A Triangle-based Unstructured Finite Volume Method for Chemically Reactive Hypersonic Flows

Enrico Bertolazzi[1] and Gianmarco Manzini[2]

[1] *Dip. Ingegneria Meccanica e Strutturale, Università di Trento,*
via Mesiano 77, I – 38050 Trento, Italy
Fax : +39 0461 882599 Voice : +39 0461 882660
E-mail: Enrico.Bertolazzi@ing.unitn.it

[2] *Istituto di Analisi Numerica IAN – CNR,*
via Ferrata 1, I – 37100 Pavia, Italy
Fax : +39 0382 548300 Voice : +39 0382 548215
E-mail: Marco.Manzini@ian.pv.cnr.it

A triangle-based unstructured finite volume method is developed for chemically reactive hypersonic calculations. The method is based on a Steger-Warming flux vector-splitting approach generalized to mixtures of thermally perfect gases. Second order-in-space and -time accuracy is provided by limited flux blending and an implicit multi-stage time marching scheme. The final stiff non-linear problem resulting from discretization presents a very peculiar block diagonal structure. This allows a decoupling of the species and gas dynamic equations in smaller sub problems. A linear algebra argument based on M-matrix theory allows also to show that the method guarantees positivity of species mass densities and vibrational energies under a reasonable *CFL*-like constraint. Finally, a set of 2-D numerical test cases illustrates the performance of the method.

Key Words: Hypersonic Flows, Unstructured Grids, Finite Volumes, Semi-implicit schemes, M-matrix.

1. INTRODUCTION

During the eighties and early nineties both in US and Europe several projects were funded to design transatmospheric flight vehicles. This event had a great impact on the fluid dynamic community and stimulated researchers to investigate hypersonic flows and correlated topics. The modelization of such compressible flow configurations and its numerical approximation soon appeared as a considerably complex task, deserving special efforts to produce accurate and reliable results [12]. Indeed, the features of hypersonic flow regimes are quite different from those of subsonic, transonic, and supersonic ones. Hypersonic flows are characterized by high speed flows in low density environments, and non-equilibrium thermodynamical effects are normally not negligible. In fact, the crucial

issue consists in the significant departure of the fluid from ideal gas conditions, which must be correctly accounted for, see Reference [26].

Moreover, strong shock patterns may appear in the flow configuration, due to the compressible nature of the fluid. The significant increase in temperature and pressure across high Mach number shocks induces the excitation of internal vibrational modes of polyatomic molecules, and dissociation and ionization reactions may occur.

The fluid, i.e. air, is described in terms of a chemically reactive mixture of thermally perfect gases. The usual inviscid gas dynamic equations applies to the conservation of fluid mass, momentum, and energy. The global mass conservation equation is replaced by a system of mass density (or mass fraction) conservation equations for each chemical species [1, 18]. In addition, a vibrational energy conservation equation is to be considered for each polyatomic – vibrating – species, see Reference [35].

Low densities, very high temperatures, and strong shocks thus produce a non-ideal thermodynamic behavior of the fluid, which demands for special care in designing efficient and accurate numerical algorithms.

High resolution approximations of shock-wave dominated flows have been obtained since the early eighties in the CFD community by means of shock-capturing techniques. Both finite difference (FD) and finite volume (FV) schemes result the natural frameworks where the shock-capturing methodology may be applied. A great amount of literature has been produced and the interested reader may be referred for instance to the collection of historical papers recently republished in Reference [25].

Both FD and FV methods may take advantage of those classes of algorithms generally referred to as upwind or flux-split schemes. In such algorithms, the estimation of the numerical flux is biased somehow in the direction determined by the signs of the characteristic flow fields. Among them, we will just mention the Steger-Warming flux-vector splitting technique [40], whose extension to non-equilibrium flows has been utilized in the present work – see also the references above. A full presentation of these topics may be found for instance in [16, 23].

It is an historical fact that the aforementioned methods were firstly developed for perfect gas calculations. In such a case, the equation of state is relatively simple and the homogeneity property of the inviscid compressible flux makes possible an almost straightforward procedure to obtain algebraic relationships for the numerical split fluxes and their associated Jacobians, [40]. The homogeneity property is not retained for a general real gas system [20], but it is still retained in the case of thermally perfect gases, [18, 19, 32].

Several successful extensions have been proposed in literature to generalize upwind and flux-split methods to the case of thermal and chemical equilibrium flows, see for instance [14, 31, 43] and subsequently to include the treatment of non-equilibrium chemistry and vibrational relaxation effects [1, 18, 19, 32, 38].

Another potential source of difficulty that numerical algorithms for high-speed flows with real gas phenomena must account for relies in the presence of *stiff* source terms, [35]. More precisely, stiffness is originated in the different time scales associated with fluid motion and non-equilibrium chemistry and thermodynamics, see [2, 35, 18]. If a conditionally stable algorithm was naively used to advance a discrete solution in time, for instance an explicit time stepping scheme, stiffness would force the time step to a value which should be dramatically smaller than the one usually required by spatial accuracy concerns. The resulting method would be too expensive and unpractical for real calculations.

Frequent and rather common cures for stiff source terms in a set of ordinary differential equations consist in adopting implicit discretization or some special splitting procedures which allow a separate integration.

Both the approaches have been exploited and largely investigated in order to include the stiff non-equilibrium chemically reactive and vibrational energy source terms in species conservation and gas dynamic equations. However, a special care is demanded in designing such solution algorithms. It has been shown in [17, 29] and more recently discussed by [30] that when a conservation law is coupled to a stiff source term which represents chemical reactions, the numerical dissipation introduced by the scheme may produce an incorrect propagation speed for a time-dependent discontinuous solution.

A decoupling of the non-equilibrium chemical and vibrational equations from the fluid dynamic system has been adopted for instance in [2]. The chemical species are firstly convected without considering chemistry and vibrational effects, and then chemical reactions are integrated using a separate incremental scheme in conjunction with frozen flow conditions. The time-split approach has been shown to be effective also in the recent work by [13]. Therein, a 2-nd order-in-time Strang splitting is utilized in order to handle source terms separately as a set of stiff non-linear ODEs.

Alternatively, fully-coupled approach with implicit treatment of stiff source terms have been shown to be capable of accounting for non-equilibrium real gas effects, see for instance References [7, 18, 19, 21].

A strongly non-linear algebraic problem is provided by the discretization method, which requires some non-linear iterative technique with (formally) large matrix inversions at each iteration. The resulting algorithm may thus be very expensive.

Intermediate between decoupled and fully-coupled approach is the partially decoupled semi-implicit scheme proposed for unsteady flows by [21]. The full set of equations is partially split in two subsets, the first one for the usual gas dynamic variables (total fluid density, momentum and energy) and some characteristic thermal quantities (specific heat ratio, the universal gas constant, assumed to be variable in space, and part of the energy), and the second one for species mass fractions and total energy. Despite the efficiency and the originality of the approach, the authors themselves report that their scheme may fail in calculating solutions dominated by very high Mach number and complex shock patterns: a kinked Mach stem may appear in the ramp problem solution. This failure is a well-known and documented numerical effect, among the ones catalogued and discussed in details by [37].

The objective of the present paper is to present an implicit finite volume approximate solution algorithm, developed in the spirit of fully-coupled methods, for time-dependent hypersonic calculations on 2-D triangle-based unstructured meshes.

The proposed method makes usage of a flux vector-splitting technique along the lines of the original Steger-Warming approach [40] and the generalization to non-equilibrium flows proposed in [32]. However, it is formulated in a very general way by introducing a suitable chemical reaction matrix to treat *any* chemical and vibrational model. That is, no model-dependent Jacobians are used to express numerical flux formulae.

Global second order-in-space and -time accuracy is provided by flux blending and implicit multi-stage time marching scheme. A bidimensional flux limiter is introduced to control numerical oscillations. The new limiter effectively extends in 2-D the family of flux limiting functions analyzed by [41].

Although some rather standard and well-known techniques in shock-capturing methodology are utilized, the method presented herein is original – at least to the authors' knowledge – in the way the species mass densities, vibrational energies, and stiff source terms are implicitly discretized in time. The approach leads to a non-linear algebraic problem with a very peculiar block structure, which may be exploited to decouple at the algebraic level the species mass density equations, the vibrational equations and the rest of the fluid dynamic system.

Thus, the global non-linear problem is decomposed in four separate and smaller size problems, which are solved iteratively via a block Gauss-Seidel-like algorithm and a standard preconditioned Krylov subspace solver [42].

Furthermore, each single block coefficient matrix is shown to be an M-matrix. This property guarantees the positivity of the species mass densities and the vibrational energies at each time step under a suitable and not too restrictive *CFL*-like constraint. A way of enforcing positivity on mass fractions and vibrational energies was proposed by [27] in the framework of approximate Riemann solver generalized to mixtures of real gases. However, as pointed out by the author, the main trade-off in this approach just consists in the *CFL*-like condition which might be impossible to fulfill in practice.

In section 2, we present the mathematical model, in 3 the derivation of the scheme, and in 4 the algebraic decomposition and its iterative solution procedure. Finally, in section 5, several numerical examples illustrate the shock-capturing capabilities and the effectiveness in describing real gas effects of the new method. The test case suite includes both classical 1-D calculations extended in 2-D and real 2-D simulations of unsteady shock wave systems on compression ramps and a steady bow shock calculation in a blunt-body problem.

2. MATHEMATICAL MODEL

The unsteady hypersonic flow of a mixture of compressible gases with chemically reactive processes can be modelled by the following set of time-dependent partial differential equations

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{U}) = \mathbf{S}(\mathbf{U}). \tag{1}$$

System (1) expresses in divergence form the conservation of the vector of unknowns $\mathbf{U}$ by balancing their time variation rate with the convective fluxes $\mathbf{F}(\mathbf{U})$ and the chemical reaction source terms $\mathbf{S}(\mathbf{U})$. These vector quantities take the form

$$\mathbf{U} = \begin{bmatrix} \boldsymbol{\rho} \\ \mathcal{E}^v \\ \rho\mathbf{v} \\ \rho E \end{bmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} \boldsymbol{\rho} \otimes \mathbf{v} \\ \mathcal{E}^v \otimes \mathbf{v} \\ \rho\mathbf{v} \otimes \mathbf{v} + p\mathbf{I} \\ \rho H \mathbf{v} \end{bmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{z} \\ \mathbf{0} \\ 0 \end{bmatrix}. \tag{2}$$

In equation (2), the components of $\mathbf{U}$ are logically collected in the species mass density vector, denoted by $\boldsymbol{\rho}$, the vibrational energy vector, $\mathcal{E}^v$, the momentum vector, $\rho\mathbf{v}$, and the total energy, ρE. The symbol $\otimes$ which appears in the flux definition denotes the dyadic product between vectors, that is $(\mathbf{a} \otimes \mathbf{b})_{ij} = a_i b_j$. The components of $\boldsymbol{\rho}$ are the chemical mass densities of the n_s different species considered in the model, ρ indicates the total mass density, the components of the vector $\mathcal{E}^v$ are the vibrational energy of the n_v diatomic

species, $\mathbf{v} = [u, v]^T$ denotes the velocity vector, and ρE the total energy. The pressure p satisfies the equation of state for a mixture of thermally perfect gases

$$p = \sum_{k=1}^{n_s} \frac{\rho_k}{\mathcal{M}_k} \mathcal{R} T,$$

where we introduced the chemical molecular weights $\mathcal{M}_k$ of the $k-$th species, the universal gas constant, $\mathcal{R} = 8.3143 J/(mole\ ^\circ\!K)$, and the translational-rotational temperature, T.

We assume that T satisfies the mechanical-thermal relation, see Reference [7],

$$\rho E = T \sum_{k=1}^{n_s} \rho_k \left(C_k^v + h_k^0 \right) + \sum_{j=1}^{n_v} \mathcal{E}_j^v + \frac{1}{2} \rho \left| \mathbf{v} \right|^2,$$

where C_k^v and h_k^0 are, respectively, the constant volume specific heat and the heat of formation at $0\,^\circ\!K$ reference temperature for the k-th species. The translational-rotational temperature represents the contribution to the internal energy by the translational and rotational modes of all molecules and atoms in the mixture, assumed to be in thermal equilibrium. It is worth noticing that the constant coefficients C_k^v do not represent the complete specific heats, but just the part due to molecular translation and rotation. They take the value $C_k^v = (3/2)(\mathcal{R}/\mathcal{M}_k)$ for monatomic species and $C_k^v = (5/2)(\mathcal{R}/\mathcal{M}_k)$ for diatomic species, [26]. The total enthalpy is denoted by H and is related to E, p and ρ by

$$H = E + \frac{p}{\rho}.$$

The source vector $\mathbf{S}(\mathbf{U})$ takes into account both dissociation-recombination and exchange reactions in a mixture of thermally perfect gases. The term $\boldsymbol{\omega}$ models the chemical reactions while the term $\mathbf{z}$ is responsible for the exchange reactions.

We assume that $\boldsymbol{\omega}$ depends on the mass densities $\boldsymbol{\rho}$ and on the thermal state of the mixture of gases. This latter one can be described in terms of the translational-rotational temperature T and a set of vibrational temperatures (or energies), T_j^v, for $j = 1 \ldots n_v$, one for each polyatomic species. Each vibrational temperature represents the contribution to the internal energy by the vibrational modes of the corresponding polyatomic species. Let us introduce the $(n_v + 1)$ component vector $\mathbf{T} = (T,\ T_1^v,\ \ldots\ T_{n_v}^v)^T$. Hence, we can indicate the functional dependence of the source term $\boldsymbol{\omega}$ as $\boldsymbol{\omega}(\boldsymbol{\rho}, \mathbf{T})$, and its generic r-th component can be expressed as in [2] by

$$\boldsymbol{\omega}_r = \sum_{j=1}^{n_s} \mathcal{M}_j (\nu''_{rj} - \nu'_{rj}) \left[\mathrm{K}_{f\,j}(\mathbf{T}) \prod_{i=1}^{n_s} \left(\frac{\rho_i}{\mathcal{M}_i} \right)^{\nu'_{ij}} - \mathrm{K}_{b\,j}(\mathbf{T}) \prod_{i=1}^{n_v} \left(\frac{\rho_i}{\mathcal{M}_i} \right)^{\nu''_{ij}} \right]. \quad (3)$$

In (3), ν'_{rj} and ν''_{rj} are stoichiometric coefficients, while $\mathrm{K}_f(\mathbf{T})$ and $\mathrm{K}_b(\mathbf{T})$ denote the forward and backward rate coefficients, respectively. Their functional dependence on $\mathbf{T}$ is provided by the chemical model.

