{1}{\rho} & 0 & 0 & 0 \\ -\dfrac{u_2}{\rho} & 0 & \dfrac{1}{\rho} & 0 & 0 \\ -\dfrac{u_3}{\rho} & 0 & 0 & \dfrac{1}{\rho} & 0 \\ \dfrac{|\boldsymbol{u}|^2 - e_1^\rho}{\rho c_v} & -\dfrac{u_1}{\rho c_v} & -\dfrac{u_2}{\rho c_v} & -\dfrac{u_3}{\rho c_v} & \dfrac{1}{\rho c_v} \end{bmatrix} \tag{A.30}$$

The Euler Jacobians with respect to $\boldsymbol{Y}$, $\boldsymbol{A}_i = \boldsymbol{F}_{i,\boldsymbol{Y}}^{\text{adv}}$, are given by

$$\boldsymbol{A}_1 = \begin{bmatrix} u_1 & \rho & 0 & 0 & 0 \\ u_1^2 + \dfrac{1}{\rho\beta_T} & 2\rho u_1 & 0 & 0 & \dfrac{\alpha_p}{\beta_T} \\ u_{12} & \rho u_2 & \rho u_1 & 0 & 0 \\ u_{31} & \rho u_3 & 0 & \rho u_1 & 0 \\ u_1 e_2^\rho & -e_3^\rho + \rho u_1^2 & \rho u_{12} & \rho u_{31} & u_1 e_4^\rho \end{bmatrix} \tag{A.31}$$

$$\boldsymbol{A}_2 = \begin{bmatrix} u_2 & 0 & \rho & 0 & 0 \\ u_{12} & \rho u_2 & \rho u_1 & 0 & 0 \\ u_2^2 + \dfrac{1}{\rho\beta_T} & 0 & 2\rho u_2 & 0 & \dfrac{\alpha_p}{\beta_T} \\ u_{23} & 0 & \rho u_3 & \rho u_2 & 0 \\ u_2 e_2^\rho & \rho u_{12} & e_3^\rho + \rho u_2^2 & \rho u_{23} & u_2 e_4^\rho \end{bmatrix} \tag{A.32}$$

$$\boldsymbol{A}_3 = \begin{bmatrix} u_3 & 0 & 0 & \rho & 0 \\ u_{31} & \rho u_3 & 0 & \rho u_1 & 0 \\ u_{23} & 0 & \rho u_3 & \rho u_2 & 0 \\ u_3^2 + \dfrac{1}{\rho\beta_T} & 0 & 0 & 2\rho u_3 & \dfrac{\alpha_p}{\beta_T} \\ u_3 e_2^\rho & \rho u_{31} & \rho u_{23} & e_3^\rho + \rho u_3^2 & u_3 e_4^\rho \end{bmatrix} \tag{A.33}$$

The diffusivity coefficient matrices $\boldsymbol{K}_{ij}$, where $\boldsymbol{K}_{ij}\boldsymbol{Y}_{,j} = \boldsymbol{F}_i^{\text{diff}}$, are

$$\boldsymbol{K}_{11} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \chi & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 \\ 0 & \chi u_1 & \mu u_2 & \mu u_3 & \kappa \end{bmatrix} \tag{A.34}$$

$$\boldsymbol{K}_{12}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & \lambda & 0 & 0\\ 0 & \mu & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & \mu u_2 & \lambda u_1 & 0 & 0\end{bmatrix} \tag{A.35}$$

$$\boldsymbol{K}_{13}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \lambda & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & \mu & 0 & 0 & 0\\ 0 & \mu u_3 & 0 & \lambda u_1 & 0\end{bmatrix} \tag{A.36}$$

$$\boldsymbol{K}_{21}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & \mu & 0 & 0\\ 0 & \lambda & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & \lambda u_2 & \mu u_1 & 0 & 0\end{bmatrix} \tag{A.37}$$

$$\boldsymbol{K}_{22}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & \mu & 0 & 0 & 0\\ 0 & 0 & \chi & 0 & 0\\ 0 & 0 & 0 & \mu & 0\\ 0 & \mu u_1 & \chi u_2 & \mu u_3 & \kappa\end{bmatrix} \tag{A.38}$$

