COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING 84 (1990) 275–326
NORTH-HOLLAND

A NEW FINITE ELEMENT METHOD FOR SOLVING COMPRESSIBLE NAVIER–STOKES EQUATIONS BASED ON AN OPERATOR SPLITTING METHOD AND *h-p* ADAPTIVITY

L. DEMKOWICZ, J.T. ODEN
The Texas Institute for Computational Mechanics, The University of Texas at Austin, Austin, TX 78712, U.S.A.

W. RACHOWICZ
Computational Mechanics Company, Inc., 7701 North Lamar, Suite 200, Austin, TX 78705, U.S.A.

Received 14 February 1990

A new finite element method for solving compressible Navier–Stokes equations is proposed. The method is based on a version of Strang's operator splitting and an *h-p* adaptive finite element approximation in space. This paper contains the formulation of the method with a detailed discussion of boundary conditions, a sample adaptive strategy and numerical examples involving compressible viscous flow over a flat plate with Reynolds numbers Re = 1000 and Re = 10,000.

1. Introduction

1.1. *Introduction and motivation: Open problems in computational fluid dynamics*

Despite the significant and well-publicized advances in Computational Fluid Dynamics (CFD) over the past decade, several fundamental and, as yet, unresolved problems remain that threaten to severely limit the scope and applicability of numerical flow simulations. Among these are some notorious paradoxes:

1. *The paradox of scale and artificial diffusion.* To stabilize high order-accurate schemes and to eliminate spurious oscillations of numerical solutions near shocks, successful numerical methods generally possess a generous amount of artificial viscosity which is typically dependent on the mesh size. (Indeed, for hyperbolic systems such as the Euler equations, it is known that schemes without artificial viscosity may not converge to physically-meaningful solutions.) As is well-known, this unphysical dissipation mechanism annihilates the true viscous effects and makes impossible a numerical simulation of viscous boundary layers unless the mesh size used is much smaller than the boundary layer thickness. But then such fine-scale modeling of the boundary layers may require many hundreds of thousands (even millions) of grid points to resolve flow features of interest, with the result that the size of the computational problem taxes or exceeds the capacity of the largest supercomputers now available.

2. *The paradox of algebraic convergence.* Virtually every numerical scheme used in general flow simulations today, essentially all classical finite difference, finite volume and finite

element methods, particle methods, and others, exhibit (asymptotically) only so-called algebraic convergence. This means that the logarithm of the global computational error E varies linearly with the logarithm of the problem size N for N sufficiently large (see Fig. 1.1). Translated into practical terms, this means that no matter how large the capacity and speed of the computational device on hand, one can always define a level of resolution of the flowfield (a size of the total error) that cannot be attained. Stated in another way, the range of applicability of contemporary methods of computational fluid dynamics is limited; one can easily define levels of accuracy that cannot be attained by merely increasing the number and density of gridpoints. In certain cases, spectral methods can exhibit non-algebraic convergence rates, but, aside from their limited use in problems with complex geometries and non-periodic boundary conditions, recent results prove that these techniques also offer no solution to the paradox; difficulties with numerical stability and ill-conditioning eliminate these approaches as alternatives for very complex three-dimensional flow simulations.

This paper describes the results of an initial study on new types of CFD algorithms which better resolve, or have the potential for resolving, the above paradoxes. The basic approach has several components:

1. *Operator splitting.* The scale paradox arises primarily because of the conflicting influences of the convective terms and the viscous terms in the Navier–Stokes equations. By separating these influences, algorithms can be devised that resolve convective effects (shocks, contact discontinuities, etc.) with controlled artificial viscosity in one step and resolve viscous boundary layers independently in another step. The challenges are to split the operator in a consistent manner, retaining high-order accuracy in time in both convective and viscous parts, to handle conflicting boundary conditions imposed for each component step, and to merge the

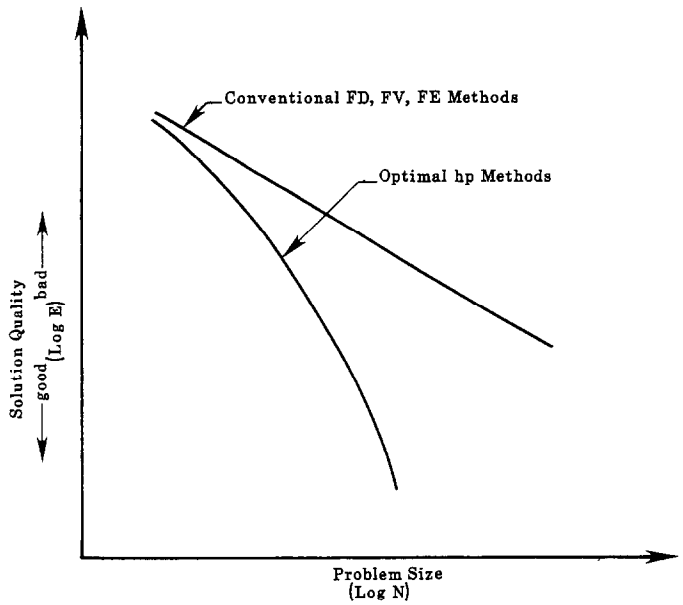

Fig. 1.1. Performance of various classes of CFD methods.

steps effectively to handle combined influence on the flow, such as shock-boundary layer interaction. Other finer-scale flow features, such as chemical reactions, can in principle also be incorporated into such splitting schemes.

2. Adaptive h-p finite element methods. A new methodology, referred to as adaptive *h-p* finite element methods, provides unprecedented flexibility in CFD modeling. The idea is to approximate the flow variables on unstrucured meshes over which different high-order spectral approximations can be employed over each gridcell. The mesh size h and the spectral order p of each cell can be changed by mesh refinement or coarsening and polynomial order enrichment or reduction. In this way, fine-grid, low-order and stable approximations can be employed to resolve shocks, for example, while large gridcells but high-order polynomial approximations can be used to resolve boundary layers with a minimum increase in problem size. Implementation of such methods in an operator-splitting framework appears to be a very natural way of exploiting the flexibility of *h-p* schemes. Challenges are to devise an efficient data structure and adaptive strategy for implementing *h-p* schemes, the adaptive scheme generally being controlled to reduce cellwise errors. This leads to another challenge: the construction of efficient and reliable element error estimation methods to control the adaptive process.

3. Optimal meshes. The distribution of mesh sizes and spectral orders in a mesh to systematically reduce and control numerical error is an exercise in mesh optimization. Aside from minimizing the number of unknowns (the problem size N) required to deliver a given level of accuracy, the use of optimal *h-p* meshes may result in a surprising property: exponential convergence. Challenges are to devise efficient strategies that produce near optimal *h-p* mesh sequences with minimal computational cost.

1.2. Introductory comments on operator splitting methods

The idea of operator splitting has been discussed in the literature for at least two decades. The time discretization schemes associated with some methods of splitting are often referred to as methods of fractional steps (see [1]) as a typical time step calculation consists of several fractioned steps dealing with the solution of different evolution problems.

The idea of operator splitting resulting in an approximation which is second order accurate in time is attributed to the pioneering work of Strang [2] and is often referred to as viscous splitting or Strang's splitting. The main idea is explained as follows. Consider an arbitrary system of partial differential equations of the form

$$\boldsymbol{u}_t = \boldsymbol{c}(\mathrm{D}^{\alpha}\boldsymbol{u}, \boldsymbol{x}, t) = \boldsymbol{a}(\mathrm{D}^{\alpha}\boldsymbol{u}, \boldsymbol{x}, t) + \boldsymbol{b}(\mathrm{D}^{\alpha}\boldsymbol{u}, \boldsymbol{x}, t)\,, \tag{1.1}$$

where $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t)$ is a vector of state variables and $\boldsymbol{c}$ represents an arbitrary function of derivatives $\mathrm{D}^{\alpha}\boldsymbol{u}$ of $\boldsymbol{u}$ (up to a certain order), including the function $\boldsymbol{u}$ itself and the independent space variables $\boldsymbol{x}$ and time t.

As a first step, we split the function $\boldsymbol{c}$ into two arbitrary contributions $\boldsymbol{a}$ and $\boldsymbol{b}$, as indicated in (1.1), and identify two fractional evolution problems of the form

$$\boldsymbol{u}_{,\tau} = \boldsymbol{a}(\mathrm{D}^{\alpha}\boldsymbol{u}, \boldsymbol{x}, t+\tau)\,, \qquad \boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0(\boldsymbol{x}, t) \tag{1.2}$$

and

$$u_{,\tau} = b(D^{\alpha}u, x, t+\tau)\,, \qquad u(x,0) = u_0(x,t)\,. \tag{1.3}$$

Introducing the corresponding evolution operators

$$(M_t(\tau)u_0)(x,t) \overset{\text{def}}{=} u(x,\tau)\,, \tag{1.4}$$

where $u(x,\tau)$ is the solution to the Cauchy problem (1.2) and

$$(N_t(\tau)u_0)(x,t) \overset{\text{def}}{=} u(x,\tau) \tag{1.5}$$

with $u(x,\tau)$ the solution of (1.3), we have the fundamental result by Strang:

$$|\{(E_t(\tau) - M_{t+\tau/2}(\tfrac{1}{2}\tau)\cdot N_t(\tau)\cdot M_t(\tfrac{1}{2}\tau))u_0\}(x,t)| \leqslant c(x)\tau^3\,, \tag{1.6}$$

where $E_t(\tau)$ is the evolution operator corresponding to the original problem (i.e., $u(x,\tau) = (E_t(\tau)u_0)(x)$). In other words, to within terms of order τ^3,

$$u(x,t) = (M_{t+\tau/2}(\tfrac{1}{2}\tau)\cdot N_t(\tau)\cdot M_t(\tfrac{1}{2}\tau)u_0)(x,t)\,. \tag{1.7}$$

This splitting scheme is illustrated in Fig. 1.2. Interpreting τ as the incremental time increase beginning at the time level t with an initial solution u_0, we progress first with half of the time step $\frac{1}{2}\tau$ using the first of the fractional step problem (1.2). Next we step a half step backwards and proceed with the full step consistent with the second fractional operator. Finally, in the third step, we return to the half-time step level and proceed with a half time step with the first operator. The result is purely formal in the sense that none of the individual fractional step solutions themselves makes sense. In Fig. 1.2, the three different time levels are introduced merely to illustrate the way in which functions a and b are to be evaluated if they depend explicitly on time. The result is also local as the 'constant' $c(x)$ on the right-hand side of (1.6) depends upon x.

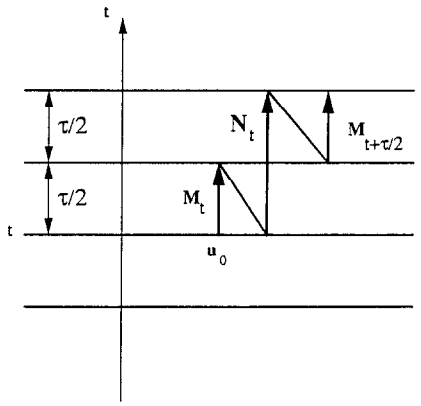

Fig. 1.2. Interpretation of Strang's operator splitting.

It is interesting that a corresponding two-step scheme,

$$|\{(\boldsymbol{E}_t(\tau) - \boldsymbol{N}_t(\tau) \cdot \boldsymbol{M}_t(\tau))\boldsymbol{u}_0\}(\boldsymbol{x})| \leq c(\boldsymbol{x})\tau^2 \tag{1.8}$$

is only of first order. The second-order Strang splitting eliminates also the dependence on the order in which the fractional step operators are composed, as the operations in (1.7) can be written in the form

$$\boldsymbol{M}_{t+\tau/2}(\tfrac{1}{2}\tau) \cdot \underbrace{\boldsymbol{N}_{t+\tau/2}(\tfrac{1}{2}\tau) \cdot \boldsymbol{N}_t(\tfrac{1}{2}\tau)}_{\boldsymbol{N}_t(\tau)} \boldsymbol{M}_t(\tfrac{1}{2}\tau)\,, \tag{1.9}$$

i.e., in the second half time step the order of fractional step operators is reversed.

We are convinced that the idea of operator splitting has great potential, especially from the practical, programming point of view. Very often operator splitting may result naturally from a generalization of a simpler problem formulation by adding new coupling terms and/or new variables resulting from physical problems with an increasing complexity (Euler Equations → Navier–Stokes equations → chemically reactive flows → radiative chemically reactive flows, etc.). In such a situation, there is a good chance that an existing method (code) can be modified by adding the additional two fractional time steps, to solve the more complex phenomena.

Another important observation is the possibility of separation of operators of conflicting nature (transport from diffusion, etc.). Once the operators are separated, highly specialized methods can (at least formally) be used to solve each of the fractional steps.

As noted earlier, the challenge in applying Strang's splitting is connected with the study of boundary conditions. As (1.7) is purely a local estimate, incorrect boundary conditions may cause $c(\boldsymbol{x})$ in (1.6) to grow toward infinity, which would lead to completely erroneous results. In general, a study of boundary conditions combined with (1.7) should lead to some global estimates, as done by Szymczak in [3] for linear, steady state convection–diffusion problems.

Second-order splitting methods were also applied by Beale and Majda [4] to study incompressible Navier–Stokes equations and by Ramos [5] in the analysis of linear reactive–diffusive systems. Most of the existing methods in the literature are based, however, on first order splitting (see [6] for a very detailed analysis of an operator splitting in the context of nonlinear advection–diffusion–reaction problems, and [7] for a study of convection-dominated problems in one dimension). Little work has been done to study the effect of boundary conditions on the performance and accuracy of splitting methods (see [5, 8] for some numerical experiments).

In this paper, we present an operator splitting method for the compressible Navier–Stokes equations. The splitting separates the Euler fluxes from the viscous contributions resulting in an Euler, transport step and a 'viscous' step including the remaining terms. Many other splittings are obviously possible, but the present approach has the advantage of allowing the use of an existing Euler equation solver to solve the transport step. Keeping this motivation in mind, our study on boundary conditions is done in such a way that the Euler step is accompanied by the proper boundary conditions for the Euler equations which implies that the Euler solver is used without any changes in implementation.

1.3. *h-p adaptive finite element methods*

Adaptive finite element techniques provide attractive approaches to many classes of problems in both computational fluid and solid mechanics. The main objectives of such methods are

— the use of an a posteriori error estimation guaranteeing the reliability of the numerical solution and providing a control parameter for an adaptive strategy,
— the use of optimal meshes, delivering the best possible results for a fixed number of degrees of freedom.

As noted earlier in the introduction, *h*-methods attempt to improve solution quality by refining (locally) the mesh, while *p*-methods attempt to improve quality by increasing the local degree of the polynomial shape functions and are, thus, intimately related to spectral methods. The *h*-methods have been successfully applied to solve several compressible flow problems [9, 10], and the *p*-methods have been used mostly in context of steady, slow incompressible viscous flow. Combined *h-p* methods have been applied to linear elliptic problems (see [10–12]) and to special classes of steady compressible flow problems only recently [13].

For elliptic equations, adaptive *h-p* methods can deliver an exponential rate of convergence (see [12]) suggesting that the highest possible increase in accuracy per increase in problem size can be attained. In this work, we address another important advantage of using the *h-p* methods, namely

— an optimal choice of an *h-p* mesh guarantees the stability of the solution for problems with small coefficients (in our case resulting from convection-dominated diffusion).

As shown in this work, a judicious construction of an *h-p* mesh along a solid-wall boundary leads to stable behavior of the solution and results in highly accurate resolution of boundary layer phenomena.

The plan of this paper is as follows. In Section 2, the initial boundary problem for the Navier–Stokes equations is formulated and the choice of boundary conditions is discussed. Section 3 presents the operator splitting approach with an emphasis placed upon boundary conditions for fractional step problems. Section 4 is devoted to the fundamental issues of adaptive methods, a posteriori error estimation, and the design of an adaptive *h-p* strategy compatible with stability requirements for problems with small parameters. In Section 5, several numerical results for the Carter flat plate problem are presented and conclusions are summarized in Section 6.

2. Preliminaries: Properties of initial boundary-value problems for the Navier–Stokes equations for compressible flow

2.1. *Notation and formulation of the equations*

Given a domain $\Omega \subset \mathbb{R}^N$ (we assume $N = 2$ for notational simplicity), we consider a system of compressible gas dynamics equations in the form

$$\boldsymbol{U}_{,t} + \sum_{i=1}^{2} \boldsymbol{F}^i(\boldsymbol{U})_{,i} = \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \boldsymbol{K}^{ij}(\boldsymbol{U}) \boldsymbol{U}_{,j} \right)_{,i}, \tag{2.1}$$

where

$$\boldsymbol{U} = \{\rho, m_1, m_2, e\}^t \tag{2.2}$$

is the vector of conservation variables with density ρ, momentum components $m_i = \rho u_i$, ($\boldsymbol{u} = \{u_i\}$ being the velocity vector), e the total energy per unit volume, and commas denoting differentiation with respect to time t and the spatial coordinates $(\ )_{,i} = \partial/\partial x_i$.

$\boldsymbol{F}^i(\boldsymbol{U})$ are the Eulerian fluxes defined as

$$\begin{aligned} \boldsymbol{F}^1 &= \left\{ m_1, \frac{m_1^2}{\rho} + p, \frac{m_1 m_2}{\rho}, (e+p)\frac{m_1}{\rho} \right\}^t , \\ \boldsymbol{F}^2 &= \left\{ m_2, \frac{m_1 m_2}{\rho}, \frac{m_2^2}{\rho} + p, (e+p)\frac{m_2}{\rho} \right\}^t , \end{aligned} \tag{2.3}$$

with p the thermodynamic pressure. Assuming the constitutive equations for a perfect gas we have

$$p = (\gamma - 1)\iota , \tag{2.4}$$

where ι is the internal energy per unit volume,

$$\iota = e - \tfrac{1}{2}\rho(u_1^2 + u_2^2) , \tag{2.5}$$

and γ is the ratio of specific heats, i.e.,

$$\gamma = c_p/c_v , \tag{2.6}$$

provided c_p is the specific heat at constant pressure and c_v is the specific heat at constant volume. Given the internal energy ι, we calculate the temperature θ as

$$\theta = \frac{1}{c_v}\,\hat{\iota} = \frac{1}{c_v}\,\frac{\iota}{\rho} , \tag{2.7}$$

$\hat{\iota} = \rho^{-1}\iota$ being the specific internal energy.

Viscous fluxes $\boldsymbol{F}_v^i = \Sigma_{j=1}^2 \boldsymbol{K}^{ij}(\boldsymbol{U})\boldsymbol{U}_{,j}$ are defined by

$$\boldsymbol{F}_v^i = \left(0; \tau_{i1}; \tau_{i2}; \sum_{j=1}^{2} \tau_{ij}u_j + q_i\right) \tag{2.8}$$

with viscous stresses τ_{ij} given by the formula

$$\tau_{ij} = \mu(u_{i,j} + u_{j,i}) + \lambda u_{k,k}\delta_{ij} \tag{2.9}$$

and the heat flux q_i defined as

$$q_i = \kappa\theta_{,i} . \tag{2.10}$$

Here κ is the thermal conductivity and μ and λ are temperature dependent viscosities.

In practical calculations, we employ Sutherland's law and Stokes' relation, introduced later. Explicit formulas for the viscous matrices $\boldsymbol{K}^{ij}(\boldsymbol{U})$ are

$$\boldsymbol{K}^{11} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -(2\mu+\lambda)\frac{m_1}{\rho^2} & (2\mu+\lambda)\frac{1}{\rho} & 0 & 0 \\ -\mu\frac{m_2}{\rho^2} & 0 & \frac{\mu}{\rho} & 0 \\ -(2\mu+\lambda)\frac{m_1^2}{\rho^3} - \mu\frac{m_2^2}{\rho^3} + \frac{\kappa}{c_v}\left(-\frac{e}{\rho} + \frac{m_1^2+m_2^2}{\rho^3}\right) & \left(2\mu+\lambda-\frac{\kappa}{c_v}\right)\frac{m_1}{\rho^2} & \left(\mu-\frac{\kappa}{c_v}\right)\frac{m_2}{\rho^2} & \frac{\kappa}{c_v}\frac{1}{\rho} \end{bmatrix},$$

$$\boldsymbol{K}^{12} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\lambda\frac{m_2}{\rho^2} & 0 & \frac{\lambda}{\rho} & 0 \\ -\mu\frac{m_1}{\rho^2} & \frac{\mu}{\rho} & 0 & 0 \\ -(\lambda+\mu)\frac{m_1 m_2}{\rho^3} & \mu\frac{m_2}{\rho^2} & \lambda\frac{m_1}{\rho^2} & 0 \end{bmatrix}, \quad \boldsymbol{K}^{21} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\mu\frac{m_2}{\rho^2} & 0 & \frac{\mu}{\rho} & 0 \\ -\lambda\frac{m_1}{\rho^2} & \frac{\lambda}{\rho} & 0 & 0 \\ -(\lambda+\mu)\frac{m_1 m_2}{\rho^3} & \lambda\frac{m_2}{\rho^2} & \mu\frac{m_1}{\rho^2} & 0 \end{bmatrix}$$

and

$$\boldsymbol{K}^{22} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\mu\frac{m_1}{\rho^2} & \frac{\mu}{\rho} & 0 & 0 \\ -(2\mu+\lambda)\frac{m_2}{\rho^2} & 0 & (2\mu+\lambda)\frac{1}{\rho} & 0 \\ -\mu\frac{m_1^2}{\rho^3} - (2\mu+\lambda)\frac{m_2^2}{\rho^3} + \frac{\kappa}{c_v}\left(-\frac{e}{\rho} + \frac{m_1^2+m_2^2}{\rho^3}\right) & \left(\mu-\frac{\kappa}{c_v}\right)\frac{m_1}{\rho^2} & \left(2\mu+\lambda-\frac{\kappa}{c_v}\right)\frac{m_2}{\rho^2} & \frac{\kappa}{c_v}\frac{1}{\rho} \end{bmatrix}.$$

Note that the equations are fully determined if the following constants are known:
— ratio of specific heats $\gamma = c_p/c_v$,
— specific heat at a constant volume c_v,
— the thermal conductivity κ,
— viscosities μ and λ prescribed as functions of temperature θ.
For most gases, these quantities are not independent of each other as

$$\kappa = \frac{c_p\mu}{\mathrm{Pr}} = \frac{\gamma c_v \mu}{\mathrm{Pr}}, \tag{2.11}$$

where Pr is the Prandtl number.