As pointed out in [5], the source term $\boldsymbol{\omega}$ can be written in the form $\mathbf{C}(\boldsymbol{\rho},\mathbf{T})\boldsymbol{\rho}$, where $\mathbf{C}(\boldsymbol{\rho},\mathbf{T})$ is a $n_s \times n_s$ matrix with continuous entries $C_{ij}(\boldsymbol{\rho},\mathbf{T})$, such that

$$\begin{aligned} &\text{(a)} && C_{ii}(\boldsymbol{\rho},\mathbf{T}) \le 0 && i=1,\dots,n_s \\ &\text{(b)} && C_{ij}(\boldsymbol{\rho},\mathbf{T}) \ge 0 && i \ne j \\ &\text{(c)} && \sum_{i=1}^{n_s} C_{ij}(\boldsymbol{\rho},\mathbf{T}) = 0 && j=1,\dots,n_s. \end{aligned}$$

The generic r-th component of $\boldsymbol{\omega}$ can be decomposed in the sum of a production and a consumption term, respectively denoted by ω_r^+ and ω_r^-. That is, it holds: $\omega_r = \omega_r^+ + \omega_r^-$. The diagonal and off-diagonal entries of the matrix $\mathbf{C}(\boldsymbol{\rho},\mathbf{T})$ respectively contribute to $\omega_r^{\pm}$ as

$$\begin{aligned} \omega_r^-(\boldsymbol{\rho},\mathbf{T}) &= C_{rr}(\boldsymbol{\rho},\mathbf{T})\rho_r\,, \\ \omega_r^+(\boldsymbol{\rho},\mathbf{T}) &= \sum_{j=1}^{n_s}{}^{(r)} C_{rj}(\boldsymbol{\rho},\mathbf{T})\rho_j\,, \end{aligned} \tag{4}$$

where the symbol $^{(r)}$ indicates that the term $j=r$ is dropped out in the sum.

The generic j-th component of the vibrational energy exchange source term can be written as

$$z_j = \frac{\mathcal{E}_j^{eq} - \mathcal{E}_j^{v}}{\tau_j} + \frac{\omega_j^+}{\rho_j}\mathcal{E}_j^{eq} + \frac{\omega_j^-}{\rho_j}\mathcal{E}_j^{v}, \qquad j=1,\dots,n_v, \tag{5}$$

where $\mathcal{E}^{eq}$ is the vector of the equilibrium energy densities and the τ_j are the relaxation times, [26]. Both constitutive relations for $\mathcal{E}_j^{eq}$ and expressions for τ_j to be used in practical computations are model-dependent. As remarked in Reference [32], the choice of an appropriate model for describing a given phenomenology may be a difficult task. In this paper the approximate solution algorithm is formulated as generally as possible by using the matrix $\mathbf{C}(\boldsymbol{\rho},\mathbf{T})$. In this way, we do not attempt to suggest which model should be used for practical calculations, but many different models could be included in the algorithm by formally changing the entries of $\mathbf{C}(\boldsymbol{\rho},\mathbf{T})$. More details about the chemical and thermodynamical model used for the numerical tests in section 5 are reported therein.

3. FINITE VOLUME FORMULATION

In this section, we present the construction of the cell-centered finite volume method for the numerical approximations on unstructured 2-D grids of unsteady hypersonic flows. An upwind semi-discrete approximation of the integral form of system (2) is first derived. A finite volume scheme can be obtained by introducing a finite difference approximation of the time derivative of the vector of unknowns $\mathbf{U}$. It is well-known that this approach yields numerical schemes which are formally first order accurate in space and time, [23]. The upwinding mechanism provides enough dissipation to ensure monotonicity, and prevents the formation of numerical oscillations even if the solution shows strong shock discontinuities, [16]. Higher order accuracy can be obtained by blending the first order upwind numerical flux with a second order accurate symmetric one, along the lines of [44, 45, 46].

This blending is *locally* driven by a suitable limiting procedure which controls the numerical oscillations, [41]. Second order accuracy in time is finally achieved by means of a two-step time-marching integration scheme, [9]. Although some rather standard ideas from shock-capturing techniques are applied, to the best of the authors' knowledge, several aspects in the proposed approach are substantially new. They are: the final functional form of the numerical flux for a mixture of thermally perfect gases; the form of the flux-limiter, the way the implicit discretization is performed and the algebraic decomposition to solve the final stiff non-linear problem. For this reason, the derivation of the scheme is presented in full herein, while the essential features of the non-linear solution procedure are discussed in details in the next section.

3.1. Preliminaries

Let us first introduce a *triangulation* which covers all the computational domain – also referred to as the *mesh* in the rest of the paper. The triangulation is defined as a collection of N triangles, conventionally denoted by K, and assumed *regular* and *conformal* in the sense specified by [8]. Triangles are conventionally labeled by an integer identifier ranging in a global numbering system. The identifier may be generically indicated by the index letters i, j or k. For the generic triangle i we indicate by $|K_i|$ the area of the triangle, by $\sigma(i)$ the set of adjacent triangles, and by $\sigma'(i)$ the subset of triangle edges located at the boundary. The internal edge shared by triangles i and j will be denoted by the pair ij. For the sake of notation consistency, a boundary edge will also be indicated by a pair of indices ik, i being in such a case the unique triangle the edge belongs to, and k a specific boundary edge identifier (like a fictitious "external" triangle). This convention allows us to refer to either internal or boundary edges by means of an index pair. For the generic edge ij, we indicate by ℓ_{ij} its length, and by $\mathbf{n}_{ij}$ the normal vector. This latter one is assumed to be oriented from cell i to cell j when the edge is internal and outward directed when the edge is on the boundary. For the sake of clearness, throughout the paper the notation $\mathbf{F}(\mathbf{U}, \mathbf{n})$ will indicate the normal projection of the flux vector, i.e. $\mathbf{F}(\mathbf{U}) \cdot \mathbf{n}$.

The semi-discrete finite volume approximation is

$$|K_i| \frac{d\mathbf{U}_i}{dt} + \sum_{j \in \sigma(i)} \ell_{ij} \mathbf{H}(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij}) + \sum_{j' \in \sigma'(i)} \ell_{ij'} \mathbf{H}^{(bc)}_{ij'} = |K_i| \, \mathbf{S}_i(\mathbf{U}_i), \tag{6}$$

with the index i running throughout all the mesh triangles, i.e $i = 1, 2, ..., N$. This formulation is obtained by integrating in a cell-wise fashion system (1), applying the divergence theorem, approximating the interface integrals with the midpoint rule, and finally introducing some suitable numerical flux function, [23].

The quantity $\mathbf{U}_i$ stands for the cell-averaged solution within the triangle i. The terms $\mathbf{H}(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij})$ and $\mathbf{H}^{(bc)}_{ij'}$ are respectively the numerical flux function at internal edges and at boundary edges. The former one depends on the cell-averaged solutions $\mathbf{U}_i$ and $\mathbf{U}_j$ within the cells sharing the given edge, while the latter one depends on the cell-averaged solution $\mathbf{U}_i$ within the unique boundary triangle i and may depend in some suitable form on a set of *external* data $\mathbf{U}^{(bc)}_{j'}$.

Remark. For boundary fluxes the notation $\mathbf{H}^{(bc)}_{ij'}$ instead of $\mathbf{H}^{(bc)}(\mathbf{U}_i, \mathbf{U}^{(bc)}_{j'}, \mathbf{n}_{ij'})$ is preferred. This is because boundary conditions may differ at distinct boundary edges, implying also a different functional form for the numerical fluxes.

3.2. Construction of the numerical flux

The construction of the numerical flux by a flux-vector splitting approach dates back to [40], where a decomposition of the form

$$\mathbf{H}(\mathbf{U},\mathbf{V},\mathbf{n}) = \mathbf{F}^{+}(\mathbf{U},\mathbf{n}) + \mathbf{F}^{-}(\mathbf{V},\mathbf{n}) \tag{7}$$

is proposed. In (7), $\mathbf{F}^{\pm}(\mathbf{U},\mathbf{n})$ are the normal projections of the vectors $\mathbf{F}^{\pm}(\mathbf{U})$, and by definition there holds $\mathbf{F}^{\pm}(\mathbf{U}) = \mathbf{J}^{\pm}(\mathbf{U})\mathbf{U}$. The matrices $\mathbf{J}^{\pm}(\mathbf{U})$ are built by diagonalizing the Jacobian matrix $\mathbf{J}(\mathbf{U})$ and splitting its eigenvalues into a positive and a negative part. This approach basically relies on the homogeneity property $\mathbf{F}(\mathbf{U}) = \mathbf{J}(\mathbf{U})\mathbf{U}$, which is satisfied by the compressible Euler flux, [23]. As discussed in [18], homogeneity is again retained in the case of a mixture of thermally perfect gases. Thus, accordingly to the form of the flux vector $\mathbf{F}(\mathbf{U})$ for a reactive multi-component system in (2), and after some algebraic manipulations, the flux split methodology yields the two partial contributions $\mathbf{F}^{\pm}$ as

$$\mathbf{F}^{\pm}(\mathbf{U},\mathbf{n}) = \mathbb{a}^{\pm}(\mathbf{U},\mathbf{n})\mathbf{U} + \mathbf{G}^{\pm}(\rho,\rho\mathbf{v},\mathbf{U},\mathbf{n}), \tag{8}$$

where

$$\mathbf{G}^{\pm}(\rho,\rho\mathbf{v},\mathbf{U},\mathbf{n}) = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \rho\,\mathbb{b}^{\pm}(\mathbf{U},\mathbf{n})\mathbf{n} \\ \rho\mathbf{v}\cdot\mathbf{n}\,\mathbb{b}^{\pm}(\mathbf{U},\mathbf{n}) + \rho\,\mathbb{c}^{\pm}(\mathbf{U},\mathbf{n}) \end{bmatrix}, \tag{9}$$

and

$$\begin{aligned} \mathbb{a}^{\pm}(\mathbf{U},\mathbf{n}) &= \frac{2\beta(\mathbf{v}\cdot\mathbf{n})^{\pm} + (\mathbf{v}\cdot\mathbf{n}+c)^{\pm} + (\mathbf{v}\cdot\mathbf{n}-c)^{\pm}}{2(1+\beta)}, \\ \mathbb{b}^{\pm}(\mathbf{U},\mathbf{n}) &= \frac{\kappa}{2}\left((\mathbf{v}\cdot\mathbf{n}+c)^{\pm} - (\mathbf{v}\cdot\mathbf{n}-c)^{\pm}\right), \\ \mathbb{c}^{\pm}(\mathbf{U},\mathbf{n}) &= \frac{\kappa^2}{2}\left((\mathbf{v}\cdot\mathbf{n}+c)^{\pm} + (\mathbf{v}\cdot\mathbf{n}-c)^{\pm} - 2(\mathbf{v}\cdot\mathbf{n})^{\pm}\right). \end{aligned} \tag{10}$$

In equations (10), for $q = (\mathbf{v}\cdot\mathbf{n})$, $(\mathbf{v}\cdot\mathbf{n}\pm c)$, the symbol $q^{\pm}$ takes the usual definition $q^{\pm} = (q \pm |q|)/2$. Finally, the parameters β, c, and κ are defined as

$$\beta = \frac{\mathcal{R}\displaystyle\sum_{i=1}^{n_s}\frac{\rho_i}{\mathcal{M}_i}}{\displaystyle\sum_{i=1}^{n_s}\rho_i C_i^v}, \qquad c = \sqrt{\frac{1+\beta}{\rho}p}, \qquad \kappa = \frac{c}{1+\beta},$$

where c is the frozen speed of sound and $1+\beta$ is the frozen specific heat ratio of the mixture, [21].

3.3. Semi-discrete FV formulation

The semi-discrete FV formulation

$$
\begin{aligned}
|K_i| \frac{d\mathbf{U}_i}{dt} &+ \sum_{j\in\sigma(i)} \ell_{ij} \left\{ \mathbb{a}^+(\mathbf{U}_i, \mathbf{n}_{ij})\mathbf{U}_i - \mathbb{a}^+(\mathbf{U}_j, \mathbf{n}_{ji})\mathbf{U}_j \right\} \\
&+ \sum_{j\in\sigma(i)} \ell_{ij} \left\{ \mathbf{G}^+(\rho_i, (\rho\mathbf{v})_i, \mathbf{U}_i, \mathbf{n}_{ij}) - \mathbf{G}^+(\rho_j, (\rho\mathbf{v})_j, \mathbf{U}_j, \mathbf{n}_{ji}) \right\} \\
&+ \sum_{j'\in\sigma'(i)} \ell_{ij'} \mathbf{H}^{(bc)}_{ij'} = |K_i|\, \mathbf{S}_i(\mathbf{U}_i),
\end{aligned}
\tag{11}
$$

is obtained by using in (6) the definition (8) for $\mathbf{F}^{\pm}$, and then exploiting the geometric identity $\mathbf{n}_{ij} = -\mathbf{n}_{ji}$, and the "local" flux consistency condition $\mathbf{F}^+(\mathbf{U}, \mathbf{n}) = -\mathbf{F}^-(\mathbf{U}, -\mathbf{n})$, which holds since the scheme must be conservative. The numerical flux $\mathbf{H}^{(bc)}_{ij'}$ at boundaries can be written as

$$
\mathbf{H}^{(bc)}_{ij'} = \theta^{solid}_{ij'} p_i \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{n}_{ij'} \\ 1 \end{bmatrix} + \theta^{free}_{ij'} \mathbf{F}(\mathbf{U}_i) \cdot \mathbf{n}_{ij'} + \theta^{inlet}_{ij'} \mathbf{F}(\mathbf{U}^{(bc)}_{j'}) \cdot \mathbf{n}_{ij'}. \tag{12}
$$

In equation (12) the parameters $\theta^{solid}_{ij'}$, $\theta^{free}_{ij'}$ and $\theta^{inlet}_{ij'}$ are three mutually exclusive switchers which may take integer values 0 or 1, in order to get the correct boundary flux expression (respectively for a *solid wall*, *free* or *supersonic inlet* boundary).

Using the definition of $\omega^{\pm}_r$ in (4), within the triangle i the source term $\mathbf{S}_i(\mathbf{U}_i)$ takes the form

$$
\mathbf{S}_i(\mathbf{U}_i) = \mathbf{S}(\boldsymbol{\rho}_i, \mathbf{T}_i) = \begin{bmatrix} \mathbf{C}(\boldsymbol{\rho}_i, \mathbf{T}_i)\boldsymbol{\rho}_i \\ \mathbf{s}(\boldsymbol{\rho}_i, \mathbf{T}_i) \otimes \mathcal{E}^{eq} - \mathbf{d}(\boldsymbol{\rho}_i, \mathbf{T}_i) \otimes \mathcal{E}^{v} \\ \mathbf{0} \\ 0 \end{bmatrix}
$$

where

$$
[\mathbf{d}(\boldsymbol{\rho}, \mathbf{T})]_r = \frac{1}{\tau_r} + \frac{\omega^+_r(\boldsymbol{\rho}, \mathbf{T})}{\rho_r}, \qquad [\mathbf{s}(\boldsymbol{\rho}, \mathbf{T})]_r = \frac{1}{\tau_r} - \frac{\omega^-_r(\boldsymbol{\rho}, \mathbf{T})}{\rho_r}.
$$

The symbol $[\cdot]_r$ indicates the component related to the r-th species.

Note that the distinct functional dependence on the density vector $\boldsymbol{\rho}$ and the temperature vector $\mathbf{T}$ in the source term $\mathbf{S}$ is still retained. This choice will be motivated in the following section.

