Adjoint-based error estimation for adaptive Petrov-Galerkin finite element methods

S. D'Angelo, M. Ricchiuto, H. Deconinck

14-17 September 2015

Contents

Abstract

The current work concerns the study and the implementation of a modern algorithm for *a posteriori* error estimation in Computational Fluid Dynamics (CFD) simulations based on partial differential equations (PDEs). This estimate involves the use of the adjoint argument. By solving the adjoint problem, it is possible to obtain important information about the transport of the error related to the quantity of interest.
Therefore, we first derive and solve the discrete primal problem in agreement with the chosen numerical method. According to consistency and compatibility conditions, we can use the same discretisation for solving the adjoint problem, simply by swapping the position of the unknowns and the test functions in the linear variational operator.
This procedure, fully developed for discontinuous Galerkin (DG) and Finite Volume (FV) methods, is here for the first time applied in a fully consistent way for Petrov-Galerkin (PG) discretisations. Some numerical schemes such as Streamline Upwind Petrov-Galerkin (SUPG), stabilized Residual Distribution (RD) and bubble stabilised FE method have been selected for implementation and testing. A scalar linear advection equation is used as a model problem for verifying the accuracy of the adjoint-based a posteriori error estimate. Next, we apply the method to a complete collection of numerical examples, starting from scalar nonlinear problem to 2D compressible Euler equations.

1 Introduction

Over the last decade, much progress has been made about *a posteriori* error estimation for a predefined target functional. This type of estimate is indeed an efficient numerical device to automatically verify the order of accuracy of the PDE discretisation. No matter how sophisticated finite element methods used to solve mathematical models are, all results involve numerical errors. The main aim is not to get an approximation of the error but rather to estimate a computable measure of the error in order to purvey refinement indicators to be used in adaptive procedures.
In this field, the *a posteriori* error analysis is one of the most used procedures to compute numerical error indicators. With *a posteriori* error study, we are able to guarantee an error bound defined by a numerical measure of the real error as

$$\text{Error} < e(u_h),$$

where $e(u_h)$ is a computable function of the numerical solution u_h, usually, involving the numerical residual, obtained by inserting the computed solution into the current problem equations. Very recently, new methods based on duality techniques have been developed for supplying the calculation of error bounds of local quantities of interest. Such estimates provide for the so-called *goal-oriented* adaptive methods which adapt meshes to yield good approximations of local quantities of interest. In engineering applications, these target quantities, $\mathcal{J}(u)$, are typically functionals of the analytical solution u such as a mean, point value or boundary flux. In fluid dynamics, they are often associated to the drag and lift of an airfoil, to punctual values on the profile, e.g. the pressure at the stagnation point or the entropy increase (supposed to be zero for subsonic Euler solutions) over the

domain.
By employing a duality argument we derive an error representation formula, where the error in the target quantity, $\mathcal{J}(u) - \mathcal{J}(u_h)$, is the sum of elementwise error indicators η_κ of the triangulation $\mathcal{T}_h$. This local error estimate consists of the finite element residual of u_h multiplied by local terms based on the solution z of the current adjoint problem. However, since the dual solution z is usually unknown, it is computed numerically by solving an adjoint approximated problem and then replaced to build an approximated error representation $\overline{\mathcal{R}}$. Although this requires a further numerical problem to be solved, the cost of this additional computation is in general cheap because the adjoint numerical problem is always a system of linear partial differential equations (even when the original problems are nonlinear).

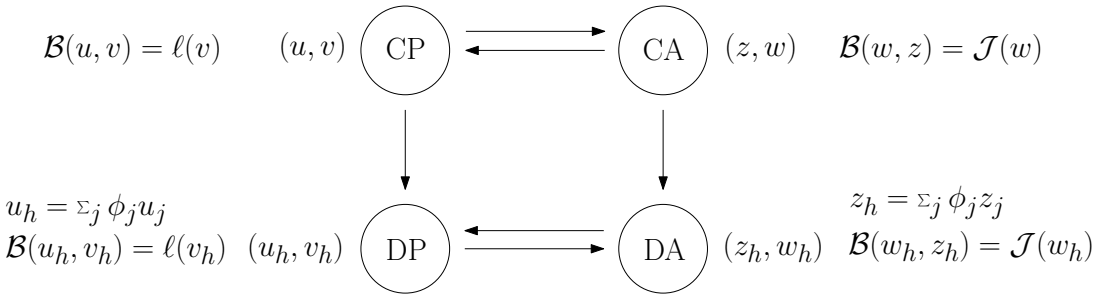


Figure 1: Relation between discrete (D) versus continuum (C), primal (P) versus adjoint (A) for Discontinuous Galerkin discretisation.

1.1 Motivation of the present work

In the present work we will limit our attention to variational Finite Element methods of Galerkin and Petrov-Galerkin type. For Finite Volumes (FV) methods see (Barth (2002), Barth and Larson (2002)). During the last decade, the above procedure has already been developed and deeply applied for Discontinuous Galerkin methods (DG), (Süli and Houston (2002), R. Hartmann (2002), R. Hartmann and P. Houston (2002)). These methods can preserve the Galerkin structure coming from the variational continuum primal (CP) problem in the corresponding discrete primal (DP) problem. Because only one discrete functional space $\mathcal{V}_h$ is used and both the primal (u_h) and adjoint (z_h) solution belong to this space, it is possible to solve a discrete adjoint (DA) problem, consistent with the continuum adjoint (CA) problem, by swapping test and trial functions from the primal discretisation. This property is called *adjoint consistency* meaning that the discrete adjoint problem obtained by swapping the arguments in the primal problem, is at the same time the consistent Galerkin discretisation of the continuum adjoint problem, which allows to close the loop in the Figure 1. In this figure, the top row denotes the continuum (C) while the bottom row stands for as the discrete (D) problem; the left column denotes the primal (P) and the right column the adjoint (A) problem. The pair (p, q) denotes the variational form, with the first argument (p) the trial function and the second argument (q) the test function. Thereby, the continuum functions are $u, v, z, w \in \mathcal{V}$ while their discrete equivalents are $u_h, v_h, z_h, w_h \in \mathcal{V}_h$.
However, this procedure has never been developed in a consistent way for Petrov-Galerkin

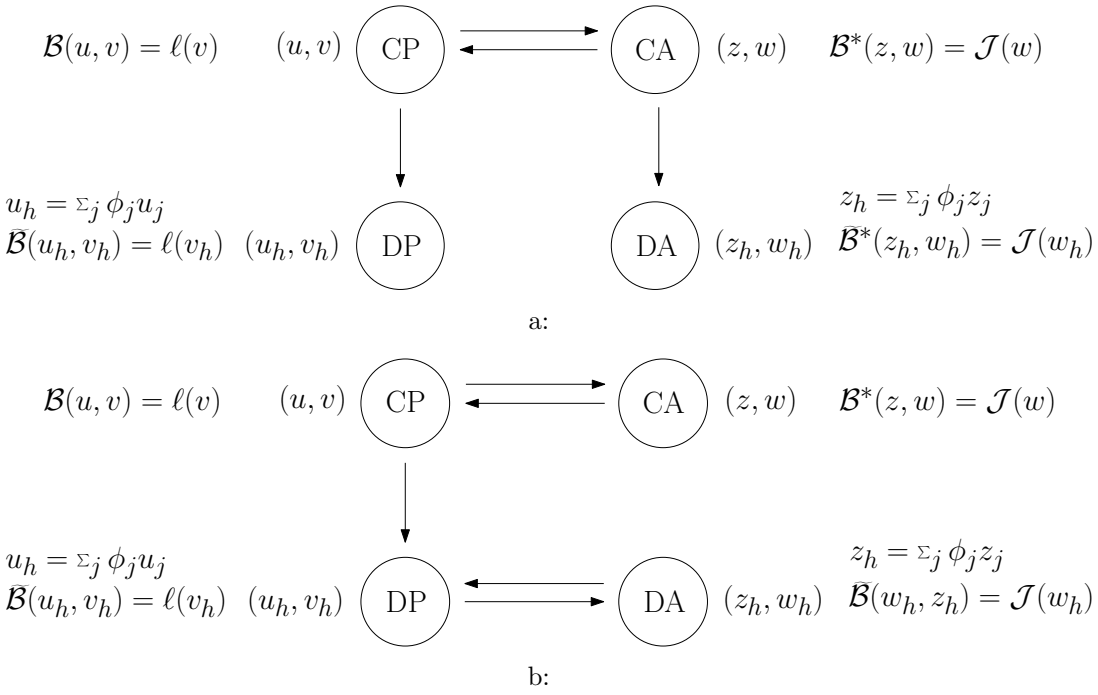


Figure 2: Relation between discrete (D) versus continuum (C), primal (P) versus adjoint (A) for classical Petrov-Galerkin discretisation. (a) Continuum adjoint, (b) discrete adjoints.