2.2. *Nondimensional form of the equations*

Following [14], we introduce the following scaling parameters:
— a reference length L,
— a reference speed V_∞,
— a reference density ρ_∞,
— a reference temperature θ_∞,
— a reference viscosity μ_∞.
The reference quantities are usually introduced as those corresponding to the free stream flow data, but this is not necessary. (In the shock tube problem, for instance, there is no free

stream data and the static speed of sound can be assumed for V_∞.) The nondimensional quantities are then introduced as follows:

$$x_i^* = \frac{x_i}{L}, \quad t^* = \frac{V_\infty}{L} t, \quad \rho^* = \frac{\rho}{\rho_\infty}, \quad u_i^* = \frac{u_i}{V_\infty}, \quad m_i^* = \frac{m_i}{\rho_\infty V_\infty},$$
$$p^* = \frac{p}{\rho_\infty V_\infty^2}, \quad \iota^* = \frac{\iota}{\rho_\infty V_\infty^2}, \quad \theta^* = \frac{\theta}{\theta_\infty}, \quad \mu^* = \frac{\mu}{\mu_\infty}, \quad \lambda^* = \frac{\lambda}{\mu_\infty}. \tag{2.12}$$

It is also convenient to introduce the reference (free stream) Reynolds number,

$$\mathrm{Re} = \rho_\infty V_\infty L / \mu_\infty, \tag{2.13}$$

and the reference (free stream) Mach number,

$$M_\infty = V_\infty / \sqrt{\gamma(\gamma - 1) c_v \theta_\infty}. \tag{2.14}$$

The nondimensionalized equations assume essentially the same form as the original formulas (2.4), (2.5), (2.7), (2.11), provided new material constants are defined as follows:

$$\tilde{\mu} = \frac{\mu^*}{\mathrm{Re}_L}, \quad \tilde{\lambda} = \frac{\lambda^*}{\mathrm{Re}_L}, \quad \tilde{c}_v = [\lambda(\lambda - 1) M_\infty^2]^{-1}. \tag{2.15}$$

Following [14], we assume the Stokes relation,

$$3\lambda + 2\mu = 0, \tag{2.16}$$

and Sutherland's law in the nondimensional form,

$$\mu^* = (\theta^*)^{1.5} (1 + c_1) / (\theta^* + c_1), \tag{2.17}$$

where

$$c_1 = \frac{110[^\circ\mathrm{K}]}{\theta_\infty[^\circ\mathrm{K}]} = \frac{198.6[^\circ\mathrm{R}]}{\theta_\infty[^\circ\mathrm{R}]} \tag{2.18}$$

(see also [14, 15]).

Throughout this paper, asterisks and tildes are omitted and it is understood that the nondimensionalized forms of the governing equations are in force. This section is concluded with a list of some nondimensional quantities of interest:

— local Mach number:

$$M = \frac{\sqrt{u_1^2 + u_2^2}}{c} = \frac{\sqrt{u_1^{*2} + u_2^{*2}}}{c^*}, \tag{2.19}$$

where the sound speed c is

$$c^2 = \gamma \frac{p}{\rho} \quad \text{and} \quad c^{*2} = \gamma \frac{p^*}{\rho^*}; \tag{2.20}$$

— pressure coefficient:

$$C_p = \frac{p - p_\infty}{0.5\rho_\infty V_\infty^2} = 2(p^* - 1) \tag{2.21}$$

($p_\infty = \rho_\infty V_\infty^2$, see (2.13));

— skin friction coefficient:

$$C_f = \frac{\tau_s}{0.5\rho_\infty V_\infty^2} = 2\tau_s^* , \tag{2.22}$$

where the nondimensional shear stress τ_s^* is calculated using the nondimensional form of (2.9) (with $\tilde{\boldsymbol{u}}$ replacing $\boldsymbol{u}$);

— heat flux coefficient:

$$C_h = \frac{q}{0.5\rho_\infty V_\infty^3} = 2q^* , \tag{2.23}$$

where

$$q = q_i n_i \quad \text{and} \quad q^* = q_i^* n_i , \tag{2.24}$$

provided $\boldsymbol{n} = \{n_i\}$ denotes a unit vector. The nondimensional heat flux vector q_i^* is evaluated using (2.10), the nondimensional temperature θ^* and the coefficient of conductivity,

$$\tilde{\kappa} = \gamma \tilde{c}_v \tilde{\mu} / \mathrm{Pr} \tag{2.25}$$

(compare with (2.11)).

2.3. *Entropy function and the symmetrization of Navier–Stokes equations*

Introducing the Euler–Jacobian Matrices,

$$\boldsymbol{A}^i(\boldsymbol{U}) = \boldsymbol{F}_{\boldsymbol{U}}^i(\boldsymbol{U}) , \tag{2.26}$$

we rewrite the Navier–Stokes equations in the quasilinear form

$$\boldsymbol{U}_{,t} + \sum_{i=1}^{2} \boldsymbol{A}^i(\boldsymbol{U})\boldsymbol{U}_{,i} = \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \boldsymbol{K}^{ij}(\boldsymbol{U})\boldsymbol{U}_{,j} \right)_{,i} . \tag{2.27}$$

It is well-known that the entropy function $H = H(\boldsymbol{U})$, defined as

$$H = -\rho \hat{s} \tag{2.28}$$

where $\hat{s}$ is the specific entropy

$$\hat{s} = \ln(p\rho^{-\gamma}) , \tag{2.29}$$

is such that the Hessian $A_0(U)$ of H evaluated at U,

$$A_0 = H_{UU} \,. \tag{2.30}$$

Then A_0 is a symmetrizer of the system (2.27) in the following sense:

$$\begin{aligned}
&(1)\quad A_0 = A_0^{\mathrm{t}} > 0 \quad \text{(i.e., } A_0 \text{ is positive definite)}\,,\\
&(2)\quad A_0 A^i = (A_0 A^i)^{\mathrm{t}} \geqslant 0 \quad \text{(positive semidefinite)}\,,\\
&(3)\quad A_0 K^{ij} = (A_0 K^{ij})^{\mathrm{t}} \quad \text{for } i = j\,,\\
&(4)\quad A_0 K^{ij} = (A_0 K^{ji})^{\mathrm{t}} \quad \text{for } i \neq j\,,\\
&(5)\quad U_{,i}^{\mathrm{t}} A_0 K^{ij} U_{,j} \geqslant 0 \quad \text{for every } U_{,j}\,.
\end{aligned} \tag{2.31}$$

This fact is crucial to most of the developments in this report. An explicit formula for A_0 is

$$A_0 = \begin{bmatrix}
\frac{1}{\rho}\left[\psi^2 + \gamma\right] & -\psi\frac{u}{\iota} & -\psi\frac{v}{\iota} & \frac{1}{\iota}(\psi - 1) \\
 & \frac{1}{\iota}\left(1 + \rho\frac{u^2}{\iota}\right) & \rho\frac{uv}{\iota^2} & -\rho\frac{u}{\iota^2} \\
 & \text{Symm.} & \frac{1}{\iota}\left(1 + \rho\frac{v^2}{\iota}\right) & -\frac{\rho v}{\iota^2} \\
 & & & \frac{\rho}{\iota^2}
\end{bmatrix},$$

where $\psi = \rho(u^2 + v^2)/(2\iota)$.

2.4. *Linearized Navier–Stokes equations—A stability condition*

As a complete mathematical theory for the full Navier–Stokes equations does not exist, (global uniqueness, existence results), some approximate tools have to be applied when studying the well-posedness of the Initial-Boundary Value Problem (IBVP) for the equations. A classical approach would be to study the linearized form of the Navier–Stokes equations accompanied by linearized boundary conditions and to insist that the resulting linear IBVP is well posed.

Let U be a solution to (2.1) and δU a 'small' perturbance of U. Then

$$F^i(U + \delta U) \approx F^i(U) + A^i(U)\,\delta U\,, \qquad K^{ij}(U + \delta U) \approx K^{ij}(U) + K_U^{ij}(U)\,\delta U\,. \tag{2.32}$$

Using the linearized formulas (2.32), substituting into (2.1), subtracting the original equations for U and neglecting the higher order terms, we get

$$\delta U_{,t} + \sum_{i=1}^{2} \left(A^i(U)\,\delta U\right)_{,i} = \sum_{i=1}^{2} \left(\sum_{j=1}^{2} K^{ij}(U)\,\delta U_{,j} + K_U^{ij}(U)\,\delta U U_{,j}\right)_{,i}\,. \tag{2.33}$$

At this point, it is customary (see [16]) to assume that the reference state $\boldsymbol{U}$ is approximately uniform so that $\boldsymbol{U}_{,j}$ is zero or of the same order as $\delta\boldsymbol{U}$, an assumption which is certainly violated at shocks or in boundary layers. Such an assumption allows us to drop the second term on the right-hand side of (2.33) as well as to simplify the Eulerian term:

$$(\boldsymbol{A}^i(\boldsymbol{U})\,\delta\boldsymbol{U})_{,i} = \boldsymbol{A}^i_{\boldsymbol{U}}(\boldsymbol{U})\boldsymbol{U}_{,i}\,\delta\boldsymbol{U} + \boldsymbol{A}^i(\boldsymbol{U})\,\delta\boldsymbol{U}_{,i} \approx \boldsymbol{A}^i(\boldsymbol{U})\,\delta\boldsymbol{U}_{,i}\,. \tag{2.34}$$

Roughly speaking, the Jacobian matrices $\boldsymbol{A}^i(\boldsymbol{U})$ are consistently treated as they were constant and this resolves practically all technical problems connected with system (2.33) which, even though linear and in conservation form, has variable coefficients.

Thus the simplified problem takes on the form

$$\delta\boldsymbol{U}_{,t} + \sum_{i=1}^{2} \boldsymbol{A}^i\,\delta\boldsymbol{U}_{,i} = \sum_{i=1}^{2}\left(\sum_{j=1}^{2} \boldsymbol{K}^{ij}\,\delta\boldsymbol{U}_{,j}\right)_{,i}\,. \tag{2.35}$$

Multiplying (2.35) by the symmetrizer $\boldsymbol{A}_0 = \boldsymbol{A}_0(\boldsymbol{U})$ and then by the perturbation $\delta\boldsymbol{U}$, we get

$$\frac{1}{2}\left\{(\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\,\delta\boldsymbol{U})_{,t} + \sum_{i=1}^{2}(\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\boldsymbol{A}^i\,\delta\boldsymbol{U})_{,i}\right\} = \delta\boldsymbol{U}^{\mathrm{t}}\sum_{i=1}^{2}\left(\sum_{j=1}^{2}\boldsymbol{A}_0\boldsymbol{K}^{ij}\,\delta\boldsymbol{U}_{,j}\right)_{,i}\,, \tag{2.36}$$

where we have consistently neglected both time and spatial derivatives of the symmetrizer $\boldsymbol{A}_0$. Integrating (2.36) by parts over Ω, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{1}{2}(\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\,\delta\boldsymbol{U})\,\mathrm{d}x + \int_{\partial\Omega}\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\boldsymbol{A}_n\,\delta\boldsymbol{U}\,\mathrm{d}s \\ = -\int_\Omega \sum_{i=1}^{2}\sum_{j=1}^{2}\delta\boldsymbol{U}^{\mathrm{t}}_{,i}\boldsymbol{A}_0\boldsymbol{K}^{ij}\,\delta\boldsymbol{U}_{,j}\,\mathrm{d}x + \int_{\partial\Omega}\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\boldsymbol{t}_n\,\mathrm{d}s\,,\end{aligned} \tag{2.37}$$

where

$$\boldsymbol{A}_n = \sum_{i=1}^{2}\boldsymbol{A}^i n_i \tag{2.38}$$

is the normal Jacobian matrix ($\boldsymbol{n} = \{n_i\}$ is the outward normal unit to $\partial\Omega$) and $\delta\boldsymbol{t}_n$ is the normal viscous flux defined as

$$\delta\boldsymbol{t}_n = \sum_{i=1}^{2}\sum_{j=1}^{2}\boldsymbol{K}^{ij}\,\delta\boldsymbol{U}_{,j}n_i\,. \tag{2.39}$$

As the first integral on the right-hand side of (2.37) is nonpositive, the linearized system of Navier–Stokes equations will be stable in the sense of the linearized entropy function (norm), i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\,\delta\boldsymbol{U}\,\mathrm{d}x \leqslant 0\,, \tag{2.40}$$

provided the following stability condition is satisfied on the boundary:

$$\tfrac{1}{2}\delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\boldsymbol{A}_n\,\delta\boldsymbol{U} - \delta\boldsymbol{U}^{\mathrm{t}}\boldsymbol{A}_0\,\delta\boldsymbol{t}_m \geqslant 0\,. \tag{2.41}$$

Stability condition (2.41) will be a basis for our discussion on boundary conditions in the next sections.

2.5. Well-posedness of incomplete parabolic problems—Number of boundary conditions

The linearized problem (2.35) falls into the category of incomplete parabolic systems studied by Strikwerda [17]. The reasoning we present in this section follows the application of Strikwerda's theory to the Navier–Stokes equations by Gustafsson and Sudström [16], though different variables were used in [16] and some of our conclusions are different as well. (See also [18] for the more recent developments.)

Following [16], by an incomplete parabolic system we mean a system of the form

$$\boldsymbol{A}_0 \boldsymbol{U}_{,t} + \sum_i \boldsymbol{A}^i \boldsymbol{U}_{,i} = \sum_i \left(\sum_j \boldsymbol{K}^{ij} \boldsymbol{U}_{,j} \right)_{,i}, \tag{2.42}$$

where $\boldsymbol{A}_0$ is symmetric and positive definite and $\boldsymbol{A}^i$ and $\boldsymbol{K}^{ij}$ are constant and symmetric and of the following form:

$$\boldsymbol{K}^{ij} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}^{ij}_{22} \end{bmatrix}, \tag{2.43}$$

where rank $[\boldsymbol{K}^{ij}_{22}] = r$ and

$$\boldsymbol{A}^i = \begin{bmatrix} \boldsymbol{A}^i_{11} & \boldsymbol{A}^i_{12} \\ \boldsymbol{A}^i_{21} & \boldsymbol{A}^i_{22} \end{bmatrix} \tag{2.44}$$

and the decomposition corresponds to the representation of the solution vector in the form

$$\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2) . \tag{2.45}$$

The conditions for the well-posedness of a Cauchy problem (no boundary conditions, the equations are solved in the whole of $\mathbb{R}^N$) are

(1) The reduced second order system

$$\boldsymbol{U}_{2,t} = \sum_{ij} (\boldsymbol{K}^{ij}_{22} \boldsymbol{U}_{2,j})_{,i} \tag{2.46}$$

is parabolic,

(2) the reduced first order system

$$\boldsymbol{U}_{1,t} + \sum_i \boldsymbol{A}^i_{11} \boldsymbol{U}_{1,i} = 0 \tag{2.47}$$

is strictly hyperbolic.

For Initial Boundary Value Problems (IBVP), Strikwerda [17] studied equations of the

form (2.42), (2.43) and (2.44) and considers boundary conditions of the form

$$\boldsymbol{R}\frac{\partial \boldsymbol{U}}{\partial n}+\boldsymbol{S}\boldsymbol{U}=\boldsymbol{g}\,, \tag{2.48}$$

where

$$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix} \tag{2.49}$$

and

$$\boldsymbol{S}=\begin{bmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{bmatrix}. \tag{2.50}$$

Thus, the coupling between $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ in both the equations and boundary conditions takes place only through the first order terms.

The number of boundary conditions n_{BC} for the well-posedness of the IVBP is

$$n_{\mathrm{BC}}=r(\text{rank of } \boldsymbol{K}_{22}^{ij})+\text{number of negative eigenvalues of } \sum_i \boldsymbol{A}_{11}^i n_i\,, \tag{2.51}$$

where $\boldsymbol{n}=\{n_i\}$ is again the outward unit vector normal to $\partial\Omega$.

In order to comply with Strikwerda's results one has to

(1) neglect the variation of $\boldsymbol{A}^i$, $\boldsymbol{K}^{ij}$ (consistent with the linearization procedure),

(2) symmetrize the equations.

The second step is to introduce the vector of entropy variables (see [19])

$$\boldsymbol{V}=\boldsymbol{A}_0\,\delta\boldsymbol{U} \tag{2.52}$$

and converting (2.35) into the form

$$\boldsymbol{A}_0^{-1}\boldsymbol{V}_{,t}+\sum_i \boldsymbol{A}^i\boldsymbol{A}_0^{-1}\boldsymbol{V}_{,i}=\sum_i\sum_j (\boldsymbol{K}^{ij}\boldsymbol{A}_0^{-1}\boldsymbol{V}_{,j})_{,i}\,. \tag{2.53}$$

One can show that $\boldsymbol{A}^i\boldsymbol{A}_0^{-1}$ and $\boldsymbol{K}^{ij}\boldsymbol{A}_0^{-1}$ are symmetric and positive semidefinite. The form of the symmetrized matrices is given in [19]. It can be verified that for the linearized Navier–Stokes equations, $r=3$ and both conditions (2.46) and (2.47) are satisfied (the second one trivially since the reduced first order system reduces to a single equation). The only eigenvalue of $\Sigma_i \boldsymbol{A}_{11}^i n_i$ turns out to be

$$\lambda=\rho u_{\mathrm{n}}/(\gamma-1)\,, \tag{2.54}$$

where u_{n} is the normal velocity on the boundary. The conclusion therefore is that for the linearized Navier–Stokes equation, four boundary conditions should be satisfied on the inflow boundary ($u_{\mathrm{n}}<0$) and only three boundary conditions on the outflow boundary ($u_{\mathrm{n}}>0$). Note that the corresponding numbers for Euler equations are

supersonic inflow	4,
subsonic inflow	3,
subsonic outflow	1,
supersonic outflow	0.

(2.55)

2.6. Discussion of boundary conditions

Let $\boldsymbol{A}_n = \Sigma_i \boldsymbol{A}^i \boldsymbol{n}_i$ denote the normal Jacobian matrix and let $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ denote the left and right eigenvectors of $\boldsymbol{A}_n$ [20]. We recall the following fundamental facts concerning $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ and their relation with the symmetrizer:

(1) $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ form two dual bases in $\mathbb{R}^4$, i.e.,

$$\boldsymbol{b}_i \cdot \boldsymbol{c}_j = \delta_{ij} ; \tag{2.56}$$

(2) the symmetrizer transforms right eigenvectors into left eigenvectors, i.e.,

$$\boldsymbol{A}_0 \boldsymbol{c}_i = \boldsymbol{b}_i . \tag{2.57}$$

Substituting

$$\delta \boldsymbol{U} = \sum_i (\delta \boldsymbol{U} \cdot \boldsymbol{b}_i)\boldsymbol{c}_i , \qquad \delta \boldsymbol{U}^{\mathrm{t}} = \sum_j (\delta \boldsymbol{U} \cdot \boldsymbol{c}_j)\boldsymbol{b}_j^{\mathrm{t}} , \tag{2.58}$$

we reduce (2.41) to the form

$$\tfrac{1}{2} \sum_i \lambda_i (\delta \boldsymbol{U} \cdot \boldsymbol{b}_i)^2 - \sum_i (\delta \boldsymbol{t}_n \cdot \boldsymbol{b}_i)(\delta \boldsymbol{U} \cdot \boldsymbol{b}_i) \geqslant 0 . \tag{2.59}$$

Recall also that

$$\lambda_1 = u_{\mathrm{n}} - c , \qquad \lambda_2 = \lambda_3 = u_{\mathrm{n}} , \qquad \lambda_4 = u_{\mathrm{n}} + c . \tag{2.60}$$

When discussing various possible versions of boundary conditions satisfying the stability condition (2.59), the following criteria are taken into account.

(1) According to the results from the previous section, there must be four boundary conditions on the inflow boundary and only three on the outflow boundary.
(2) On an open boundary, the boundary conditions for the Navier–Stokes equations should converge to the right boundary conditions for Euler equations (both in number and type), when the Reynolds number Re approaches infinity.
(3) As the incremental equations were obtained by linearizing the full nonlinear Navier–Stokes equations, the boundary conditions for the linearized problem should result from a linearization of some nonlinear boundary conditions for the original equations.

A few useful algebraic results must be recorded before we proceed with the discussion of specific boundary conditions.

The scalar products of the incremental vector $\delta \boldsymbol{U}$ with the left eigenvectors $\boldsymbol{b}_i$ can be represented in the form

$$\begin{aligned} &\delta \boldsymbol{U} \cdot \boldsymbol{b}_1 = \frac{1}{2c}\left(\frac{\delta p}{c} - \rho\, \delta u_{\mathrm{n}}\right), \qquad \delta \boldsymbol{U} \cdot \boldsymbol{b}_2 = -\frac{\rho}{c}\, \delta u_{\mathrm{s}} , \\ &\delta \boldsymbol{U} \cdot \boldsymbol{b}_3 = -\frac{\delta p}{c^2} + \delta \rho = -\frac{\rho}{\gamma}\, \delta \hat{s} , \qquad \delta \boldsymbol{U} \cdot \boldsymbol{b}_4 = \frac{1}{2c}\left(\frac{\delta p}{c} + \rho\, \delta u_{\mathrm{n}}\right). \end{aligned} \tag{2.61}$$

In the above, δp, δu_n, δu_s, $\delta \hat{s}$ are the increments in pressure, normal and tangential velocity, and specific entropy. For instance, as

$$p = (\gamma - 1)(e - \tfrac{1}{2}(m_1^2 + m_2^2)/\rho)\,, \tag{2.62}$$

the increment in the pressure is

$$\delta p = (\gamma - 1)\left[\delta e + \left(\frac{m_1^2 + m_2^2}{\rho^2}\right)\delta\rho - \left(\frac{m_1}{\rho}\right)\delta m_1 - \left(\frac{m_2}{\rho}\right)\delta m_2\right], \tag{2.63}$$

etc. Various other interpretations are possible corresponding to different choices of the state variables.