3.4. Base first order semi-implicit scheme

The base time-marching scheme is obtained by approximating the time derivative of $\mathbf{U}_i$ – which appears in the first term in the semi-discrete formulation (11) – by first-order finite differences

$$
\left.\frac{d\mathbf{U}_i(t)}{dt}\right|_{t=t^n} \approx \frac{\mathbf{U}^{n+1}_i - \mathbf{U}^n_i}{\Delta t},
$$

where $\mathbf{U}_i^{n+1}$ and $\mathbf{U}_i^n$ are the cell-averaged solutions within the triangle i at times t^{n+1} and t^n, and $\Delta t = t^{n+1} - t^n$. Let us also introduce the symbols a_{ij}^n and $\mathbf{G}_{ij}^n$ defined as

$$\mathrm{a}_{ij}^n = \mathbb{a}^+(\mathbf{U}_i^n, \mathbf{n}_{ij}), \qquad \mathbf{G}_{ij}^{n,\,n+1} = \mathbf{G}^+(\rho_i^{n+1}, (\rho\mathbf{v})_i^{n+1}, \mathbf{U}_i^n, \mathbf{n}_{ij}).$$

Thus, the base time-marching scheme takes the form

$$\begin{aligned} |K_i|\,\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} \; &+ \sum_{j\in\sigma(i)} \ell_{ij}\left(\mathrm{a}_{ij}^n \mathbf{U}_i^{n+1} - \mathrm{a}_{ji}^n \mathbf{U}_j^{n+1}\right) \\ &+ \sum_{j\in\sigma(i)} \ell_{ij}\left(\mathbf{G}_{ij}^{n,\,n+1} - \mathbf{G}_{ji}^{n,\,n+1}\right) \\ &+ \sum_{j'\in\sigma'(i)} \ell_{ij'} \mathbf{H}_{ij'}^{(bc)n} = |K_i|\,\mathbf{S}(\boldsymbol{\rho}_i^{n+1}, \mathbf{T}_i^n). \end{aligned} \tag{13}$$

The term $(\mathrm{a}_{ij}^n \mathbf{U}_i^{n+1} - \mathrm{a}_{ji}^n \mathbf{U}_j^{n+1})$, which appears in the first summation in (13), is originated by considering the first — zero-th order — term of the Taylor expansion in time of $\mathbb{a}^+(\mathbf{U}(t), \mathbf{n}_{ij})$ around t^n. This strategy will also produce a block-diagonal matrix operator, whose block-matrix components are M-matrices, see section 4.

The term $(\mathbf{G}_{ij}^{n,\,n+1} - \mathbf{G}_{ji}^{n,\,n+1})$, which appears in the second summation in (13), is instead originated by using $\mathbf{U}^n$ in $\mathbb{b}^{\pm}(\mathbf{U}, \mathbf{n})$ and $\mathbb{c}^{\pm}(\mathbf{U}, \mathbf{n})$, and ρ^{n+1} and $(\rho\mathbf{v})^{n+1}$ for the terms ρ and $\rho\mathbf{v}$ which appears explicitly in the definition of $\mathbf{G}^{\pm}(\rho, \rho\mathbf{v}, \mathbf{U}, \mathbf{n})$ given in (9) and (10).

The discretization of the source term $\mathbf{S}(\boldsymbol{\rho}_i^{n+1}, \mathbf{T}_i^n)$ is implicitly dependent on the density vector $\boldsymbol{\rho}$ and explicitly dependent on the temperature vector $\mathbf{T}$. The implicit dependence on $\boldsymbol{\rho}$ is chosen for stability reasons to cope with the stiffness introduced by the chemical reactions, [2]. Instead, an explicit evaluation on $\mathbf{T}$ is preferred, because an implicit evaluation of the temperature would results in a strong non-linear system due to the exponential nature of the Arrhenius equation, see the appendix.

Scheme (13) is globally first-order accurate in space and time.

3.5. Second order in space accuracy

The accuracy in space may be improved by blending the first-order accurate upwind numerical flux in (8-10) with a central symmetric flux, i.e.

$$\mathbf{H}(\mathbf{U}, \mathbf{V}, \mathbf{n}) = (1-\theta)\left(\mathbf{F}^+(\mathbf{U}, \mathbf{n}) - \mathbf{F}^+(\mathbf{V}, -\mathbf{n})\right) + \theta\frac{\mathbf{F}(\mathbf{U}, \mathbf{n}) - \mathbf{F}(\mathbf{V}, -\mathbf{n})}{2}, \tag{14}$$

where the parameter $\theta \in [0,\,1]$ can be locally estimated. The choice $\theta = 0$ returns the original first-order upwind flux, while the one $\theta = 1$ the second order central numerical flux. The issue of how to estimate an appropriate value for the parameter θ is addressed in section 3.7. Using the definition of blended numerical flux given in (14), with the normal flux $\mathbf{F}(\mathbf{U}, \mathbf{n})$ written as

$$\mathbf{F}(\mathbf{U}, \mathbf{n}) = (\mathbf{v}\cdot\mathbf{n})\mathbf{U} + \mathbf{G}(\mathbf{U}, \mathbf{n}), \qquad \mathbf{G}(\mathbf{U}, \mathbf{n}) = \frac{c^2}{1+\beta}\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \rho\mathbf{n} \\ \rho\mathbf{v}\cdot\mathbf{n} \end{bmatrix},$$

the computational scheme (13) becomes

$$
\begin{aligned}
|K_i|\,\frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^{n}}{\Delta t} &+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{a}_{ij}^{n}\mathbf{U}_i^{n+1}-\hat{a}_{ji}^{n}\mathbf{U}_j^{n+1}\right\} \\
&+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{\mathbf{G}}_{ij}^{n,\,n+1}-\hat{\mathbf{G}}_{ji}^{n,\,n+1}\right\} + \sum_{j'\in\sigma'(i)} \ell_{ij'}\mathbf{H}_{ij'}^{(bc)n} \qquad (15)\\
&= |K_i|\,\mathbf{S}(\boldsymbol{\rho}_i^{n+1},\mathbf{T}_i^{n}) - \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{v}_{ij}^{n}\mathbf{U}_i^{n}-\hat{v}_{ji}^{n}\mathbf{U}_j^{n}\right\}.
\end{aligned}
$$

For the sake of clearness, in equation (15) we have introduced the θ-dependent symbols

$$
\begin{aligned}
\hat{a}_{ij} &= (1-\theta_{ij})\mathbb{a}^{+}(\mathbf{U}_i,\mathbf{n}_{ij}),\\
\hat{v}_{ij} &= \frac{\theta_{ij}}{2}(\mathbf{v}_i\cdot\mathbf{n}_{ij}),\\
\hat{\mathbf{G}}_{ij}^{n,\,n+1} &= (1-\theta_{ij})\mathbf{G}_{ij}^{n,\,n+1} + \frac{\theta_{ij}}{2}\mathbf{G}(\rho_i^{n+1},(\rho\mathbf{v})_i^{n+1},\mathbf{U}_i^{n},\mathbf{n}_{ij}).
\end{aligned}
$$

Note that a new term appears in the r.h.s. of equation (15), due to the blending of numerical fluxes.

3.6. Second order in time accuracy

The semi-discrete formulation in (6) actually consists in a large system of ordinary differential equations in the unknowns $\mathbf{U}_{ij}$. The time-marching scheme presented in section 3.4 basically consists in the application of a semi-implicit version of the Euler method for ODE. A formally second-order accurate-in-time scheme may be obtained by applying a slightly modified version of the two-step Collatz algorithm, see [9]. The first step evaluates a preliminary solution at the intermediate time $t^{n+\frac{1}{2}}$. The second step evaluates the numerical fluxes by using the approximate solution at $t^{n+\frac{1}{2}}$ and advances the solution from t^n to t^{n+1}. An implicit treatment of the source vector term $\mathbf{S}(\boldsymbol{\rho},\mathbf{T})$ must always be devised because of the stiffness introduced in the equations by the chemical reactions. Computational stability can be ensured by treating implicitly the dependence on $\boldsymbol{\rho}$ and explicitly the one on $\mathbf{T}$ in the source term $\mathbf{S}$ in both steps, [2, 13].

First step:

$$
\begin{aligned}
|K_i|\,\frac{\mathbf{U}_i^{n+\frac{1}{2}}-\mathbf{U}_i^{n}}{\Delta t/2} &+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{a}_{ij}^{n}\mathbf{U}_i^{n+\frac{1}{2}}-\hat{a}_{ji}^{n}\mathbf{U}_j^{n+\frac{1}{2}}\right\} \\
&+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{\mathbf{G}}_{ij}^{n,\,n+\frac{1}{2}}-\hat{\mathbf{G}}_{ji}^{n,\,n+\frac{1}{2}}\right\} + \sum_{j'\in\sigma'(i)} \ell_{ij'}\mathbf{H}_{ij'}^{(bc)n} \qquad (16)\\
&= |K_i|\,\mathbf{S}(\boldsymbol{\rho}_i^{n+\frac{1}{2}},\mathbf{T}_i^{n}) - \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{v}_{ij}^{n}\mathbf{U}_i^{n}-\hat{v}_{ji}^{n}\mathbf{U}_j^{n}\right\}
\end{aligned}
$$

Second step:

$$
\begin{aligned}
|K_i|\,\frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^n}{\Delta t} &+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{a}_{ij}^{n+\frac12}\,\frac{\mathbf{U}_i^{n+1}+\mathbf{U}_i^n}{2} - \hat{a}_{ji}^{n+\frac12}\,\frac{\mathbf{U}_j^{n+1}+\mathbf{U}_j^n}{2}\right\} \qquad (17)\\
&+ \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{\mathbf{G}}_{ij}^{n+\frac12,\,n+\frac12} - \hat{\mathbf{G}}_{ji}^{n+\frac12,\,n+\frac12}\right\} + \sum_{j'\in\sigma'(i)} \ell_{ij'}\mathbf{H}_{ij'}^{(bc)n+\frac12}\\
&= |K_i|\,\mathbf{S}(\boldsymbol{\rho}_i^{n+1},\mathbf{T}_i^{n+\frac12}) - \sum_{j\in\sigma(i)} \ell_{ij}\left\{\hat{v}_{ij}^{n+\frac12}\mathbf{U}_i^{n+\frac12} - \hat{v}_{ji}^{n+\frac12}\mathbf{U}_j^{n+\frac12}\right\}
\end{aligned}
$$

3.7. 2-D limiting procedure

In this section we present our strategy for estimating the coefficients θ_{ij}. The estimation is local in the sense that a different factor θ_{ij} is computed for each internal edge, depending on the approximate solution value within adjacent cells. Remark that no θ-factor is needed for boundary edges, since the numerical flux is specified by the boundary conditions. Each θ_{ij} takes a value in the range $[0,\,1]$, and, thus, plays the role of a flux limiter. In the last two decades, the limiters have been extensively studied in the framework of high resolution finite volume schemes, and a considerable amount of literature has been produced; we just mention the general review given in [23]. For the sake of convenience, limiting strategies are grouped in two great families, the one of flux-limiters and the one of slope-limiters. The *slope-limiters* are designed to ensure properties such as the preservation within each control volume of the integral cell average. *Flux-limiters*, while ensuring conservation, may not be expected to preserve the latter property. Our limiting strategy is based on a simple heuristic 2-D extension of the 1-D class of limiters which compare consecutive variations of the approximate numerical solution. Those 1-D limiters have been widely experienced by a number of authors and theoretically analyzed by [41]. The proposed face-based limiting strategies has proved to be successful in all of the present calculations. It does allow an edge-based implementation and does not need the storage of an accumulated limiter for each triangle. Nevertheless, while ensuring conservation, to the authors' knowledge it is not known whether the presented limiting strategy can preserve the integral cell average. The above issue is beyond the scope of this work.

In order to detect shock discontinuities in the approximate solution, we compare the slopes of some given indicator q, such as the pressure or the total density. Slopes are estimated on both sides of an internal edge as normal projections of the gradient of a linear interpolant of q.

Let us denote by $\mathbf{X}_i$ the centroid of the i-th triangle and by $\mathbf{X}_{ij}$ the point defined as

$$
\mathbf{X}_{ij} = \begin{cases} \dfrac{\mathbf{X}_i+\mathbf{X}_j}{2} & \text{if } ij \text{ is an internal edge,}\\ \text{midpoint of edge } ij & \text{if } ij \text{ is a boundary edge.}\end{cases}
$$

To compute the values of θ, we use the following procedure

(i) define

$$
q_{ij} = \begin{cases} \dfrac{q_i+q_j}{2} & \text{if } ij \text{ is an internal edge,}\\ q_i & \text{if } ij \text{ is a boundary edge,}\end{cases}
$$

where q_i is the value taken by q within triangle i, as illustrated by Figure 1;

(ii) for each triangle i, build the operator π_i, which linearly interpolates the three nodes $(\mathbf{X}_{ij}, q_{ij})$, $j \in \sigma(i)$ – in case of a boundary edge, take $j \in \sigma'(i)$ in place of the missing triangle;

(iii) for each internal edge ij, define the slopes s_i and s_j as the gradient of π_i and π_j projected along the direction $\mathbf{n}_{ij}$;

(iv) for each internal edge ν_{ij} is defined as

$$\nu_{ij} = \begin{cases} 0 & \text{if } s_i s_j \leq 0, \\ \dfrac{\min\{|s_i|, |s_j|\}}{\max\{|s_i|, |s_j|\}} & \text{otherwise.} \end{cases}$$

For time-dependent calculations θ_{ij} is set to ν_{ij}. For steady ones, the limiter proceeds throughout the further steps:

(v) for each triangle define θ_i as

$$\theta_i = \sqrt[3]{\nu_{ij}\nu_{ik}\nu_{il}},$$

(vi) for the internal edge ij define θ_{ij} as

$$\theta_{ij} = \min\{\nu_{ij}, \theta_i, \theta_j\}.$$

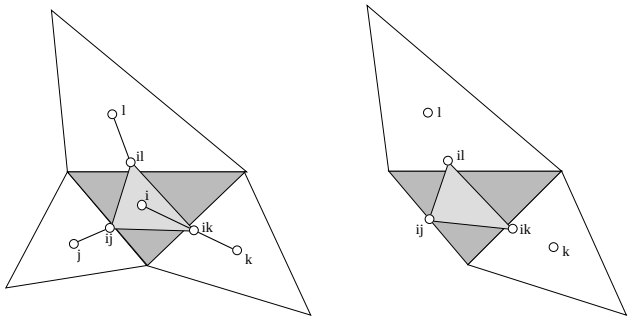

FIG. 1. Construction of the limiter

In the computation done in the paper $q = |\mathbf{v}|, c$.