(PG) methods or any other global discretisation where no numerical elementwise fluxes are modelled. Indeed, in the classical PG approach, the numerical discretisation is simplified by adding a stabilising term, such that the two different functional spaces, $\mathcal{V}$ and $\widetilde{\mathcal{V}}$, usually reduce to one discrete space $\mathcal{V}_h$ to which both functions u_h and z_h belong, Figure 2. However, this results is a loss of the Galerkin structure on the discrete problem and destroys the possibility of swapping arguments to obtain a consistent dual discrete problem. Then, either we gather the discrete dual problem from the discretisation of the continuum adjoint problem independently of the discrete primal problem (Figure a) or we compute a discrete solution from the discrete primal problem that a priori is not a consistent discretisation of the continuum adjoint problem, (Figure b). Therefore, in both cases, we lose the adjoint consistency.
Hence, in order to obtain a more general and flexible tool, the aim of this work is to extend the DG procedure to a Petrov-Galerkin discretisation. In order to do that, we preserve the Galerkin structure in the primal discrete problem by using PG trial functions, which typically belong to piecewise discontinuous spaces, $\widetilde{\mathcal{V}}_h$. Therefore, bound by the numerical consistency condition with respect to the continuum adjoint problem, the primal solution u_h is sought in $\mathcal{V}_h$, while the corresponding discrete adjoint solution $\tilde{z}_h$ will belong to $\widetilde{\mathcal{V}}_h$, see Figure 3. This constraint is more demanding and harder to work out than for the DG method where both primal and adjoint discrete solutions belong to the same Galerkin space $\mathcal{V}_h$. On the other hand, using stabilised finite element schemes for solving hyperbolic problems, brings all the advantages of these methods being more naturally adapted to advection dominated problems and with the number of degrees of freedom kept more restraint. Therefore, here we apply the adjoint consistency approach

on numerical schemes with compact stencil typically used for hyperbolic problems, such as streamline upwind (SUPG), bubble stabilizing function (BUBBLE) method and a new version of high order Residual Distribution schemes (RD).

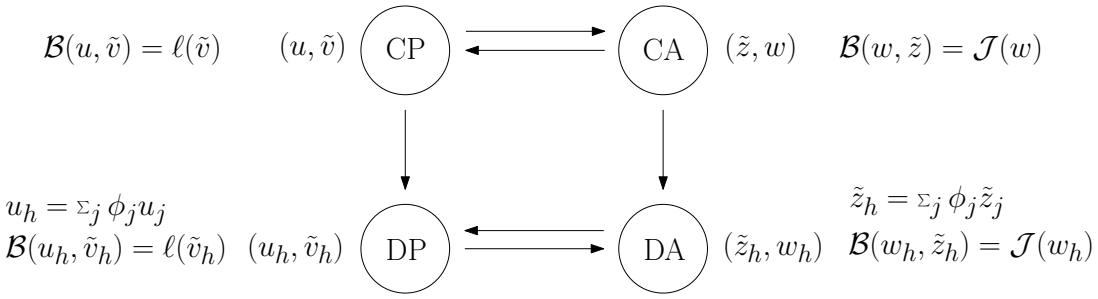


Figure 3: Relation between discrete (D) versus continuum (C), primal (P) versus adjoint (A) for present Petrov-Galerkin discretisation.

2 Definition in continuum setting

2.1 Linear primal and adjoint problems

Primal problem model We begin by introducing first some notation. So following the framework provided in Hartmann (2007), let us consider the general linear problem

$$Lu = f \quad \text{in } \Omega, \qquad\qquad Bu = g \quad \text{on } \Gamma, \tag{1}$$

where $f \in \mathcal{L}^2(\Omega)$ and $g \in \mathcal{L}^2(\Gamma)$, denoting L as a linear differential operator on the domain Ω and B as a linear boundary operator defined on the boundary Γ.

The functional $\mathcal{J}(\cdot)$ In many physical problems the quantity of interest is an output or target functional of the solution rather the solution itself. This target functional is defined as $\mathcal{J}(\cdot)$. Depending on the problem, it can be a different quantity, for example the outflow flux, the drag or the lift coefficient or a point value of the solution.
According to the theory, the linear functional is defined by

$$\mathcal{J}(u) = \big(u, j_\Omega\big)_\Omega + \big(Cu, j_\Gamma\big)_\Gamma \equiv \int_\Omega j_\Omega u \, d\boldsymbol{x} + \int_\Gamma j_\Gamma C u \, ds, \tag{2}$$

where $j_\Omega \in \mathcal{L}^2(\Omega)$ and $j_\Gamma \in \mathcal{L}^2(\Gamma)$, while C is an operator on Γ.

Associated adjoint problem The target functional is said to be *compatible* with the primal problem (1) if there are linear operators L^*, B^* and C^* such that the so-called *compatibility condition* holds

$$\big(Lu, z\big)_\Omega + \big(Bu, C^* z\big)_\Gamma = \big(u, L^* z\big)_\Omega + \big(Cu, B^* z\big)_\Gamma, \tag{3}$$

and $\big(\cdot,\cdot\big)_\Omega$ and $\big(\cdot,\cdot\big)_\Gamma$ are the inner product in $\mathcal{L}^2(\Omega)$ and $\mathcal{L}^2(\Gamma)$, respectively. In general the RHS of (3) will be obtained by applying partial integration to the weak formulation of

the primal problem, which is the LHS of (3). The new operators L^*, B^* and C^* are named *adjoint operators* to L, B and C, respectively and z will be called the adjoint solution. Moreover, from (2) and (3) we can see that the C term makes the target functional $\mathcal{J}$ compatible or not to the primal problem. Indeed, if (3) holds, we can define the adjoint associated problem as follows

$$L^*z = j_\Omega \quad \text{in } \Omega, \qquad\qquad B^*z = j_\Gamma \quad \text{on } \Gamma, \tag{4}$$

and, combining (2), (4) and (3) this yields, (see Giles and Pierce (1997))

$$\begin{aligned}\mathcal{J}(u) &= \big(u, j_\Omega\big)_\Omega + \big(Cu, j_\Gamma\big)_\Gamma = \big(u, L^*z\big)_\Omega + \big(Cu, B^*z\big)_\Gamma \\ &= \big(Lu, z\big)_\Omega + \big(Bu, C^*z\big)_\Gamma = \big(f, z\big)_\Omega + \big(g, C^*z\big)_\Gamma.\end{aligned}$$

This proves the fundamental result that the target functional $\mathcal{J}$ can be computed from the adjoint solution z and the data f and g.

2.2 Example: Linear advection-reaction problem

Once again, following the presentation in Hartmann (2008), let us consider the linear advection-reaction equation

$$Lu := \nabla \cdot (\boldsymbol{b}u) + cu = f \quad \text{in} \quad \Omega, \qquad\qquad Bu := u = g \quad \text{on} \quad \Gamma_-, \tag{5}$$

where the $\Omega \in \mathbb{R}^d$, $d \geq 1$ and Γ_- denotes the inflow part of the boundary $\Gamma = \partial\Omega$

$$\Gamma_- = \big\{\boldsymbol{x} \in \Gamma, \boldsymbol{b}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) < 0\big\},$$

besides $f \in \mathcal{L}^2(\Omega)$, $\boldsymbol{b} \in \big[C^1(\Omega)\big]^d$, $c \in \mathcal{L}^\infty(\Omega)$ and $g \in \mathcal{L}^2(\Gamma_-)$, while $\boldsymbol{n}$ is the outward normal to the computational domain boundary. In (5) the boundary operator is only defined on Γ_-, according to characteristic theory.
Next, we derive the continuum adjoint problem, by reconsidering the variational formulation of (5) with respect to z. We first deduce by partial integration

$$\big(\nabla \cdot (\boldsymbol{b}u) + cu, z\big)_\Omega + \big(u, -\boldsymbol{b} \cdot \boldsymbol{n}z\big)_{\Gamma_-} = \big(u, -\boldsymbol{b} \cdot \nabla z + cz\big)_\Omega + \big(u, \boldsymbol{b} \cdot \boldsymbol{n}z\big)_{\Gamma_+}. \tag{6}$$

Comparing each entry with (3), we extract

$$Lu = \nabla \cdot (\boldsymbol{b}u) + cu, \qquad \text{in} \quad \Omega$$

$$\begin{aligned} Bu &= u, & Cu &= 0, & &\text{on} \quad \Gamma_- \\ Bu &= 0, & Cu &= u, & &\text{on} \quad \Gamma_+ \end{aligned}$$

while the corresponding adjoint operators are

$$L^*z = -\boldsymbol{b} \cdot \nabla z + cz, \qquad \text{in} \quad \Omega$$

$$\begin{aligned} B^*z &= 0, & C^*z &= -\boldsymbol{b} \cdot \boldsymbol{n}z, & &\text{on} \quad \Gamma_- \\ B^*z &= \boldsymbol{b} \cdot \boldsymbol{n}z, & C^*z &= 0. & &\text{on} \quad \Gamma_+ \end{aligned}$$

Thereby, the strong form of the continuum adjoint problem is defined as follows

$$\begin{aligned} -\boldsymbol{b}\cdot\nabla z + cz &= j_\Omega \qquad \text{in} \quad \Omega, \\ \boldsymbol{b}\cdot\boldsymbol{n} z &= j_\Gamma \qquad \text{on} \quad \Gamma_+. \end{aligned} \tag{7}$$

where the j_Ω and j_Γ are provided by a target functional as (2).
On the other hand, the weak form of the continuum adjoint problem follows immediately from (6): find $z \in \mathcal{V}$ such that

$$\big(w, -\boldsymbol{b}\cdot\nabla z + cz\big)_\Omega + \big(w, \boldsymbol{b}\cdot\boldsymbol{n} z\big)_{\Gamma_+} = \mathcal{J}(w) \qquad \forall w \in \mathcal{V}.$$

2.3 Variational formulation of linear advection equation

For further reference, we reformulate (5) for a scalar conservation law, setting the reaction term to zero

$$\nabla\cdot\mathcal{F}(u) = f \quad \text{in} \quad \Omega, \qquad\qquad \mathcal{F}(u)\cdot\boldsymbol{n} = \mathcal{F}(g)\cdot\boldsymbol{n} \quad \text{on} \quad \Gamma_-, \tag{8}$$

where $\mathcal{F}(u)$ is the flux of the conservative quantity u and the easiest case is $\mathcal{F}(u) = \boldsymbol{b}u$. In order to derive a variational formulation, we multiply by a bounded test function v and integrate over the domain Ω,

$$\int_\Omega Lu\, v\, d\boldsymbol{x} = \int_\Omega \nabla\cdot\mathcal{F}(u) v\, d\boldsymbol{x} = \int_\Omega f v\, d\boldsymbol{x}.$$