The combined (incremental) viscous and heat normal flux vector $\delta \boldsymbol{t}$ satisfies the relations

$$\delta \boldsymbol{t}_n \cdot \boldsymbol{b}_2 = -\delta\tau_{ns} = -\mu\left(\frac{\partial\,\delta u_n}{\partial s} + \frac{\partial\,\delta u_s}{\partial n}\right), \tag{2.64}$$

$$\delta \boldsymbol{t}_n \cdot (\boldsymbol{b}_1 + \boldsymbol{b}_4) = -\delta \boldsymbol{t}_n \cdot \boldsymbol{b}_3 = -\kappa\,\frac{\partial\,\delta\theta}{\partial n}\,, \tag{2.65}$$

$$\delta \boldsymbol{t}_n \cdot (\boldsymbol{b}_1 + \boldsymbol{b}_3 + \boldsymbol{b}_4) = 0\,. \tag{2.66}$$

Here τ_{ns} are the components of shear stress, δu_n and δu_s denote the normal and tangential components of the incremental velocity vector $\delta \boldsymbol{u}$, $\partial/\partial s$ and $\partial/\partial n$ are the tangential and normal derivatives, respectively, and $\delta\theta$ denotes the increment in temperature θ.

Equipped with these auxiliary results, we can now discuss various boundary conditions.

Case 1. Open boundary—supersonic inflow $u_n + c < 0$

All eigenvalues λ_i in (2.60) are negative. To satisfy (2.59) we assume $\delta \boldsymbol{U} \cdot \boldsymbol{b}_i = 0$ for all $\boldsymbol{b}_i$, which is equivalent to the condition

$$\delta \boldsymbol{U} = \boldsymbol{0}\,. \tag{2.67}$$

This would correspond to the boundary conditions for the original equation in the form

$$\boldsymbol{U} = \boldsymbol{U}^{\text{in}}\,, \tag{2.68}$$

where $\boldsymbol{U}^{\text{in}}$ is an inflow vector.

Remark. To avoid a common point of confusion, we emphasize that the analysis here answers the question 'what is the right form of boundary conditions' from the point of view of the energy (linearized entropy) estimate and the resulting stability for the linearized problem. It does not answer the question of how various data, such as $\boldsymbol{U}^{\text{in}}$ in (2.68), are to be obtained. Usually, the free-stream $\boldsymbol{U}^{\infty}$ is taken for $\boldsymbol{U}^{\text{in}}$ in (2.68).

Case 2. Open boundary—subsonic inflow $u_n < 0$, $u_n + c > 0$

An obvious choice to satisfy (2.59) would be

$$\delta \boldsymbol{U} \cdot \boldsymbol{b}_i = 0 \quad \text{for } i = 1, 2, 3\,, \qquad \delta \boldsymbol{t}_n \cdot \boldsymbol{b}_4 = 0\,. \tag{2.69}$$

As $\mathrm{Re}\to 0$, the last condition is satisfied automatically and the first three conditions are the correct boundary conditions for the linearized Euler equations. The last boundary condition can be directly applied to the original equations in the form

$$\boldsymbol{t}_n \cdot \boldsymbol{b}_4 = 0 \tag{2.70}$$

as

$$\delta(\boldsymbol{t}_n \cdot \boldsymbol{b}_4) = \delta \boldsymbol{t}_n \cdot \boldsymbol{b}_4 + \boldsymbol{t}_n \cdot \delta \boldsymbol{b}_4 , \tag{2.71}$$

where the second term can be neglected (consistent with the linearization as it includes products of derivatives of $\boldsymbol{U}$ with $\delta \boldsymbol{U}$). The first three conditions, however, cannot be immediately generalized to the nonlinear case as

$$\delta(\boldsymbol{U} \cdot \boldsymbol{b}_i) = \delta \boldsymbol{U} \cdot \boldsymbol{b}_i + \boldsymbol{U} \cdot \delta \boldsymbol{b}_i \tag{2.72}$$

and the second term on the right-hand side of (2.72) is of the same order as the first.

It is a simple algebraic exercise to check that the first three conditions in (2.69) follow from the following nonlinear essential boundary conditions of the form (cf. [16]):

$$u_n - 2c/(\gamma - 1) , \quad u_s , \quad \hat{s} \text{ are prescribed} \tag{2.73}$$

(the second and third of these conditions follow immediately from (2.61)).

We emphasize that (2.69) and the resulting (2.73) are by no means the only subsonic inflow boundary conditions. For instance, the first three essential boundary conditions can be replaced with conditions of the form

$$\delta \boldsymbol{t}_n \cdot \boldsymbol{b}_i = \varepsilon \tfrac{1}{2} \lambda_i (\delta \boldsymbol{U} \cdot \boldsymbol{b}_i) . \tag{2.74}$$

As long as $\varepsilon > 1$, this would imply (2.59) and with $\mathrm{Re}\to\infty$ would converge to the original essential boundary conditions. Boundary conditions of this type resemble those obtained when prescribing acoustical impedance in linear acoustics. It is not clear, however, how they can be generalized to the nonlinear case.

Case 3. Open boundary—subsonic outflow $u_n > 0$, $u_n - c < 0$

Only three boundary conditions are to be prescribed. An immediate difficulty arises in the necessity of satisfying (2.59), which involves four independent terms $\delta \boldsymbol{U} \cdot \boldsymbol{b}_i$ using only three conditions. This is resolved by observing that once

$$\delta u_n = 0 \tag{2.75}$$

(implying that u_n is fixed), the first term in (2.59) is always nonnegative. Indeed, according to (2.61), (2.75) implies that

$$\delta \boldsymbol{U} \cdot \boldsymbol{b}_1 = \delta \boldsymbol{U} \cdot \boldsymbol{b}_4 , \tag{2.76}$$

and therefore

$$(u_n - c)(\delta \boldsymbol{U} \cdot \boldsymbol{b}_1)^2 + (u_n + c)(\delta \boldsymbol{U} \cdot \boldsymbol{b}_4)^2 = 2u_n (\delta \boldsymbol{U} \cdot \boldsymbol{b}_1)^2 \geq 0 . \tag{2.77}$$

Two remaining boundary conditions are specified of the form

$$\delta t_n \cdot (b_1 + b_4) = 0\,, \qquad \delta t_n \cdot (b_2) = 0\,. \tag{2.78}$$

Orthogonality condition (2.66) implies that the remaining condition

$$\delta t_n \cdot b_3 = 0 \tag{2.79}$$

is satisfied automatically.

Conditions (2.78) are equivalent to

$$\delta \tau_{ns} = \frac{\partial\, \delta u_s}{\partial n} = 0\,, \qquad \frac{\partial\, \delta \theta}{\partial n} = 0\,, \tag{2.80}$$

with the corresponding nonlinear boundary conditions

$$u_n\,, \quad \tau_{ns}\left(\text{or } \frac{\partial u_s}{\partial n}\right), \quad \frac{\partial \theta}{\partial n} \text{ are prescribed}\,. \tag{2.81}$$

Case 4. Open boundary—supersonic outflow $u_n - c > 0$

This is the simplest case as the first term in (2.59) is always nonnegative. In order to make the second term vanish, we assume

$$\delta t_n \cdot b_i = 0\,, \quad i = 1, 2, 4\,. \tag{2.82}$$

Since the orthogonality condition (2.66) implies also that

$$\delta t_n \cdot b_3 = 0\,, \tag{2.83}$$

this is equivalent to the condition

$$\delta t_n = \mathbf{0}\,. \tag{2.84}$$

(Note that, due to (2.66), only three independent conditions are prescribed.)

The corresponding nonlinear boundary condition will be of the form

$$t_n \text{ is prescribed}\,. \tag{2.85}$$

Note that all three criteria listed in the beginning of this section have been met. As the Reynolds number Re approaches infinity, the natural boundary conditions in the limit are satisfied automatically and the essential boundary conditions (possibly conditions of mixed type) converge to the correct boundary conditions (also in number) for the Euler equations.

Case 5. Solid wall—heat flux prescribed

Formally, the solid wall boundary can be classified as a special case of subsonic outflow ($u_n = 0$) which implies that only three boundary conditions should be specified. The major

difference is only in prescribing the tangential velocity as zero, i.e., the two essential boundary conditions are

$$\delta u_n = 0\,, \qquad \delta u_s = 0\,. \tag{2.86}$$

As in the case of subsonic outflow, this implies that the first term in (2.59) is always nonnegative. Once δu_s is eliminated, it follows from (2.61) that

$$\delta \boldsymbol{U} \cdot \boldsymbol{b}_2 = 0 \tag{2.87}$$

and, finally, by prescribing one extra condition of the form

$$\partial\, \delta\theta / \partial n = 0\,, \tag{2.88}$$

we force the second term in (2.59) to vanish.

The corresponding nonlinear boundary conditions are

$$u_n = 0\,, \quad u_s = 0\,, \quad \frac{\partial \theta}{\partial n} \text{ (heat flux) is prescribed}\,. \tag{2.89}$$

Case 6. Solid wall—temperature prescribed

Surprisingly, this case presents some unexpected difficulties. Replacing (2.87) with the condition

$$\delta\theta = 0 \quad (\theta \text{ prescribed})\,, \tag{2.90}$$

completes the number of boundary conditions (only three) but does not imply the vanishing of the second term in (2.59). That this is the case is seen by representing the scalar products in (2.61) in a form including an increment of temperature $\delta\theta$. With (2.86) satisfied, we obtain (using the relation $p = (\gamma - 1)c_v \rho\theta$)

$$\begin{aligned} \delta \boldsymbol{U} \cdot \boldsymbol{b}_1 &= \frac{1}{2\gamma}\,\delta\rho + \frac{\rho}{2\gamma}\,\frac{\delta\theta}{\theta}\,, \qquad \delta \boldsymbol{U} \cdot \boldsymbol{b}_3 = -\frac{1}{\gamma}\,\delta\rho - \frac{\rho}{\gamma}\,\frac{\delta\theta}{\theta} + \delta\rho\,, \\ \delta \boldsymbol{U} \cdot \boldsymbol{b}_4 &= \frac{1}{2\gamma}\,\delta\rho + \frac{\rho}{2\gamma}\,\frac{\delta\theta}{\delta\theta}\,. \end{aligned} \tag{2.91}$$

If (2.89) holds, the second term in (2.59) still takes the form

$$\left[\frac{1}{2\gamma}\,\delta \boldsymbol{t}_n \cdot (\boldsymbol{b}_1 + \boldsymbol{b}_4) + \left(1 - \frac{1}{\gamma}\right)\delta \boldsymbol{t}_n \cdot \boldsymbol{b}_3\right] \delta\rho\,, \tag{2.92}$$

which, in accordance with (2.65) and (2.66), will reduce to zero only if $\delta\rho = 0$ or

$$1/2\gamma = (1 - 1/\gamma)\,, \tag{2.93}$$

i.e., for $\gamma = 1.5$. In other words, prescribing temperature on the solid wall boundary does not guarantee the stability of the linearized problem. Of course, such a conclusion may be questionable as it is based on the linearized Navier–Stokes equations and the assumption that

the derivatives of the original solution are small, which may be invalid in regions near boundary layers or shocks.

We shall see in the next section that this difficulty disappears in the context of an operator splitting method.

3. Operator splitting of the Navier–Stokes equations

3.1. The structure of the operator splitting method

Using notations from the previous section, we define

(1) *The convection operator*

$$(\boldsymbol{E}(t)\boldsymbol{U}_0)(\boldsymbol{x}) \overset{\text{def}}{=} \boldsymbol{U}(\boldsymbol{x}, t)\,, \tag{3.1}$$

where $\boldsymbol{U}(\boldsymbol{x}, t)$ is the solution to the system of Euler equations (transport step)

$$\boldsymbol{U}_{,t} + \sum_{i=1}^{2} \boldsymbol{F}^i(\boldsymbol{U})_{,i} = \boldsymbol{0}\,, \tag{3.2}$$

with the initial condition

$$\boldsymbol{U}(\boldsymbol{x}, 0) = \boldsymbol{U}_0(\boldsymbol{x}) \tag{3.3}$$

and boundary conditions to be specified later,

(2) *The diffusion operator*

$$(\boldsymbol{H}(t)\boldsymbol{U}_0)(\boldsymbol{x}) \overset{\text{def}}{=} \boldsymbol{U}(\boldsymbol{x}, t)\,, \tag{3.4}$$

where $\boldsymbol{U}(\boldsymbol{x}, t)$ is the solution to the system of equations (viscous step)

$$\boldsymbol{U}_{,t} = \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \boldsymbol{K}^{ij}(\boldsymbol{U})\boldsymbol{U}_{,j} \right)_{,i}\,, \tag{3.5}$$

with initial condition (3.3) and boundary conditions to be discussed later.

Two compositions of the operators $\boldsymbol{H}$ and $\boldsymbol{E}$ are considered, a two-step splitting of the form

$$\boldsymbol{G}(t) = \boldsymbol{H}(t)\boldsymbol{E}(t) \tag{3.6}$$

and a three-step Strang procedure of the form

$$\boldsymbol{S}(t) = \boldsymbol{H}(\tfrac{1}{2}t)\boldsymbol{E}(t)\boldsymbol{H}(\tfrac{1}{2}t)\,. \tag{3.7}$$

It can be shown (see [2]) that the first procedure is of first order while the second is of second-order in time, i.e.,

$$|\boldsymbol{G}(t)\boldsymbol{U}_0(\boldsymbol{x}) - \boldsymbol{U}(\boldsymbol{x}, t)| \leq c(\boldsymbol{x})t^2\,, \qquad |\boldsymbol{S}(t)\boldsymbol{U}_0(\boldsymbol{x}) - \boldsymbol{U}(\boldsymbol{x}, t)| \leq c(\boldsymbol{x})t^3\,, \tag{3.8}$$

where $U(x, t)$ is a solution to the full Navier–Stokes equations with initial condition (3.3) and $c(x)$ is an unknown function of x. When proving (3.8), no boundary conditions are taken into account. With improper boundary conditions used to identify convection and diffusion operators, $c(x)$ with 'blow up' as x approaches the boundary and the procedure fails.

In the two subsequent sections, we discuss an example of the methods used to solve both fractional steps. At this point, selection of the methods is completely arbitrary, any two particular procedures to solve the Euler equations and the viscous step can be put together by means of (3.6) or (3.7).

The second-order accuracy of procedure (3.7) suggests that second-order (in time) methods should be used, especially if we are interested in solving transient problems.

Another obvious requirement is that both methods should use a compatible approximation in the space variables. In our case, this will be accomplished by means of an *h-p* FE approximation.

3.2. *Solution of the Euler equations by the Taylor–Galerkin method—A summary*

For definiteness, the Taylor–Galerkin method presented in [20] is used to solve the Euler step in the operator splitting method. A brief summary of pertinent results is given in this section.

As a starting point, the following Taylor series expansion in time is used:

$$U(t+\Delta t) - \alpha \tfrac{1}{2}\Delta t^2 U_{,tt}(t+\Delta t) = U(t) + \Delta t U_{,t} + (1-\alpha)\tfrac{1}{2}\Delta t^2 U_{,tt}(t) + \mathrm{O}(\Delta t^3)\,, \tag{3.9}$$

where the parameter of implicitness $\alpha \in [0, 1]$. Replacing the time derivatives with spatial derivatives calculated using the Euler equations (3.2), we arrive at a second-order system of differential equations of the form

$$U^{n+1} - \alpha \frac{\Delta t^2}{2}\left[\sum_{i,j=1}^{2} A^i A^j U^{n+1}_{,j}\right] = U^n - \Delta t\left[\sum_{i=1}^{2} F^i_{,i}\right] + (1-\alpha)\frac{\Delta t^2}{2}\left[\sum_{i,j=1}^{2} A^i A^j U^n_{,j}\right]_{,i}. \tag{3.10}$$

Evaluating Jacobian matrices on the left-hand side as functions of U^n does not destroy the second-order time accuracy and makes the problem linear.

A weak formulation of (3.10) is obtained in the usual way: we first multiply (3.10) by a test function W, integrate over Ω and integrate by parts arriving at the problem

Find U^{n+1} such that

$$\begin{aligned}
&\int_\Omega W^{\mathrm{t}} U^{n+1}\,\mathrm{d}x + \alpha\frac{\Delta t^2}{2}\int_\Omega \sum_{i,j=1}^{2} W^{\mathrm{t}}_{,i} A^i A^j U^{n+1}_{,j}\,\mathrm{d}x + \alpha\frac{\Delta t^2}{2}\int_{\partial\Omega} W^{\mathrm{t}} A^n U^{n+1}_{t}\,\mathrm{d}s \\
&\quad = \int_\Omega W^{\mathrm{t}} U^n\,\mathrm{d}x + \Delta t\int_\Omega \sum_{i=1}^{2} W^{\mathrm{t}}_{,i} F^i(U^n)\,\mathrm{d}x - (1-\alpha)\frac{\Delta t^2}{2}\int_\Omega \sum_{i,j=1}^{2} W^{\mathrm{t}}_{,i} A^i A^j U^n_{,j}\,\mathrm{d}x \\
&\qquad - \Delta t\int_{\partial\Omega} W^{\mathrm{t}} F^i(U^n) n_i\,\mathrm{d}s - (1-\alpha)\frac{\Delta t^2}{2}\int_{\partial\Omega} W^{\mathrm{t}} A^n U^n_{t}\,\mathrm{d}s \\
&\qquad\quad \text{for every } W\,,
\end{aligned} \tag{3.11}$$

where A^n is the normal Jacobian matrix and U_t stands for

$$U_t = -\sum_{i=1}^{2} A^i U_{,i} . \tag{3.12}$$

Evaluation of the second-order boundary term on the left-hand side of the equation will depend upon the type of boundary conditions imposed. The various cases are discussed next.

CASE 1. Open boundary—supersonic inflow

As in Section 2.6, we use the right and left eigenvectors of the normal Jacobian matrix to represent the boundary term in the form

$$\alpha \tfrac{1}{2}\Delta t^2 \sum_{i=1}^{4} \lambda_i (U_t \cdot b_i)(W \cdot c_i) . \tag{3.13}$$

The essential inflow boundary conditions

$$U^{n+1} = U^{\text{in}} \tag{3.14}$$

are applied by means of a penalty method, replacing (3.14) by

$$(U^{n+1} \cdot b_i) - \varepsilon \frac{\Delta t^2}{2} \lambda_i (U_t \cdot b_i) = (U^{\text{in}} \cdot b_i) , \tag{3.15}$$

where $\varepsilon > 0$ is a (small) penalty parameter. Use of this condition results in contributions to the left and right sides of (3.14) (the stiffness matrix and load vector) of terms

$$\frac{1}{\varepsilon} \int_{\partial\Omega} \sum_{i=1}^{4} (U^{n+1} \cdot b_i)(W \cdot c_i) \, ds \tag{3.16}$$

and

$$\frac{1}{\varepsilon} \int_{\partial\Omega} \sum_{i=1}^{4} (U^{\text{in}} \cdot b_i)(W \cdot c_i) \, ds , \tag{3.17}$$

respectively.

CASE 2. Open boundary—Supersonic outflow

The term U_t in (3.12) is interpreted as the time derivative and is approximated with the finite difference formula

$$U_t^{n+1} = (U^{n+1} - U^n)/\Delta t . \tag{3.18}$$

Note that use of approximation (3.18) does not destroy the second-order accuracy in time.

The corresponding contributions to the stiffness matrix and load vector are

$$\frac{\alpha \, \Delta t}{2} \int_{\partial\Omega} \sum_{i=1}^{4} \lambda_i (U^{n+1} \cdot b_i)(W \cdot c_i) \, ds \tag{3.19}$$

and

$$\frac{\alpha \Delta t}{2} \int_{\partial\Omega} \sum_{i=1}^{4} \lambda_i (\boldsymbol{U}^n \cdot \boldsymbol{b}_i)(\boldsymbol{W} \cdot \boldsymbol{c}_i)\, \mathrm{d}s\,, \tag{3.20}$$

respectively.

CASE 3. Open boundary—subsonic inflow and outflow

Formally, the approximation (3.18) is applied for positive eigenvalues λ_i and the penalty formulation (3.15) for negative eigenvalues. The essential difference is in the choice of penalty parameters ε which, for the ith negative eigenvalue, is selected so that

$$1/\varepsilon = -\alpha \tfrac{1}{2} \Delta t \lambda_i\,, \tag{3.21}$$

i.e., the resulting terms in the stiffness matrix and load vector are

$$\alpha \frac{\Delta t}{2} \sum_{i=1}^{4} |\lambda_i| (\boldsymbol{U}^{n+1} \cdot \boldsymbol{b}_i)(\boldsymbol{W} \cdot \boldsymbol{c}_i) \tag{3.22}$$

and

$$\alpha \frac{\Delta t}{2} |\lambda_i| (\boldsymbol{U}^{\text{in}} \cdot \boldsymbol{b}_i)(\boldsymbol{W} \cdot \boldsymbol{c}_i) \quad \text{for } \lambda_i < 0\,, \qquad \alpha \frac{\Delta t}{2} |\lambda_i| (\boldsymbol{U}^n \cdot \boldsymbol{b}_i)(\boldsymbol{W} \cdot \boldsymbol{c}_i) \quad \text{for } \lambda_i > 0\,. \tag{3.23}$$

In all the cases discussed thus far, for Δt approaching zero, the boundary conditions corresponding to $\lambda_i > 0$ are satisfied automatically while those for $\lambda_i < 0$ converge to Dirichlet boundary conditions. For $\lambda_i = 0$, the boundary coincides with a characteristic surface for the Euler equations and no boundary condition need be specified.