$$\boldsymbol{K}_{23}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \lambda & 0\\ 0 & 0 & \mu & 0 & 0\\ 0 & 0 & \mu u_3 & \lambda u_2 & 0\end{bmatrix} \tag{A.39}$$

$$\boldsymbol{K}_{31}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \mu & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & \lambda & 0 & 0 & 0\\ 0 & \lambda u_3 & 0 & \mu u_1 & 0\end{bmatrix} \tag{A.40}$$

$$\boldsymbol{K}_{32}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \mu & 0\\ 0 & 0 & \lambda & 0 & 0\\ 0 & 0 & \lambda\mu_3 & \mu u_2 & 0\end{bmatrix} \tag{A.41}$$

$$\boldsymbol{K}_{33}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & \mu & 0 & 0 & 0\\ 0 & 0 & \mu & 0 & 0\\ 0 & 0 & 0 & \chi & 0\\ 0 & \mu u_1 & \mu u_2 & \chi u_3 & \kappa\end{bmatrix} \tag{A.42}$$

Also,

$$\boldsymbol{Y}_{,V}=\boldsymbol{V}_{,Y}^{-1}=\begin{bmatrix}\dfrac{1}{s_1} & \dfrac{u_1}{s_1} & \dfrac{u_2}{s_1} & \dfrac{u_3}{s_1} & \dfrac{(|\boldsymbol{u}|^2-T^2s_2)}{s_1}\\ 0 & T & 0 & 0 & Tu_1\\ 0 & 0 & T & 0 & Tu_2\\ 0 & 0 & 0 & T & Tu_3\\ 0 & 0 & 0 & 0 & T^2\end{bmatrix} \tag{A.43}$$

$$\boldsymbol{A}_0^{\mathrm{DC}} = \boldsymbol{V}_{,Y}^{\mathrm{T}} \boldsymbol{A}_0 = \begin{bmatrix} s_1 & 0 & 0 & 0 & 0 \\ 0 & \dfrac{\rho}{T} & 0 & 0 & 0 \\ 0 & 0 & \dfrac{\rho}{T} & 0 & 0 \\ 0 & 0 & 0 & \dfrac{\rho}{T} & 0 \\ 0 & 0 & 0 & 0 & \dfrac{\rho c_v}{T^2} \end{bmatrix} \tag{A.44}$$

A.2. Coefficient matrices for $(p, \boldsymbol{u}, T)$

Let the vector of primitive variables be

$$\boldsymbol{Y} = \begin{Bmatrix} p \\ u_1 \\ u_2 \\ u_3 \\ T \end{Bmatrix} \tag{A.45}$$

The matrix $\boldsymbol{A}_0 = \boldsymbol{U}_{,Y}$ and its inverse $\boldsymbol{A}_0^{-1} = \boldsymbol{Y}_{,U}$ can be written as

$$\boldsymbol{A}_0 = \begin{bmatrix} \rho\beta_T & 0 & 0 & 0 & -\rho\alpha_p \\ \rho\beta_T u_1 & \rho & 0 & 0 & -\rho\alpha_p u_1 \\ \rho\beta_T u_2 & 0 & \rho & 0 & -\rho\alpha_p u_2 \\ \rho\beta_T u_3 & 0 & 0 & \rho & -\rho\alpha_p u_3 \\ e_1^p & \rho u_1 & \rho u_2 & \rho u_3 & e_4^p \end{bmatrix} \tag{A.46}$$

and

$$\boldsymbol{A}_0^{-1} = \frac{1}{\rho} \begin{bmatrix} \dfrac{(e_4^p + \rho\alpha_p |\boldsymbol{u}|^2)}{\rho\beta_T c_v} & -\dfrac{\alpha_p u_1}{\beta_T c_v} & -\dfrac{\alpha_p u_2}{\beta_T c_v} & -\dfrac{\alpha_p u_3}{\beta_T c_v} & \dfrac{\alpha_p}{\beta_T c_v} \\ -u_1 & 1 & 0 & 0 & 0 \\ -u_2 & 0 & 1 & 0 & 0 \\ -u_3 & 0 & 0 & 1 & 0 \\ \dfrac{\rho\beta_T |\boldsymbol{u}|^2 - e_1^p}{\rho\beta_T c_v} & -\dfrac{u_1}{c_v} & -\dfrac{u_2}{c_v} & -\dfrac{u_3}{c_v} & \dfrac{1}{c_v} \end{bmatrix} \tag{A.47}$$