By now, the function space where the solution u is to be sought in, is not defined yet, since it must depend on how we impose the boundary conditions. However, as already v and $f \in \mathcal{L}^2(\Omega)$, the integral of the left side will exist only if the operator $Lu \in \mathcal{L}^2(\Omega)$ as well, i.e. we define the Hilbert function space

$$\mathcal{H}^{1,L}(\Omega) = \big\{u \in \mathcal{L}^2(\Omega) : Lu = \nabla\cdot(\boldsymbol{b}u) \in \mathcal{L}^2(\Omega)\big\}.$$

Therefore, in order to settle the *weak* variational formulation for equation (5), we multiply it by a test function $v \in \mathcal{H}^{1,L}(\Omega)$, we integrate back and forth by parts replacing u by g on Γ_- and we end up with the following problem: find $u \in \mathcal{H}^{1,L}(\Omega)$ such that, for $\forall v \in \mathcal{H}^{1,L}(\Omega)$

$$\int_\Omega \nabla\cdot\mathcal{F}(u) v\, d\boldsymbol{x} - \int_{\Gamma_-} \mathcal{F}(u)\cdot\boldsymbol{n}\, v\, dl = \int_\Omega f v\, d\boldsymbol{x} - \int_{\Gamma_-} \mathcal{F}(g)\cdot\boldsymbol{n}\, v\, dl,$$

which incorporates explicitly the boundary conditions, J. A. Nitsche (1968). So let us finally define the bilinear form $\mathcal{B}(\cdot,\cdot)$ and the functional $\ell(\cdot)$ such that the variational operator of the advection problem (5) can be shortly written as follows,

$$\mathcal{B}(u,v) = \ell(v) \qquad \forall v \in \mathcal{H}^{1,L}(\Omega), \tag{9}$$

and where

$$\begin{aligned} \mathcal{B}(u,v) &= \int_\Omega \nabla\cdot\mathcal{F}(u) v\, d\boldsymbol{x} - \int_{\Gamma_-} \mathcal{F}(u)\cdot\boldsymbol{n}\, v\, dl, \\ \ell(v) &= \int_\Omega f v\, d\boldsymbol{x} - \int_{\Gamma_-} \mathcal{F}(g)\cdot\boldsymbol{n}\, v\, dl. \end{aligned}$$

According to this definition, we can rewrite the compatibility identity (3) in terms of bilinear forms

$$\mathcal{B}(u,\tilde{z}) = \mathcal{B}^*(\tilde{z},u), \tag{10}$$

with

$$\mathcal{B}(u,\tilde{z}) = \big(Lu,\tilde{z}\big)_\Omega + \big(Bu,C^*\tilde{z}\big)_\Gamma,$$
$$\mathcal{B}^*(\tilde{z},u) = \big(u,L^*\tilde{z}\big)_\Omega + \big(Cu,B^*\tilde{z}\big)_\Gamma.$$

such that the boundary conditions are included inside the operators. Hence, we can now define the weak formulation of the corresponding adjoint problem as follows: find $\tilde{z}$ such that

$$\mathcal{B}^*(\tilde{z},w) = \mathcal{J}(w) \qquad \forall w \in \mathcal{H}^{1,L}. \tag{11}$$

Besides, because of (10), we are also able to redefine the weak adjoint problem by using the primal bilinear form, thereby we have to find $\tilde{z}$ such that

$$\mathcal{B}(w,\tilde{z}) = \mathcal{J}(w) \qquad \forall w \in \mathcal{H}^{1,L}. \tag{12}$$

Remark 2.1. *It is important to notice that if for* (11)*, according to* (7)*, the adjoint solution has to be at least once derivable, i.e.* $\tilde{z} \in \mathcal{H}^1$*, in* (12)*, no derivability constraint has been imposed such that we can simply assume* $\tilde{z} \in \mathcal{L}^2$*.*

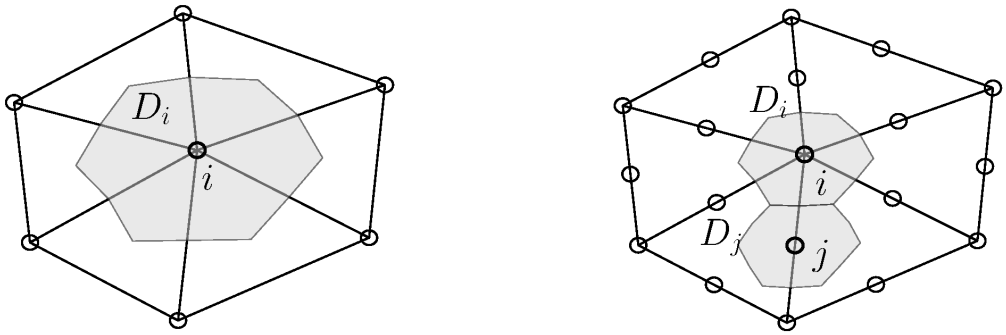


Figure 4: Dual cell for P1 and P2 order triangles.

3 Numerical discretization

3.1 Numerical approximation

Let us consider a 2D spatial discretisation of the domain Ω by non-overlapping triangular elements. We denote the grid by $\mathcal{T}_h$, h being a reference element size. In the following we define some broken (mesh related) function spaces on $\mathcal{T}_h$

Definition 3.1 (Broken Sobolev space $\mathcal{H}^m(\mathcal{T}_h)$)**.** *By* $\mathcal{H}^m(\mathcal{T}_h)$ *we denote the space of* $\mathcal{L}^2$ *functions on* Ω *whose restriction to each element* κ *belongs to the Sobolev space* $\mathcal{H}^m(\kappa)$*, i.e.*

$$\mathcal{H}^m(\mathcal{T}_h) = \big\{v \in \mathcal{L}^2(\Omega) : v|_\kappa \in \mathcal{H}^m(\kappa), \kappa \in \mathcal{T}_h\big\}$$

We denote by κ the generic triangle in $\mathcal{T}_h$ and $|\kappa|$ its area. For all the grids used here, the following regularity is assumed

$$0 < C_1 \leq \sup_{\kappa \in \mathcal{T}_h} \frac{h^2}{|\kappa|} \leq C_2 < \infty,$$

where C_1 and C_2 are positive and finite constants. This condition ensures no vanishing area elements on the grid and no very acute or obtuse angles for any triangle. Further, for every node i in the mesh, $\mathcal{D}_i$ denotes the subset of triangles which i belongs to and we state by S_i the median dual cell created by joining the barycenters of the triangles in $\mathcal{D}_i$ with the midpoints of the edges meeting i, as shown in Figure 4. We assimilate a numerical state with each node and for high order discretisation, the state is considered as a node of an imaginary subtriangulation. Besides, we use the notation χ_S, $S \subset \Omega$, to indicate the characteristic function of a subset S as follows

$$\chi_S(x,y) = \begin{cases} 1 & \text{if} \quad (x,y) \in S \\ 0 & \text{otherwise.} \end{cases}$$

Once the spatial domain has been discretized, we introduce a discrete representation of the unknowns. This representation is constructed starting from the knowledge of the nodal values of the unknown variables whose representation on the mesh is analytically known.

Definition 3.2 (Discrete space $\mathcal{V}_{h,p}^c$)**.** *For $p \geq 1$ we define the space of continuous piecewise polynomials of degree p by*

$$\mathcal{V}_{h,p}^c = \left\{ v_h \in C^0(\Omega) : v_h|_\kappa \circ m_\kappa \in P_p(\hat{\kappa}) \text{ if } \hat{\kappa} \text{ is the unit simplex}, \kappa \in \mathcal{T}_h \right\} \tag{13}$$

with P_p the space of polynomials of degree p.

Let us remind that $\mathcal{V}_{h,p}^c \subset \mathcal{V}$ where $\mathcal{V}$ is the continuous and infinite-dimensional function space where the exact solution u is to be sought in. So let $\left\{\phi_i\right\}_{i \in \mathcal{T}_h}$ denote the continuous piecewise nodal basis functions typically used in finite element methods, defined piecewise for each element κ such that $\phi_i = \sum_{\kappa \in \mathcal{D}_i} \chi_\kappa \phi_{\kappa,i}$ with $\phi_{\kappa,i}$ the locally defined basis function on element κ. These basis functions satisfy

$$\phi_i(\boldsymbol{x}_j) = \delta_{ij} \quad \forall i,j \in \mathcal{T}_h, \qquad \sum_{j \in \kappa} \phi_{\kappa,j} = 1 \quad \forall \kappa \in \mathcal{T}_h.$$

where δ_{ij} is the Kroenecker's delta. We denote with $\phi_j^p(\boldsymbol{x})$, $1 \leq j \leq N_h$, the N_h linearly independent continuous basis functions of degree p in $\mathcal{V}_{h,p}^c$ and the nodal values, u_i, are defined as $u_i = u_h(\boldsymbol{x}_i)$. We introduce the following continuous numerical approximation

$$u_h(\boldsymbol{x}) = \sum_{i \in \mathcal{T}_h} \phi_i^p(\boldsymbol{x}) u_i,$$

in the discrete space $\mathcal{V}_{h,p}^c$ here defined by

$$\mathcal{V}_{h,p}^c = \text{span}\{\phi_j^p(\boldsymbol{x})\}_{j=1}^{N_h} \subset \mathcal{V}.$$

In the following, for sake of conciseness, when the polynomial degree of the current function space is clear, we will use the short notation $\mathcal{V}_h := \mathcal{V}_{h,p}^c$ for a continuous finite element space.