CASE 4. Solid wall boundary

The only boundary condition to be specified is

$$u_{\text{n}} = 0\,. \tag{3.24}$$

It can be shown (see [20]) that once (3.24) is satisfied, the entire boundary term vanishes. In practice, (3.24) is enforced by a penalty method adding to the stiffness matrix a term of the form

$$\frac{1}{\varepsilon} \int_{\partial\Omega} u_{\text{n}} \cdot \boldsymbol{P} \boldsymbol{A}_0^{-1} \cdot \boldsymbol{W}\, \mathrm{d}s\,, \tag{3.25}$$

where $\boldsymbol{A}_0^{-1}$ is the inverse of the symmetrizer $\boldsymbol{A}_0$ and P is a projection of the form

$$\boldsymbol{P}\boldsymbol{W} = P(W_1, W_2, W_3, W_4) = W_2 n_1 + W_3 n_2\,. \tag{3.26}$$

This form of the penalty term guarantees that, after symmetrization, (3.25) will result in a positive definite and symmetric contribution of the penalty term.

For curved boundaries, in order to avoid locking phenomena (creation of artificial, discrete stagnation points), a Gauss–Lobatto underintegration combined with a continuous regularization of the outward normal unit vector must be used. Details on these issues are given in [20].

CASE 5. Open boundary. Symmetry boundary conditions II (of the second type)

Suppose a flow has stabilized across an open boundary $\partial\Omega$, i.e., using the normal–tangential system of coordinates, we have

$$\boldsymbol{U}(-n, s) = \boldsymbol{U}(n, s)\,. \tag{3.27}$$

Assuming additionally that the flow is smooth, (3.27) results in a boundary condition

$$\frac{\partial \boldsymbol{U}}{\partial n} = 0\,. \tag{3.28}$$

Rewriting the implicit boundary term in the form

$$-\alpha\,\frac{\Delta t^2}{2}\int_{\partial\Omega} \boldsymbol{W}^{\mathrm{t}}\boldsymbol{A}_{\mathrm{n}}\left(\boldsymbol{A}_{\mathrm{n}}\,\frac{\partial \boldsymbol{U}}{\partial n} + \boldsymbol{A}_{\mathrm{s}}\,\frac{\partial \boldsymbol{U}}{\partial s}\right)\mathrm{d}s\,, \tag{3.29}$$

we see that once (3.28) is satisfied, one of the terms in (3.29) vanishes while the other remains. The remaining term,

$$-\frac{\alpha\,\Delta t^2}{2}\int_{\partial\Omega} \boldsymbol{W}^{\mathrm{t}}\boldsymbol{A}_{\mathrm{n}}\boldsymbol{A}_{\mathrm{s}}\,\frac{\partial \boldsymbol{U}}{\partial s}\,\mathrm{d}s \tag{3.30}$$

must be left as an extra contribution to the stiffness matrix. For details see [21].

Boundary conditions of this type have proved to be very successful in modeling both subsonic and supersonic outflow boundary conditions, both for Euler and Navier–Stokes equations. There remains open the issue of interpreting this condition in the context of the stability analysis presented earlier.

CASE 6. Open boundary. Symmetry boundary conditions I (of the first type)

In the case of physical symmetry of the flow, the resulting boundary conditions are

$$\frac{\partial \rho}{\partial n} = m_{\mathrm{n}} = \frac{\partial m_{\mathrm{s}}}{\partial n} = \frac{\partial e}{\partial n} = 0\,, \tag{3.31}$$

where m_{n} and m_{s} are the normal and tangential components of the momentum. Vanishing of the normal momentum (velocity) is enforced as on the solid wall boundary, by adding the penalty term (3.25). Once again, vanishing of the normal velocity eliminates all remaining boundary terms. The only practical difference between these conditions and the solid wall case is that no special integration is needed here, as the symmetry conditions are applied along a segment of a straight line. An additional slight difference is encountered in treating the boundary terms resulting from artificial viscosity (see [20]).

Linear stability

The Taylor–Galerkin method for solving the Euler equations is unconditionally stable for $\alpha \geq 0.5$ independently of the approximation in the space variables. The second-order terms present on the left-hand side modify the L^2-projection and contribute to the stabilization of the method. The remaining details are given in [20].

3.3. Solution of the viscous step

The form of the differential equations (3.5) defining the viscous step leads to two simple observations:

— The density ρ remains unchanged in the viscous step, i.e.,

$$\rho^{n+1} = \rho^n \,. \tag{3.32}$$

— The remaining three equations can be decoupled. We first solve a system of two equations for the momentum components

$$m_{i,t} = \sum_{j=1}^{2} \tau_{ij,j} \tag{3.33}$$

and then a single equation for the total energy

$$e_{,t} = \sum_{j=1}^{2} \tau_{ij} u_j + q_i \,. \tag{3.34}$$

To affect the decoupling, boundary conditions for (3.33) must not contain energy terms.

As a starting point to solve both (3.33) and (3.34), we employ a first-order finite difference formula of the form

$$\boldsymbol{U}(t+\Delta t) - \beta\,\Delta t\,\boldsymbol{U}_{,t}(t+\Delta t) = \boldsymbol{U}(t) + (1-\beta)\,\Delta t\,\boldsymbol{U}_{,t}(t) + \mathrm{O}(\Delta t^2) \,. \tag{3.35}$$

Note that for $\beta = \frac{1}{2}$, (3.35) is of second-order and reduces to a Crank–Nicholson scheme.

Replacing the time derivatives with spatial derivatives in (3.35), (3.33) is approximated by a system of two equations of the form

$$m_j^{n+1} - \beta\,\Delta t \sum_{i=1}^{2} \tau_{ij,i}^{n+1} = m_j^n + (1-\beta)\,\Delta t \sum_{i=1}^{2} \tau_{ij,i}^n \,. \tag{3.36}$$

Equations (3.36), if rewritten in terms of the velocity components, reduce to a system of two symmetric, elliptic, equations. Unfortunately, in order to comply with the conservative form of the equations, (3.36) must be solved in momentum components.

Next, (3.36) are linearized by evaluating the viscosities μ and λ (not the whole viscous matrices $\boldsymbol{K}^{ij}$) as functions of $\boldsymbol{U}^n$ rather than $\boldsymbol{U}^{n+1}$. For $\beta = \frac{1}{2}$, second-order accuracy in time is formally lost, but this is often considered to be an acceptable simplification. Once the system (3.36) is solved, one proceeds with the energy equation. Combining (3.35) with (3.34) gives a single equation of the form

$$e^{n+1} - \beta\,\Delta t \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \tau_{ij}^{n+1} u_j^{n+1} + \kappa \theta_{,i}^{n+1} \right)_{,i} = e^n + (1-\beta)\,\Delta t \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \tau_{ij}^n u_j^n + \kappa \theta_{,i}^n \right)_{,i} , \tag{3.37}$$

which, when rewritten in terms of temperature, takes the form of a single, elliptic equation.

3.4. Variational formulation for the momentum equations. Boundary conditions

Multiplying (3.36) with a test function δu_j, integrating over Ω and integrating by parts, we arrive at the variational formulation:

Find m_j^{n+1}, $j = 1, 2$ such that

$$\int_\Omega \sum_{i=1}^{2} m_i^{n+1}\,\delta u_i\,\mathrm{d}x + \beta\,\Delta t \int_\Omega \sum_{i,j=1}^{2} \tau_{ij}^{n+1}\,\delta u_{i,j}\,\mathrm{d}x - \beta\,\Delta t \int_{\partial\Omega} \sum_{i,j=1}^{2} \tau_{ij}^{n+1} n_j\,\delta u_i\,\mathrm{d}s$$
$$= \int_\Omega \sum_{i=1}^{2} m_i^n\,\delta u_i\,\mathrm{d}x - (1-\beta)\,\Delta t \int_\Omega \sum_{i,j=1}^{2} \tau_{ij}^n\,\delta u_{i,j}\,\mathrm{d}x$$
$$+ (1-\beta)\,\Delta t \int_{\partial\Omega} \sum_{i,j=1}^{2} \tau_{ij}^n n_j\,\delta u_i\,\mathrm{d}s\,, \tag{3.38}$$

where the viscous stresses τ_{ij} are evaluated using (2.9) with viscosities μ and λ evaluated as functions of U^n. Derivatives $u_{i,j}$ in (2.9) are calculated in terms of derivatives of the momentum components and density,

$$\frac{\partial u_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{m_i}{\rho}\right) = -\frac{m_i}{\rho^2}\frac{\partial \rho}{\partial x_j} + \frac{1}{\rho}\frac{\partial m_i}{\partial x_j}\,. \tag{3.39}$$

This results in the presence of first-order terms in (3.38) and makes the bilinear form on the left-hand side of (3.38) unsymmetric. In order to recover symmetry, we represent the test function δu_i in terms of 'virtual momentum components', i.e., make the substitution

$$\delta u_i = \delta m_i/\rho\,. \tag{3.40}$$

The resulting bilinear form in terms of m_j^{n+1} and δm_j is symmetric and positive definite. The evaluation of the boundary integral on the implicit side of (3.38) depends on the boundary conditions. The following cases are taken into account:

CASE 1. Open boundary—supersonic inflow

Full Dirichlet boundary conditions are prescribed,

$$m_i^{n+1} = m_i^{\mathrm{in}}\,, \tag{3.41}$$

where m_i^{in} are the momentum components of the same inflow vector as used in the supersonic inflow boundary conditions for the Euler step.

CASE 2. Open boundary—subsonic inflow

The same Dirichlet boundary conditions are used, but with the inflow vector replaced with the solution from the Euler step, i.e.,

$$m_i^{n+1} = m_i^n\,. \tag{3.42}$$

CASE 3. Open boundary—subsonic outflow
Mixed boundary conditions are used,

$$m_{\mathrm{n}}^{n+1} = m_{\mathrm{n}}^{n}\,, \qquad \tau_{\mathrm{s}}^{n+1} = \tau_{\mathrm{s}}^{n}\,, \tag{3.43}$$

where m_{n} is the normal component of the momentum,

$$m_{\mathrm{n}} = m_1 n_1 + m_2 n_2\,, \tag{3.44}$$

and τ_{s} is the tangential viscous stress vector component,

$$\tau_{\mathrm{s}} = (\tau_{22} - \tau_{11}) n_1 n_2 + \tau_{12}(n_1^2 - n_2^2)\,. \tag{3.45}$$

CASE 4. Open boundary—supersonic outflow
Full Neuman boundary conditions are applied,

$$\tau_{ij}^{n+1} n_j = \tau_{ij}^{n} n_j\,. \tag{3.46}$$

CASE 5. Solid wall boundary conditions
Full Dirichlet boundary conditions are used,

$$m_i^{n+1} = 0\,. \tag{3.47}$$

CASE 6. Symmetry boundary conditions (of the first kind)
Mixed boundary conditions are applied,

$$m_{\mathrm{n}}^{n+1} = 0\,, \qquad \frac{\partial m_{\mathrm{s}}^{n+1}}{\partial n} = 0\,, \tag{3.48}$$

where m_{n} and m_{s} are the normal and tangential components of the momentum vector.

CASE 7. Symmetry boundary conditions (of the second kind)
Full Neuman boundary conditions are applied,

$$\partial m_i^{n+1} / \partial n = 0\,. \tag{3.49}$$

As in the case of Euler equations, substituting (3.49) into the boundary integral in (3.38), results in some non-zero terms which must be included in the stiffness matrix calculations (see [21]).

In practice, all essential boundary conditions are implemented using the penalty approach, i.e., replacing the full Dirichlet boundary conditions (3.42) with

$$m_i^{m+1} + \varepsilon\beta\,\Delta t \sum_{j=1}^{2} \tau_{ij}^{n+1} n_j = m_i^{\mathrm{in}}\,, \tag{3.50}$$

and the first of (3.42) conditions with

$$m_{\mathrm{n}}^{n+1} + \varepsilon\beta\,\Delta t\,\tau_{\mathrm{n}}^{n+1} = m_{\mathrm{n}}^{n}\,, \tag{3.51}$$

where τ_{n}^{n+1} is the normal viscous stress,

$$\tau_{\mathrm{n}} = \sum_{i,j=1}^{2} \tau_{ij} n_i n_j\,. \tag{3.52}$$

3.5. Variational formulation for the energy equation. Boundary conditions

We proceed along the same lines with the energy equations (3.37). The resulting variational formulation is then

Find e^{n+1} such that

$$\begin{aligned}
&\int_\Omega e^{n+1}\,\delta\theta\,\mathrm{d}x + \beta\,\Delta t \int_\Omega \kappa \sum_{i=1}^{2} \theta_{,i}^{n+1}\,\delta\theta_{,i}\,\mathrm{d}x - \beta\,\Delta t \int_{\partial\Omega} \kappa \sum_{i-1}^{2} \theta_{,i}^{n+1} n_i\,\delta\theta\,\mathrm{d}s \\
&\quad = \int_\Omega e^{n}\,\delta\theta\,\mathrm{d}x - (1-\beta)\,\Delta t \int_\Omega \kappa \sum_{i=1}^{2} \theta_{,i}^{n}\,\delta\theta_{,i}\,\mathrm{d}x + (1-\beta)\,\Delta t \int_{\partial\Omega} \kappa \sum_{i-1}^{2} \theta_{,i}^{n} n_i\,\delta\theta\,\mathrm{d}s \\
&\qquad + \Delta t \int_\Omega \left\{ \beta \sum_{i,j=1}^{2} \tau_{ij}^{n+1} u_j^{n+1} + (1-\beta) \sum_{i,j=1}^{2} \tau_{ij}^{n} u_j^{n} \right\} \mathrm{d}x \quad \text{for arbitrary } \delta\theta\,.
\end{aligned} \tag{3.53}$$

The temperature θ is evaluated using (2.5), (2.7), which imply the following formula for the derivatives:

$$\theta_{,i} = \frac{1}{c_{\mathrm{v}}} \left\{ \frac{1}{\rho}\, e_{,i} - \frac{e}{\rho^2} \sum_{j=1}^{2} m_j m_{j,i} + \frac{1}{\rho^3} \sum_{j=1}^{2} m_j^2 \rho_{,i} \right\}. \tag{3.54}$$

As in the case of the momentum equations, the resulting formulation is unsymmetric and the symmetry of the problem can be recovered by representing $\delta\theta$ in terms of δe,

$$\delta\theta = \frac{1}{c_{\mathrm{v}}}\,\frac{1}{\rho}\,\delta e\,. \tag{3.55}$$

The evaluation of the boundary integral on the implicit side of (3.53) depends again upon boundary conditions.

CASE 1. Temperature (energy) prescribed

A single Dirichlet boundary condition is applied,

$$e^{n+1} = \bar{e}\,. \tag{3.56}$$

The choice of $\bar{e}$ will vary with the particular kind of boundary:
— for supersonic inflow $\bar{e}$ corresponds to the inflow vector;
— for subsonic inflow $\bar{e}$ is evaluated using the solution from the previous step;

— for a solid wall with temperature prescribed $\bar{e}$ corresponds to the prescribed temperature on the wall and the density ρ (unchanged in the viscous step).

Note that the operator splitting eliminates the contradictions resulting from the discussion of the solid wall boundary conditions with temperature prescribed, as the density remains unchanged (recall (2.92)).

CASE 2. Heat flux prescribed

A single Neumann boundary condition is applied,

$$\kappa \, \partial\theta^{n+1}/\partial n = \bar{q} \, . \tag{3.57}$$

The heat flux $\bar{q}$ is calculated in the following way:

— for subsonic and supersonic outflow, $\bar{q}$ is evaluated using the solution from the previous step,

— for an adiabatic wall and symmetry boundary conditions of both kinds, $\bar{q}$ is assumed to be zero.

3.6. Boundary conditions—concluding remarks

We conclude this section with a brief discussion on the implemented boundary conditions in conjunction with the analysis presented in Section 2. The main idea is that, as the time step Δt converges to zero, the boundary conditions implemented for various fractional time steps should result in well-posed boundary conditions for the original problem.

Supersonic inflow

There is no problem with the supersonic inflow boundary conditions. The full Dirichlet boundary conditions used in both the Euler and viscous steps are independent of the time step and are of the same form as those for the original equations. The free stream vector $\boldsymbol{U}^{\infty}$ is implemented as an inflow vector.

Solid wall

In both cases of temperature and heat flux prescribed, the solution to the Navier–Stokes equations has to be sampled after the viscous step. The essential boundary conditions implemented in the viscous step are then the same as for the original equations. The same concerns the heat flux boundary condition.

Subsonic inflow and outflow

For negative eigenvalues λ_i, the combination of mixed boundary conditions in the Euler step and the Dirichlet boundary conditions in the viscous step guarantees that in the limit the Dirichlet boundary conditions in the form of the linearized Navier–Stokes equations are recovered. This is the first inconsistency, as these boundary conditions cannot be obtained by linearizing some nonlinear conditions for the original equations.

The situation is much worse in the case of natural boundary conditions. As all natural boundary conditions for $\boldsymbol{U}^{n+1}$ in both the Euler and viscous steps are expressed in terms of the previous step solution $\boldsymbol{U}^{n}$, in the limit $(\Delta t \to 0)$ they are satisfied automatically. This formally means that no natural boundary conditions need be prescribed. This certainly resolves

practically the problem of specifying the viscous fluxes in the natural boundary conditions presented in Section 2, but it is inconsistent with the linear theory presented earlier.

A partial justification of such a procedure is as follows. As the viscous flow effects are of interest only in a small neighborhood of the solid wall boundary, viscosity constants (functions of temperature in fact) could be assumed to vanish (converge to zero) away from the solid wall boundary. As a result, at a far enough distance, solutions to the Navier–Stokes equations closely approximate solutions to the Euler equations. Therefore, as long as the open boundary is far enough away from the solid wall, the implemented boundary conditions should be consistent with the analysis for the Euler equations only.

This type of reasoning cannot, of course, be justified if the open boundary intersects the boundary layer.

Symmetry boundary conditions

The symmetry boundary conditions, of both the first and second kinds, are defined in exactly the same way for both the Euler and viscous time steps. As they are independent of the time step Δt, in the limit, the same boundary conditions are applied for the Navier–Stokes equations. The symmetry boundary conditions of the first kind fall into the category of open boundary subsonic outflow conditions (the normal velocity is zero) and, therefore, are consistent with the analysis presented in Section 2. It is not clear how one should analyze the symmetry boundary conditions of the second kind. From the practical point of view, as will be discussed in the next section, they have been successfully used to model not only open boundaries (as an alternative for the open boundary subsonic inflow and both outflow boundary conditions) but also the part of the open boundary crossing the boundary layer.

4. *h-p* adaptivity

4.1. A posteriori error estimation—A summary

The two essential ingredients of adaptive computational methods are
— an a posteriori error estimation technique,
— an adaptive strategy.

The goal of the a posteriori error estimation is at least two-fold. First, it assures that a certain preassigned accuracy of the solution has been achieved, thus guaranteeing the reliability of the numerical solution in hand. At the same time, in most practical implementations, the estimated error distribution is used as a basis for control of possible mesh modifications. Both functions, the error estimation and the choice of an adaptive strategy, are challenging research areas, especially in the context of the *h-p* approximations discussed in this paper. As many possible variations of *h-p* meshes may be allowed to occur during the solution, the error estimation method must be sufficiently stable to handle variations in element size h and spectral order of approximation p. In turn, an adaptive strategy must not only resolve the question of where to refine the mesh, but also determine how to refine it, i.e., whether an *h*- or a *p*-refinement should be selected.

In [22], an extensive study on various error estimation techniques for arbitrary *h-p* meshes and scalar elliptic equations is presented. A result of that investigation was that the Element

Residual Method (ERM) provides an excellent choice of a reliable and inexpensive error estimation technique. Since a typical Euler step in the operator splitting method with the regularization resulting from the Taylor–Galerkin discretization can be viewed as a system of (elliptic-like) second order linear equations, the ERM can be used to estimate the error resulting from the spatial approximation of a single Euler step. The same reasoning applies to both viscous fractional steps as they are identified as a system of two elliptic equations and a single elliptic (heat) equation. Obviously, such an error estimation does not take into account the error resulting from either the operator splitting procedure or the discretization in time. Practically, this means that the time step Δt must be small enough to guarantee that the dominating part of the error is due to the spatial discretization. A full, complete analysis taking both temporal and spatial discretizations into account remains an open task.

Another nontrivial issue concerning the application of ERM is that it was originally developed in the context of elliptic operators with a dominating symmetric part (see [22]), while neither the Euler step nor the viscous step in our formulations are symmetric. This problem has been successfully resolved using the idea of a symmetrizer and symmetrizable problems. The whole procedure and the corresponding theory are to be described in detail in a forthcoming paper [23].

The basic idea can be explained as follows. Let $B(U, W)$ denote the bilinear form corresponding to the left-hand sides of (3.11) for the Euler step, or (3.38) and (3.53) for fractional viscous time steps. None of these forms is symmetric, i.e.

$$B(U, W) \neq B(W, U)\,, \tag{4.1}$$

or positive definite. In each of the cases, however, we can introduce a positive definite symmetric matrix $A_0 = A_0(x)$, called the symmetrizer, such that a new bilinear form

$$\tilde{B}(U, W) \overset{\text{def}}{=} B(U, A_0 W) \tag{4.2}$$

is both symmetric and positive definite.