The Euler Jacobians with respect to $\boldsymbol{Y}$, $\boldsymbol{A}_i = \boldsymbol{F}_{i,\boldsymbol{Y}}^{\mathrm{adv}}$, are given by

$$\boldsymbol{A}_1 = \begin{bmatrix} \rho\beta_T u_1 & \rho & 0 & 0 & -\rho\alpha_p u_1 \\ \rho\beta_T u_1^2 + 1 & 2\rho u_1 & 0 & 0 & -\rho\alpha_p u_1^2 \\ \rho\beta_T u_{12} & \rho u_2 & \rho u_1 & 0 & -\rho\alpha_p u_{12} \\ \rho\beta_T u_{31} & \rho u_3 & 0 & \rho u_1 & -\rho\alpha_p u_{31} \\ u_1 e_2^p & e_3^p + \rho u_1^2 & \rho u_{12} & \rho u_{31} & u_1 e_4^p \end{bmatrix} \tag{A.48}$$

$$\boldsymbol{A}_2 = \begin{bmatrix} \rho\beta_T u_2 & 0 & \rho & 0 & -\rho\alpha_p u_2 \\ \rho\beta_T u_{12} & \rho u_2 & \rho u_1 & 0 & -\rho\alpha_p u_{12} \\ \rho\beta_T u_2^2 + 1 & 0 & 2\rho u_2 & 0 & -\rho\alpha_p u_2^2 \\ \rho\beta_T u_{23} & 0 & \rho u_3 & \rho u_2 & -\rho\alpha_p u_{23} \\ u_2 e_2^p & \rho u_{12} & e_3^p + \rho u_2^2 & \rho u_{23} & u_2 e_4^p \end{bmatrix} \tag{A.49}$$

$$\boldsymbol{A}_3 = \begin{bmatrix} \rho\beta_T u_3 & 0 & 0 & \rho & -\rho\alpha_p u_3 \\ \rho\beta_T u_{31} & \rho u_3 & 0 & \rho u_1 & -\rho\alpha_p u_{31} \\ \rho\beta_T u_{23} & 0 & \rho u_3 & \rho u_2 & -\rho\alpha_p u_{23} \\ \rho\beta_T u_3^2 + 1 & 0 & 0 & 2\rho u_3 & -\rho\alpha_p u_3^2 \\ u_3 e_2^p & \rho u_{31} & \rho u_{23} & e_3^p + \rho u_3^2 & u_3 e_4^p \end{bmatrix} \tag{A.50}$$

The diffusivity coefficient matrices $\boldsymbol{K}_{ij}$, where $\boldsymbol{K}_{ij}\boldsymbol{Y}_{,j} = \boldsymbol{F}_i^{\text{diff}}$, are

$$\boldsymbol{K}_{11} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \chi & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 \\ 0 & \chi u_1 & \mu u_2 & \mu u_3 & \kappa \end{bmatrix} \tag{A.51}$$

$$\boldsymbol{K}_{12} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mu u_2 & \lambda u_1 & 0 & 0 \end{bmatrix} \tag{A.52}$$

$$\boldsymbol{K}_{13} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & \mu u_3 & 0 & \lambda u_1 & 0 \end{bmatrix} \tag{A.53}$$

$$\boldsymbol{K}_{21} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda u_2 & \mu u_1 & 0 & 0 \end{bmatrix} \tag{A.54}$$

$$\boldsymbol{K}_{22} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & \chi & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 \\ 0 & \mu u_1 & \chi u_2 & \mu u_3 & \kappa \end{bmatrix} \tag{A.55}$$

$$\boldsymbol{K}_{23} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & \mu u_3 & \lambda u_2 & 0 \end{bmatrix} \tag{A.56}$$

$$\boldsymbol{K}_{31} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & \lambda u_3 & 0 & \mu u_1 & 0 \end{bmatrix} \tag{A.57}$$

$$\boldsymbol{K}_{32} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda u_3 & \mu u_2 & 0 \end{bmatrix} \tag{A.58}$$

$$\boldsymbol{K}_{33}=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & \mu & 0 & 0 & 0\\ 0 & 0 & \mu & 0 & 0\\ 0 & 0 & 0 & \chi & 0\\ 0 & \mu u_1 & \mu u_2 & \chi u_3 & \kappa\end{bmatrix} \tag{A.59}$$