3.2 Petrov-Galerkin method

When trial and test functions belong to different function spaces, i.e. $u_h \in \mathcal{V}_h$ and $\tilde{v}_h \in \widetilde{\mathcal{V}}_h$ with $\mathcal{V}_h \neq \widetilde{\mathcal{V}}_h$, the finite element discretisation is called a *Petrov-Galerkin* method.
Thereby let us now consider the model problem (5) and write the discrete formulation of (9) for a Petrov-Galerkin method as follows: find $u_h \in \mathcal{V}_h \subset \mathcal{H}^{1,L}(\mathcal{T}_h)$ such that

$$\mathcal{B}(u_h, \tilde{v}_h) = \ell(\tilde{v}_h) \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h, \tag{14}$$

where the operator $\mathcal{B}(\cdot,\cdot)$ maintains exactly the same expression as given in (9). Then, as it was pointed out in Remark 2.1, no derivability constraint will be imposed on the test function space $\widetilde{\mathcal{V}}_h$ and this is the reason why it could also be discontinuous over the discrete domain $\mathcal{T}_h$. The only limitation imposed is to be integrable and bounded. Thereby, we construct a particular discrete space where its basis functions $\tilde{\phi}_i$ are defined as follows,

Definition 3.3 (Discrete space $\widetilde{\mathcal{V}}^d_{h,p}$)**.** *For every local solution basis function $\phi_{\kappa,i}$ on element κ we construct a local kernel basis function $\tilde{\phi}_{\kappa,i}$ and the global basis function is as usual given by $\tilde{\phi}_i = \sum_{\kappa\in\mathcal{D}_i} \chi_\kappa \tilde{\phi}_{\kappa,i}$. Within an element κ, the $\phi_{\kappa,i}$ are uniformly bounded functions of the local solution basis functions on the element, their corresponding gradients and the Jacobian of the flux $\mathcal{F}_u$, i.e.*

$$\tilde{\phi}_{\kappa,i} = \phi\big(\{\phi_{\kappa,j}, \nabla\phi_{\kappa,j}, j = 1,\dots,N_\kappa\}, \mathcal{F}_u\big),$$

where N_κ is the number of degrees of freedom on the element. Finally, the local basis functions must satisfy the partition of unity argument, i.e.

$$\tilde{\phi}_{\kappa,i}(\boldsymbol{x}_j) \neq \delta_{ij} \quad \forall i,j \in \mathcal{T}_h, \qquad \textit{but} \qquad \sum_{j\in\kappa} \tilde{\phi}_{\kappa,j}(\boldsymbol{x}) = 1 \quad \forall \boldsymbol{x}.$$

Therefore any function $\tilde{v}_h$ belonging to this space can be written as a linear combination of the $\widetilde{\mathcal{V}}^d_{h,p}$ basis functions as follows,

$$\tilde{v}_h(\boldsymbol{x}) = \sum_{i\in\mathcal{T}_h} \phi^p_i(\boldsymbol{x})\tilde{v}_i \tag{15}$$

As for the previous discrete function space, in the following we also use the short notation $\widetilde{\mathcal{V}}_h := \widetilde{\mathcal{V}}^d_{h,p}$ unless possible misunderstandings or ambiguities.
The use of the test function $\tilde{v}_h \in \widetilde{\mathcal{V}}_h$ could generate some problems of compatibility condition along some boundary types. In order to overcome this problem, a Lagrangian function v_h can be used as a trace of $\tilde{v}_h$ on the boundary. The functional space $\widetilde{\mathcal{V}}_h$ is then redefined as follows

$$\widetilde{\mathcal{V}}_h = \text{span}\big\{\tilde{\phi}_i(\boldsymbol{x}), i = 1,\dots,N_h \,\big|\, \tilde{\phi}_i|_\Gamma = \tilde{\phi}^+_i \equiv \phi_i\big\}.$$

Galerkin orthogonality Because the discrete test function $\tilde{v}_h \in \widetilde{\mathcal{V}}_h \subset \widetilde{\mathcal{V}}$, and since the operators for both discrete and continuum problems are the same, the continuum solution satisfies

$$\mathcal{B}(u, \tilde{v}_h) = \ell(\tilde{v}_h) \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h.$$

Then subtracting this from (14) and using the linearity of the operator $\mathcal{B}(\cdot,\cdot)$, we deduce the important so-called *Galerkin orthogonality* property

$$\mathcal{B}(u - u_h, \tilde{v}_h) = 0 \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h. \tag{16}$$

Hence, the error $e = u - u_h$ is orthogonal to the function space $\widetilde{\mathcal{V}}_h$.

Discrete primal problem After constructing a basis for $\mathcal{V}_h$ and $\widetilde{\mathcal{V}}_h$, given respectively by $\{\phi_i, i = 1, \ldots, N_h\}$ and $\{\tilde{\phi}_j, j = 1, \ldots, N_h\}$, the discrete linear system to be solved for the unknowns u_j is given by

$$\sum_{j \in \mathcal{T}_h} \mathcal{B}(\phi_j, \tilde{\phi}_i) u_j = \ell(\tilde{\phi}_i), \qquad j = 1, \ldots, N_h, \tag{17}$$

where $\mathcal{B}(\cdot,\cdot)$ and $\ell(\cdot)$ have been defined in (9). This numerical method defines a wide family of numerical schemes, depending on the space $\widetilde{\mathcal{V}}_h$ defined by the test function $\tilde{v}_h$.

Discrete adjoint problem If a numerical discretisation takes place on the bilinear operator, (14), the following discrete adjoint problem is defined: find $\tilde{z}_h \in \widetilde{\mathcal{V}}_h$ such that

$$\mathcal{B}(w_h, \tilde{z}_h) = \mathcal{J}(w_h) \qquad \forall w_h \in \mathcal{V}_h, \tag{18}$$

Hence based on (15), the solution $\tilde{z}_h$ must be built as a linear combination of the basis functions $\tilde{\phi}_i$ of the primal test space

$$\tilde{z}_h = \sum_{i \in \mathcal{T}_h} \tilde{z}_i \tilde{\phi}_i^p(\boldsymbol{x}),$$

with the nodal shape function $\tilde{\phi}_i^p = \sum_{\kappa \in \mathcal{D}_i} \chi_\kappa \tilde{\phi}_{\kappa,i}$ and where the coefficients $\tilde{z}_i$ have no more a physical meaning.

3.3 Stabilized finite element methods

It is well known, Gresho and Lee (1979), that for advection-diffusion problems which are dominated by the advection, Galerkin formulations perform well only if the grid is severely refined (cell Péclet number < 1) in order to capture the possible strong gradients arising on the domain. Otherwise spurious oscillations appear and they are carried all over the domain. For linear advection-diffusion equations, this instability is explained in terms of the lack of a suitable functional space coercivity which might control the directional derivatives, see Hartmann (2008).
Hence, only by adding diffusion in the streamline direction we gain control of this contribution and stabilize the scheme leading to a Petrov-Galerkin formulation. In fact, these schemes present in the test function a stabilizer which depends on the local directional derivatives and therefore on the advection part of the current operator L, as follows

$$\tilde{v}_h = \tilde{v}_h(\tau, L_{\text{adv}} v_h), \tag{19}$$

where $\tau \geq 0$ is a properly elementwise defined parameter and L_{adv} the linear (or linearised) advection operator, such that e.g. $L_{\text{adv}} = \boldsymbol{b} \cdot \nabla$ for (5).

Among the different stabilized schemes in the literature, we choose three particular cases which set three Petrov-Galerkin schemes: Streamline Upwind Petrov-Galerkin, PG with stabilising bubble function and stabilized Residual Distribution LDA. Here below, we briefly describe the function $\tilde{v}_h$ belonging to the corresponding functional space $\widetilde{\mathcal{V}}_h$ which defines each of these schemes. For more information and details about them, we suggest to have a look on D'Angelo (2014), here we just notice that α is a scaling parameter and the index l loops over the current triangle nodes..

Streamline Upwind Petrov-Galerkin (SUPG) This is a full Petrov-Galerkin scheme, where we stabilize the central term v_h by adding the directional derivative scaled by all the outflow derivatives in the current element. So the basis function of the node i over the element κ is defined as follows

$$\tilde{\phi}_{\kappa,i} = \phi_{\kappa,i} + \alpha \frac{k_i}{\sum_l k_l^+}, \qquad k_i = L_{\text{adv}}\phi_{\kappa,i},$$
$$k_i^+ = \text{MAX}(k_i, 0) = \gamma_i^+ k_i, \qquad \gamma_i^+ = \frac{1 + \text{SIGN}(k_i)}{2}.$$

Because of the stabilising term, the basis function $\tilde{\phi}_i$ is discontinuous between two adjacent triangles of $\mathcal{T}_h$.

Bubble function (BUBBLE) Bubble function stabilised schemes have since long been developed as an alternative of GLS-stabilized finite element methods for stabilizing the numerical solution provided by the Galerkin method. Therefore,

$$\tilde{\phi}_{\kappa,i} = \phi_{\kappa,i} + \alpha b_\kappa \left(\frac{k_i^+}{\sum_l k_l^+} - \phi_{\kappa,i} \right), \qquad k_i = L_{\text{adv}}\phi_{\kappa,i},$$
$$k_i^+ = \text{MAX}(k_i, 0) = \gamma_i^+ k_i, \qquad \gamma_i^+ = \frac{1 + \text{SIGN}(k_i)}{2},$$

with b_κ a bubble function which holds the condition $b_\kappa = 0$ on $\partial\kappa$ and α a scaling constant such that the integral of αb_κ over the triangle κ is unit. This scheme leads to a continuous function $\tilde{v}_h$ along the element boundary.