4. SOLUTION ALGORITHM

By inspecting the structure of the two-stage scheme formulated in section 3.6, we note that equations (16)–(17) can be written more compactly in matricial form

$$\begin{aligned} &\textbf{first step:} && \boldsymbol{\mathcal{L}}^n(\boldsymbol{\rho}^{n+\frac{1}{2}})\mathbf{U}^{n+\frac{1}{2}} + \boldsymbol{\mathcal{G}}^{n,\,n+\frac{1}{2}} = \boldsymbol{\mathcal{B}}^n, \\ &\textbf{second step:} && \boldsymbol{\mathcal{L}}^{n+\frac{1}{2}}(\boldsymbol{\rho}^{n+1})\mathbf{U}^{n+1} + \boldsymbol{\mathcal{G}}^{n+\frac{1}{2},\,n+\frac{1}{2}} = \boldsymbol{\mathcal{B}}^{n+\frac{1}{2}}, \end{aligned}$$

where $\mathfrak{L}^{n+\alpha}(\boldsymbol{\rho})$ is the block-diagonal matrix operator:

$$\left(\begin{array}{c|c|c|c} \mathbf{D}_{\rho}^{n+\alpha}(\boldsymbol{\rho}) + \mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_s} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{D}_{e}^{n+\alpha}(\boldsymbol{\rho}) + \mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_v} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{M}^{n+\alpha} \otimes \mathbf{I}_2 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}^{n+\alpha} \end{array}\right) \tag{18}$$

A unified time index notation was introduced such that $\alpha = 0$ and $\alpha = 1/2$ respectively denote the first and the second stage of the method. The terms involved in $\mathfrak{L}^{n+\alpha}$ are defined as

$$\begin{aligned} \mathbf{D}_{\rho,\,ij}^{n+\alpha}(\boldsymbol{\rho}) &= \delta_{ij}\Delta t \left(\alpha + \frac{1}{2}\right) \mathbf{C}(\boldsymbol{\rho}, \mathbf{T}_i^{n+\alpha}), \\ \mathbf{D}_{e,\,ij}^{n+\alpha}(\boldsymbol{\rho}) &= \delta_{ij}\Delta t \left(\alpha + \frac{1}{2}\right) \mathbf{d}(\boldsymbol{\rho}, \mathbf{T}_i^{n+\alpha}), \end{aligned} \tag{19}$$

and

$$M_{ij}^{n+\alpha} = \begin{cases} |K_i| + \dfrac{\Delta t}{2} \displaystyle\sum_{k \in \sigma(i)} \ell_{ik} \hat{a}_{ki}^{n+\alpha} & \text{if } i = j, \\ -\dfrac{\Delta t}{2} \ell_{ij} \hat{a}_{ij}^{n+\alpha} & \text{if } i \neq j, \end{cases} \tag{20}$$

where δ_{ij} in (19) is the Dirac-Kronecker symbol. The $\mathfrak{G}$ operator

$$\mathfrak{G}^{n+\alpha,\,n+\frac{1}{2}} = \left(\begin{array}{c|c|c|c} \mathbf{0} & \mathbf{0} & \mathfrak{G}_{\rho\mathbf{v}}^{n+\alpha,\,n+\frac{1}{2}} & \mathfrak{G}_{\rho E}^{n+\alpha,\,n+\frac{1}{2}} \end{array}\right)^T \tag{21}$$

and the r.h.s. vector

$$\mathfrak{B}^{n+\alpha} = \left(\begin{array}{c|c|c|c} \mathbf{b}_{\rho}^{n+\alpha} & \mathbf{b}_{\mathcal{E}^v}^{n+\alpha} & \mathbf{b}_{\rho\mathbf{v}}^{n+\alpha} & \mathbf{b}_{\rho E}^{n+\alpha} \end{array}\right)^T \tag{22}$$

are expressed by some rather complicated formulae reported in the final appendix for the sake of completeness. Let us first notice that $\mathbf{D}_{\rho}^{n+\alpha}$ and $\mathbf{D}_{e}^{n+\alpha}$ are respectively a block-diagonal and a diagonal non-negative matrix.

A remarkable property holds for all the diagonal blocks forming the operator (18). We formally state it in the next proposition. The proof is given in the appendix.

PROPOSITION 4.1. *The four diagonal blocks defined in* (18) *are M-matrices.*

The formal inverse of an M-matrix has only non negative entries. Thus, if $\mathbf{b}_{\rho}^{n+\alpha} \geq \mathbf{0}$ and $\mathbf{b}_{e}^{v,\,n+\alpha} \geq \mathbf{0}$, then $\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}}, \mathcal{E}^{n+\alpha+\frac{1}{2}} \geq \mathbf{0}$, that is the scheme prevents by construction that negative – i.e. unphysical – densities or vibrational energies appear during the solution process. A simple inspection of the r.h.s. terms – see the appendix for details – shows that positivity of species mass densities and vibrational energies generally hold under a *CFL*-like constraint on Δt. This can be formally stated as follows.

PROPOSITION 4.2. *Let us denote by h the minimum height of the triangles in the mesh, and suppose that*

$(i)\boldsymbol{\rho}^n \geq 0$ *and* $\mathcal{E}^{v,\,n} \geq 0$*;*
$(ii)\dfrac{\Delta t}{h}\max\{u, c\} < 1/6$*, where* $u = \max\{|\mathbf{v}_i|\}$ *and* $c = \max\{c_i\}$.

Then, $\boldsymbol{\rho}^{n+1} \geq 0$ *and* $\mathcal{E}^{v,\,n+1} \geq 0$.

The proof is given in the appendix. Condition (Prop. 4.2-ii) leads to an important reduction of the time step size. The time step becomes smaller but comparable to the one typically associated with an explicit time marching scheme. Numerical experience shows that positivity of species mass densities and vibrational energies still holds when greater time steps, corresponding to *CFL* numbers of about one, are used. It is reasonable to conjecture that proposition 4.2 states a somewhat crude estimate which is far from optimality.

The block-diagonal structure of the operator $\mathfrak{L}^{n+\alpha}$ suggests a block Gauss-Seidel-like decoupling into four separate sub problems to be solved sequentially. The solution algorithm proceeds as follows

(i) solve the non-linear system for species mass densities $\boldsymbol{\rho}$,

$$\left(\mathbf{D}_{\rho}^{n+\alpha}(\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}}) + \mathbf{M}^{n+\alpha}\otimes\mathbf{I}_{n_s}\right)\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}} = \mathbf{b}_{\rho}^{n+\alpha},$$

(ii) solve the linear system for vibrational energies $\boldsymbol{\mathcal{E}}^v$,

$$\left(\mathbf{D}_{e}^{n+\alpha}(\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}}) + \mathbf{M}^{n+\alpha}\otimes\mathbf{I}_{n_v}\right)\mathcal{E}^{v,\,n+\alpha+\frac{1}{2}} = \mathbf{b}_{\mathcal{E}}^{n+\alpha},$$

(iii) solve the linear system for momenta $\rho\mathbf{v}$,

$$\left(\mathbf{M}^{n+\alpha}\otimes\mathbf{I}_2\right)(\rho\mathbf{v})^{n+\alpha+\frac{1}{2}} = \mathbf{b}_{\rho\mathbf{v}}^{n+\alpha} - \mathfrak{G}_{\rho\mathbf{v}}^{n+\alpha,\,n+\frac{1}{2}},$$

(iv) solve the linear system for total energy ρE,

$$\mathbf{M}^{n+\alpha}(\rho E)^{n+\alpha+\frac{1}{2}} = \mathbf{b}_{\rho E}^{n+\alpha} - \mathfrak{G}_{\rho E}^{n+\alpha,\,n+\frac{1}{2}}.$$

Stage (i) requires the solution of a non-linear problem. The non-linearity is due to the implicit treatment of the chemical reaction source terms. If chemical reaction processes were absent, a linear problem should be solved instead. Let us introduce the map $\Phi(\cdot) : R^{N\times n_s} \to R^{N\times n_s}$ which is formally written as

$$\Phi(\boldsymbol{\rho}) := \left(\mathbf{D}_{\rho}^{n+\alpha}(\boldsymbol{\rho}) + \mathbf{M}^{n+\alpha}\otimes\mathbf{I}_{n_s}\right)^{-1}\mathbf{b}_{\rho}^{n+\alpha},$$

and whose domain of definition is the convex compact set

$$K = \left\{\boldsymbol{\rho}\in R^{N\times n_s} \mid \boldsymbol{\rho}_i \geq 0, \;\text{ for all } i,\; \sum_{i=1}^{N}|K_i|\,\|\boldsymbol{\rho}_i\|_1 = \left\|\mathbf{b}_{\rho}^{n+\alpha}\right\|_1\right\}.$$

The solution of (i) is a fixed point $\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}} = \Phi(\boldsymbol{\rho}^{n+\alpha+\frac{1}{2}})$. As already noted in [5], Φ is a continuous map from K into itself and the Brouwer fixed point theorem [48] implies that at least one non-negative fixed point must exist. Furthermore, if the time step Δt is sufficiently small, the map is also contractive and convergence of the iterative fixed point scheme is guaranteed. In practice, it has been noticed that iterations easily converge when the initial guess solution is obtained by first solving a problem without chemistry – i.e. with null source terms. The upgraded values of $\boldsymbol{\rho}_{\rho}^{n+\alpha+\frac{1}{2}}$ are then substituted in $\mathbf{D}_{e}^{n+\alpha+\frac{1}{2}}$. Stage (ii) is thus linearized and may be solved by a diagonally preconditioned Bi-CGSTAB method, see Reference [42]. Stage (iii) is also linearized by using the upgraded values of the species mass densities in $\mathfrak{G}_{\rho\mathbf{v}}^{n+\alpha,\, n+\frac{1}{2}}$. Similarly, stage (iv) can be linearized by substituting in $\mathfrak{G}_{\rho E}^{n+\alpha,\, n+\frac{1}{2}}$ the upgraded values for $\mathbf{D}_{\rho}^{n+\alpha+\frac{1}{2}}$ and also for $(\rho\mathbf{v})^{n+\alpha+\frac{1}{2}}$ previously estimated in stage (iii).

Stages (iii) and (iv) are solved by an iterative Bi-CGSTAB method, preconditioned by an incomplete LDU factorization. Since in stages (iii) and (iv) the system matrix is the same, the preconditioner is computed only once. As reported earlier in this section, the coefficient matrix of the resulting linear problem in (iii) and (iv) is an M-matrix. It is possible to show that the LDU factorization of an M-matrix does not require any pivoting (see [3]), thus resulting in a simplification of the incomplete factorization algorithm for preconditioning.

Remark. The cost of solving the non-linear system does not depend directly on the number of reactions included in the model, see e.g. [4]. However, it is affected by the stiffness of each reaction. That is, if one more reaction is included into the model, which is very stiff, the iteration matrix of the final non-linear system may become more ill conditioned and the number of iteration steps be increased.

In our implementation of the method, each iteration step basically requires the direct solution of a linear system of n_s equations in n_s unknowns per triangular cell, where n_s is the number of species. Then, the cost of such an iteration is roughly proportional to ${n_s}^3 \times N$, where N is the number of triangles, if the factorization is performed at each iterative step. This cost could be reduced in order to be proportional to ${n_s}^2 \times N$ if the system was first linearized. Nevertheless, this strategy would imply a great amount of storage for the memorization of the linearized factors, and the loss of the non-negativity of the computed mass fractions.

We finally mention that in literature some efficient techniques have been proposed to solve similar non-linear systems, see References [10, 33].

5. NUMERICAL TESTS

Seven different test cases from literature illustrate the capabilities of the present method in hypersonic flow calculations with non-equilibrium chemical reaction and vibrational energy effects. The implementation is based on the freely available software library `p2mesh`, see Reference [6].

The first three test cases are bidimensional transpositions of some rather "classical" 1-D shock-tube problems. They include the Lax problem [28] and the Sod problem [39] for inert air, and the chemically reactive shock tube problem for a mixture of gases reported in [1].

The second three test cases consist in different shock wave systems propagating in quiescent inert air on a two-dimensional compression ramp. Inert air is assumed to be a mixture of about 76.71% N_2, 23.29% O_2, and null fractions for N, O and NO. In all these test cases, complex unsteady shock wave patterns are characteristic of the solution and must be correctly captured by the scheme.

The final test case consists in a hypersonic blunt-body flow in pure nitrogen, where a steady bow shock forms.

In the first two 1-D shock tube calculations – the ones with Lax and Sod initial data – the calculations is performed for all the chemical species considered in the model, but neither reactions nor vibrational energy exchange occur. These calculations just show that the scheme is capable of solving "pure" gas dynamic problems. In all the other test cases, chemical reactions and vibrational energy exchanges are present.

The chemical reaction processes herein considered utilize the Dunn-Kang or the Park air model, see Reference [36, 47] for a detailed presentation. Both models – in the version with no free electrons and ions and associated reactions – describe air as a mixture of five species, N_2, O_2, NO, N and O, which are assumed to be thermally perfect gases. Both models consider fifteen elementary dissociation–recombination reactions and two exchange reactions

$$\begin{aligned} N_2 + M &\rightleftharpoons 2N + M\,, \\ O_2 + M &\rightleftharpoons 2O + M\,, \\ NO + M &\rightleftharpoons N + O + M\,, \\ N_2 + O &\rightleftharpoons NO + N\,, \\ NO + O &\rightleftharpoons O_2 + N. \end{aligned}$$

M represents a collision partner or *catalytic molecule* and may be any one of the previous species. The equilibrium energy $\mathcal{E}_r^{eq}$ and the vibrational energy $\mathcal{E}_r^v$ respectively depend on the translational temperature T and the vibrational temperature T_r^v by

$$\mathcal{E}_r^{eq} = \frac{\rho_r \theta_r \mathcal{R}}{\mathcal{M}_r} \left(\exp\left(\frac{\theta_r}{T}\right) - 1 \right)^{-1}, \qquad \mathcal{E}_r^{v} = \frac{\rho_r \theta_r \mathcal{R}}{\mathcal{M}_r} \left(\exp\left(\frac{\theta_r}{T_r^v}\right) - 1 \right)^{-1},$$

with $r = N_2,\, O_2,\, NO$. Their characteristic vibrational temperatures take the values $\theta_{N_2} = 3395\,\mathcal{K}$, $\theta_{O_2} = 2239\,\mathcal{K}$ and $\theta_{NO} = 2817\,\mathcal{K}$, which were obtained via spectroscopic measurements and are reported for example in [7].

The Landau-Teller relaxation times τ_r in equation (5) are given by

$$\tau_r = \frac{\sum_{s=1}^{n_s} \frac{\rho_s}{\mathcal{M}_s} \exp\left(A_r \left(T^{-\frac{1}{3}} - 0.015\, \mu_{rs}^{\frac{1}{4}} \right) - 18.42 \right)}{\frac{p}{101325} \sum_{s=1}^{n_s} \frac{\rho_s}{\mathcal{M}_s}}, \qquad \mu_{rs} = 1000 \frac{\mathcal{M}_s \mathcal{M}_s}{\mathcal{M}_s + \mathcal{M}_s}.$$

The coefficients A_r take the values $A_{N_2} = 220$, $A_{O_2} = 129$, $A_{NO} = 168$, see e.g. [7]. These semi-empirical relations are known to be valid over a temperature range from $300\,\mathcal{K}$ through $9000\,\mathcal{K}$, see [7, 34].

5.1. Shock-tube problems in 2-D

The computational domain for the shock-tube problems is the bidimensional strip $[-0.5,\, 0.5] \times [-0.025,\, 0.025]$. A diaphragm, located along the vertical line $x = 0$, separates the initial left state, denoted by *(L)*, and the right one, denoted by *(R)*. In Lax and Sod problems, the shock-tube is filled by inert air. In this case, the usual value of the specific heat ratio holds, i.e. $\gamma = 1 + \beta = 1.4$.

In the shock-tube problem with chemistry, only the right side of the shock-tube domain contains quiescent inert air. At the left of the diaphragm, the temperature is high enough to induce chemical and vibrational effects. The air composition is thus assumed at equilibrium at the given temperature,which corresponds to a mixture composition of $44.2\%\ N_2$, $7.36 \times 10^{-2}\%\ O_2$, $2.3\%\ NO$, $31.4\%\ N$, and $22\%\ O$.