Also,

$$\boldsymbol{Y}_{,V}=\boldsymbol{V}_{,Y}^{-1}=\begin{bmatrix}\rho T & \rho T u_1 & \rho T u_2 & \rho T u_2 & \rho T\left(h+\frac{1}{2}|\boldsymbol{u}|^2\right)\\ 0 & T & 0 & 0 & T u_1\\ 0 & 0 & T & 0 & T u_2\\ 0 & 0 & 0 & T & T u_3\\ 0 & 0 & 0 & 0 & T^2\end{bmatrix} \tag{A.60}$$

and

$$\boldsymbol{A}_0^{\mathrm{DC}}=\boldsymbol{V}_{,Y}^{\mathrm{T}}\boldsymbol{A}_0=\begin{bmatrix}(\rho\beta_T)^2 s_1 & 0 & 0 & 0 & -\rho\alpha_p\rho\beta_T s_1\\ 0 & \frac{\rho}{T} & 0 & 0 & 0\\ 0 & 0 & \frac{\rho}{T} & 0 & 0\\ 0 & 0 & 0 & \frac{\rho}{T} & 0\\ -\rho\alpha_p\rho\beta_T s_1 & 0 & 0 & 0 & (-\rho\alpha_p)^2 s_1+\frac{\rho c_v}{T^2}\end{bmatrix} \tag{A.61}$$

A.3. Coefficient matrices for conservation variables

The matrices were computed with the help of a change of variables between primitive variables including density, $\boldsymbol{Y}=(\rho,\boldsymbol{u},T)$, and conservation variables

$$\boldsymbol{U}=\begin{Bmatrix}\rho\\ \rho u_1\\ \rho u_2\\ \rho u_3\\ \rho e^{\mathrm{tot}}\end{Bmatrix} \tag{A.62}$$

Then

$$\boldsymbol{A}_i=\boldsymbol{F}_{i,U}^{\mathrm{adv}}=\boldsymbol{F}_{i,Y}^{\mathrm{adv}}\boldsymbol{Y}_{,U}$$

and similarly for the diffusive matrices. The following constants were used when defining the coefficient matrices

$$e_1^c=\left(\frac{\partial T}{\partial\rho}\right)_{\rho\boldsymbol{u},e^{\mathrm{tot}}}=\frac{|\boldsymbol{u}|^2-e_1^\rho}{\rho c_v},\qquad e_2^c=\left(\frac{\partial p}{\partial\rho}\right)_T-e_3^c+e_1^c\left(\frac{\partial p}{\partial T}\right)_\rho \tag{A.63}$$

$$e_3^c=h+\frac{1}{2}|\boldsymbol{u}|^2,\qquad e_4^c=1+\frac{1}{\rho c_v}\left(\frac{\partial p}{\partial T}\right)_\rho \tag{A.64}$$

which after substitution simplify to

$$e_1^c=\frac{T\bar{\gamma}}{\rho}-\frac{\bar{e}_1}{\rho c_v},\qquad e_2^c=a^2-\bar{e}_1\bar{\gamma}-e_1 \tag{A.65}$$

$$e_3^c=e_1,\qquad e_4^c=\bar{\gamma}+1 \tag{A.66}$$

where

$$v=\frac{1}{\rho},\qquad a^2=\frac{vc_p}{c_v\beta_T},\qquad \bar{\gamma}=\frac{v\alpha_p}{\beta_T c_v}$$
$$k=\frac{|\boldsymbol{u}^2|}{2},\qquad e_1=h+k,\qquad \bar{e}_1=h-k \tag{A.67}$$

The metric tensor $\boldsymbol{A}_0=\boldsymbol{U}_{,Y}$ and its inverse are the identity matrix:

$$\boldsymbol{A}_0\text{–}\boldsymbol{A}_0^{-1}=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix} \tag{A.68}$$