Residual Distribution-Low Diffusion A (RD-LDA) For linear operators, new RD techniques can be seen as PG schemes (Ricchiuto (2010), Vymazal et al. (2014)), as for example the LDA scheme defined by,

$$\tilde{\phi}_{\kappa,i} = \frac{k_i^+}{\sum_l k_l^+}, \qquad k_i = L_{\text{adv}}\phi_{\kappa,i},$$
$$k_i^+ = \text{MAX}(k_i, 0) = \gamma_i^+ k_i, \qquad \gamma_i^+ = \frac{1 + \text{SIGN}(k_i)}{2}.$$

Once again, the complete function will not be continuous between two adjacent triangles. Despite the lack of a center function on the local basis, by the energy balance shown in Ricchiuto (2005), we are able to prove how the energy production of the LDA scheme can be split into a stabilizing term related to the dissipative mechanism of the multidimensional upwinding plus a centered term as for the other PG schemes.

3.4 Consistency and adjoint consistency analysis

One of the most important properties of a numerical discretisation is its consistency with respect to the corresponding continuum differential equation. Indeed, according to the Lax-Wendrov theorem sfor conservative methods Lax and Wendroff (1960), a finite element model requests mainly consistency and stability for the convergence of the discrete solution. Furthermore, in finite element methods and consequently for a Petrov-Galerkin discretisation, when consistency holds, it implies directly the Galerkin orthogonality condition. For this reason it is important to check this property when a numerical solution is sought from a discretisation.
Since here the same discretisation is also used for solving the adjoint problem, the consistency of the corresponding discrete adjoint problem has to be proved with respect to the continuum equation. Hence, let us recall the discrete primal problem (14)

$$\mathcal{B}(u_h, \tilde{v}_h) = \ell(\tilde{v}_h) \quad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h,$$

where $\mathcal{B}(\cdot,\cdot)$ is the original bilinear form and $\ell(\cdot)$ the numerical linear form of the source and boundary data, see e.g. (14). This discretisation is said to be *consistent* if the exact primal solution $u \in \mathcal{V}$ satisfies

$$\mathcal{B}(u, \tilde{v}) = \ell(\tilde{v}) \quad \forall \tilde{v} \in \widetilde{\mathcal{V}}.$$

Similarly, the same discretisation is said to be *adjoint consistent* if the exact adjoint solution $\tilde{z} \in \widetilde{\mathcal{V}}$ satisfies (18)

$$\mathcal{B}(w, \tilde{z}) = \mathcal{J}(w) \quad \forall w \in \mathcal{V}. \tag{20}$$

This signifies that for an adjoint consistent discretisation, the discrete adjoint problem represents a consistent discretisation of the continuous adjoint problem. As illustrated in Hartmann (2007), this property plays a key role for the optimal order estimates in finite element methods.
In order to verify the adjoint consistency of the model problem under the current Petrov-Galerkin discretisation, we first rewrite the weak discrete variational formulation (14) in residual form

$$\int_\Omega R(u_h)\tilde{v}_h \, d\boldsymbol{x} + \int_\Gamma r(u_h)\tilde{v}_h^+ \, dl = 0 \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h(\Omega),$$

with $\tilde{v}_h^+ = v_h$ the trace of $\tilde{v}_h$ along the boundary and where $R(u_h)$ and $r(u_h)$ denote the inner and boundary residual, respectively, defined by

$$\begin{aligned} R(u_h) = &\quad f - \nabla\cdot(\boldsymbol{b}u_h) - cu_h && \text{in} \quad \Omega, \\ r(u_h) = &\quad \boldsymbol{b}\cdot\boldsymbol{n}(g - u_h) && \text{on} \quad \Gamma_-, \\ &\quad 0 && \text{on} \quad \Gamma_+. \end{aligned}$$

and which easily verifies the consistency property of the discrete problem, because, if the exact solution $u \in \mathcal{H}^{1,L}$, then $R(u) = 0$ in Ω and $r(u) = 0$ on Γ.
Let us now verify also the adjoint consistency. To do this, we integrate by parts the volume integral in (14)

$$\int_\Omega \big(\nabla\cdot(\boldsymbol{b}u_h) + cu_h\big)\tilde{v}_h \, d\boldsymbol{x} = \int_\Gamma \boldsymbol{b}\cdot\boldsymbol{n}u_h\tilde{v}_h \, dl + \int_\Omega u_h\big(-\boldsymbol{b}\cdot\nabla\tilde{v}_h + c\tilde{v}_h\big) \, d\boldsymbol{x}$$

and applying this on (18) we formulate

$$\int_\Omega w_h R^*(\tilde{z}_h)\, d\boldsymbol{x} + \int_\Gamma w_h r^*(\tilde{z}_h)\, dl = 0 \qquad \forall w_h \in \mathcal{V}_h(\Omega),$$

where

$$\begin{aligned} R^*(\tilde{z}_h) = &\quad j_\Omega + \boldsymbol{b}\cdot\nabla\tilde{z}_h - c\tilde{z}_h \qquad \text{in} \quad \Omega, \\ r^*(\tilde{z}_h) = &\quad j_\Gamma - \boldsymbol{b}\cdot\boldsymbol{n}\,\tilde{z}_h \qquad \text{on} \quad \Gamma_+, \\ &\quad 0 \qquad \text{on} \quad \Gamma_-. \end{aligned}$$

with j_Ω and j_Γ the smooth specific functions defining the target functional $\mathcal{J}(\cdot)$. So, it is simple to validate that $R^*(z) = 0$ and $r^*(z) = 0$ for any exact adjoint solution $z \in \mathcal{H}^{1,L}$, see (7). We highlight that the adjoint residuals depend on the target functional by j_Ω and j_Γ, but as long as this functional is linear, the adjoint consistency is not affected by the definition of $\mathcal{J}(\cdot)$. However, more complicated (and nonlinear) problems and quantities of interest can harm the consistency of the discrete adjoint problem. In those cases, in order to obtain an adjoint consistent discretisation, it might be necessary to apply a consistent modification, see §6.3.

4 *A priori* error estimation

The general objective of this section is to investigate for the new formulation what properties can be proved and how much justification of the numerical results can be given. In particular we aim to give some (albeit limited) information of the underlying functional spaces and on the convergence properties expected for the solutions.
In order to provide the *a priori* error estimation of the numerical solution u_h with respect to the exact u, i.e. $e = u - u_h$, we must introduce and recall some analytical properties of the current operators for a more general framework (see Bochev (2005)) than the usal Galerkin discretisation and where in fact, trial and test spaces are not the same, i.e. $\mathcal{V} \neq \widetilde{\mathcal{V}}$.
First of all, according to (9), let us define the continuous problem as: find $u \in \mathcal{V}$ such that

$$\mathcal{B}(u, \tilde{v}) = \ell(\tilde{v}) \qquad \forall \tilde{v} \in \widetilde{\mathcal{V}}, \tag{21}$$

where $\mathcal{B}(\cdot,\cdot)$ is a bilinear form defined over $\mathcal{V} \times \widetilde{\mathcal{V}}$ and $\ell(\cdot)$ is a linear form defined over $\widetilde{\mathcal{V}}$. Furthermore, recalling Bochev (2005), we know that

Definition 4.1 (Continuity and weak-coercivity properties)**.** *If we introduce the norms* $\|\cdot\|_{\mathcal{V}}$ *and* $\|\cdot\|_{\widetilde{\mathcal{V}}}$ *for the two Hilbert spaces* $\mathcal{V}$ *and* $\widetilde{\mathcal{V}}$*, respectively, then*

- *the bilinear form* $\mathcal{B}(\cdot,\cdot)$ *is* ***continuous*** *on* $\mathcal{V} \times \widetilde{\mathcal{V}}$*, if there exists* $C_c > 0$ *such that*

$$\mathcal{B}(u, v) \leq C_c \| u \|_{\mathcal{V}} \| v \|_{\widetilde{\mathcal{V}}} \qquad \forall u \in \mathcal{V}, v \in \widetilde{\mathcal{V}}.$$

- *the bilinear form* $\mathcal{B}(\cdot,\cdot)$ *is* ***weak-coercive*** *on* $\mathcal{V} \times \widetilde{\mathcal{V}}$*, if there exists* $C_s > 0$ *such that*

$$\sup_{u \in \mathcal{V}} \frac{\mathcal{B}(u, \tilde{v})}{\| u \|_{\mathcal{V}}} \geq \widetilde{C}_s \| \tilde{v} \|_{\widetilde{\mathcal{V}}} \qquad \forall \tilde{v} \in \widetilde{\mathcal{V}},$$

and

$$\sup_{\tilde{v}\in\widetilde{\mathcal{V}}} \frac{\mathcal{B}(u,\tilde{v})}{\|\,\tilde{v}\,\|_{\widetilde{\mathcal{V}}}} \geq C_s \|\,u\,\|_{\mathcal{V}} \qquad \forall u \in \mathcal{V}.$$

In the present work, we consider solution $u \in \mathcal{H}^1(\Omega)$, hence $\mathcal{V} \equiv \mathcal{H}^1$, while for the test space the requirement is weaker, namely $\tilde{v} \in \mathcal{L}^2(\Omega)$, because the test functions will be discontinuous in general. Therefore, the two norms become $\|\cdot\|_{\mathcal{V}} \equiv \|\cdot\|_{\mathcal{H}^1(\Omega)}$ and $\|\cdot\|_{\widetilde{\mathcal{V}}} \equiv \|\cdot\|_{\mathcal{L}^2(\Omega)}$.

Existence and uniqueness of solution We recall the *Necas theorem* to remind the condition under which the problem (21) is well-posed and a solution $u \in \mathcal{V}$ of (21) exists and this solution is unique.