Lax problem:

$(L)\ \rho = 0.445\, Kg/m^3,\ u = 0.698\, m/s,\ p = 3.528\, Pa.$
$(R)\ \rho = 0.5\, Kg/m^3,\ u = 0\, m/s,\ p = 0.571\, Pa.$

Sod problem:

$(L)\ \rho = 1\, Kg/m^3,\ u = 0\, m/s,\ p = 1\, Pa.$
$(R)\ \rho = 0.125\, Kg/m^3,\ u = 0\, m/s,\ p = 0.1\, Pa.$

Shock-tube problem with chemistry:

$(L)\ \rho = 2.532\, Kg/m^3,\ u = 0\, m/s,\ T = 9000\ \mathcal{K}.$
$(R)\ \rho = 1.156\, Kg/m^3,\ u = 0\, m/s,\ T = 300\ \mathcal{K}.$

At time $t = 0$ the diaphragm is instantly removed and well-known patterns of interacting rarefaction waves, contact and shocks discontinuities begin to form, [1, 23]. All the

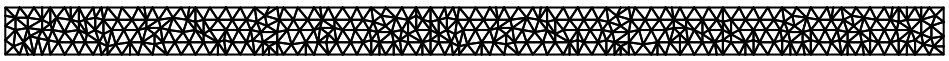

FIG. 2. Computational Mesh.

results reported here were calculated on an unstructured irregular mesh of about 700–1500 triangles, see Figure 2. Figure 3 shows the density solution of the Lax problem computed at $t = 0.15s$. The calculation was performed by using a fixed time step $\Delta t = 0.15s/100 = 1.5ms$. Figure 4 shows the density solution of the Sod problem computed at $t = 0.24s$. The calculation was performed by using a fixed time step $\Delta t = 0.24s/100 = 2.4ms$. Figures 5–6 show the solution of the shock-tube problem with chemistry at $t = 0.16s$, computed by using a fixed time step $\Delta t = 0.16ms/200 = 0.8\mu s$. In order to compare the results obtained in this test case with the ones reported by [1], the Park air model was included in the solver. It is informative to say that in the three cases the 2-D solver was tested on several kind of meshes, both regular and irregular ones. Irregular meshes were generated by `triangle`, with the requirement that a maximum angle constraint be satisfied, see the documentation reported within the software distribution package and also Reference [8]. Regular meshes were built by simply partitioning regular square-shaped cells in four sub-triangles. No orientation effects were noticed in these computed solutions. However, we

noticed that triangles may produce a slight distortion in the front of an advancing shock wave. This effect seems to be influenced by chemical and vibrational relaxation processes, being more apparent in the reactive calculations. As far as shock resolution accuracy is concerned, we remark that the discontinuity is generally well resolved in these problems with an average "width" of three-five triangles. This result is also in accord with the behavior observed in the two-dimensional calculations reported in the following section.

5.2. Two dimensional test cases

Three numerical solutions of shock wave propagation problems are presented in order to illustrate the performance of the method in predicting 2-D unsteady hypersonic flows. The examples consists in calculating the single, complex and double Mach reflection of a planar shock wave incident on a compression ramp. An unsteady complex shock pattern forms on the ramp and evolves during the reflection process. In these calculations we used the limiter described in steps (i)–(iv), section 3.7.

The final test case concerns with the numerical approximation of a blunt-body hypersonic flow around a two-dimensional circular cylinder and illustrates how the method performs in a steady state calculation. The longitudinal axis of the cylinder is orthogonal to the free-stream flow direction and a steady bow shock forms above the cylinder. Across the steady shock wave the flow temperature raises and thermodynamic equilibrium is reached by a strong non-equilibrium dissociation process. In this calculation we used the limiter described in steps (i)–(vi), section 3.7. The steady state solution is achieved by relaxing the initial free-stream solution by using (16), which is a first-order-in-time marching scheme. No acceleration techniques – such as local time stepping, residual smoothing, . . . – have been introduced.

The cost of solving the non-linear system for the species mass densities was experimentally measured and varies throughout 60% to 70% of the total cost of the computation, depending on the test case, see Table 2.

A detailed presentation of all of the test cases can be found in References [11, 15, 21, 24], where both results from laboratory experiences and numerical simulations are reported. A comparison to the mentioned literature results shows that our method generally performs well.

TABLE 1

Initial values for compression ramp problems

	Single	Complex	Double
$\rho^{in}[Kg/m^3]$	1.05702	0.70678	0.382407
$u^{in}[m/s]$	447.077	3205.8	2646.49
$T^{in}[°K]$	510.147	4080.08	3373.17
$\rho^{out}[Kg/m^3]$	0.387	0.0777	0.0476
$T^{out}[°K]$	299.2	299	299.2

5.2.1. Single Mach reflection in air

Table 1 summarizes the inlet and initial quiescent states of the problem. The moving shock wave Mach number is $M_s = 2.03$ and the compression angle is 27^0. In this reflection process, the temperature never attains values capable of inducing non-equilibrium reactions. Hence, the test case mainly allows to check the shock-capturing capability of the method

TABLE 2

Relative chemical computational costs

	Complex	Double	Blunt
cost percentage	68%	67%	59%

as a compressible gas dynamic solver. Figures 7–8 depict the density contours and the wall density distribution at time $t = 100\mu s$. The solution is computed by using a fixed time step $\Delta t = 100\mu s/800 \approx 0.125\mu s$ on a mesh with 52742 triangles. The *CFL* ranges throughout 0.79 to 0.99 .

5.2.2. Complex Mach reflection in air

Table 1 summarizes the inlet and initial quiescent states of the problem. The moving shock wave Mach number is $M_s = 10.37$ and the compression angle is 10^0. Figures 9–10 depict the density contours and the wall density distributions at time $t = 20\mu s$. The approximate solution is computed by using a fixed time step $\Delta t = 20\mu s/800 = 0.025\mu s$ on a mesh with 34833 triangles The *CFL* is stable at 0.87.

5.2.3. Double Mach reflection in air

Table 1 summarizes the inlet and initial quiescent states of the problem. The moving shock wave Mach number is $M_s = 8.7$ and the compression angle is 27^0. Figures 11–12 depict the density contours and the wall density distributions at time $t = 24\mu s$. The approximate solution is computed by using a fixed time step $\Delta t = 24\mu s/1000 = 0.024\mu s$ on a mesh with 64435 triangles. The *CFL* ranges throughout 0.67 to 0.82.

5.2.4. Blunt-body flow in nitrogen

Table 3 summarizes the free-stream conditions of the problem. The free-stream gas is 7% dissociated nitrogen, the frozen-flow Mach number is about 6.9. The radius of the cylinder is 2.54 cm.

Figures 13–14 depict the density contours over all the computational domain and three zoomed views of the region across the bow shock.

This steady state solution is reached in 6000 iterations with a time step $\Delta t = 0.05\mu s$ corresponding to *CFL*$= 1.3$ on a mesh with 5277 triangles.

TABLE 3

Free-stream values for the blunt-body problem

$p_\infty[Pa]$	$\rho_\infty[Kg/m^3]$	$u_\infty[m/s]$	$T_\infty[^\circ K]$
2.445	$5.5\,10^{-3}$	$5.5\,10^3$	1400

6. CONCLUSIONS

We have presented here a finite volume scheme for hypersonic flows on 2-D unstructured triangle-based meshes. The scheme makes usage of a numerical flux obtained by blending a first-order upwind flux with a second-order central one. A new 2-D flux limiter

is presented in order to prevent the formation of numerical oscillations near solution discontinuities. Then, a two-stage time-stepping scheme formally guarantees a global second order accuracy. The way we propose to perform implicitation in time of the stiff chemical source terms and of the numerical flux produces a non-linear fully-coupled algorithm with a very peculiar algebraic block structure. We suggest a resolution strategy based on a block Gauss-Seidel-like decomposition into four smaller size problems which may be solved sequentially. A linear algebra argument based on M-matrix theory ensures the positivity of species mass densities and vibrational energies under a not too restrictive *CFL*-like constraint. Several numerical examples illustrate the capabilities of our methodologies in accurately solving strong shock dominated hypersonic flows. Finally, it is worth mentioning that the method is quite general and can be easily extended in three dimensions, where reduction of computational costs are essential.

REFERENCES

1. R. Abgrall, L. Fezoui, and J. Talandier, *An extension of Osher's Riemann solver for chemical and vibrational non-equilibrium gas flow*, International Journal of Numerical Methods in Fluids **14** (1992), 935–960.
2. J. Argyris, I.St. Doltsinis, H. Friz., and J. Urban, *An exploration of chemically reacting viscous hypersonic flow*, Computer Methods in Applied Mechanics and Engineering **89** (1991), 85–128.
3. A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, SIAM, Philadelphia, 1994, (republished in "Classics in Applied Mathematics").
4. E. Bertolazzi, *Positive and conservative schemes for mass action kinetics*, Computers & Mathematics with Applications **32** (1996), 29–43.
5. ———, *A finite volume scheme for two dimensional chemically reactive hypersonic flows*, International Journal of Numerical Methods for Heat & Fluid Flow **8** (1998), 888–933.
6. E. Bertolazzi and G. Manzini, *Template classes for PDE solvers on 2-D unstructured meshes*, Tech. Report 1124, IAN-CNR, 1998.
7. G. Candler, *The computation of weakly ionized hypersonic flows in thermo-chemical nonequilibrium*, PhD Thesis.– Dept. of Aeronautics and Astronautics Stanford University, 1988.
8. P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland Publishing Company, Amsterdam, Holland, 1980.
9. L. Collatz, *The numerical treatment of differential equations*, Springer, Berlin, 1960.
10. Y. D'Angelo and B. Larrouturou, *Comparison and analysis of some numerical schemes for stiff complex chemistry problems*, Mathematical Modelling and Numerical Analysis **29** (1995), 259–301.
11. R. L. Deschambault and I. I. Glass, *An update on non-stationary oblique shock-wave reflections: actual isopicnics and numerical experiments*, Journal of Fluid Mechanics **131** (1983), 27–57.
12. J. A. Desideri, R. Glowinsky, and J. Périaux (eds), *Hypersonic flows for reentry problems.*, Springer-Verlag, Berlin, 1991, Proceedings of a workshop held in Antibe (France), 22-25 January 1990.
13. R. P. Fedkiw, B. Merriman, and S. Osher, *High accuracy numerical methods for thermally perfect gas flows with chemistry*, Journal of Computational Physics **132** (1997), 175–190.
14. P. Glaister, *An approximate linearised Riemann solver for the Euler equations for real gases*, Journal of Computational Physics **77** (1988), 361–383.
15. H. M. Glaz, P. Colella, I. I. Glass, and R. L. Deschambault, *A numerical study of oblique shock-wave reflections with experimental comparisons*, Proc. of the Royal Society of London **398** (1985), 117–140.
16. E. Godlewski and P.-A. Raviart, *Numerical approximation of hyperbolic systems of conservation laws*, Springer, Berlin/New York, 1996.
17. D. F. Griffiths, A. M. Stuart, and H. C. Yee, *Numerical wave propagation in an advection equation with a non-linear source term*, SIAM, Journal on Numerical Analysis **29** (1992), 1244–1260.
18. B. Grossman and P. Cinnella, *The computation of non-equilibrium, chemically-reacting flows*, Computers & Structures **30** (1988), no. 1/2, 79–93.
19. ———, *Flux split algorithms for flows with non-equilibrium chemistry and vibrational relaxation*, Journal of Computational Physics **88** (1990), 131–168.

20. B. Grossman and R.W. Walters, *An analysis of flux-split algorithms for Euler's equations with real gases*, 8^{th} Computational fluid dynamic conference, no. AIAA-87-1117, AIAA, 1987, pp. 177–186.
21. C. P. T. Groth and J. J. Gottlieb, *TVD finite-difference methods for computing high-speed thermal and chemical non-equilibrium flows with strong shocks*, International Journal of Numerical Methods for Heat & Fluid Flow **3** (1993), 483–516.
22. P. R. Halmos, *Finite-dimensional vector spaces*, Van Nostrand, Princeton, N.J., 1958.
23. C. Hirsch, *Numerical computation of internal and external flows*, J. Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex PO19 1UD, England, 1990.
24. H. G. Hornung, *Non-equilibrium dissociating nitrogen flow over spheres and circular cylinders*, Journal of Fluid Mechanics **53** (1972), 149–176.
25. M. Y. Hussaini, B. van Leer, and J. Van Rosendale (eds), *Upwind and high-resolution schemes*, Springer, Berlin/New York, 1997.
26. Jr. J. Anderson, *Hypersonic and high temperature gas dynamics*, Mc Graw-Hill, 1990.
27. B. Larrouturou, *How to preserve mass fraction positivity when computing compressible multi-component flows*, Journal of Computational Physics **95** (1991), 59–84.
28. P. D. Lax, *Weak solutions of nonlinear hyperbolic equations and their numerical computations*, CPAM **7** (1954), 159–193.
29. R. J. LeVeque and H. C. Yee, *A study of the numerical methods for hyperbolic conservation laws with stiff source terms*, Journal of Computational Physics **86** (1990), 187–210.
30. D. Lindstrom, *Effects of numerical dissipation on the speed of combustion waves*, Tech. Report 174/1996, Uppsala University, 1996.
31. M. S. Liou, B. Van Leer, and J. S. Shuen, *Splitting of inviscid fluxes for real gases*, Journal of Computational Physics **87** (1990), 1–24.
32. Y. Liu and M. Vinokur, *Nonequilibrium flow computations. I. An analysis of numerical formulations of conservation laws*, Journal of Computational Physics **83** (1989), 373–397.
33. K. Meintjes and A. P. Morgan, *Performance of algorithms for calculating the equilibrium composition of a mixture of gases*, Journal of Computational Physics **60** (1985), 219–234.
34. R. C. Millikan and D. R. White, *Systematics of vibrational relaxation*, The Journal of Chemical Physics **39** (1963), 3209–3212.
35. E. S. Oran and J. P. Boris, *Numerical simulation of reactive flow*, Elsevier, Amsterdam/New York, 1987.
36. C. Park, *Problems of rate chemistry in the flight regimes of aeroassisted orbital transfer vehicles*, 19 Thermophysics Conference, no. AIAA-84-1730, AIAA, June 1984, pp. 1–11.
37. J. J. Quirk, *A contribution to the great Riemann solver debate*, International Journal for Numerical Methods in Fluids **18** (1994), 555–574.
38. J. S. Shuen, M. S. Liou, and B. van Leer, *Inviscid flux-splitting algorithms for real gases with non-equilibrium chemistry*, Journal of Computational Physics **90** (1990), 371–395.
39. G. A. Sod, *A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws*, Journal of Computational Physics **27** (1978), 1–31.
40. J. L. Steger and R. F. Warming, *Flux-vector splitting of the inviscid gas dynamic equations with application to finite-difference methods*, Journal of Computational Physics **40** (1981), 263–293.
41. P. K. Sweby, *High resolution schemes using flux limiters for hyperbolic conservation laws*, SIAM, Journal on Numerical Analysis **21** (1984), 995–1011.
42. H. A. van der Vorst, *Bi-CGSTAB: a fast and smoothly converging variant for Bi-CG for the solution of nonsymmetric linear systems*, SIAM, J. Sci. Stat. Computing **13** (1992), 631–644.
43. M. Vinokur and J.-L. Montagné, *Generalized flux-vector splitting and Roe average for an equilibrium real gas*, Journal of Computational Physics **89** (1990), 276–300.
44. H. C. Yee, *Construction of explicit and implicit symmetric TVD schemes and their applications*, Journal of Computational Physics **68** (1987), 151–179.
45. H. C. Yee, G. H. Klopfer, and J.-L. Montagné, *High-resolution shock capturing schemes for inviscid and viscous hypersonic flows*, Journal of Computational Physics **88** (1990), 31–61.
46. H. C. Yee, R. F. Warming, and A. Harten, *Implicit Total Variation Diminishing (TVD) schemes for steady-state calculations*, Journal of Computational Physics **57** (1985), 327–360.
47. S. T. Yu, B. J. McBridge, K. C. Hsiesh, and J. S. Shuen, *Numerical simulation of hypersonic inlet flows with equilibrium or finite rate chemistry*, 26th Aerospace Sciences Meeting, no. AIAA-88-0273, AIAA, January 1988, pp. 1–13.
48. E. Zeidler, *Nonlinear functional analysis and its applications*, Springer-Verlag, New York, 1986.