The choice of a symmetrizer for both viscous steps is very natural: in both cases the unsymmetry of the problem is a result of a discrepancy between the solution variables (momentum components, total energy) and the test functions (velocity components, temperature). As the solution variables are fixed (use of the conservative variables is one of the main issues in the present approach), the symmetry must be recovered by changing the test variables, i.e., representing the virtual velocities and temperature in terms of a prescribed density (fixed in the viscous step) and virtual momentum components or virtual energy. Formally this is done by introducing the symmetrizer of the form

$$A_0 = (1/\rho)\mathbf{I}\,, \tag{4.3}$$

where ρ is the density function and $\mathbf{I}$ is the identity operator.

In the Euler step, the symmetrizer A_0 introduced in Section 2 is used with the same consistent assumption that its derivatives are neglected. Note that this not only simplifies the calculation of the symmetrized bilinear form, but in fact only then is the new bilinear form

symmetric and positive definite. (This also pertains to the discrete problem, independent of the numerical quadrature rule used.)

Having introduced the notion of symmetrized problems, we proceed now with a short summary of the ERM. The procedure consists of two steps. The first step is independent of the form of the symmetrizer (which, incidentally, is not unique) and consists in solving for each finite element K a local variational problem of the form

$$\text{Find } \varphi_K \in X^0_{h,p+1}(K) \text{ such that}$$

$$B_K(\varphi_K, W) = L_K(W) - B_K(U_{h,p}, W) + \int_{\partial K \setminus \partial\Omega} \left[\left[\sum_{i,j=1}^{n} \sum_{K,L=1}^{2} a^{ij}_{KL} \frac{\partial u^j}{\partial x_L} \right] \right] w^i n_K \, ds \tag{4.4}$$

$$\forall W \in X^0_{h,p+1}(K).$$

Here $X^0_{h,p+1}(K)$ is the space of bubble functions defined as a subspace of the space of element shape functions $X_{h,p+1}(K)$ obtained by raising the order of approximation by one (from p to $p+1$), B_K and L_K are identified as element contributions to the global bilinear and linear forms B and L, respectively, from element K, $U_{h,p}$ is the h-p FE solution and the remaining interelement boundary integral includes the average fluxes corresponding to the first order derivatives. Details are given in [22] and, for the Euler equations, in [23].

The relatively low cost of implementing this technique should be noted. As the dimension of the space of bubble functions (for a scalar problem) is $2p+3$, where p is the order of approximation, the time needed to solve the local problem is minimal. From the point of view of coding, the procedure is also comparatively straightforward since the same routines used to solve the global problem are used to construct the local problems as well.

The functions φ_K, the solutions to the local problem, are known as error indicator functions, and form the basis for the calculation of the error estimate. Globally, the error bound takes the form

$$\|U_{h,p} - U_{h,p+1}\|_E \leq \left[\sum_K B_K(\varphi_K, A_0 \varphi_K) \right]^{1/2}, \tag{4.5}$$

where A_0 is the particular symmetrizer and the energy norm on the left-hand side corresponds to the symmetrized bilinear form, i.e.,

$$\|U\|^2_E = B(U, A_0 U) = \tilde{B}(U, U). \tag{4.6}$$

As indicated in (4.5), the ERM technique estimates the relative error between the given h-p solution $U_{h,p}$ and another h-p solution $U_{h,p+1}$ obtained by raising the order of approximation p in every element uniformly by one. Note that

(1) the difference between $U_{h,p+1}$ and the exact solution to a fractional step is of an order less than that of the relative error, and

(2) it makes little sense to consider the exact solution to the fractional step problem neglecting the time discretization error.

Thus, the choice of the relative error as an indicator seems to be justified. The introduction of symmetrizers, especially in the Euler step, also seems to be a practical answer to the problem of defining a 'natural' norm to measure the error.

The element contribution to the right-hand side of (4.5) is defined as the element error indicator

$$\eta_K^2 = B_K(\varphi_K, A_0\varphi_K)\,. \tag{4.7}$$

4.2. Stability of h-p approximations for problems with small parameters

Most refinements and the introduction of new degrees of freedom will enhance the quality of the numerical solution only if the scheme is stable. Error estimation techniques can produce completely erroneous indications of error whenever the solution is oscillatory or ill-behaved near singularities, shocks or irregularities in the boundary conditions. The proper notion of stability for a problem is notoriously difficult to identify. Unfortunately, both the Euler and viscous steps in the present approach belong to classes of problems for which the notions of stability are not firmly established. Research on systems of conservation laws, mainly in the context of finite difference and finite volume methods, has resulted in several different criteria of stability, e.g., discrete maximum principles, monotone schemes, total variation diminishing schemes, etc. The stability of schemes for nonlinear systems in several space dimensions remains an area of active research.

The situation with stability of the viscous fractional step is a bit better. Both problems are identified as symmetrizable elliptic problems with zeroth order terms (L^2-projections) dominating those of second order (H^1-projections). Neglecting the issue of symmetrization, we choose as a model problem a single elliptic equation of the form

$$u - \Delta t\varepsilon u_{xx} = 0 \quad \text{in } (0, l)\,, \tag{4.8}$$

with Dirichlet boundary conditions

$$u(0) = 1\,, \qquad u(l) = 0\,. \tag{4.9}$$

For l large enough (asymptotically for $l \to \infty$), the exact solution is of the form

$$u(x) = \exp(-x/\sqrt{\Delta t\,\varepsilon})\,. \tag{4.10}$$

This solution exhibits a boundary layer near the origin vanishing quickly to zero away from $x = 0$ and it does satisfy the maximum principle in the sense that the range of u coincides with the interval $[0, 1]$.

Now, suppose that one attempts to solve (4.9) using classical h finite element (or finite difference) methods. It is well-known that in order to satisfy the discrete maximum principle the mesh size h must be (uniformly throughout the mesh) of the order of the small constant in front of the second order derivatives, i.e.,

$$h^2 \sim \Delta t\,\varepsilon \tag{4.11}$$

(see, e.g., [24] for details).

Roughly speaking, both terms contributing to the stiffness matrix must be of the same order. Assuming additionally that Δt is of order h (consistent with truncation error analysis and linear stability estimates for the convection step), we arrive at the conclusion that h must be of order ε. This is very pessimistic as we are interested in the range of $\varepsilon \sim 10^{-3}$–10^{-7}, identifying ε with the reciprocal of the Reynolds number.

In general, derivation of sufficient conditions for the discrete maximum principle to hold

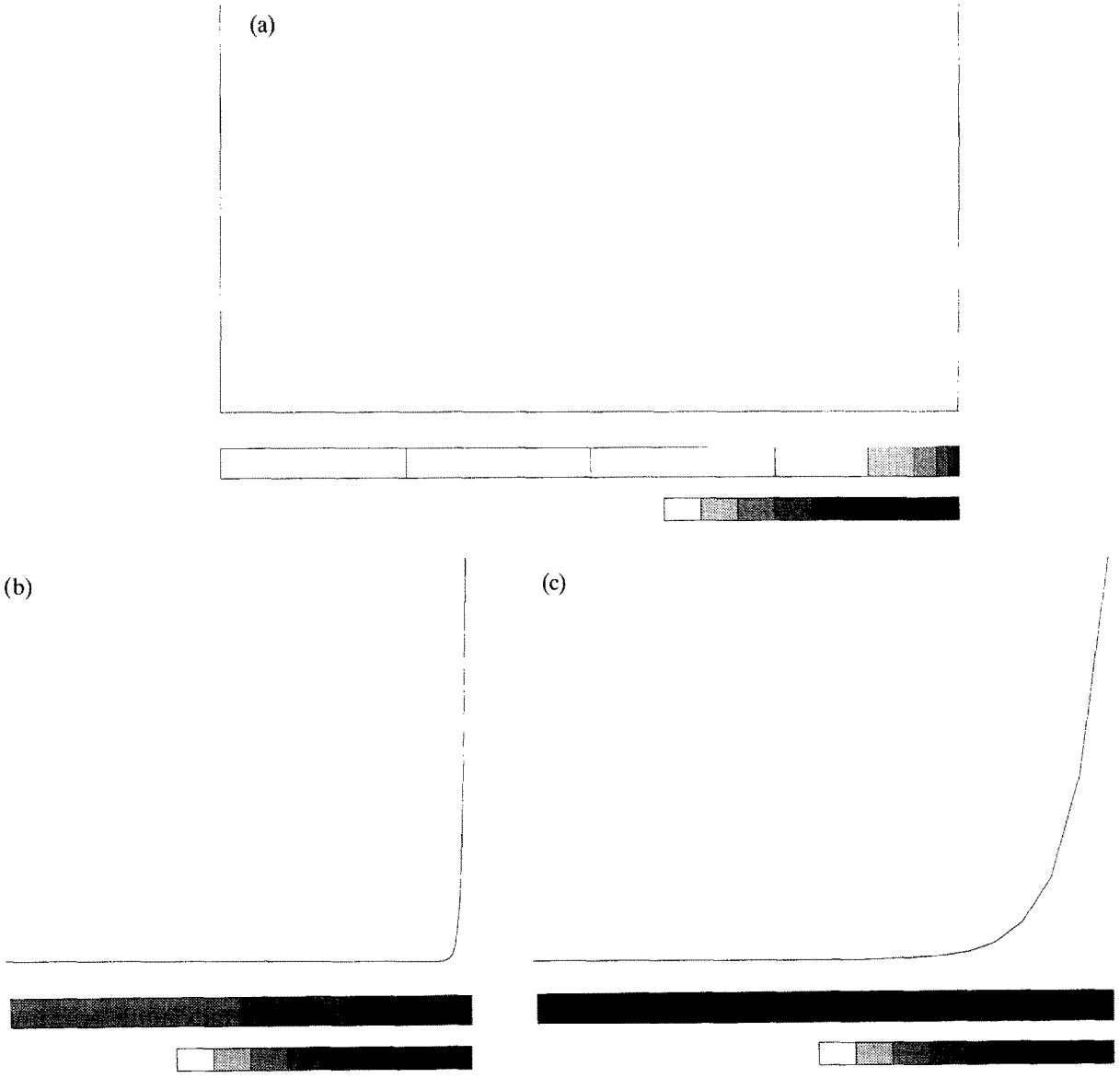

Fig. 4.1. (a) Solution of the problem $-\varepsilon\,\Delta t u'' + u = 0$, $u(0) = 0$, $u(1) = 1$ and the h-p mesh; $\varepsilon = 10^{-6}$, $\Delta t = h_{\min}/(p_{\max} + 1)$, $h_{\min} = 0.25 \times (0.5)^6$, $p_{\max} = 7$. (b) Zoom of (a), magnification 10 times. (c) Zoom of (a), magnification 100 times. The shaded legend indicates various spectral orders, $p = 1$ (white), $p = 2$ (next grey), $p = 3, \ldots, p = 8$ (black).

(see [12]) assumes a uniform mesh in h. It is interesting to note, however, that in order to avoid oscillations of the numerical solution to (4.9), it is sufficient to gradually refine the mesh only in the neighborhood of the origin, reaching the minimum h of order ε only in the elements near the boundary. For $\varepsilon \sim 10^{-6}$, this still results in twenty levels of refinement, but global refinements of this size are not necessary. A similar effect can be obtained by increasing the order of approximation. It is noteworthy that no supportive theory exists for these simple observations.

This is the key point: neither the twenty levels of refinements nor the order of approximation exceeding $p = 10$ are acceptable from the practical point of view (especially in the context of three-dimensional problems). We claim that the optimal computational choice is a special combination of both h- and p-adaptive techniques. By decreasing the mesh size and increasing at the same time the order of approximation gradually toward the boundary, both the level of refinements and the order of approximation are kept within reasonable limits. Figure 4.1 shows experimental results obtained using an ad hoc assumption that every h-refinement of the element adjacent to the boundary is accompanied by a subsequent p-refinement of the element next to the boundary. As a result, an experimental estimate for the minimum element size is obtained in the form

$$h \sim \varepsilon^{1/3} . \tag{4.12}$$

For $\varepsilon \sim 10^{-6}$ this limits both the level of refinements and the order of approximation to $p = 7$ (see Fig. 4.1).

This simple observation will be used as a basis for an adaptive strategy proposed in the next section.

4.3. *Examples of h-p adaptive strategies*

Two different adaptive strategies are investigated here, both for steady state problems. As the whole method and accompanying FE code can, in principle, handle transient problems, design of an efficient space–time adaptive strategy for transient problems remains beyond the scope of this paper.

The two algorithms differ in the way the p-refinements are used. In the first algorithm, an initial mesh geometrically gradated towards the solid wall boundary is implemented using the technique described in the previous section. Thus, both h- and p-refinements are used to generate the initial mesh. In the mesh modifications following convergence to the steady state solution on the current mesh, only h-refinements are used. The use of p-refinements is therefore limited to only the initial mesh generation. This strategy is very much in the spirit of the h-p method presented in [13] based on the SUPG algorithm.

The second algorithm consists of three steps. In the first step, starting with an initial mesh of linear elements and prescribing a maximum level of refinement, an optimal h-mesh of linear elements is established. At this stage, the artificial viscosity present in the Euler solver produces such viscous phenomena as boundary layers and stabilizes the equations.

In this step, a simple h unrefinement/refinement strategy is used. Once a steady-state solution on a current mesh is reached, all elements whose error contributions sum up to less than a fixed percent c_1 of the global error are unrefined and those which contribute altogether

with c_2 percent of the global error are refined. Obviously $c_1 < c_2$. Typical values are $c_1 = 1$ and $c_2 = 90$. Such a strategy is aimed at obtaining a solution with the quality close to the solution obtained on a uniform mesh corresponding to the specified level of refinements.

In the second step, all elements in the boundary layer are p-enriched; that is, the order of approximation is raised. The new order is determined by the element size and the Reynolds number. Simply speaking, the new mesh must capture effects of the real viscosity.

In the third step, p-refinements are used based on the error corresponding to the viscous step only. The same strategy as the one for h-refinements is used.

5. Numerical experiments

5.1. Computational details

Operator splitting

Numerical experiments using both first order and second order operator splittings are investigated. As the steady state solutions are of interest in this paper, attention is first given to the cheaper and faster first order splitting formula (3.6). Formally, once two second order steps (3.7) are put together,

$$\begin{aligned} \boldsymbol{U}^{n+2} &= \boldsymbol{H}(\tfrac{1}{2}\Delta t)\boldsymbol{E}(\Delta t)\boldsymbol{H}(\tfrac{1}{2}\Delta t)\boldsymbol{U}^{n+1} \\ &= \boldsymbol{H}(\tfrac{1}{2}\Delta t)\boldsymbol{E}(\Delta t)\underbrace{\boldsymbol{H}(\tfrac{1}{2}\Delta t)\boldsymbol{H}(\tfrac{1}{2}\Delta t)}_{\boldsymbol{H}(\Delta t)}\boldsymbol{E}(\Delta t)\boldsymbol{H}(\tfrac{1}{2}\Delta t)\boldsymbol{U}^{n}\,, \end{aligned} \tag{5.1}$$

the last and the first viscous fractional steps of the first and second global steps, respectively, can be combined. As a result, second order splitting differs from first order splitting only in the presence of the half-time diffusion step to initiate the process and in the way the solution is sampled at a particular time (second order splitting terminates with a half-time diffusion step). For steady state problems, the initial half-time diffusion step is neglected.

Euler step

The h-p Euler equations solver based on the Taylor–Galerkin formulation was used to solve the Euler step. One of the essential ideas concerning the operator splitting approach from the practical point of view is that the coding should not require any changes in the existing Euler solver. Once an efficient solver for the Euler equations is identified, the coding should consist only in developing the viscous step and merging it with the existing Euler equations solver within an existing data structure. This approach does not guarantee the highest efficiency of the code, but it offers great flexibility for extending the existing code to more complex problems. For instance, the present method can be viewed as one fractional step of the operator splitting procedure for chemically reacting flows.

To be consistent with this point of view, no modifications are made in the Euler equations solver described fully in [20]. There are several nontrivial computational issues concerning the practical implementation of the Taylor–Galerkin method for the Euler equations, including the handling of penalty methods in h-p approximations, error estimation, etc. These details are given in [20].

Linear equation solver

An overlapping block Jacobi iterative solver was used to solve the system of linear equations resulting from the Euler step. The idea of the solver is explained in Fig. 5.1. With every active node A of first order (an unconstrained vertex), a patch of elements including the node is constructed. This includes not only the elements physically containing the vertex but also the elements with constrained nodes for which the considered node is one of the 'parent nodes'. The activated block includes then all degrees of freedom corresponding to the nodes belonging to sides containing the considered node and the central nodes of involved elements.

First, notice that for higher order approximations and the use of a penalty method to handle boundary conditions, preconditioning is often necessary. The conditioning of an element mass matrix, for instance, grows at the rate $\sim 10^p$ with the order of approximation p. The presence of penalty terms in the contributions to the stiffness matrix may cause the standard conjugate gradient algorithm without preconditioning to fail for moderate values of the penalty parameter ($\varepsilon \sim 10^{-4}$, 10^{-5}).

As the principal idea behind the construction of a preconditioner is that the calculation of an inverse should be rather inexpensive, it is reasonable to assume that a preconditioner should have a block-diagonal form. If each block corresponds to all degrees of freedom associated with a single node, it becomes clear that the use of such a preconditioner would affect only the numerical stability of the local patch problems. The particular method of block Jacobi iteration described here turns out to converge reasonably quickly, is not sensitive to high p and use of a penalty approximation does not appear to require the use of an accelerator.

The convergence tolerance constant *tol* for the iterative solver is generally related to the constant ε in the context of the convergence to the steady state solution. With ε guaranteeing convergence to four significant digits and an order smaller *tol*, the number of iterations, as the solution converges to steady state, stabilizes at around 7–8 iterations. Setting *tol* equal to ε may result in oscillations in the last phase of the solution preventing convergence to the steady state.

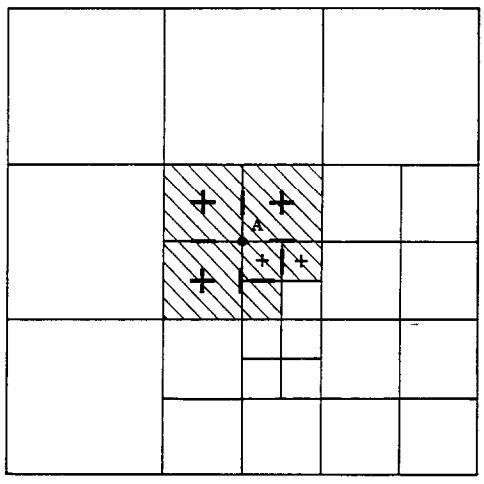

Fig. 5.1. A patch of elements corresponding to node A and activated nodes for the patch.

Regularization

Use of operator splitting procedures for steady state problems is effectively equivalent to the solution of a regularized problem. The construction of the regularized problem depends upon three factors:

— the operator splitting procedure,
— the methods used to solve the fractional steps problems (in our case the Taylor–Galerkin and implicit/explicit Euler methods), and
— the choice of the time step Δt.

Obviously, once the splitting procedure is fixed and the methods are identified, the only parameter of interest is the time step Δt. There are at least two factors limiting the maximum value of Δt. The first is a linear stability criterion: most of the explicit or implicit/explicit methods are only conditionally stable, i.e., the integration in time method is stable under a time step limitation which, for *h-p* elements, usually takes on the form

$$\Delta t \leq c \min_K h_K^r / p_K \,, \tag{5.2}$$

with the maximum taken over all elements K and c depending on local velocity and sound speed. The exponent $r = 1$ for first order problems (Euler step), or $r = 2$ for the second order problems (viscous step). Estimates of this type are not readily available, especially for nonuniform meshes (see comments in [20]). This limitation does not formally exist if unconditionally stable methods are used. In our case, this limits the use of implicitness parameters for both methods to $\alpha, \beta \geq 0.5$.

Another important factor results from the truncation error analysis (a priori error estimates for finite elements). As we are interested in solving the original problem as opposed to a regularized problem, the time step Δt should stay comparable with the number of degrees of freedom corresponding to the spatial approximation (h or p in h- or p-approximations). In other words, a very fine solution obtained with small h and/or high p but with a large time step Δt may have little relevance to the solution to the original problem.

5.2. Sample results

The following examples were solved using the first order splitting method with both parameters of implicitness α and β equal to one. A uniform time step Δt was selected (for the initial mesh and after every refinement) using the formula

$$\Delta t = \mathrm{CFL} \min_K (h_K / \lambda_K) \,, \tag{5.3}$$

where h_K is the element size and λ_K is the maximum eigenvalue of the Euler Jacobian matrices in both x- and y-directions.

Carter's Flat Plate Problem (see [15])

We consider the Carter's Flat Plate Problem with the following data

$$M_\infty = 3 \,, \qquad \mathrm{Re}_L = 1000 \,, \qquad \mathrm{Pr} = 0.72 \,, \qquad \gamma = 1.4 \,, \qquad \theta_\infty = 390[^\circ\mathrm{R}] \,. \tag{5.4}$$

The geometry and boundary conditions are shown in Fig. 5.2. The following flow features are recognized:

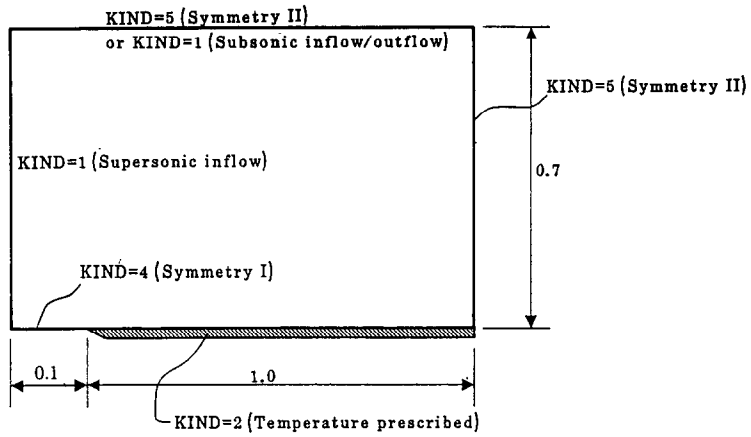

Fig. 5.2. Carter's Flat Plate Problem. Geometry and boundary conditions.