The Euler Jacobians with respect to $\boldsymbol{U}$, $\boldsymbol{A}_i=\boldsymbol{F}_{i,Y}^{\text{adv}}$, are given by

$$\boldsymbol{A}_1=\begin{bmatrix}0&1&0&0&0\\ a^2-u_1^2-\bar{e}_1\bar{\gamma}&u_1(2-\bar{\gamma})&-u_2\bar{\gamma}&-u_3\bar{\gamma}&\bar{\gamma}\\ -u_{12}&u_2&u_1&0&0\\ -u_{31}&u_3&0&u_1&0\\ u_1e_2^c&e_3^c-u_1^2\bar{\gamma}&-u_{12}\bar{\gamma}&-u_{31}\bar{\gamma}&u_1e_4^c\end{bmatrix} \tag{A.69}$$

$$\boldsymbol{A}_2=\begin{bmatrix}0&0&1&0&0\\ -u_{12}&u_2&u_1&0&0\\ a^2-u_2^2-\bar{e}_1\bar{\gamma}&-u_1\bar{\gamma}&u_2(2-\bar{\gamma})&-u_3\bar{\gamma}&\bar{\gamma}\\ -u_{23}&0&u_3&u_2&0\\ u_2e_2^c&-u_{12}\bar{\gamma}&e_3^c-u_2^2\bar{\gamma}&-u_{23}\bar{\gamma}&u_2e_4^c\end{bmatrix} \tag{A.70}$$

$$\boldsymbol{A}_3=\begin{bmatrix}0&0&0&1&0\\ -u_{31}&u_3&0&u_1&0\\ -u_{23}&0&u_3&u_2&0\\ a^2-u_3^2-\bar{e}_1\bar{\gamma}&-u_1\bar{\gamma}&-u_2\bar{\gamma}&u_3(2-\bar{\gamma})&\bar{\gamma}\\ u_3e_2^c&-u_{31}\bar{\gamma}&-u_{23}\bar{\gamma}&e_3^c-u_3^2\bar{\gamma}&u_3e_4^c\end{bmatrix} \tag{A.71}$$

The diffusivity coefficient-matrices $\boldsymbol{K}_{ij}$, where $\boldsymbol{K}_{ij}\boldsymbol{U}_{,j}=\boldsymbol{F}_i^{\text{diff}}$, are

$$\boldsymbol{K}_{11}=\frac{1}{\rho}\begin{bmatrix}0&0&0&0&0\\ -\chi u_1&\chi&0&0&0\\ -\mu u_2&0&\mu&0&0\\ -\mu u_3&0&0&\mu&0\\ -\chi u_1^2-\mu(u_2^2+u_3^2)+\kappa\rho e_1^c&\left(\chi-\frac{\kappa}{c_v}\right)u_1&\left(\mu-\frac{\kappa}{c_v}\right)u_2&\left(\mu-\frac{\kappa}{c_v}\right)u_3&\frac{\kappa}{c_v}\end{bmatrix} \tag{A.72}$$

$$\boldsymbol{K}_{12}=\frac{1}{\rho}\begin{bmatrix}0&0&0&0&0\\ -\lambda u_2&0&\lambda&0&0\\ -\mu u_1&\mu&0&0&0\\ 0&0&0&0&0\\ -(\mu+\lambda)u_{12}&\mu u_2&\lambda u_1&0&0\end{bmatrix} \tag{A.73}$$

$$\boldsymbol{K}_{13}=\frac{1}{\rho}\begin{bmatrix}0&0&0&0&0\\ -\lambda u_3&0&0&\lambda&0\\ 0&0&0&0&0\\ -\mu u_1&\mu&0&0&0\\ -(\mu+\lambda)u_{31}&\mu u_3&0&\lambda u_1&0\end{bmatrix} \tag{A.74}$$

$$\boldsymbol{K}_{21}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -\mu u_2 & 0 & \mu & 0 & 0 \\ -\lambda u_1 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -(\mu+\lambda)u_{12} & \lambda u_2 & \mu u_1 & 0 & 0 \end{bmatrix} \tag{A.75}$$

$$\boldsymbol{K}_{22}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -\mu u_1 & \mu & 0 & 0 & 0 \\ -\chi u_2 & 0 & \chi & 0 & 0 \\ -\mu u_3 & 0 & 0 & \mu & 0 \\ -\chi u_2^2-\mu(u_1^2+u_3^2)+\kappa\rho e_1^c & \left(\mu-\frac{\kappa}{c_v}\right)u_1 & \left(\chi-\frac{\kappa}{c_v}\right)u_2 & \left(\mu-\frac{\kappa}{c_v}\right)u_3 & \frac{\kappa}{c_v} \end{bmatrix} \tag{A.76}$$