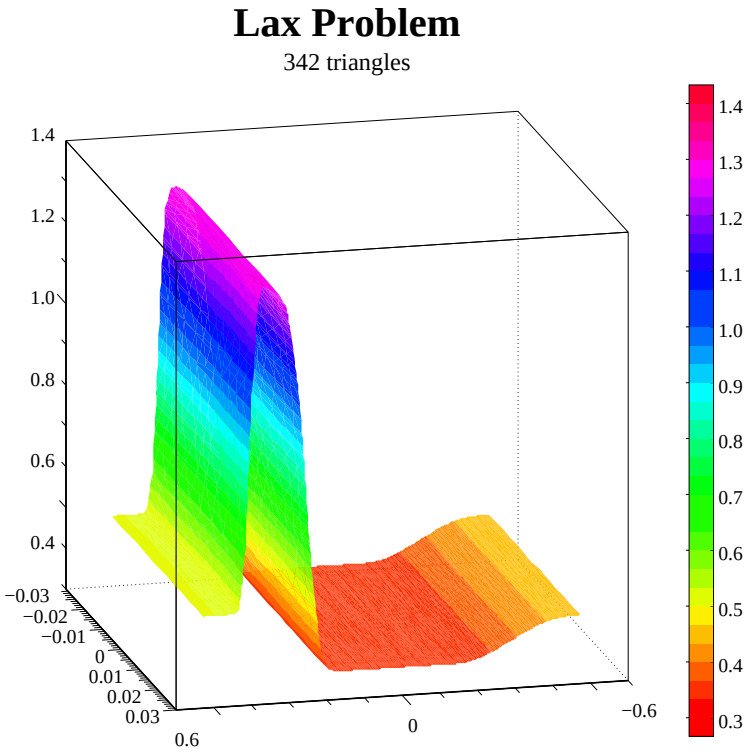

FIG. 3. Solution of Lax problem at time $t = 0.15s$; mass density distribution.

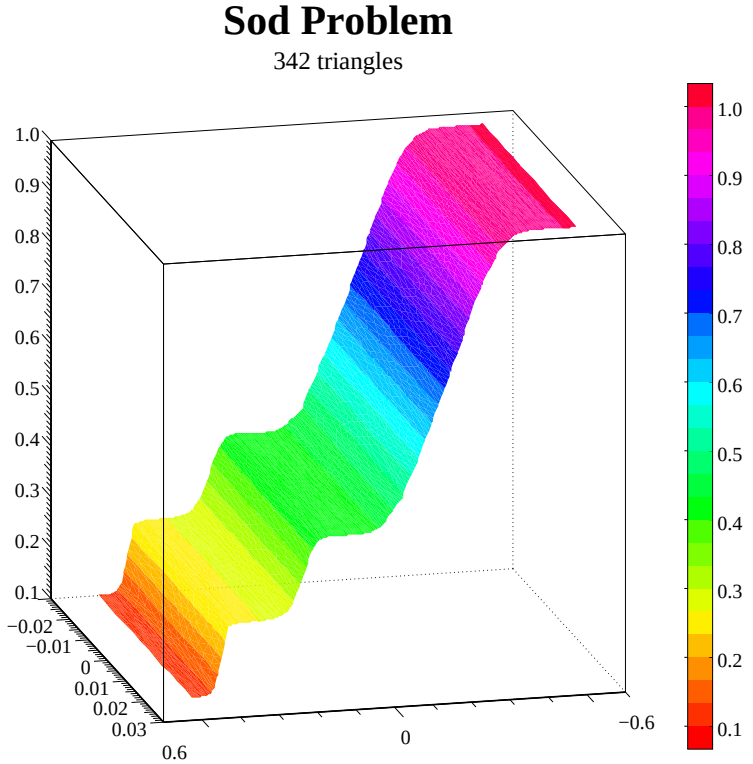

FIG. 4. Solution of Sod problem at time $t = 0.24s$; mass density distribution.

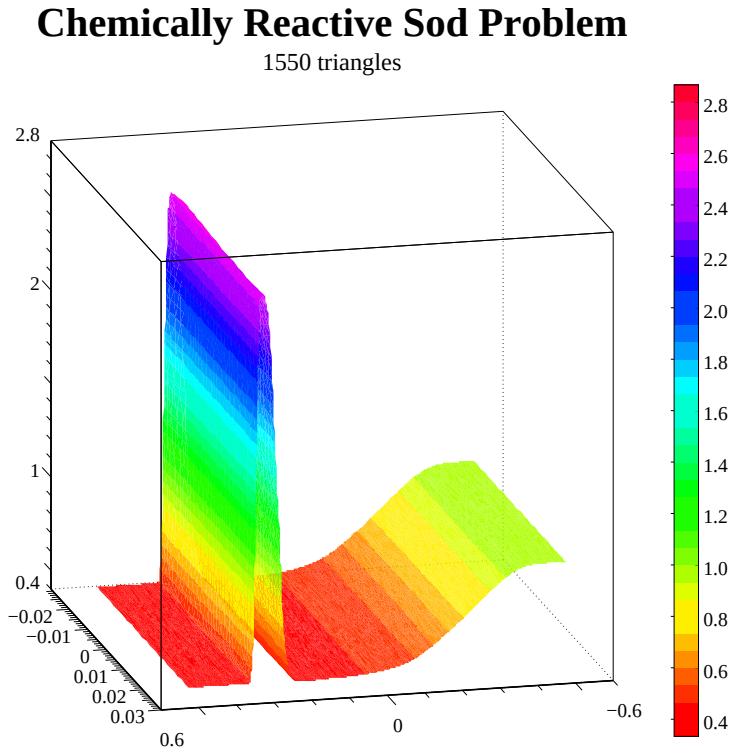

FIG. 5. Solution of Shock Tube problem at time $t = 160\,\mu s$; mass density distributions.

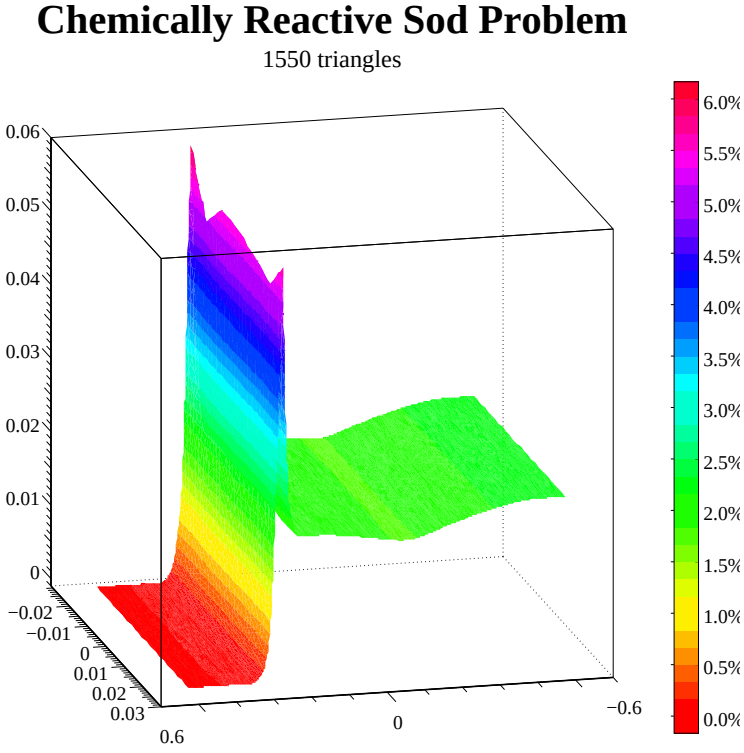

FIG. 6. Solution of Shock Tube problem at time $t = 160\,\mu s$; NO mass fraction distribution.

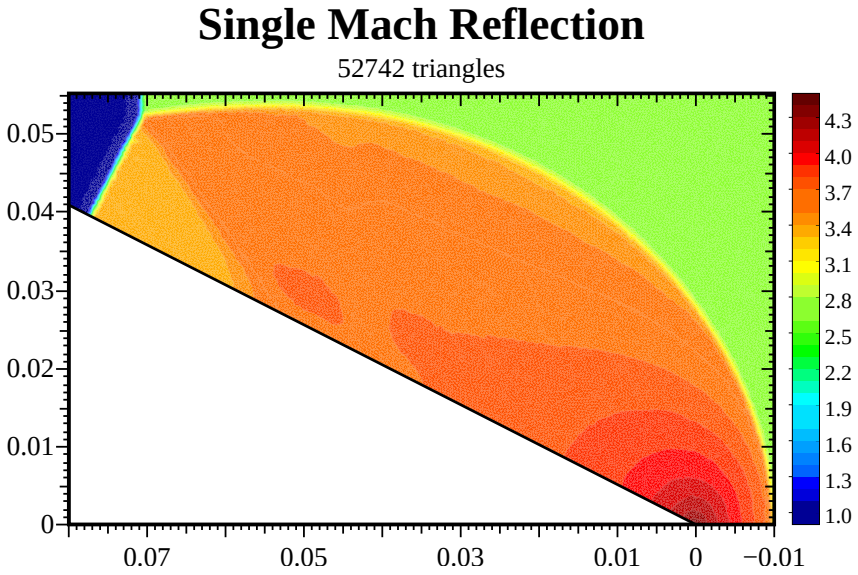

FIG. 7. Single mach reflection in air; mass density ratio (ρ/ρ_0) distribution, where ρ_0 is the density of quiescent air. Axes units are in meters.

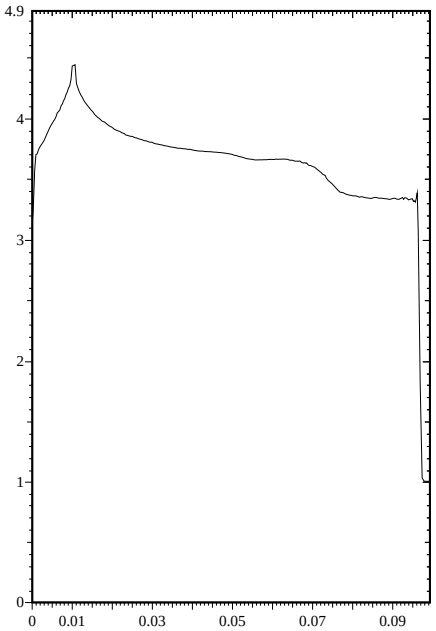

FIG. 8. Single mach reflection in air; mass density ratio wall distribution. Axes units are in meters.

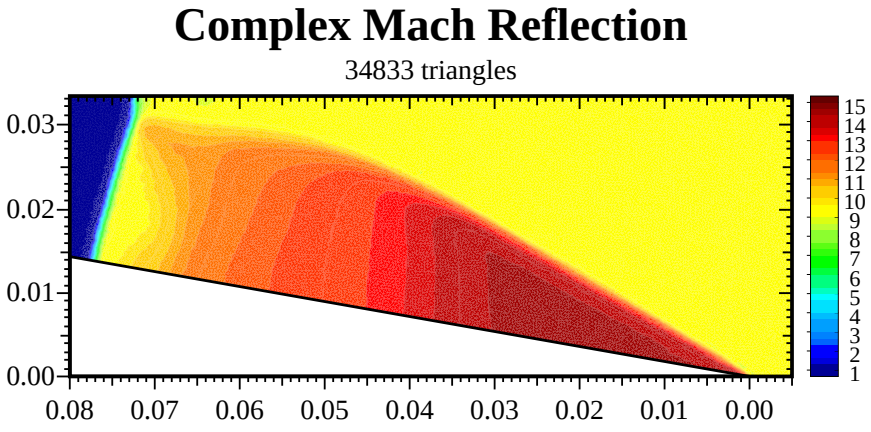

FIG. 9. Complex mach reflection in air; mass density ratio (ρ/ρ_0) distribution, where ρ_0 is the density of quiescent air. Axes units are in meters.

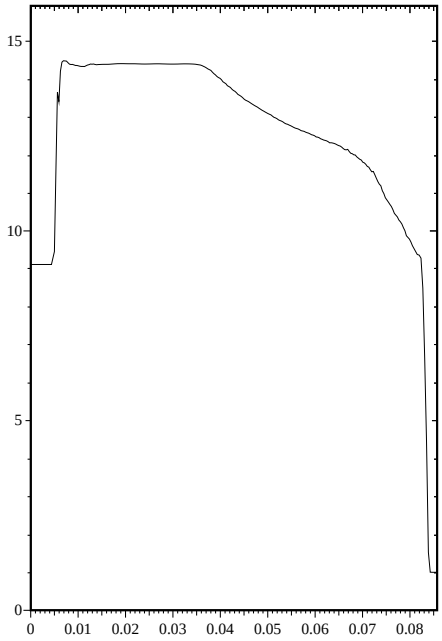

FIG. 10. Complex mach reflection in air; mass density ratio wall distribution. Axes units are in meters.

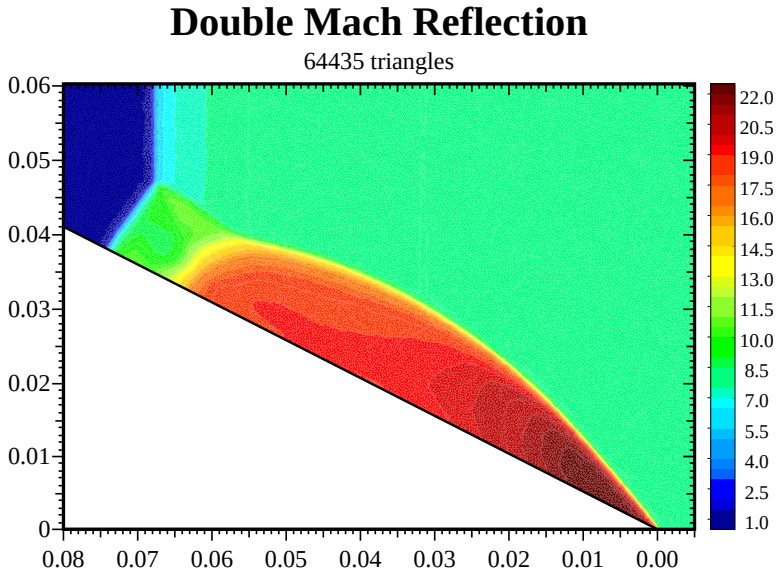

FIG. 11. Double mass reflection in air; mass density ratio (ρ/ρ_0) distribution, where ρ_0 is the density of quiescent air. Axes units are in meters.