Theorem 4.1 (Necas theorem)**.** *Let $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ be two Hilbert spaces, if the bilinear operator $\mathcal{B} : \mathcal{V} \times \widetilde{\mathcal{V}} \to \mathbb{R}$ is continuous and weak-coercive and the linear functional $\ell : \widetilde{\mathcal{V}} \to \mathbb{R}$ is also bounded, then there is a unique solution $u \in \mathcal{V}$ such that*

$$\mathcal{B}(u,\tilde{v}) = \ell(\tilde{v}) \qquad \forall \tilde{v} \in \widetilde{\mathcal{V}}.$$

Proof. Any textbook, see e.g. Aziz (1972) and Braess (1997), on linear functional analysis or finite element methods and where in case of a pure Galerkin method (i.e. $B(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$) and a coercive bilinear operator, we end up to the famous and so-called *Lax-Milgram theorem.* □

In finite dimensions this theorem corresponds to setting a linear system with a non-singular matrix. Therefore, let us denote a finite dimensional subspace $\mathcal{V}_h$ of $\mathcal{V}$ and also $\widetilde{\mathcal{V}}_h$ for $\widetilde{\mathcal{V}}$, then the discrete counterpart of (21) reads as follows: find $u_h \in \mathcal{V}_h$ such that

$$\mathcal{B}(u_h,\tilde{v}_h) = \ell(\tilde{v}_h) \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h. \tag{22}$$

It is well known that because the discrete problem is defined in terms of exactly the same bilinear and linear forms as the problem it approximates, its well-posedness is governed by exactly the same rules as the well-posedness of the original problem. Thereby, as long as the two discrete spaces, $\mathcal{V}_h$ and $\widetilde{\mathcal{V}}_h$, keep the continuity and weak-coercivity properties of the form $\mathcal{B}(\cdot,\cdot)$, the discrete problem (14) is automatically stable and uniquely solvable. However, the current Petrov-Galerkin discretisation uses stabilised test functions $\tilde{v}_h \in \widetilde{\mathcal{V}}_h \subset \widetilde{\mathcal{V}}$ that cannot assure *a priori* and for all the schemes the stability property of weak-coercivity. In particular, if we consider a conforming discretisation, i.e. $u_h \in \mathcal{V}_h \subset \mathcal{V}$, we can assume

$$\sup_{\tilde{v}_h\in\widetilde{\mathcal{V}}_h} \frac{\mathcal{B}(u_h,\tilde{v}_h)}{\|\,\tilde{v}_h\,\|_{\widetilde{\mathcal{V}}}} \geq C_{s,h} \|\,u_h\,\|_{\mathcal{V}} \qquad \forall u_h \in \mathcal{V}_h,$$

However, we cannot assure the condition

$$\sup_{u_h\in\mathcal{V}} \frac{\mathcal{B}(u_h,\tilde{v}_h)}{\|\,u_h\,\|_{\mathcal{V}}} \geq \widetilde{C}_{s,h} \|\,\tilde{v}_h\,\|_{\widetilde{\mathcal{V}}} \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h,$$

for arbitrary test space $\widetilde{\mathcal{V}}_h$.

Remark 4.1. *For SUPG test space, the weak-coercivity and even coercivity property have been proved, see e.g. Hartmann (2008), while for Residual Distribution schemes (RD-LDA and BUBBLE) weak-coercivity has not been proved so far, despite several efforts. However, in Abgrall et al. (2009) and Abgrall and Roe (2003), it has been shown that even though these schemes cannot be characterised with an algebraic stability estimate, the solution converges and it does with the same rates observed in practice for SUPG. So motivated by that and by years of numerical experience, in the present work, we will postulate that weak-coercivity also holds for the RD-LDA and BUBBLE schemes and rely on the numerical results to confirm this conjecture.*

Best approximation property Now, let us remind that the solution $u \in \mathcal{H}^1$ and apply the following best approximation property, proved in Bochev (2005),

Lemma 4.1 (Best approximation Lemma)**.** *Let the bilinear form $\mathcal{B}$ be continuous and at least weakly-coercive while the linear functional ℓ is bounded. Hence, let $u \in \mathcal{V}$ and $u_h \in \mathcal{V}_h \subset \mathcal{V}$ be the solution to* (21) *and* (22)*, respectively and let w_h denote an arbitrary element of $\mathcal{V}_h$. Then,*

$$\| u - u_h \|_{\mathcal{V}} \leq C \inf_{w_h \in \mathcal{V}_h} \| u - w_h \|_{\mathcal{V}},$$

and the constant $C = (1 + C_c/C_{s,h})$. Hence u_h is the best approximation of u in the space $\mathcal{V}_h$.

So, the discrete error $e = u - u_h$ is bounded by the difference $u - w_h$ for any discrete function $w_h \in \mathcal{V}_h$. Thereby, if we choose $w_h \in \mathcal{V}_h$ as an interpolant of u, $w_h = I_h u$, we obtain

$$\| u - u_h \|_{\mathcal{V}} \leq C \| u - I_h u \|_{\mathcal{V}}, \tag{23}$$

where, according to what defined above, $\mathcal{V} \equiv \mathcal{H}^1$ and hence we can replace the error norm of space $\mathcal{V}$ by the $\mathcal{H}^1$-norm. Consequently, the discrete error $e = u - u_h$ can be bounded by the interpolation error $u - I_h u$ apart from a constant. Therefore, the order of the discrete error is also limited by the order of the interpolation error into $\mathcal{V}_h$.

4.1 Primal error estimate

According to the best approximation lemma and the inequality (23), we assure that if we approximate the solution $u \in \mathcal{V}$ with a certain polynomial degree, then the convergence rate of the discrete FEM error will be bounded by the corresponding interpolation rate. This is supported by the well known *a priori* error estimate analysis for a smooth function $w \in \mathcal{H}^{s+1}(\Omega) \subset \mathcal{V}$ (see e.g. Ern and Guermond (2004)), that states

Corollary 4.1 (Interpolation estimate)**.** *For a shape regular mesh $\mathcal{T}_h$ of Ω, let $p \geq 1$ and I_h be an interpolant operator onto $\mathcal{V}^c_{h,p}$. Suppose that $w \in \mathcal{H}^{s+1}(\Omega)$ with $0 \leq s \leq p$, then there exists a positive constant C, independent of w and the mesh function h, such that*

$$\| w - I_{h,p} w \|_{\mathcal{H}^m(\Omega)} \leq C\, h^{s+1-m} |w|_{\mathcal{H}^{s+1}(\Omega)}, \tag{24}$$

where $0 \leq m \leq s+1$ while $\| \cdot \|_{\mathcal{H}^m(\Omega)}$ and $| \cdot |_{\mathcal{H}^{s+1}(\Omega)}$ are the norm and the semi-norm with respect to the space $\mathcal{H}^m$ and $\mathcal{H}^{s+1}$ over the domain Ω, respectively.

In particular, for $m = 0, 1$, the convergence rate reduces to

$$\| w - I_{h,p} w \|_{\mathcal{L}^2(\Omega)} \leq C\, h^{p+1} |w|_{\mathcal{H}^{p+1}(\Omega)} \qquad \text{for} \qquad m = 0,$$
$$\| w - I_{h,p} w \|_{\mathcal{H}^1(\Omega)} \leq C\, h^{p} |w|_{\mathcal{H}^{p+1}(\Omega)} \qquad \text{for} \qquad m = 1,$$

so the interpolant error is of $\mathcal{O}(h^{p+1})$ in $\mathcal{L}^2(\Omega)$-norm and $\mathcal{O}(h^p)$ in $\mathcal{H}^1(\Omega)$-norm.

According to (9), we remind that the current solution u is sought in the functional space $\mathcal{H}^{1,L}$ and then, for any $w_h \in \mathcal{H}^{1,L}(\Omega)$, let us define the graph norm $\| \cdot \|_{\mathcal{H}^{1,L}}$ as follows

$$\| w_h \|^2_{\mathcal{H}^{1,L}(\Omega)} = h \| \boldsymbol{b} \cdot \nabla w_h \|^2_{\mathcal{L}^2(\Omega)} + \| w_h \|^2_{\mathcal{L}^2(\Omega)} + \int_{\Gamma_-} |\boldsymbol{b} \cdot \boldsymbol{n}| w_h^2 \, dl,$$

As already proposed by Süli and Houston (2002), passing through the classical Galerkin approach of SUPG scheme, it is then possible to define the *a priori* error convergence rate of the scheme with respect to this norm. However, we try to extend this for any Petrov-Galerkin scheme by the following

Conjecture 4.1 ($\mathcal{H}^{1,L}$-error estimate)**.** *Let us assume that the corollary 4.1 holds and consider the Galerkin coercive bilinear form $\widetilde{\mathcal{B}} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$. Then we conjecture that there exists a positive constant C, independent of the mesh function h, such that*

$$\| u - u_h \|_{\mathcal{H}^{1,L}(\Omega)} \leq C\, h^{s+1/2} |u|_{\mathcal{H}^{s+1}(\Omega)}, \tag{25}$$

The idea behind this conjecture is first that (25) can be proven to hold for SUPG, in fact it has been proven (Hartmann (2008)) that the discretisation converges with $\mathcal{O}(p + 1/2)$ where p is the best interpolant order of the discrete solution. This result is possible since the SUPG scheme allows to obtain a discrete Galerkin problem with same test and solution space, (D'Angelo (2014)), and the bilinear operator is coercive with respect to a norm equivalent to the natural norm of these Hilbert spaces, the $\mathcal{H}^{1,L}$-norm. Unfortunately, we cannot obtain the same for the other schemes. In fact, for BUBBLE for example, although we can ideally construct a standard Galerkin form with as a test space a subset of $\mathcal{H}^1$, (see Villedieu (2009)), we cannot show a properly modified bilinear operator inducing a norm equivalent to $\mathcal{H}^{1,L}$ or similar, hence coercivity fails. Similarly, for RD, a Galerkin form is possible as already mentioned, however, no coercivity condition has been provided so far.
So, we have no general proof available as we lack a theoretical background to characterize a general algebraic stability for the Petrov-Galerking solution and thus, we are unable to give a theoretical error estimate. Nevertheless, it is observed that the practical convergence rate is up to the $\mathcal{O}(p + 1/2)$ for all the schemes (see e.g. Villedieu (2009) for BUBBLE and Abgrall et al. (2009) for RD scheme).