$$\boldsymbol{K}_{23}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\lambda u_3 & 0 & 0 & \lambda & 0 \\ -\mu u_2 & 0 & \mu & 0 & 0 \\ -(\mu+\lambda)u_{23} & 0 & \mu u_3 & \lambda u_2 & 0 \end{bmatrix} \tag{A.77}$$

$$\boldsymbol{K}_{31}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -\mu u_3 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\lambda u_1 & \lambda & 0 & 0 & 0 \\ -(\mu+\lambda)u_{31} & \lambda u_3 & 0 & \mu u_1 & 0 \end{bmatrix} \tag{A.78}$$

$$\boldsymbol{K}_{32}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\mu u_3 & 0 & 0 & \mu & 0 \\ -\lambda u_2 & 0 & \lambda & 0 & 0 \\ -(\mu+\lambda)u_{23} & 0 & \lambda u_3 & \mu u_2 & 0 \end{bmatrix} \tag{A.79}$$

$$\boldsymbol{K}_{33}=\frac{1}{\rho}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -\mu u_1 & \mu & 0 & 0 & 0 \\ -\mu u_2 & 0 & \mu & 0 & 0 \\ -\chi u_3 & 0 & 0 & \chi & 0 \\ -\chi u_3^2-\mu(u_1^2+u_2^2)+\kappa\rho e_1^c & \left(\mu-\frac{\kappa}{c_v}\right)u_1 & \left(\mu-\frac{\kappa}{c_v}\right)u_2 & \left(\chi-\frac{\kappa}{c_v}\right)u_3 & \frac{\kappa}{c_v} \end{bmatrix} \tag{A.80}$$

Also,

$$\boldsymbol{U}_{,V}=\tilde{\boldsymbol{A}}_0=\frac{\beta_T T}{v^2}\begin{bmatrix} 1 & u_1 & u_2 & u_3 & e_2 \\ & c_1 & u_{12} & u_{31} & u_1 e_3 \\ & & c_2 & u_{23} & u_2 e_3 \\ & \text{symm} & & c_3 & u_3 e_3 \\ & & & & e_5 \end{bmatrix} \tag{A.81}$$

$$\boldsymbol{A}_0^{\mathrm{DC}}=\boldsymbol{V}_{,U}^{\mathrm{T}}\boldsymbol{A}_0=\tilde{\boldsymbol{A}}_0^{-1}=\frac{v}{c_v T^2}\begin{bmatrix} \bar{e}_5 & u_1\bar{e}_3 & u_2\bar{e}_3 & u_3\bar{e}_3 & -\bar{e}_2 \\ & \bar{c}_1 & u_{12} & u_{31} & -u_1 \\ & & \bar{c}_2 & u_{23} & -u_2 \\ & \text{symm} & & \bar{c}_3 & -u_3 \\ & & & & 1 \end{bmatrix} \tag{A.82}$$

where

$$
\begin{aligned}
&k = \frac{|\boldsymbol{u}^2|}{2}, \qquad d = \frac{v\alpha_p T}{\beta_T}, \qquad \bar{\gamma} = \frac{v\alpha_p}{\beta_T c_v} \\
&c_1 = u_1^2 + \frac{v}{\beta_T} \\
&c_2 = u_2^2 + \frac{v}{\beta_T}, \qquad c_3 + u_3^2 + \frac{v}{\beta_T}, \qquad \bar{c}_1 = u_1^2 + c_v T \\
&\bar{c}_2 = u_2^2 + c_v T, \qquad \bar{c}_3 = u_3^2 + c_v T, \qquad e_1 = h + k \\
&e_2 = e_1 - d, \qquad e_3 = e_2 + \frac{v}{\beta_T}, \qquad e_4 = e_2 + 2\frac{v}{\beta_T} \\
&\bar{e}_1 = h - k, \qquad \bar{e}_2 = \bar{e}_1 - d, \qquad \bar{e}_3 = \bar{e}_2 - c_v T \\
&\bar{e}_4 = \bar{e}_2 - 2c_v T, \qquad u_{12} = u_1 u_2, \qquad u_{23} = u_2 u_3 \\
&u_{31} = u_3 u_1 \\
&e_5 = e_1^2 - 2e_1 d + \frac{v(2k + c_p T)}{\beta_T}, \qquad \bar{e}_5 = \bar{e}_1^2 - 2\bar{e}_1 d + 2kc_v T + \frac{vc_p T}{\beta_T}
\end{aligned}
\tag{A.83}
$$