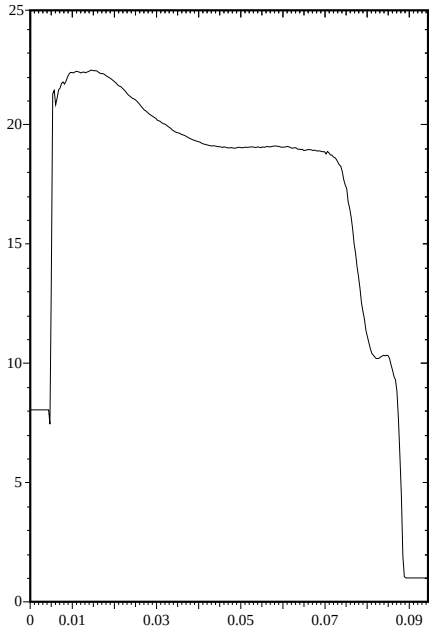

FIG. 12. Double mass reflection in air; mass density ratio wall distribution. Axes units are in meters.

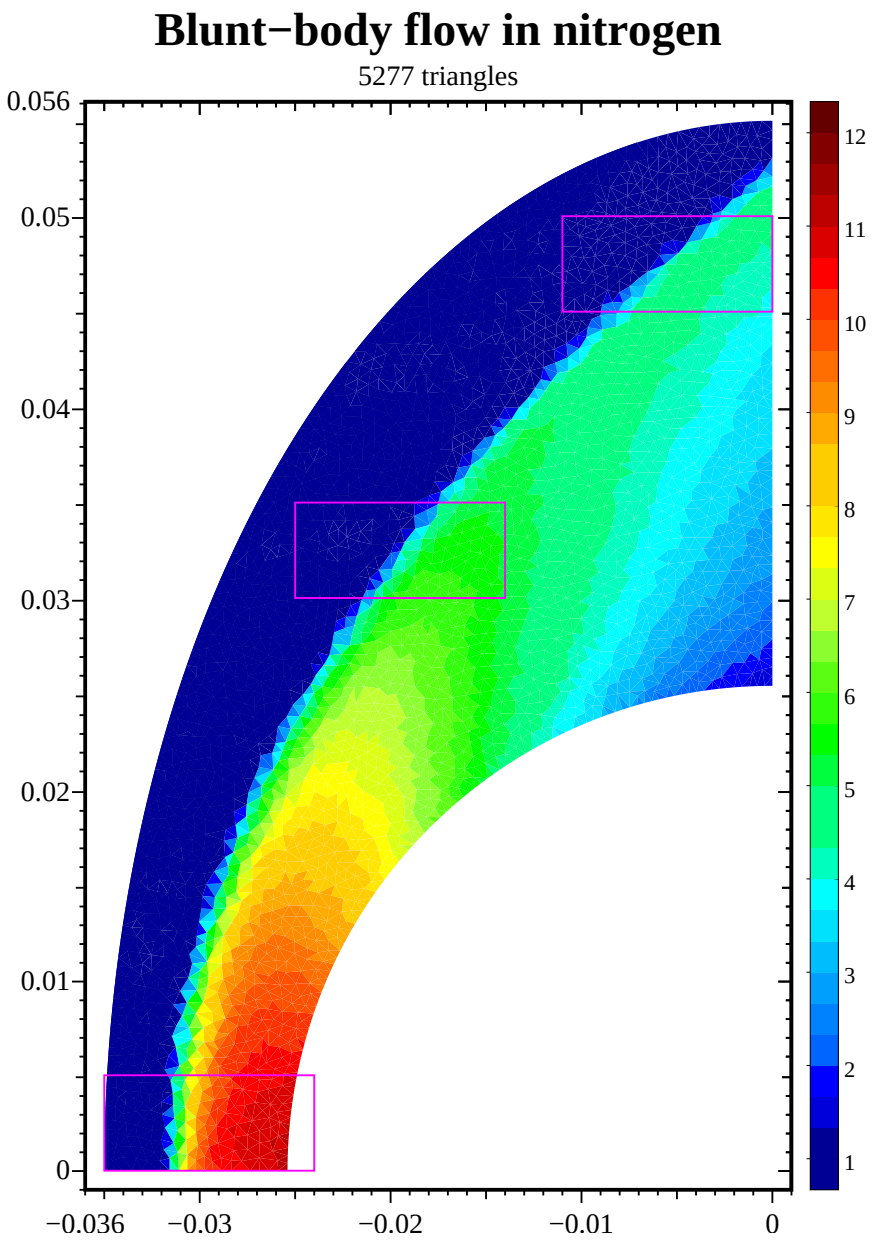

FIG. 13. Blunt-body flow in nitrogen; mass density ratio (ρ/ρ_∞) distribution, where ρ_∞ is the density free-stream gas. Axes units are in meters.

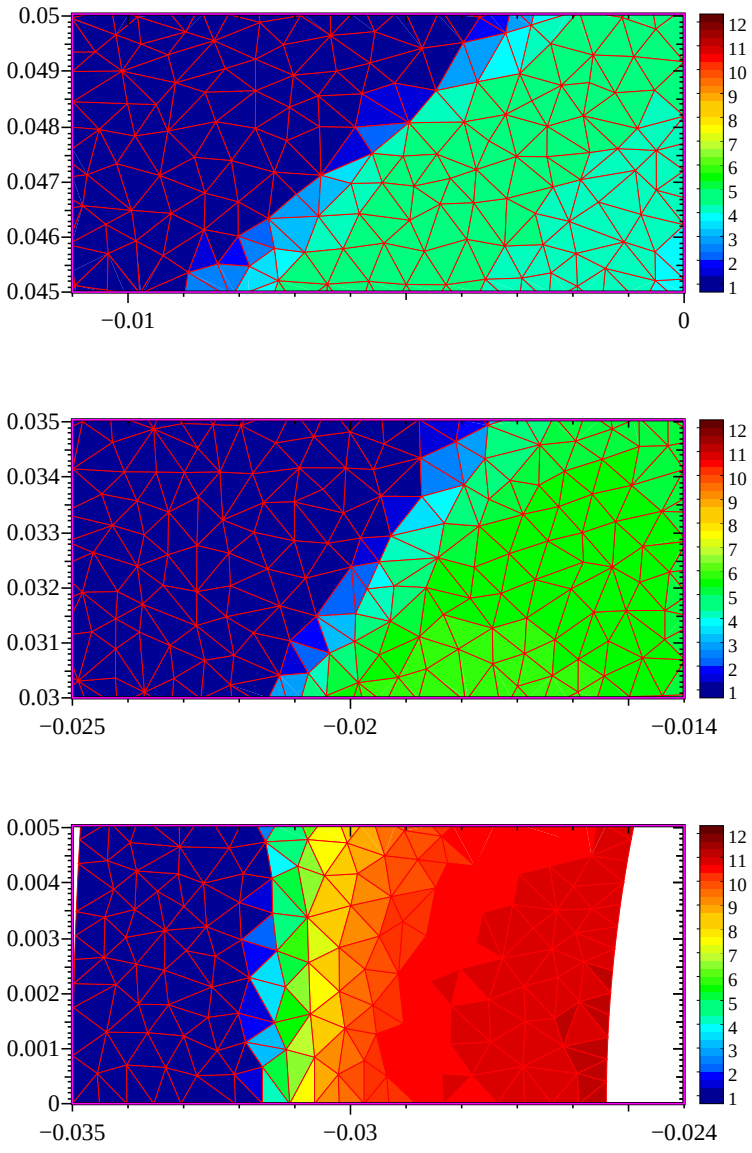

FIG. 14. Blunt-body flow in nitrogen; zoomed views from Figure 13

TABLE 4
The reaction matrix $\mathbf{C}(\boldsymbol{\rho}, \mathbf{T})$ for the five species Dunn and Kang model

$-\Gamma_{N_2} - r_O \mathrm{K}_{f\,N}$	0	$c_1 r_N \mathrm{K}_{b\,N}$	$2\Omega_{N_2} r_N + c_1 r_{NO} \mathrm{K}_{b\,N}$	0
0	$-\Gamma_{O_2} - r_N \mathrm{K}_{b\,O}$	$b_1 r_O \mathrm{K}_{f\,O}$	0	$2\Omega_{O_2} r_O + b_1 r_{NO} \mathrm{K}_{f\,O}$
$a_1 r_O \mathrm{K}_{f\,N}$	$d_1 r_N \mathrm{K}_{b\,O}$	$-\Gamma_{NO} - r_O \mathrm{K}_{f\,O} - r_N \mathrm{K}_{b\,N}$	$\Omega_{NO} r_O + d_1 r_{O_2} \mathrm{K}_{b\,O}$	$\Omega_{NO} r_N + a_1 r_{N_2} \mathrm{K}_{f\,N}$
$\Gamma_{N_2} + a_2 r_O \mathrm{K}_{f\,N}$	0	$\alpha\Gamma_{NO} + b_2 r_O \mathrm{K}_{f\,O}$	$-2\Omega_{N_2} r_N - \Omega_{NO} r_O$ $-r_{NO} \mathrm{K}_{b\,N} - r_{O_2} \mathrm{K}_{b\,O}$	$a_2 r_{N_2} \mathrm{K}_{f\,N} + b_2 r_{NO} \mathrm{K}_{f\,O}$
0	$\Gamma_{O_2} + d_2 r_N \mathrm{K}_{b\,O}$	$(1-\alpha)\Gamma_{NO} + c_2 r_N \mathrm{K}_{b\,N}$	$c_2 r_{NO} \mathrm{K}_{b\,N} + d_2 r_{O_2} \mathrm{K}_{b\,O}$	$-2\Omega_{O_2} r_O - \Omega_{NO} r_N$ $-r_{N_2} \mathrm{K}_{f\,N} - r_{NO} \mathrm{K}_{f\,O}$

APPENDIX

The explicit form of the reaction matrix $\mathbf{C}(\boldsymbol{\rho}, \mathbf{T})$ is showed in Table 4 where

$$
\begin{aligned}
r_m &= \frac{\rho_m}{\mathcal{M}_m}, \qquad \alpha = \frac{\mathcal{M}_N}{\mathcal{M}_N + \mathcal{M}_O}, \\
a_1 &= \frac{\mathcal{M}_N + \mathcal{M}_O}{2\mathcal{M}_N + \mathcal{M}_O}, \qquad a_2 = 1 - a_1, \\
b_1 &= \frac{2\mathcal{M}_O}{\mathcal{M}_N + 2\mathcal{M}_O}, \qquad b_2 = 1 - b_1, \\
c_1 &= \frac{2\mathcal{M}_N}{2\mathcal{M}_N + \mathcal{M}_O}, \qquad c_2 = 1 - c_1, \\
d_1 &= \frac{\mathcal{M}_N + \mathcal{M}_O}{\mathcal{M}_N + 2\mathcal{M}_O}, \qquad d_2 = 1 - d_1,
\end{aligned}
$$

and

$$
\Gamma_j = \sum_m \mathrm{K}_{f\,jm} r_m, \qquad \Omega_j = \sum_m \mathrm{K}_{b\,jm} r_m,
$$

where $j = N_2, O_2, NO$ and $m = N_2, O_2, NO, N, O$. The reaction rates K_f and K_b are assumed to be functions of T, T_1^v, T_2^v, T_3^v and they are described by the modified forms of the Arrhenius equation. These reaction rates K_f, K_b take the functional form

$$
\kappa T_s^\theta \exp\left(-\frac{T_a}{T_s}\right),
$$

where T_s is a function of the translational and vibrational temperatures and κ, θ, T_a are coefficients that depend on the specific reaction. The values of the coefficients κ, θ, T_a and the precise form of the function T_s for each reaction can be found in [7, 36, 47].

A.0.1. Operator $\mathfrak{G}$ in (21).

Let us set

$$C_{ij}^{n+\alpha} = \left(\alpha+\frac{1}{2}\right)\Delta t\ell_{ij}\left\{\frac{\theta_{ij}c_i^2}{2+2\beta_i}+(1-\theta_{ij})\mathbb{b}^+(\mathbf{U}_i,\mathbf{n}_{ij})\right\}^{n+\alpha},$$

$$D_{ij}^{n+\alpha} = \left(\alpha+\frac{1}{2}\right)\Delta t\ell_{ij}(1-\theta_{ij}^{n+\alpha})\mathbb{c}^+(\mathbf{U}_i^{n+\alpha},\mathbf{n}_{ij}).$$

Then we have

$$\mathfrak{G}_{\rho\mathbf{v},i}^{n+\alpha,\,n+\frac{1}{2}} = \sum_{j\in\sigma(i)}\left(C_{ij}^{n+\alpha}\rho_i^{n+\frac{1}{2}}+C_{ji}^{n+\alpha}\rho_j^{n+\frac{1}{2}}\right)\mathbf{n}_{ij},$$

$$\mathfrak{G}_{\rho E,i}^{n+\alpha,\,n+\frac{1}{2}} = \sum_{j\in\sigma(i)}\left(C_{ij}^{n+\alpha}(\rho_i\mathbf{v}_i)^{n+\frac{1}{2}}\cdot\mathbf{n}_{ij}-C_{ji}^{n+\alpha}(\rho_j\mathbf{v}_j)^{n+\frac{1}{2}}\cdot\mathbf{n}_{ji}\right) + \sum_{j\in\sigma(i)}\left(D_{ij}^{n+\alpha}\rho_i^{n+\frac{1}{2}}-D_{ji}^{n+\alpha}\rho_j^{n+\frac{1}{2}}\right).$$

A.0.2. R.H.S. in (22).