4.2 Adjoint error estimate

As for the primal solution let us now obtain the convergence rate for the adjoint numerical error $z - \tilde{z}_h$ and later for the error in the target quantity $\mathcal{J}$. Thereby, based on the analysis

provided in Bochev (2005), if we assume an adjoint consistent discretisation (20), we can obtain the best approximation of the adjoint solution as

$$\| z - \tilde{z}_h \|_{\widetilde{\mathcal{V}}} \leq \widetilde{C} \inf_{\tilde{w}_h \in \widetilde{\mathcal{V}}_h} \| z - \tilde{w}_h \|_{\widetilde{\mathcal{V}}},$$

where $\widetilde{C} = (1 + C_c/\widetilde{C}_{s,h})$. So, similarly to the primal solution, let us remind the interpolant operator I_h, and then we consider $\tilde{w}_h \in \widetilde{\mathcal{V}}_h$ as an interpolant of $\tilde{z} \in \widetilde{\mathcal{V}}$, such that $\tilde{w}_h = I_h \tilde{z} \in \widetilde{\mathcal{V}}_h$.

$$\| z - \tilde{z}_h \|_{\widetilde{\mathcal{V}}} \leq C \| z - I_h \tilde{z} \|_{\widetilde{\mathcal{V}}},$$

However, differently to the primal solution, the best interpolant order depends on the applied scheme and it is not *a priori* the same as the discrete adjoint solution. Hence, if the discrete solution is $\tilde{z}_h \in \mathcal{V}^d_{h,\tilde{p}}$, its interpolant $I_h \tilde{z} \in \mathcal{V}^d_{h,p}$, with a priori $\tilde{p} \neq p$. There, we end up with the following result (D'Angelo (2014)),

Lemma 4.2 (Best interpolant on $\widetilde{\mathcal{V}}_h$)**.** *For SUPG scheme, for a smooth adjoint solution $z \in \mathcal{H}^{p+1}$, the order of the best interpolant $I_h \tilde{z}$ is p. Whilst, for RD-LDA and BUBBLE scheme, the order in $\mathcal{H}^1$-norm of the best interpolant $I_h \tilde{z}$ is simply a constant, i.e. $p = 0$.*

Furthermore, in order to compute possible discrete derivatives of test space norm, we introduce the $\mathcal{L}^2$-projection, applied from the adjoint solution $\tilde{z}$ towards the continuous discrete space $\mathcal{V}_h$, i.e.

$$\int_\Omega \left(\tilde{z} - P^c \tilde{z} \right) v_h \, d\boldsymbol{x} = 0 \qquad \forall v_h \in \mathcal{V}_h.$$

Next, if $z_{\mathcal{L}} = P^c \tilde{z}$ is the continuum adjoint solution computed by the $\mathcal{L}^2$-projection, we define its interpolation onto $\mathcal{V}^c_{h,p}$ as follows

$$I^c_{h,p} z_{\mathcal{L}}(\boldsymbol{x}) = \sum_i^{N_h} \phi_i(\boldsymbol{x}) z_{\mathcal{L}}(\boldsymbol{x}_i). \tag{26}$$

Now, based on primal convergence rate, (24), we are ready to state the following corollary,

Corollary 4.2 (Approximation estimates)**.** *Let $p \geq 0$ be the order of the best interpolant $I_h z_{\mathcal{L}}$ defined in* (26)*. Suppose that $z \in \mathcal{H}^{s+1}(\Omega)$ with also $s \geq 0$, then*

$$\| z - I_{h,p} z_{\mathcal{L}} \|_{\mathcal{H}^m(\Omega)} \leq \tilde{C} \, h^{t+1-m} |z|_{\mathcal{H}^{t+1}(\Omega)}, \tag{27}$$

where $t = \min(s, p)$.

Therefore the rate of the *a priori* adjoint error estimate depends not only on the smoothness of the exact solution z but also on the best order of the interpolant of $z_{\mathcal{L}}$. However, the latter depends in turn on the starting functional space $\widetilde{\mathcal{V}}$ and consequently on the discrete Petrov-Galerkin scheme used. Therefore, the convergence rate expected in $\mathcal{H}^1$-norm will be $\mathcal{O}(p)$ for SUPG and only $\mathcal{O}(0)$ for the other two while in $\mathcal{L}^2$-norm it will be respectively $\mathcal{O}(p+1)$ and $\mathcal{O}(1)$.

4.3 Target functional estimate

For the convergence rates of the target functional $\mathcal{J}(\cdot)$, we recall the *a priori* results for both primal (24) and adjoint solution (27) and based on the analysis given in Süli and Houston (2002) we state the following theorem

Theorem 4.2 (Target estimates)**.** *We suppose the following estimates for the primal and adjoint solution*

$$\| u - u_h \|_{\mathcal{H}^m(\Omega)} \leq C h^q |u|_{\mathcal{H}^{p+1}(\Omega)} \quad \forall u \in \mathcal{H}^{p+1}(\Omega),$$

and

$$\| z - I_h z_{\mathcal{L}} \|_{\mathcal{H}^m(\Omega)} \leq \tilde{C} h^{\tilde{q}} |z|_{\mathcal{H}^{p+1}(\Omega)} \quad \forall z \in \mathcal{H}^{p+1}(\Omega),$$

with $q(p)$ and $\tilde{q}(p)$ positive values and where $z_{\mathcal{L}}$ denotes the adjoint $\mathcal{L}^2$-projected solution from the discrete space $\widetilde{\mathcal{V}}_h$. Let us further assume the bilinear operator (9) *and a linear target quantity as in* (2) *with j_Ω and j_Γ smooth functions on Ω and Γ, respectively.*
If the assumptions in §4 about the properties of $\mathcal{B}(\cdot,\cdot)$ hold, then for a smooth adjoint solution, $z \in \mathcal{H}^{p+1}$ and an adjoint consistent discretisation,

$$\mathcal{B}(w, z) = \mathcal{J}(w) \qquad \forall w \in \mathcal{V},$$

there is a positive constant $\hat{C}$ such that

$$|\mathcal{J}(u) - \mathcal{J}(u_h)| \leq \hat{C} h^{q+\tilde{q}} |u|_{\mathcal{H}^{p+1}} |z|_{\mathcal{H}^{p+1}}. \tag{28}$$

Proof. If we set $e = u - u_h \in \mathcal{V}$, then

$$\begin{aligned}
|\mathcal{J}(u) - \mathcal{J}(u_h)| &= |\mathcal{J}(e)| && \text{(linearity } \mathcal{J}) \\
&= |\mathcal{B}(e, z)| && \text{(adjoint consistency)} \\
&= |\mathcal{B}(u - u_h, z - \tilde{z}_h)| && \text{(Galerkin orthogonality)} \\
&\leq C_c \| u - u_h \|_{\mathcal{V}_h} \| z - \tilde{z}_h \|_{\widetilde{\mathcal{V}}_h} && \text{(continuity } \mathcal{B}) \\
&\leq \hat{C} h^q |u|_{\mathcal{H}^{p+1}(\Omega)} \, h^{\tilde{q}} |z|_{\mathcal{H}^{p+1}(\Omega)}. && \text{(solution convergence rates)}
\end{aligned}$$

□

Based on the properties of primal and adjoint solutions dealt with here and in the previous chapter, the best *a priori* suitable and conservative norms for the functional spaces $\mathcal{V}_h$ and $\widetilde{\mathcal{V}}_h$ are respectively $\mathcal{H}^1$ and $\mathcal{L}^2$. Indeed, according to (9), primal solution must be at least once differentiable and integrable, i.e. $u_h \in \mathcal{H}^1(\mathcal{T}_h)$, while the space $\widetilde{\mathcal{V}}_h$ and thus the adjoint solution in (18), is asked being only integrable; thereby, $z_h \in \mathcal{L}^2$. Thereby, based on their convergence rates, it is possible to estimate the *a priori* error rate of the target functional $\mathcal{J}$.

Remark 4.2. *As mentioned in D'Angelo (2014), for the SUPG scheme, we can avail ourselves of a $\mathcal{H}^{1,L}$-norm and besides, in case of a pure advection problem with no source, the discretisation of this scheme becomes also self-adjoint; therefore the same norm can be taken for the adjoint solution as well. So only for this case and scheme, the taken suitable norm for both solutions is $\mathcal{H}^{1,L}$-norm while in all the other cases, $\mathcal{H}^1$ and $\mathcal{L}^2$-norm are considered respectively for the u and z solution.*