— a singularity exists near the leading edge of the plate,
— a curved bow shock is developed from the tip of the plate, and
— a boundary layer is formed along the plate.
The flags in the figure (KIND = 1, . . . , 5) correspond to various boundary conditions incorporated in the code. The solid wall temperature

$$\theta_{\text{wall}} = 1092[^{\circ}\text{R}] . \tag{5.5}$$

The non-uniform initial *h-p* finite element mesh shown in Fig. 5.3 with second and third order elements defined along the solid wall boundary is selected. The *h*- and *p*-refinements were combined in the fashion described in the previous section to guarantee the stability of the method along the solid wall boundary. In order to reduce the numbers of degrees of freedom in most of the boundary layer region (except near the stagnation point), anisotropic elements were used in which a higher order of approximation in the direction normal to the plate is used and linear approximation is used along the plate.

Using the pure *h*-adaptive strategy described in the previous section with constants $c_1 = 0$ (no unrefinements) and $c_2 = 90$ (which corresponded to the initial refinements in the boundary layer) the mesh shown in Fig. 5.4 was obtained. The corresponding density contours are presented in Fig. 5.5. Starting with this mesh, constant c_2 was set to 30, and a sequence of consecutive meshes was produced with the restriction on the level of refinement increased. Two of those meshes are shown in Figs. 5.6 and 5.8 with the corresponding density contours in Figs. 5.7 and 5.9. After the mesh shown in Fig. 5.8 was obtained, the process was stopped. In all steps, the CFL constant was set to 0.5.

Figures 5.10 and 5.11 compare the computed profiles of pressure and skin friction distributions along the solid wall with the original results of Carter, showing good agreement away from the stagnation point and the ability of the method to also resolve the stagnation point itself. Figure 5.12 shows the computed heat flux along the solid wall (not presented in Carter's report). Note the positive sign in the vicinity of the tip of the plate (the gas heating plate). The stabilizing effect of higher order elements near the solid wall can also be observed in the pressure contours shown in Fig. 5.13 corresponding to the final mesh.

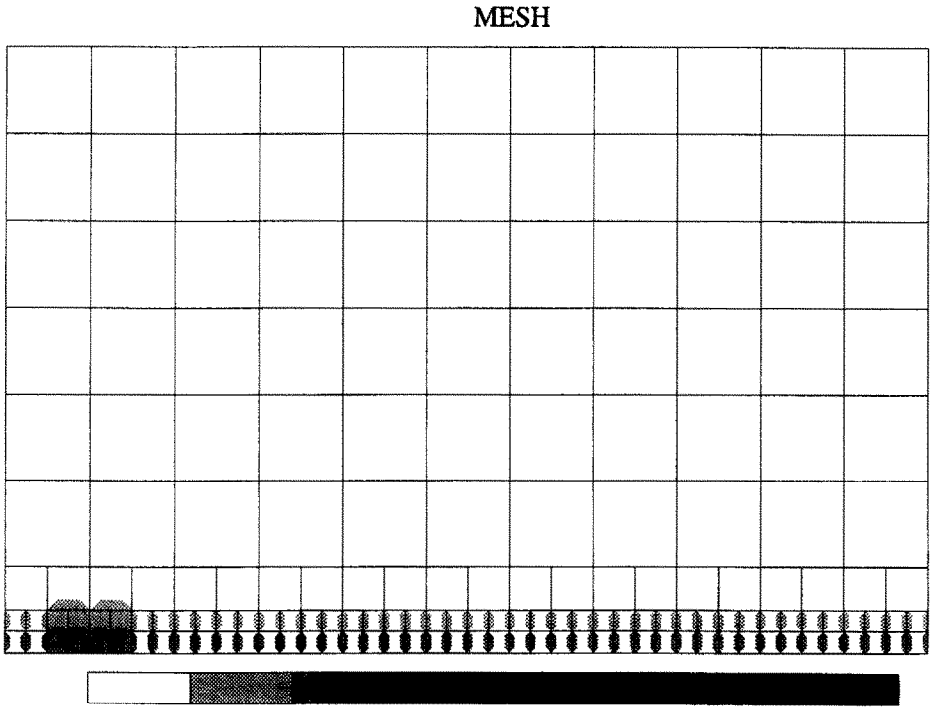

Fig. 5.3. Carter's Flat Plate Problem with $Re = 10^3$. An *h-p* FE initial mesh.

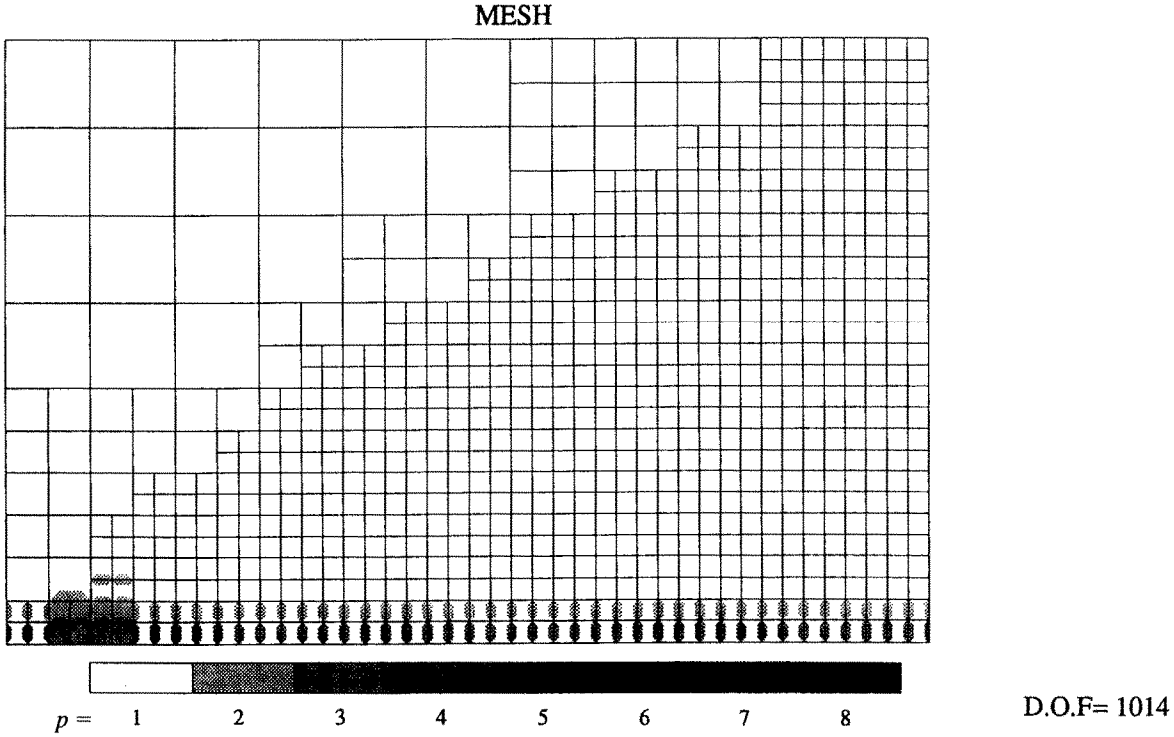

Fig. 5.4. Carter's Flat Plate Problem. An optimal mesh with maximum level of refinement equal 2 (Mesh 1).

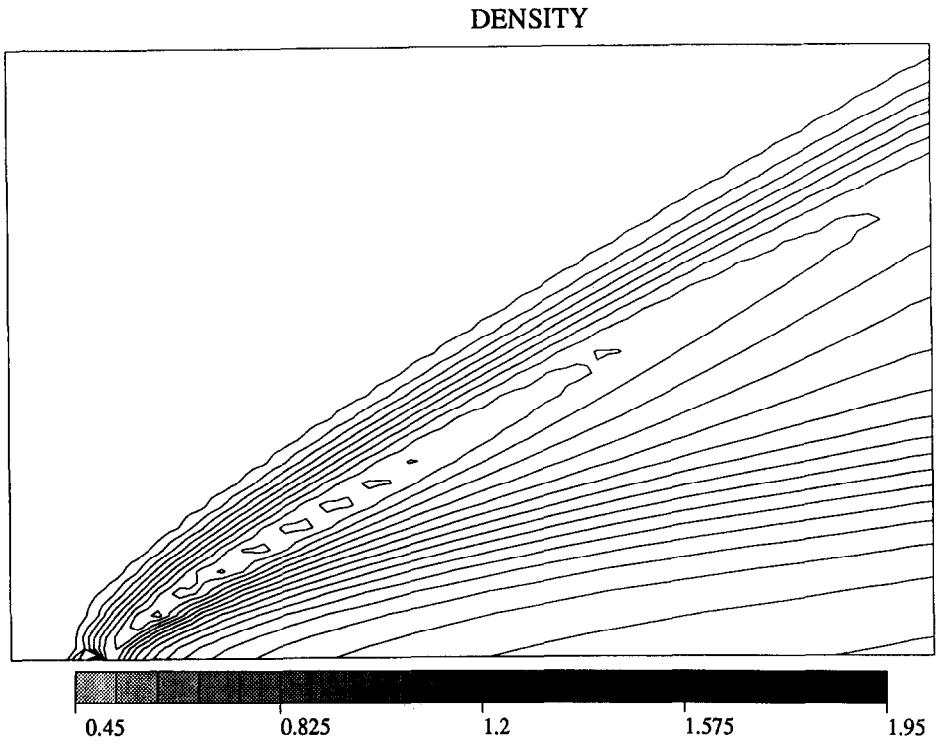

Fig. 5.5. Carter's Problem. Mesh 1. Density contours.

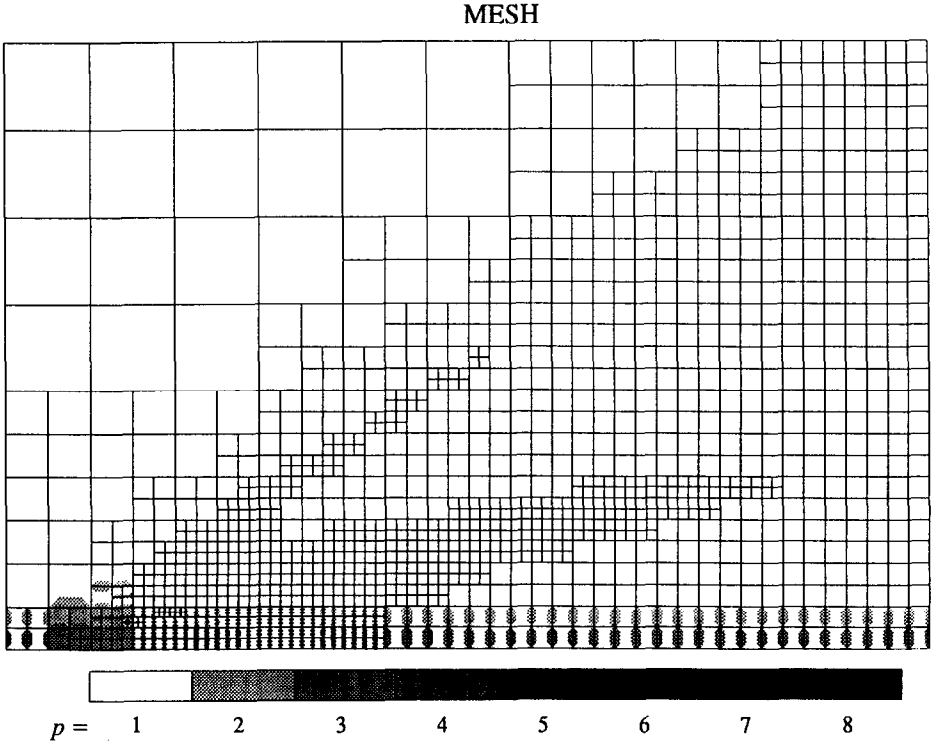

Fig. 5.6. Carter's Problem. Mesh 2.

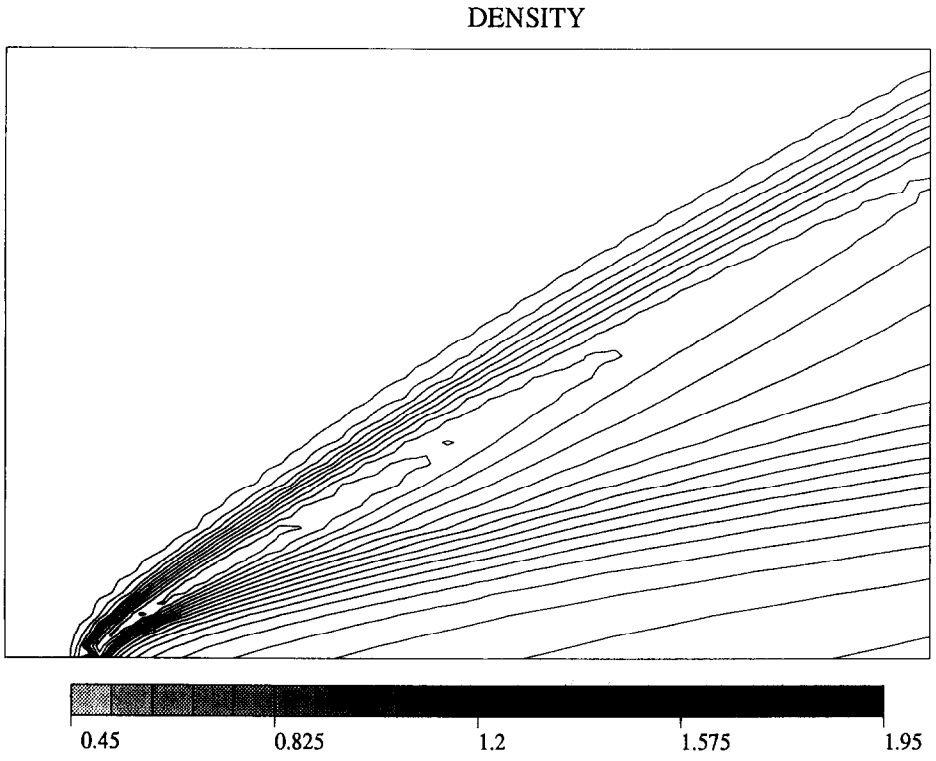

Fig. 5.7. Carter's Problem. Mesh 2. Density contours.

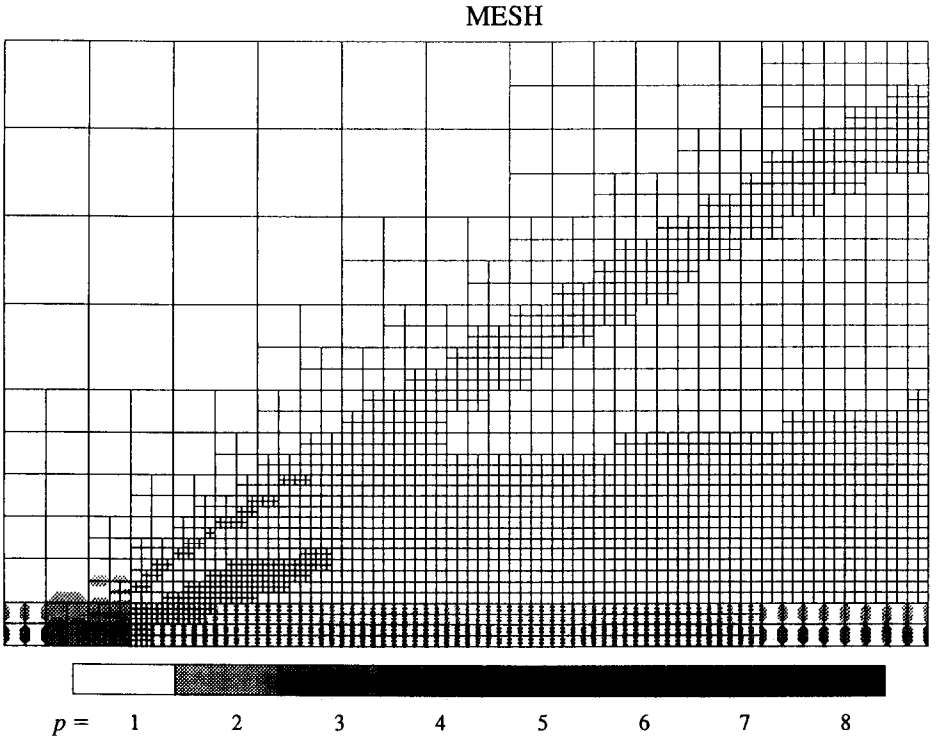

Fig. 5.8. Carter's Problem. Mesh 3.

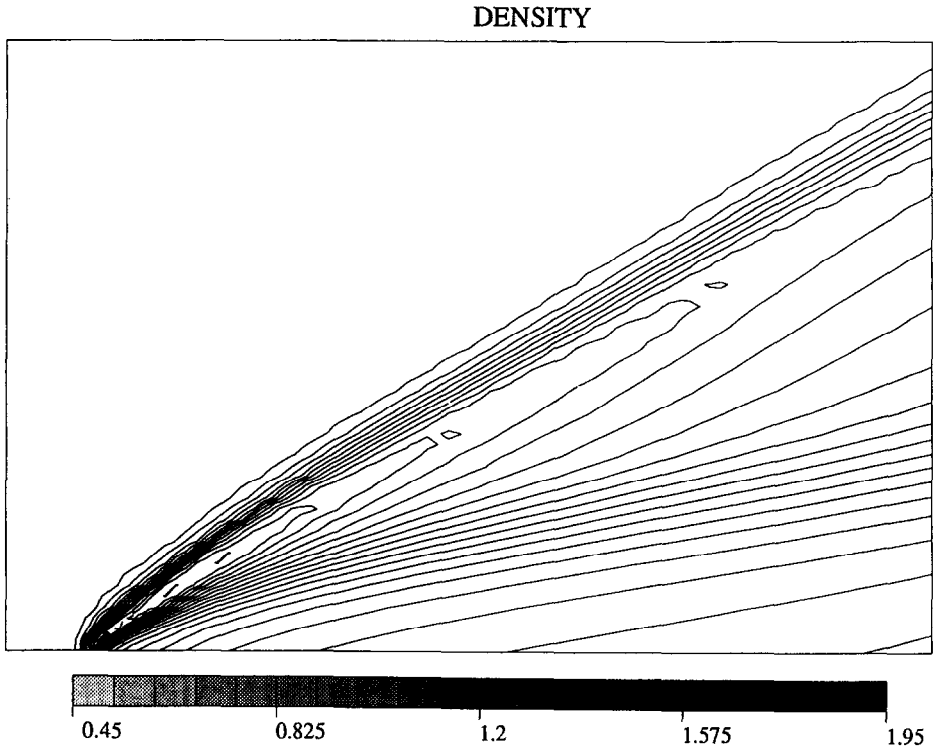

Fig. 5.9. Carter's Problem. Mesh 3. Density contours.

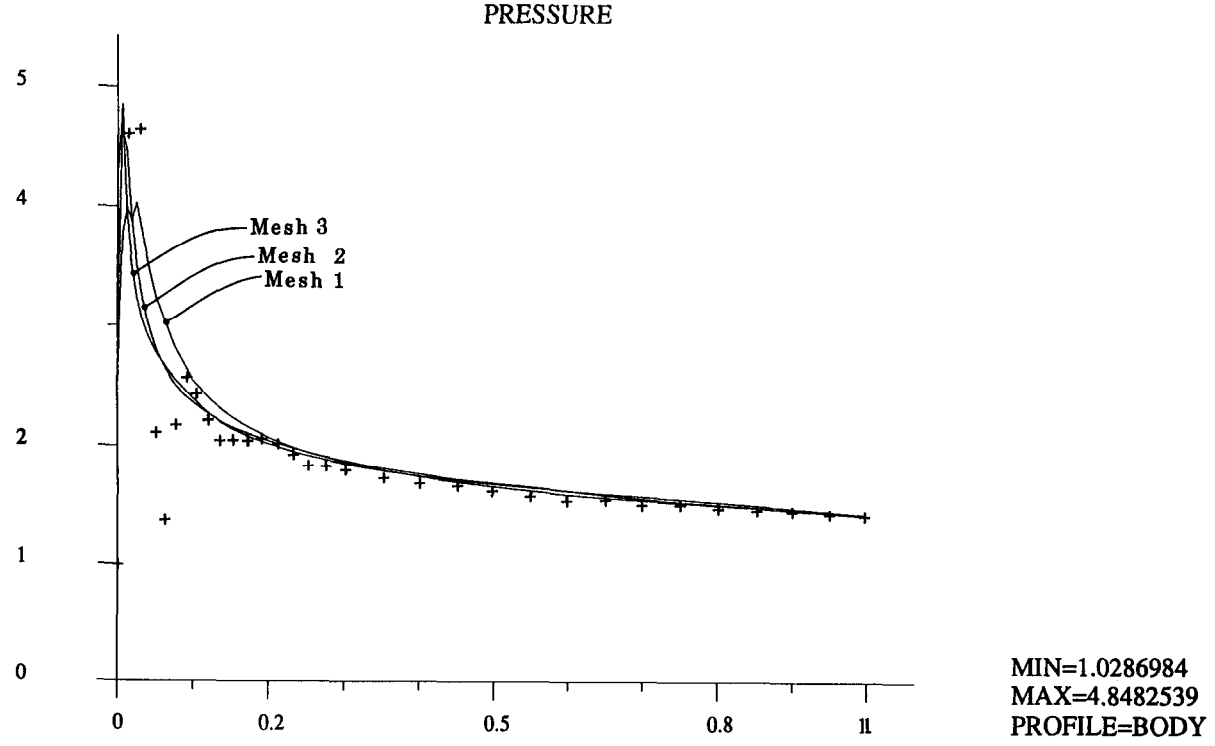

Fig. 5.10. Carter's Problem. Pressure profiles along the plate.