Let us first introduce

$$A_{ij}^{n+\alpha}=\frac{\Delta t}{2}\begin{cases}0 & \text{if } ij \text{ is not an edge,}\\ -\ell_{ij}\hat{a}_{ji}^{n+\alpha} & \text{if } i\neq j,\\ \displaystyle\sum_{k\in\sigma(i)}\ell_{ik}\hat{a}_{ik}^{n+\alpha} & \text{if } i=j,\end{cases}$$

and

$$V_{ij}^{n+\alpha}=\Delta t\left(\alpha+\frac{1}{2}\right)\begin{cases}0 & \text{if } ij \text{ is not an edge,}\\ -\ell_{ij}\hat{v}_{ji}^{n+\alpha} & \text{if } i\neq j,\\ \displaystyle\sum_{k\in\sigma(i)}\ell_{ik}\hat{v}_{ik}^{n+\alpha} & \text{if } i=j.\end{cases}$$

The r.h.s. terms for densities and vibrational energies take the form

$$\mathbf{b}_{\rho,i}^{n+\alpha} = |K_i|\,\boldsymbol{\rho}_i^n-\sum_j\left\{2\alpha A_{ij}^{n+\alpha}\boldsymbol{\rho}_j^n+V_{ij}^{n+\alpha}\boldsymbol{\rho}_j^{n+\alpha}\right\} \tag{A.1}$$

$$-\left(\alpha+\frac{1}{2}\right)\Delta t\sum_{j'\in\sigma'(i)}\ell_{ij'}\left[\theta_{ij'}^{inlet}(\mathbf{v}_{j'}^{(bc)}\cdot\mathbf{n}_{ij'})\boldsymbol{\rho}_{j'}^{(bc)}+\theta_{ij'}^{free}(\mathbf{v}_i\cdot\mathbf{n}_{ij'})\boldsymbol{\rho}_i\right]^{n+\alpha},$$

$$\mathbf{b}_{\mathcal{E},i}^{n+\alpha} = |K_i|\left(\mathcal{E}_i^{vn}+\mathbf{s}(\boldsymbol{\rho}_i^{n+\alpha+\frac{1}{2}},\mathbf{U}_i^{n+\alpha})\otimes\mathcal{E}_j^{eqn}\right)-\sum_j\left\{2\alpha A_{ij}^{n+\alpha}\mathcal{E}_j^{v,\,n}+V_{ij}^{n+\alpha}\mathcal{E}_j^{v,\,n+\alpha}\right\}$$

$$-\left(\alpha+\frac{1}{2}\right)\Delta t\sum_{j'\in\sigma'(i)}\ell_{ij'}\left[\theta_{ij'}^{inlet}(\mathbf{v}_{j'}^{(bc)}\cdot\mathbf{n}_{ij'})\mathcal{E}_{j'}^{v,\,(bc)}+\theta_{ij'}^{free}(\mathbf{v}_i\cdot\mathbf{n}_{ij'})\mathcal{E}_i^v\right]^{n+\alpha}.$$

The r.h.s. term for momentum equation is

$$\begin{aligned}
\mathbf{b}^{n+\alpha}_{\rho\mathbf{v},i} \;=\; & |K_i|\,(\rho\mathbf{v})^n_i - \sum_j \left\{2\alpha A^{n+\alpha}_{ij}(\rho\mathbf{v})^n_j + V^{n+\alpha}_{ij}(\rho\mathbf{v})^{n+\alpha}_j\right\} \\
& - \left(\alpha+\frac{1}{2}\right)\Delta t \sum_{j'\in\sigma'(i)} \ell_{ij'}\theta^{inlet}_{ij'}\left[(\mathbf{v}^{(bc)}_{j'}\cdot\mathbf{n}_{ij'})(\rho\mathbf{v})^{(bc)}_{j'} + p^{(bc)}_{j'}\mathbf{n}_{ij'}\right]^{n+\alpha} \\
& - \left(\alpha+\frac{1}{2}\right)\Delta t \sum_{j'\in\sigma'(i)} \ell_{ij'}\theta^{free}_{ij'}\left[(\mathbf{v}_i\cdot\mathbf{n}_{ij'})(\rho\mathbf{v})_i + p_i\mathbf{n}_{ij'}\right]^{n+\alpha} \\
& - \left(\alpha+\frac{1}{2}\right)\Delta t \sum_{j'\in\sigma'(i)} \ell_{ij'}\theta^{solid}_{ij'}p^{n+\alpha}_i\mathbf{n}_{ij'}.
\end{aligned}$$

The r.h.s. term for total energy equation is

$$\begin{aligned}
\mathbf{b}^{n+\alpha}_{\rho E,i} \;=\; & |K_i|\,(\rho E)^n_i - \sum_j \left\{2\alpha A^{n+\alpha}_{ij}(\rho E)^n_j + V^{n+\alpha}_{ij}(\rho E)^{n+\alpha}_j\right\} \\
& - \left(\alpha+\frac{1}{2}\right)\Delta t \sum_{j'\in\sigma'(i)} \ell_{ij'}\theta^{inlet}_{ij'}\left\{(\mathbf{v}^{(bc)}_{j'}\cdot\mathbf{n}_{ij'})[(\rho E)^{(bc)}_{j'} + p^{(bc)}_{j'}]\right\}^{n+\alpha} \\
& - \left(\alpha+\frac{1}{2}\right)\Delta t \sum_{j'\in\sigma'(i)} \ell_{ij'}\theta^{free}_{ij'}\left\{(\mathbf{v}_i\cdot\mathbf{n}_{ij'})[(\rho E)_i + p_i]\right\}^{n+\alpha}.
\end{aligned}$$

A.0.3. Proof of Proposition 4.1

If $\mathbf{A}$ and $\mathbf{B}$ are two matrices of order $m\times n$ and $p\times q$, the *tensor product* $\mathbf{A}\otimes\mathbf{B}$ is the block matrix of order $mp\times nq$ whose block i,j is given by $(\mathbf{A}\otimes\mathbf{B})_{i,j} = A_{ij}\mathbf{B}$. The tensor product have some noteworthy properties, see for instance [22]. We just mention the one used in the following proof, that is $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$, with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ four generic matrices (with compatible dimensions).

Let us introduce the definition of an M-matrix.

DEFINITION A.1. Any matrix $\mathbf{A}$ of the form $\mathbf{A} = s\mathbf{I} - \mathbf{B}$, with $s > 0$ and $\mathbf{B}$ a non-negative matrix, for which $s \geq \rho(\mathbf{B})$, the spectral radius of $\mathbf{B}$, is called an *M-matrix*. When $s > \rho(\mathbf{B})$, $\mathbf{A}$ is called a *non-singular* M-matrix.

We recall in the following technical lemma some well-known results without proof. A detailed exposition of the matter can be found in [3].

LEMMA A.1. *The following statements are equivalent:*

(i) $\mathbf{A}$ *is a non-singular M-matrix;*
(ii) $\mathbf{A}^T$ *is a non-singular M-matrix;*
(iii) *there exists a positive vector* $\mathbf{x}$ *such that* $\mathbf{Ax}$ *is also a positive vector;*
(iv) $\mathbf{A}^{-1}$ *is a non-negative matrix.*

Remark. Statement (iv) in Lemma A.1 is not needed in the present proof, but is used in section 4, and is mentioned here for the sake of completeness.

A positive vector $\mathbf{x}$ is a vector all of whose entries are positive. As usual, the property will be denoted by $\mathbf{x} > \mathbf{0}$, which stands for $x_i > 0$ for all i. The symbol $\mathbf{e}_k$ will also be used, which indicates a vector in $\mathbb{R}^k$ whose components are all equal to unity.

Proof of Proposition 4.1. Here we will show in reverse order that each diagonal block in (18) is a non-singular M-matrix.

(i) Since there holds

$$\begin{aligned}(\mathbf{e}_N^T \mathbf{M}^{n+\alpha})_j &= \sum_{i=1}^{N} M_{ij}^{n+\alpha} = M_{jj}^{n+\alpha} + \sum_{i \neq j} M_{ij}^{n+\alpha} \\ &= |K_j| + \frac{\Delta t}{2} \sum_{k \in \sigma(j)} \ell_{jk} \hat{a}_{kj}^{n+\alpha} - \frac{\Delta t}{2} \sum_{k \in \sigma(j)} \ell_{kj} \hat{a}_{kj}^{n+\alpha} \\ &= |K_j| > 0,\end{aligned}$$

it immediately follows from Lemma (A.1) with $\mathbf{x} = \mathbf{e}_N$ that the *fourth* diagonal block in (18), i.e. the matrix $\mathbf{M}^{n+\alpha}$ defined in (20), is an M-matrix. It is also worth noticing that $\mathbf{M}^{n+\alpha}$ is a strictly column diagonally dominant matrix.

(ii) Let us take the vector $\mathbf{x} = \mathbf{e}_N \otimes \mathbf{e}_k$. We have

$$\begin{aligned}\mathbf{x}^T (\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_k) &= (\mathbf{e}_N \otimes \mathbf{e}_k)^T (\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_k) \\ &= (\mathbf{e}_N^T \otimes \mathbf{e}_k^T)(\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_k) \\ &= \mathbf{e}_N^T \mathbf{M}^{n+\alpha} \otimes \mathbf{e}_k^T \mathbf{I}_k \\ &= \mathbf{e}_N^T \mathbf{M}^{n+\alpha} \otimes \mathbf{e}_k^T > \mathbf{0},\end{aligned}$$

from the definition of the dyadic product and $\mathbf{e}_N^T \mathbf{M}^{n+\alpha} > \mathbf{0}$. Thus, Lemma (A.1) yields that the matrix $\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_k$ is a non-singular M-matrix. Choosing $k = 2$ it follows that the *third* diagonal block in (18) is a non-singular M-matrix.

(iii) The same argument used in (ii) with $k = n_v$ yields that $\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_v}$ is a non-singular M-matrix. Since $\mathbf{D}_e^{n+\alpha}(\boldsymbol{\rho})$ is a diagonal non-negative matrix, Lemma (A.1) with $\mathbf{x} = \mathbf{e}_N \otimes \mathbf{e}_{n_v}$ and a direct calculation yield that the *second* diagonal block in (18), i.e. $\mathbf{D}_e^{n+\alpha}(\boldsymbol{\rho}) + \mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_v}$, is also a non-singular M-matrix.

(iv) The same argument used in (ii) with $k = n_s$ yields that $\mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_s}$ is a non-singular M-matrix. Let us take $\mathbf{x} = \mathbf{e}_N \otimes \mathbf{e}_{n_s}$. From the properties of the matrix $\mathbf{C}$ – see section 2 –and the definition in (19) we have that $\mathbf{x}^T \mathbf{D}_\rho^{n+\alpha}(\boldsymbol{\rho}) = \mathbf{0}$. Lemma (A.1) and a direct calculation yield that the *first* diagonal block in (18), i.e. $\mathbf{D}_\rho^{n+\alpha}(\boldsymbol{\rho}) + \mathbf{M}^{n+\alpha} \otimes \mathbf{I}_{n_v}$, is also a non-singular M-matrix.

A.0.4. Proof of proposition 4.2

To simplify the notation, let us introduce in (A.1) the symbols

$$\begin{aligned}F_{ij}^{(u)} &= \mathrm{a}_{ij}^{n+\alpha} q_i - \mathrm{a}_{ji}^{n+\alpha} q_j \,, \\ F_{ij}^{(c)} &= \frac{1}{2} (q_i \mathbf{v}_i + q_j \mathbf{v}_j)^{n+\alpha} \cdot \mathbf{n}_{ij},\end{aligned}$$

where q_i is anyone of the components of $\boldsymbol{\rho}_i$ within cell i. Boundary condition terms will also be neglected. That is, for the sake of clearness we will consider only internal edges. A generalization of the following argument which includes also boundary edges is trivial.

The r.h.s term $\mathbf{b}_{\rho,\,i}^{n+\alpha}$ associated to the i-th triangle cell can be expressed as

$$|K_i|\, q_i - \sum_{j\in\sigma(i)} \alpha\Delta t \ell_{ij}(1-\theta_{ij}^{n+\alpha})F_{ij}^{(u)} - \sum_{j\in\sigma(i)} \left(\frac{1}{2}+\alpha\right)\Delta t\ell_{ij}\theta_{ij}^{n+\alpha}F_{ij}^{(c)}.$$

We seek for a sufficient condition which implies $\mathbf{b}_{\rho i}^{n+\alpha} \geq 0$. Let us split the first term into three contributions to be attributed to each (internal) edge of cell i. Inequality surely holds when we have

$$\frac{\min\left\{q_i\,|K_i|\,, q_j\,|K_j|\right\}}{3} - \alpha\Delta t\ell_{ij}F_{ij}^{(u)} \geq \Delta t\ell_{ij}\theta_{ij}^{n+\alpha}\left[\left(\frac{1}{2}+\alpha\right)F_{ij}^{(c)} - \alpha F_{ij}^{(u)}\right] \quad \text{(A.2)}$$

for each $j\in\sigma(i)$.

Consider first the second stage, i.e. $\alpha = 1/2$, we get

$$\frac{2\min\left\{q_i\,|K_i|\,, q_j\,|K_j|\right\}}{3\Delta t\ell_{ij}} - F_{ij}^{(u)} \geq \theta_{ij}^{n+\frac{1}{2}}\left[2F_{ij}^{(c)} - F_{ij}^{(u)}\right]. \quad \text{(A.3)}$$

We must necessarily require that there holds

$$\frac{2\min\left\{q_i\,|K_i|\,, q_j\,|K_j|\right\}}{3\Delta t\ell_{ij}} - F_{ij}^{(u)} \geq 0,$$

which is ensured by assuming that Δt satisfies a *CFL*-like constraint of the form

$$\Delta t \leq \frac{2\,|K_i|}{3\ell_{ij}\max\{\mathrm{a}_{ij}^{n+\frac{1}{2}}, \mathrm{a}_{ji}^{n+\frac{1}{2}}\}}. \quad \text{(A.4)}$$

After this condition is assumed, $\theta_{ij}^{n+\frac{1}{2}}$ can be computed from (A.3). Since for every cell i and every edge $j\in\sigma(i)$ there holds

$$\max\left\{\mathrm{a}_{ij}^{n+\frac{1}{2}}, \mathrm{a}_{ji}^{n+\frac{1}{2}}\right\} \leq \max\left\{\left|\mathbf{v}_i^{n+\frac{1}{2}}\right|, \left|\mathbf{v}_j^{n+\frac{1}{2}}\right|, \left|c_i^{n+\frac{1}{2}}\right|, \left|c_j^{n+\frac{1}{2}}\right|\right\} \leq \max\left\{|\mathbf{v}|\,, |c|\right\},$$

and $\dfrac{|K_i|}{\ell_{ij}} \geq \dfrac{h}{2}$, we obtain

$$\frac{\Delta t}{h}\max\{|\mathbf{v}|\,,\,|c|\} \leq \frac{1}{6}\,, \quad \text{(A.5)}$$

which is the constraint on Δt given in Proposition 4.2. A similar argument for the vibrational energy source term $\mathbf{b}_{e,i}^{n+\alpha-\frac{1}{2}}$ gives the same constraint.

At the first stage, i.e. $\alpha = 0$, condition (A.2) reduces to the following condition on θ:

$$\frac{2\min\left\{q_i\,|K_i|\,, q_j\,|K_j|\right\}}{3\Delta t\ell_{ij}} \geq \theta_{ij}^{n}F_{ij}^{(c)}.$$

but if condition (A.4) is satisfied there are no restriction on θ_{ij}^n.

Remark. It is worth mentioning that (A.5) does not prevent $\theta_{ij}^{n+\frac{1}{2}}$ from reaching unity in smooth solution regions, where second-order accuracy should be (formally) attained. In fact, in smooth region we can have approximatively $q_i \approx q_j$ and $F_{ij}^{(u)} \approx F_{ij}^{(c)}$. Then we can approximate (A.2) as

$$\frac{2}{3}\frac{q_i \min\{|K_i|\,,\,|K_j|\}}{\Delta t \ell_{ij}} - q_i \mathbf{v}\cdot\mathbf{n} \geq (1+\theta_{ij}^{n+\frac{1}{2}}) q_i \mathbf{v}\cdot\mathbf{n},$$

Hence, we have

$$\frac{2}{3}\frac{\min\{|K_i|\,,\,|K_j|\}}{\Delta t \ell_{ij}} \geq (1+\theta_{ij}^{n+\frac{1}{2}}) \mathbf{v}\cdot\mathbf{n},$$

if we want $\theta_{ij}^{n+\frac{1}{2}} \approx 1$ we finally have

$$\frac{\Delta t}{6h}\mathbf{v}\cdot\mathbf{n} < 1. \tag{A.6}$$

Hence, $\theta_{ij}^{n+\frac{1}{2}}$ is not constrained in smooth region by the *CFL*-like condition (A.6)

ACKNOWLEDGMENTS

All the unstructured grids used for the calculations presented here were generated by the mesh generator `triangle` developed by Shewchuck, see the *URL*: `http://almond.srv.cs.cmu.edu/afs/cs/project/-quake/public/www/triangle.html`. We would like to thank Mario Arioli, Francesco Bassi and Stefano Rebay for some useful suggestions and discussions. We acknowledge the suggestions of the anonymous referees.