References

[1] G. Hauke and T.J.R. Hughes, A unified approach to compressible and incompressible flows, Comput. Methods Appl. Mech. Engrg. 113 (1994) 389–396.

[2] L.P. Franca and S.L. Frey, Stabilized finite element methods: II. The incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Engrg. 89 (1992) 141–219.

[3] F. Shakib, T.J.R. Hughes and Z. Johan, A new finite element formulation for computational fluid dynamics: X. The compressible Euler and Navier–Stokes equations, Comput. Methods Appl. Mech. Engrg. 89 (1991) 141–219.

[4] F. Chalot and T.J.R. Hughes, A consistent equilibrium chemistry algorithm for hypersonic flows, Comput. Methods Appl. Mech. Engrg. 112 (1994) 25–40.

[5] T.J.R. Hughes, L.P. Franca and M. Mallet, A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier–Stokes equations and the second law of thermodynamics, Comput. Methods Appl. Mech. Engrg. 54 (1986) 223–234.

[6] F. Chalot, T.J.R. Hughes and F. Shakib, Symmetrization of conservation laws with entropy for high-temperature hypersonic computations, Comput. Syst. Engrg. 1 (1990) 495–521.

[7] T.J.R. Hughes and M. Mallet, A new finite element formulation for computational fluid dynamics: III. The generalized streamline operator for multidimensional advective–diffusive systems, Comput. Methods Appl. Mech. Engrg. 58 (1986) 305–328.

[8] T.J.R. Hughes, L.P. Franca and G. Hulbert, A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective–diffusive equations, Comput. Methods Appl. Mech. Engrg. 73 (1989) 173–189.

[9] A. Naïm, M. Mallet, P. Rostand and J.M. Hasholder, Local aerothermal problems during HERMES re-entry, AGARD Conference Proceedings 514, pp. 42-1 to 42-16, AGARD, 7 Rue Ancelle, 92200 Neuilly sur Seine, France, 1993.

[10] T.J.R. Hughes and M. Mallet, A new finite element formulation for computational fluid dynamics: IV. A discontinuity-capturing operator for multidimensional advective–diffusive systems, Comput. Methods Appl. Mech. Engrg. 58 (1986) 329–336.

[11] T.J.R. Hughes, The Finite Element Method: Linear Static and Dynamic Finite Element Analysis (Prentice-Hall, Englewood Cliffs, NJ, 1987).

[12] R.M. Ferencz, Element-by-element preconditioning techniques for large-scale, vectorized finite element analysis in nonlinear solid and structural mechanics, Ph.D. Thesis, Division of Applied Mechanics, Stanford University, 1989.

[13] G.J. Le Beau, S.E. Ray, S.K. Aliabadi and T.E. Tezduyar, SUPG finite element computation of compressible flows with the entropy and conservation variables formulations, Comput. Methods Appl. Mech. Engrg. 104 (1993) 397–422.

[14] S.K. Aliabadi, S.E. Ray and T.E. Tezduyar, SUPG finite element computation of viscous compressible flows based on the conservation and entropy variables formulation, University of Minnesota Supercomputer Institute Research Report UMSI 92/136, 1992.

[15] U. Ghia, K.N. Ghia and C.T. Shin, "High-Re solutions for incompressible flow using the Navier–Stokes equations and a multigrid method, J. Comput. Phys. 48 (1982) 387–441.

[16] M. Coutanceau and R. Bouard, Experimental determination of the viscous flow in a wake of a circular cylinder in uniform translation. Part I. Steady flow, J. Fluid Mech. 79 (1977) 231–256.

[17] R. Schreiber and H. Keller, Driven cavity flows by efficient numerical techniques, J. Comput. Phys. 49 (1983) 310–333.