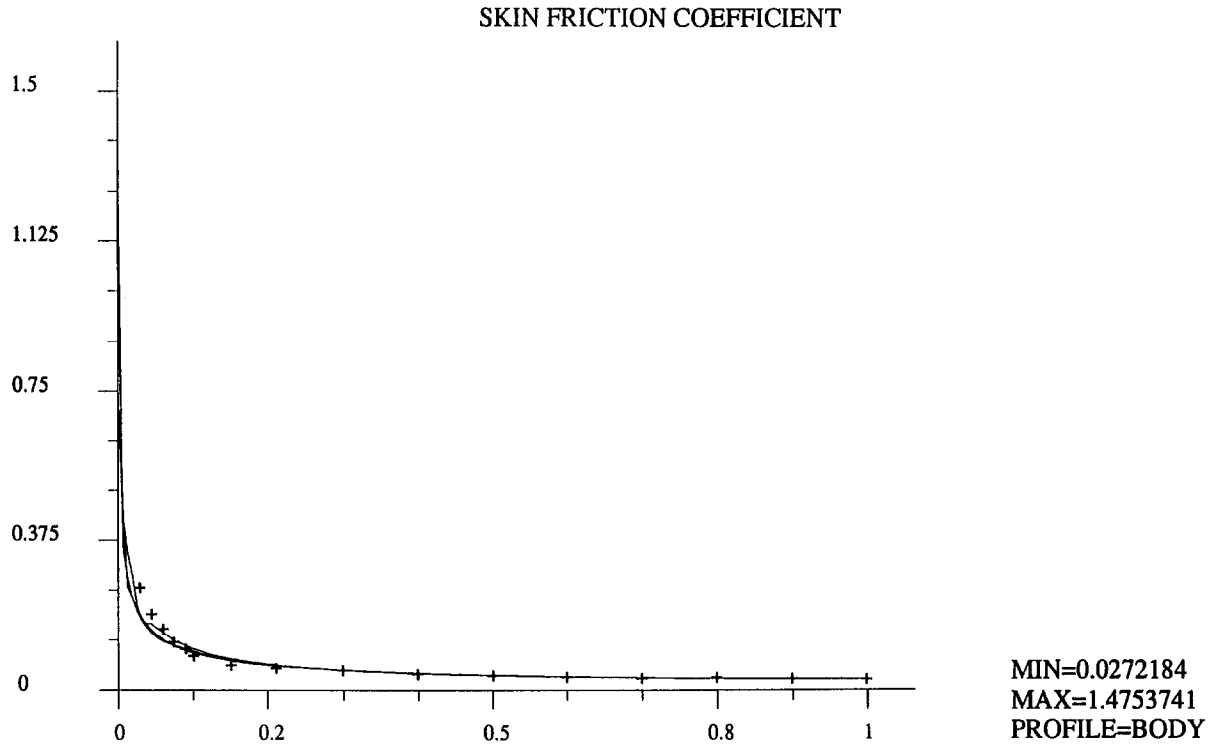

Fig. 5.11. Carter's Problem. Skin friction profiles along the plate.

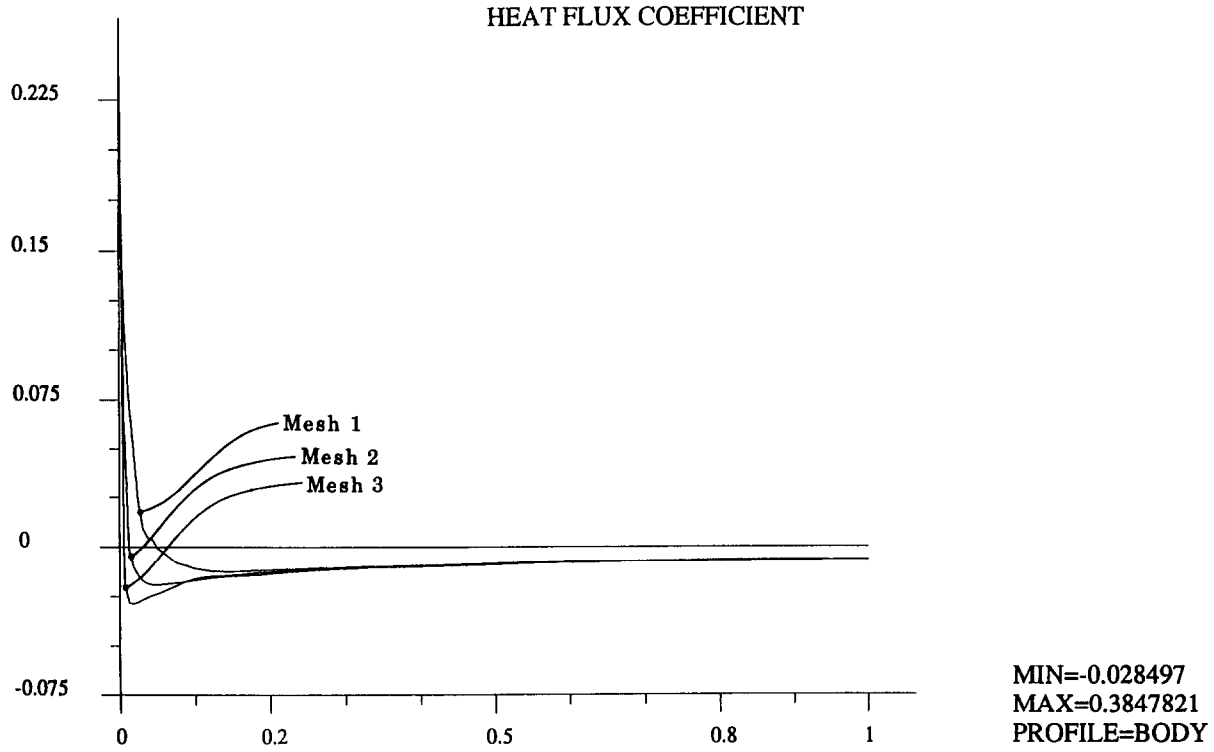

Fig. 5.12. Carter's Problem. Heat flux coefficient along the plate.

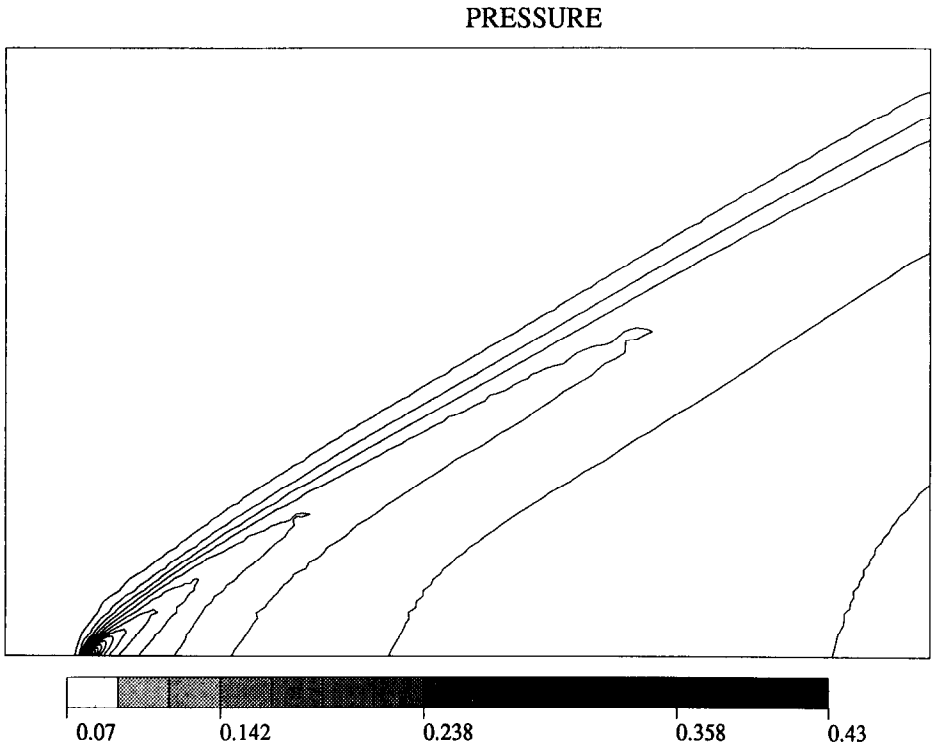

Fig. 5.13. Carter's Problem. Mesh 3. Pressure contours.

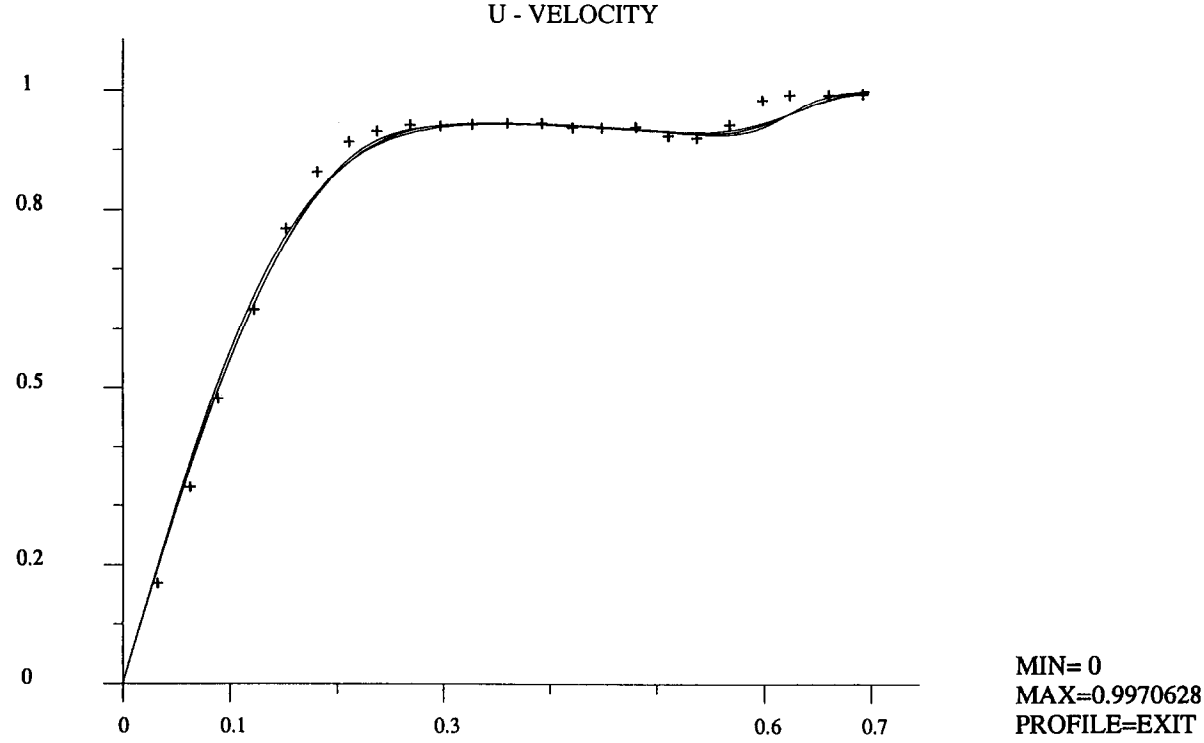

Fig. 5.14. Carter's Problem. Exit profiles of the horizontal velocity.

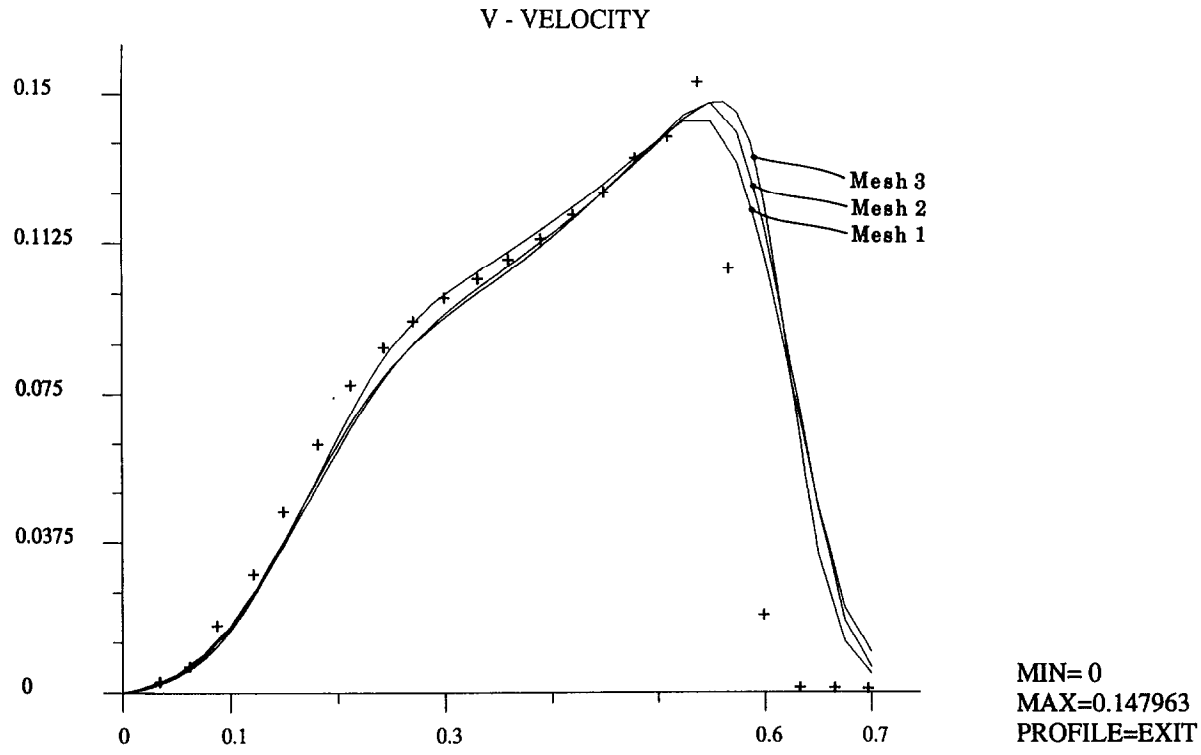

Fig. 5.15. Carter's Problem. Exit profiles of the vertical velocity.

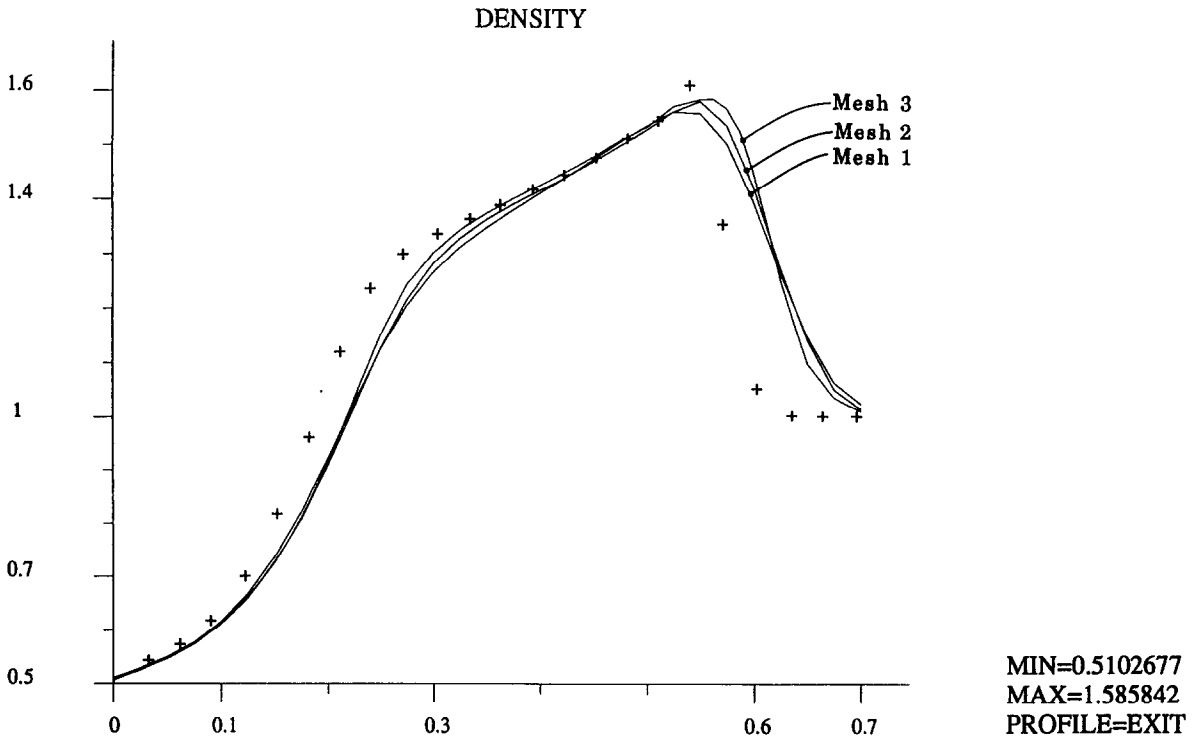

Fig. 5.16. Carter's Problem. Exit profiles of the density.

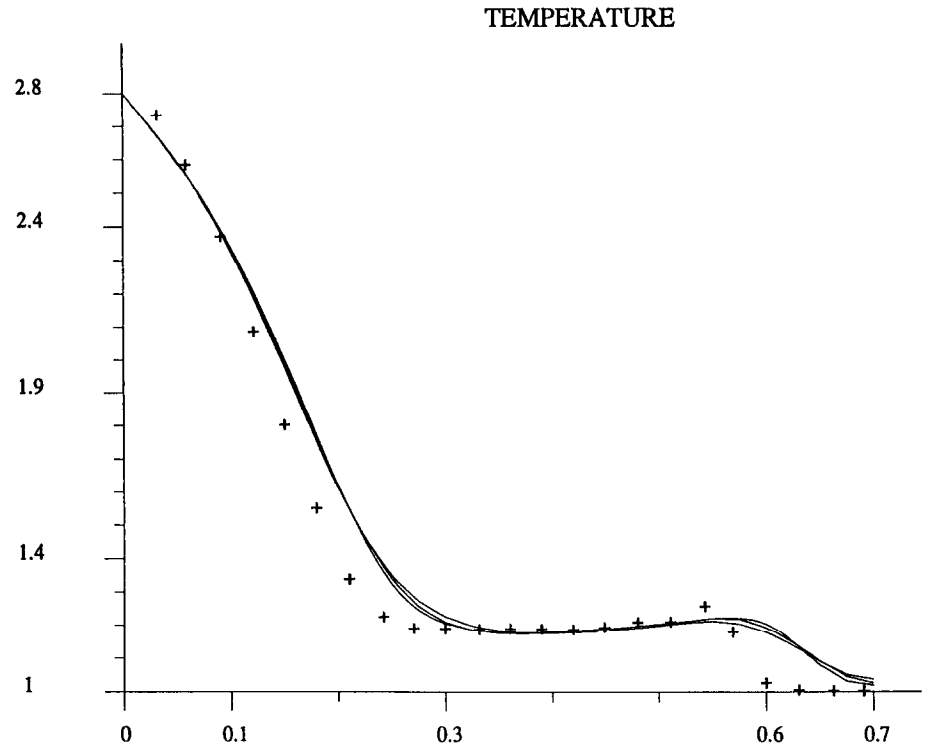

Fig. 5.17. Carter's Problem. Exit profiles of the temperature.

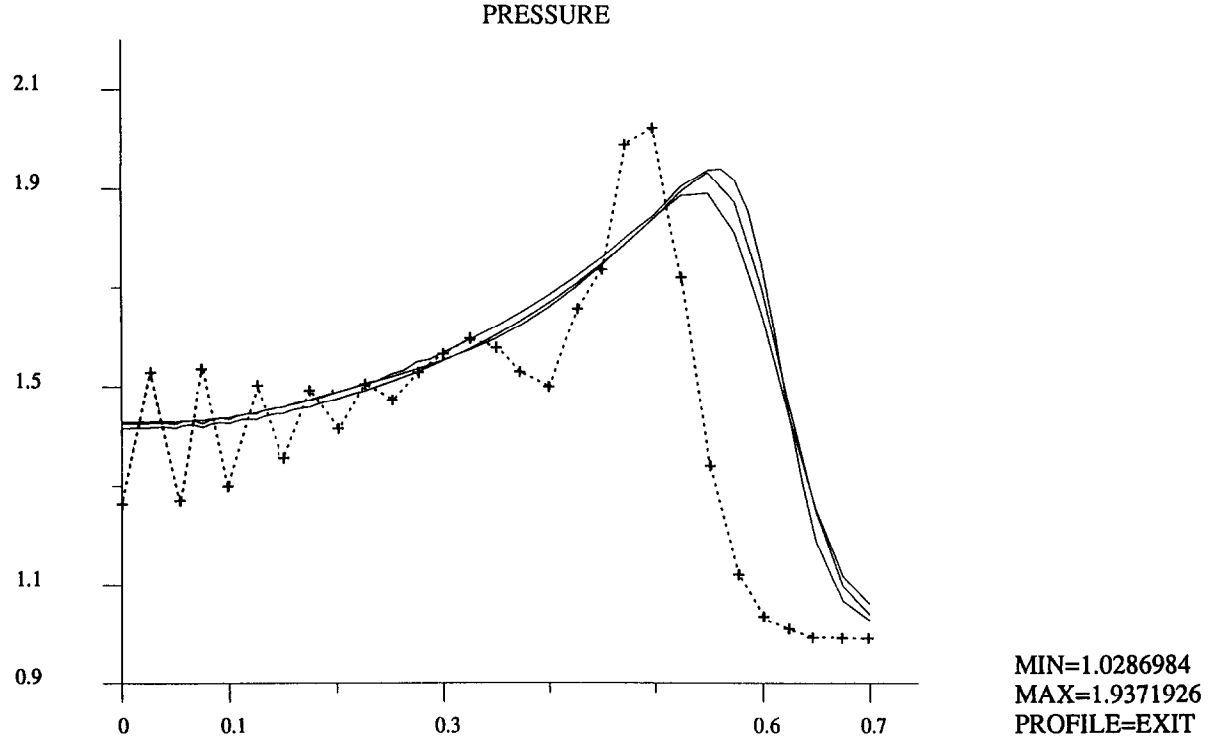

Fig. 5.18. Carter's Problem. Exit profiles of the pressure.

Finally, Figs. 5.14–5.17 show the exit profiles of both velocity components, density and temperature showing good agreement with Carter's results. The corresponding exit profile of the pressure shown in Fig. 5.18 is seen to be free of oscillations.

All of these results were obtained directly from the FE representations with no post-processing of any kind.

Flat plate problem with Re = *10,000*

As a second example, an analysis of the same flat plate problem but with a Reynolds number of one order larger was attempted. This required that the problem with the previous data must be solved in a domain roughly ten times larger than the previous example. In order to avoid a large number of degrees-of-freedom in the initial meshes resulting from higher order elements in the initial stage of the solution procedure, the second adaptive strategy described earlier was employed.

Starting with an initial mesh of only 4×8 elements, a sequence of consecutive optimal linear meshes was obtained using the adaptive strategy discussed before with constants $c_1 = 1$ and $c_2 = 90$. A total of 7 refinements were generated, each computed only when the optimal mesh for a particular level of refinement had been obtained. The final mesh and the corresponding density contours are shown in Fig. 5.19. Figure 5.20 presents a three-dimensional perspective of the same density function showing a clear separation of the shock from the boundary layer.

For coarse meshes (low levels of refinement), the computed boundary layer is primarily due to numerical viscosity, the corresponding viscous quantities, especially near the leading edge, being far from those obtained in the previous example. Figure 5.21 presents, for instance, the profile of heat flux along the solid wall (compare with Fig. 5.11).

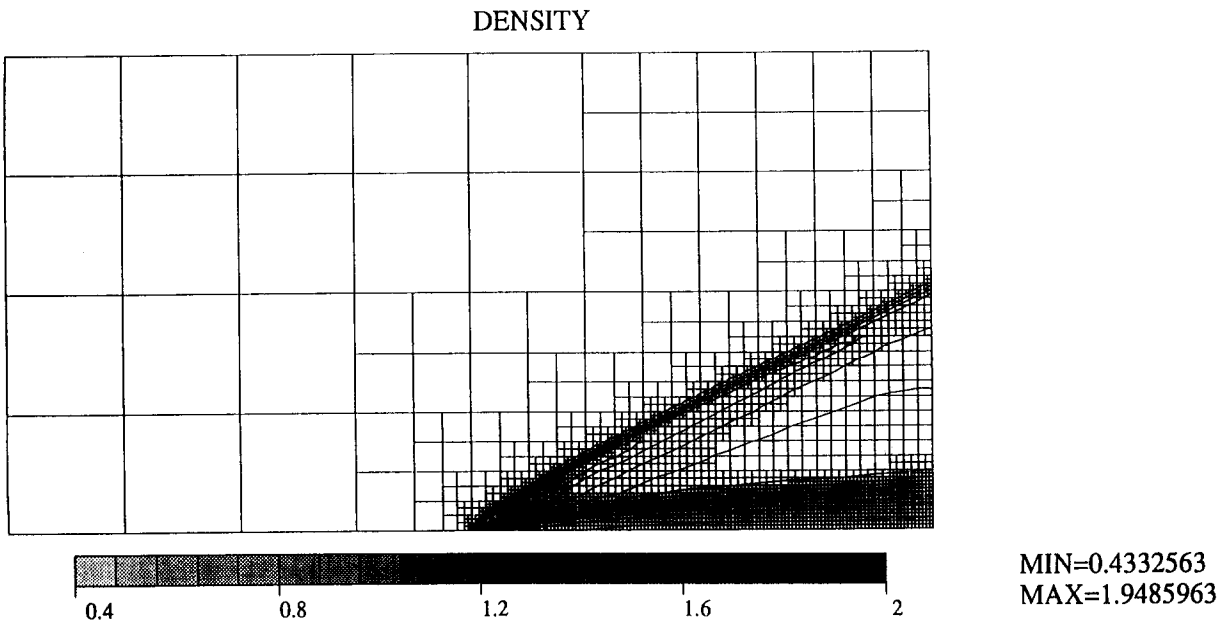

Fig. 5.19. Flat plate problem with Re = 10,000. Density contours and optimal mesh of linear elements with the maximum level of refinement = 7.

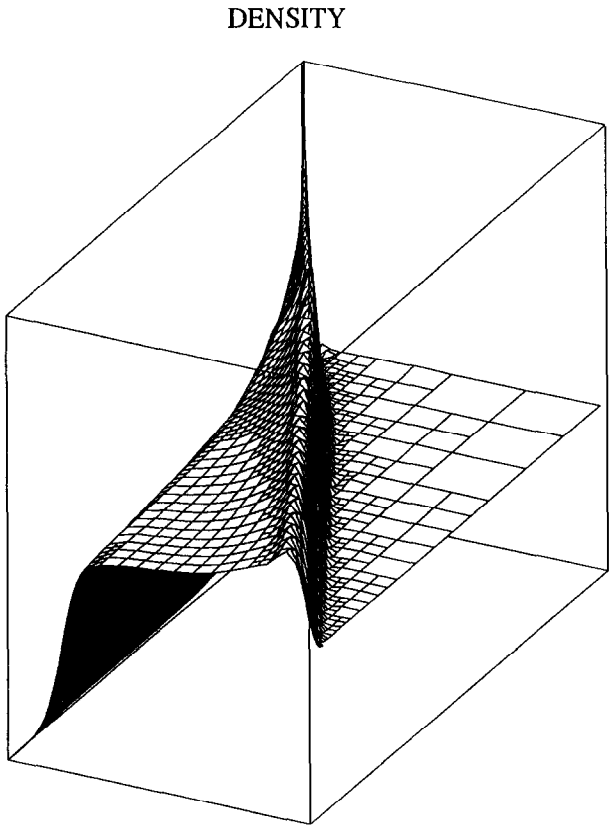

Fig. 5.20. Flat plate problem with Re = 10,000. A three-dimensional perspective of the density function on the optimal mesh of linear elements.

The situation drastically changes when a layer of second, third and fourth order elements is introduced in the mesh (see Fig. 5.22). The corresponding heat flux profile, shown in Fig. 5.23, agrees qualitatively with the results obtained in the previous example.

6. Some concluding comments

The theory and numerical examples presented in this study show the potential of the proposed method in resolving stationary Navier–Stokes equations with effective means for the modeling boundary layers and the resolution of viscous effects.

Results of sample calculations show that despite the fact that both schemes used to solve the fractional steps in the operator-splitting technique are unconditionally stable (in the linear sense only, see [20]), a nonlinear loss of stability in the vicinity of the stagnation point was frequently encountered (negative pressures and densities) when higher elements were used there. This suggests that for the particular algorithms employed the use of higher order elements should be restricted to the boundary layer. In early stages of the calculation as the solution gradually converges from initial conditions to a steady state, the use of linear elements and h-refinement seems to be advantageous. Due to changes in the evolving structure of the solution, both unrefinements and h-refinements are necessary.

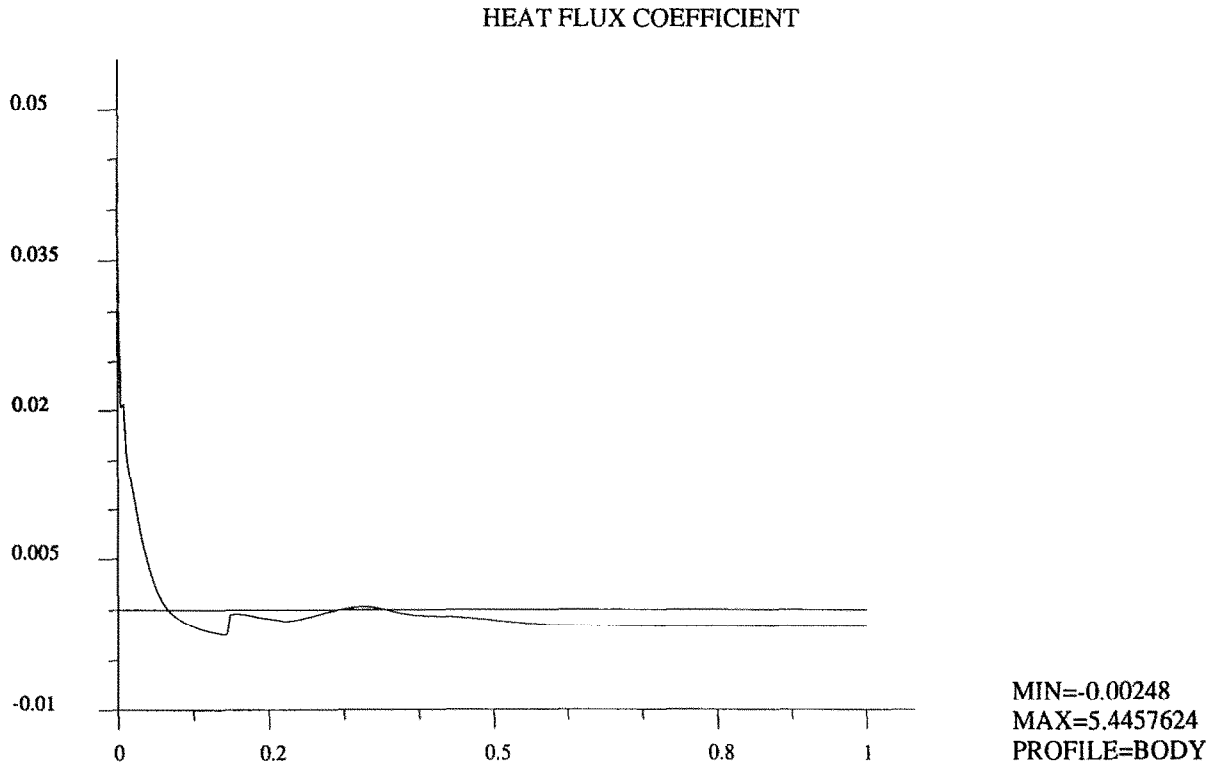

Fig. 5.21. Flat plate problem with Re = 10,000. Heat flux coefficient profile along the plate for the mesh of linear elements.

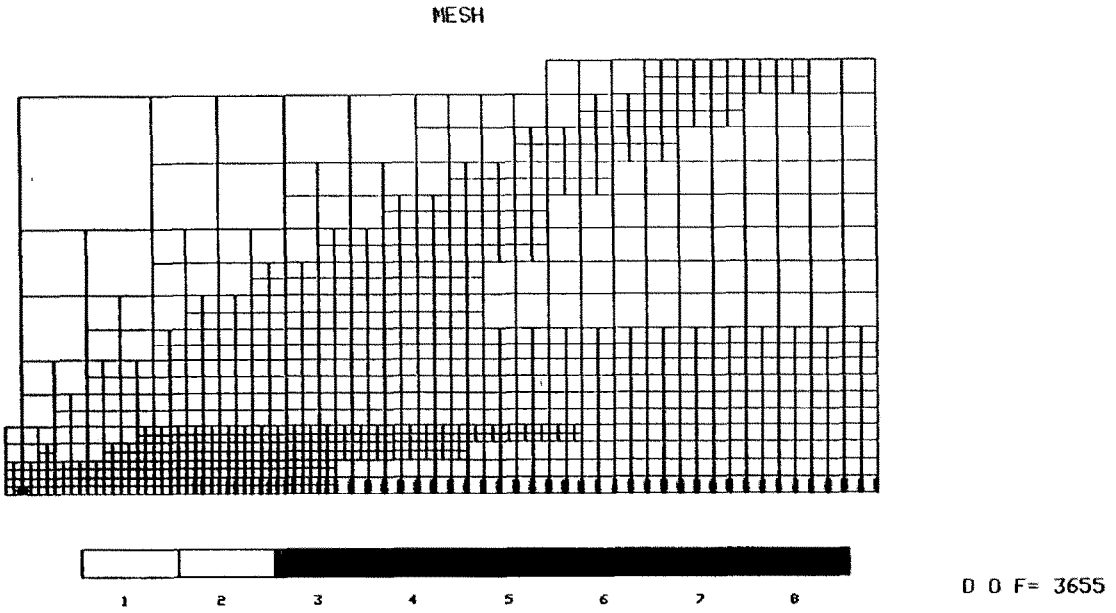

Fig. 5.22. Flat plate problem with Re = 10,000. The optimal mesh enriched in the boundary layer.

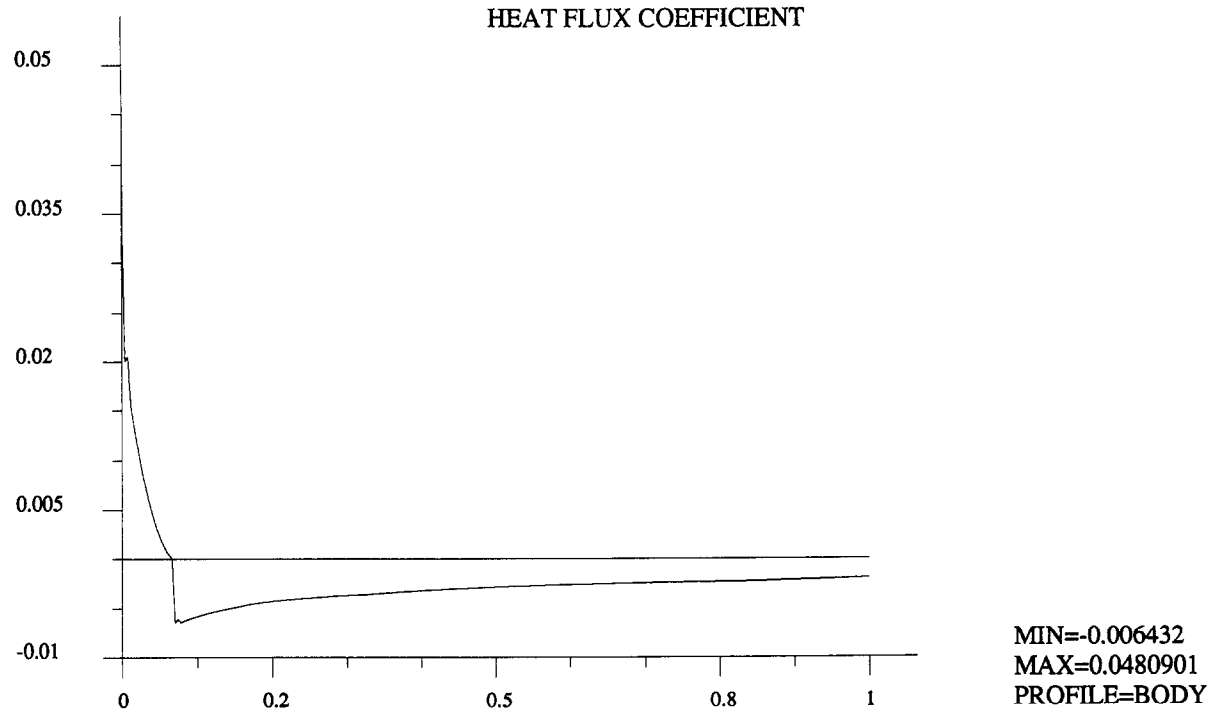

Fig. 5.23. Flat plate problem with Re = 10,000. Heat flux coefficient profile along the plate for the enriched mesh.

Later in the adaptive process, as a large number of small elements are generated in the boundary layer, one observes little change in the flowfield tangent to the boundary—as expected in classical thin-layer Navier–Stokes models. Then the use of anisotropic p-approximations normal to the boundary is effective while h-refinement is less important there. These results also suggest that a new adaptive data structure that accommodates thin-high-aspect-ratio elements, with anisotropic p-capability normal to the boundary would be effective. This should result not only in great savings in the number of degrees-of-freedom but, in the case of higher order anisotropic elements, result in better resolution of the layer itself. We note, however, that type of boundary-layer treatment can be done only a posteriori, when the convective phenomena are sufficiently resolved, as the location of the shocks changes with refinements.

Acknowledgment

The authors gratefully acknowledge support of this work by the National Aeronautics and Space Administration (NASA) under contracts NAS1-18746 and NAS2-13000 from the Langley Research Center and the Ames Research Center.

References

[1] N.N. Yanenko, The Method of Fractional Steps (Springer, Berlin, 1971).
[2] G. Strang, On the construction and comparison of difference schemes, SIAM J. Numer. Anal. 5 (3) (September 1968) 506–517.

[3] W.G. Szymczak, An analysis of viscous splitting and adaptivity for steady state convection-diffusion problems, Comput. Methods Appl. Mech. Engrg. 67 (1988) 311–354.
[4] J.T. Beale and A. Majda, Rates of convergence for viscous splitting of the Navier–Stokes equations, Math. Comp. 37 (156) (October 1981) 243–259.
[5] T.I. Ramos, Numerical solution of reactive-diffusive systems. Part 3: Time linearization and operator-splitting techniques, Internat. J. Comput. Math. 18 (1986) 289–309.
[6] M.F. Wheeler and C.N. Dawson, An operator-splitting method for advection–diffusion–reaction problems, in: Mathematics of Finite Elements and Applications VI (Academic Press, New York, 1988).
[7] L. Demkowicz and J.T. Oden, An adaptive characteristic Petrov–Galerkin finite element method for convection-dominated linear and nonlinear parabolic problems in one space variable, J. Comput. Phys. 68 (1) (1986) 188–273.
[8] K.O. Aiyesimoju and R.J. Sobey, Process splitting of the boundary conditions for the advection–dispersion equation, Internat. J. Numer. Methods Fluids 9 (1989) 235–244.
[9] Ph. Devloo, J.T. Oden and T. Strouboulis, Implementation of fan adaptive refinement technique for the SUPG algorithm, Comput. Methods Appl. Mech. Engrg. 61 (1987) 339–358.
[10] J.T. Oden, T. Strouboulis, Ph. Devloo and T. Howe, Recent advances in error estimation and adaptive improvement of finite element calculations, in: A.K. Noor, ed., Computational Mechanics, Advances and Trends (ASME, New York, 1986) 369–410.
[11] L. Demkowicz, J.T. Oden, W. Rachowicz and O. Hardy, Toward a universal *h-p* adaptive finite element strategy, Part 1. Constrained approximation and data structure, Comput. Methods Appl. Mech. Engrg. 77 (1989) 79–112.
[12] W. Rachowicz, J.T. Oden and L. Demkowicz, Toward a universal *h-p* adaptive finite element strategy, Part 3. Design of *h-p* meshes, Comput. Methods Appl. Mech. Engrg. 77 (1989) 181–212.
[13] Ph. Devloo, J.T. Oden and P. Pattani, An *h-p* adaptive finite element method for the numerical simulation of compressible flow, Comput. Methods Appl. Mech. Engrg. 70 (1988) 203–235.
[14] D.A. Anderson, J.C. Tannehill and R.H. Fletcher, Computational Fluid Mechanics and Heat Transfer (McGraw-Hill, New York, 1984).
[15] J.E. Carter, Numerical solutions of the Navier–Stokes equations for the supersonic laminar flow over a two-dimensional compression corner, NASA Technical Report TR R-385, July 1972.
[16] B. Gustafsson and A. Sudström, Incompletely parabolic problems in fluid dynamics, SIAM J. Appl. Math. 35 (2) (September 1978) 343–357.
[17] J.C. Strikwerda, Initial boundary value problems for incompletely parabolic systems, Ph.D. Thesis, Department of Mathematics, Stanford University, Stanford, CA, 1976.
[18] P. Dutt, Stable boundary conditions and difference schemes for Navier–Stokes equations, SIAM J. Numer. Anal. 25 (2) (1988) 245–267.
[19] 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.
[20] L. Demkowicz, J.T. Oden, W. Rachowicz and O. Hardy, An *h-p* Taylor–Galerkin finite element method for compressible Euler equations, Comput. Methods Appl. Mech. Engrg. (to appear).
[21] L. Demkowicz, A note on symmetry boundary conditions (in preparation).
[22] J.T. Oden, L. Demkowicz, W. Rachowicz and T.A. Westermann, Toward a universal *h-p* adaptive finite element strategy, Part 2. A posteriori error estimation, Comput. Methods Appl. Mech. Engrg. 77 (1989) 113–180.
[23] L. Demkowicz, J.T. Oden, W. Rachowicz and T. Westermann, An a posteriori error estimation technique for the Euler and the Navier–Stokes equations (in preparation).
[24] I. Tsutomu, Maximum principle in finite element models for convection–diffusion phenomena, in: H. Fujita and M. Vamaguti, eds., Lecture Notes in Numerical and Applied Analysis, Vol. 4, No. 76, Mathematics Studies (North-Holland, Amsterdam, 1983).
[25] L. Demkowicz and W. Rachowicz, On a characteristic finite element method for compressible gas dynamics, J. Engrg. Sci. 25 (10) (1987) 1259–1281.
[26] J.T. Oden and L. Demkowicz, A survey of adaptive finite element methods in computational mechanics, in: A.K. Noor and J.T. Oden, eds., State-of-the-Art Surveys on Computational Mechanics (ASME, New York, 1989).