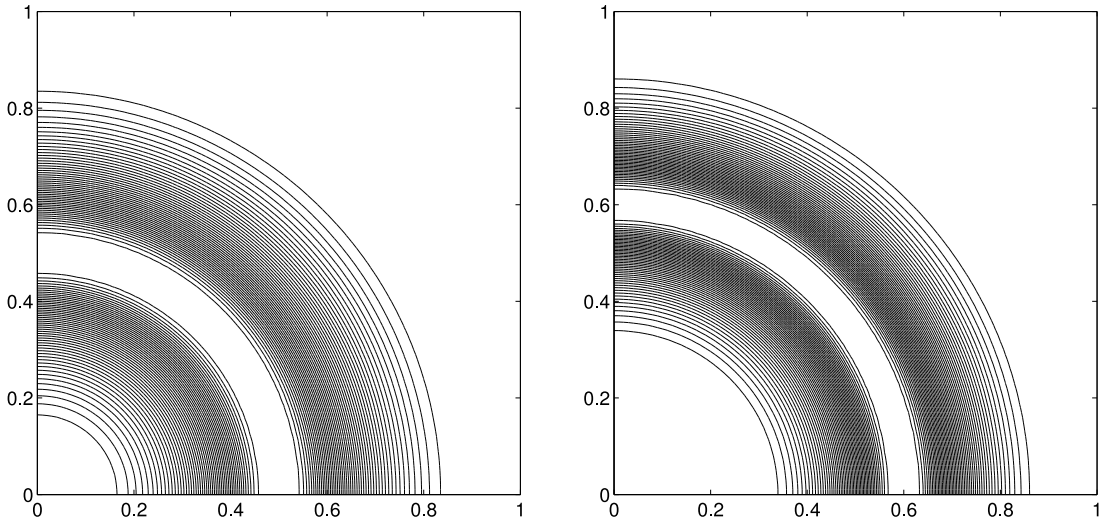


Figure 5: Linear advection problem Barth (2002): (a) primal and (b) adjoint solution

4.4 Numerical example

To numerically verify the convergence rate for smooth solutions and target data, we develop here below the example taken from Barth (2002) of a two-dimensional pure advection problem where the target quantity is a weighted outflow functional.
So let us consider the following problem

$$\begin{aligned} \boldsymbol{b}\cdot\nabla u &= 0 \quad \text{in } \Omega, \\ u &= g \quad \text{on } \Gamma_-. \end{aligned}$$

with circular advection field $\boldsymbol{b} = (-y, x)$ and boundary conditions at the inflow boundary $y = 0$ are given such that the exact solution over the domain is defined as

$$u_{\text{exact}}(x, y) = g(r),$$

with $r = \sqrt{x^2 + y^2}$ and where

$$g(r) = \begin{cases} \widetilde{\psi}(9/20; |r - 1/2|)\big(1 - \widetilde{\psi}(9/20; |r - 1/20|)\big) & r \le 1/2 \\ \widetilde{\psi}(9/20; |r - 1/2|)\big(1 - \widetilde{\psi}(9/20; |r - 19/20|)\big) & r > 1/2. \end{cases} \tag{29}$$

Here, $\widetilde{\psi}(\cdot;\cdot)$ is a C^∞ mollifier function

$$\widetilde{\psi}(r_0; r) = \begin{cases} 0 & r \ge r_0 \\ e^{r^2/(r^2 - r_0^2)} & r < r_0 \end{cases}$$

The target quantity is the weighted outflow flux functional

$$\mathcal{J}(u) = \int_0^1 (\boldsymbol{b}\cdot\boldsymbol{n})\psi_{\text{outflow}}(y)\, u(x, y)\, dy,$$

with the weighting function

$$\psi_{\text{outflow}}(y) = \begin{cases} \widetilde{\psi}(7/20; |y-3/5|)\big(1-\widetilde{\psi}(7/20; |y-1/4|)\big) & y \leq 3/5 \\ \widetilde{\psi}(7/20; |y-3/5|)\big(1-\widetilde{\psi}(7/20; |y-19/20|)\big) & y > 3/5. \end{cases}$$

Thereby, the exact target value equals $\mathcal{J}(u) = 0.09243028358703$. Figure 5 shows the primal and adjoint solutions, while Table 1-4 tabulate values of the global solution error using a sequence of five nested meshes by the three Petrov-Galerkin schemes, SUPG, RD-LDA and BUBBLE, for different norms, i.e. $\mathcal{L}^2$, $\mathcal{H}^1$ and $\mathcal{H}^{1,L}$-norm. Finally, in order to be able to compute the derivative contribution on the norms, we provide the adjoint solution z_h by the $\mathcal{L}^2$ projection of the discrete adjoint solution $\tilde{z}_h$ onto the continuous space $\mathcal{V}_h$.
So, from Table 1 we can prove a order p rate in $\mathcal{H}^1$-norm for the primal solution u in all the schemes, as estimated by (24). However, in the P2 case, maybe due to rounding effects, RD-LDA and in particular BUBBLE seem to achieve an one half lower order. In Table 2, except for BUBBLE that reach $\mathcal{O}(p)$, the primal solution always converges with $\mathcal{O}(p+1/2)$ as it has been mentioned in (25). Table 3 resumes the error rates for the adjoint $\mathcal{L}^2$-norm. There, we notice a different behaviour from SUPG with respect to the other two schemes. The first reduces with $p + 3/2$ order while RD-LDA and BUBBLE show a constant unit order apart from the P1 case, where they achieve $p = 2$. This attitude is also confirmed in Table 4 where SUPG converges in $\mathcal{L}^2$-norm with a $\mathcal{O}(p+1/2)$ rate while the other schemes show $\mathcal{O}(1)$ for P1 and $\mathcal{O}(1/2)$ otherwise. As noticed in D'Angelo (2014), the SUPG scheme presents a self-adjoint discretisation for pure advection problem. This justifies the higher orders of convergence for the adjoint solution in these norms. However, in the P1 case, all these schemes bear close similarities to SUPG (Ricchiuto (2005) or Villedieu (2009)), and that explains the alike order for all of them.
Finally, we can observe the error rate in the target functional $\mathcal{J}$. Table 5 tabulates the final results. Hence, SUPG gives higher rates achieving the *superconvergence* rate with order $2p + 1$, whilst RD-LDA and BUBBLE can achieve only $p + 1$. Therefore, according to (28) and what we remarked beforehand, it is easy to verify how the target error rate is simply the sum of the two solution rates in their suitable norms, both $\mathcal{H}^{1,L}$-norm for SUPG and $\mathcal{H}^1$ and $\mathcal{L}^2$-norm for RD-LDA and BUBBLE scheme. Only in P1 case, the adjoint solution order seems higer than what we can expect but as already noted, this must be due to a simplification of the scheme that makes all similar to the SUPG space. Indeed, using a $\mathcal{H}^{1,L}$ also for them in P1, we get back a consistent estimate.

5 Error representation formula

Let us now denote by $\tilde{z}_h$ the adjoint solution belonging to the discrete test space $\widetilde{\mathcal{V}}_h$, used in the discrete primal problem. This solution may be computed directly by an adjoint consistent discrete problem of (12) or by any suitable projection operator (i.e. interpolation, $\mathcal{L}_2$ projection) whose image belongs to $\widetilde{\mathcal{V}}_h$.
In the wake of Barth (2002), recalling the Galerkin orthogonality (16) and the linearity

h	p	$\|u-u_h\|^S_{\mathcal{H}^1}$ (rates)	$\|u-u_h\|^R_{\mathcal{H}^1}$ (rates)	$\|u-u_h\|^B_{\mathcal{H}^1}$ (rates)
.0625	1	$5.1884\,10^{-1}$	$5.4313\,10^{-1}$	$5.3124\,10^{-1}$
.0312	1	$2.8781\,10^{-1}$ (0.85)	$2.9418\,10^{-1}$ (0.88)	$2.6888\,10^{-1}$ (0.98)
.0156	1	$1.2741\,10^{-1}$ (1.18)	$1.3428\,10^{-1}$ (1.31)	$1.2482\,10^{-1}$ (1.11)
.0078	1	$6.0689\,10^{-2}$ (1.07)	$6.3045\,10^{-2}$ (1.09)	$6.0719\,10^{-2}$ (1.04)
.0039	1	$2.9810\,10^{-2}$ (1.03)	$3.0388\,10^{-2}$ (1.05)	$2.9954\,10^{-2}$ (1.02)
.0625	2	$1.3751\,10^{-1}$	$1.2584\,10^{-1}$	$1.2402\,10^{-1}$
.0312	2	$3.8527\,10^{-2}$ (1.84)	$4.4294\,10^{-2}$ (1.51)	$4.5823\,10^{-2}$ (1.44)
.0156	2	$9.3208\,10^{-3}$ (2.05)	$1.4327\,10^{-2}$ (1.63)	$1.6999\,10^{-2}$ (1.43)
.0078	2	$2.3014\,10^{-3}$ (2.02)	$4.4832\,10^{-3}$ (1.68)	$6.4688\,10^{-3}$ (1.39)
.0039	2	$5.9959\,10^{-4}$ (1.94)	$1.3289\,10^{-3}$ (1.75)	$2.3309\,10^{-3}$ (1.47)
.0625	3	$4.0925\,10^{-2}$	$4.5317\,10^{-2}$	$5.2523\,10^{-2}$
.0312	3	$7.7465\,10^{-3}$ (2.40)	$9.2820\,10^{-3}$ (2.29)	$8.8897\,10^{-3}$ (2.56)
.0156	3	$9.8986\,10^{-4}$ (2.97)	$1.6032\,10^{-3}$ (2.53)	$1.4566\,10^{-3}$ (2.61)
.0078	3	$1.1032\,10^{-4}$ (3.17)	$1.7786\,10^{-4}$ (3.17)	$1.3096\,10^{-4}$ (3.48)
.0039	3	$1.3314\,10^{-5}$ (3.05)	$1.7392\,10^{-5}$ (3.35)	$1.4036\,10^{-5}$ (3.22)

Table 1: Convergence rates primal $\mathcal{H}^1$-norm for SUPG, RD-LDA and BUBBLE scheme on the linear advection problem.

h	p	$\|u-u_h\|^S_{\mathcal{H}^L}$ (rates)	$\|u-u_h\|^R_{\mathcal{H}^L}$ (rates)	$\|u-u_h\|^B_{\mathcal{H}^L}$ (rates)
.0625	1	$2.790\,10^{-2}$	$2.940\,10^{-2}$	$2.877\,10^{-2}$
.0312	1	$1.071\,10^{-2}$ (1.38)	$1.131\,10^{-2}$ (1.38)	$1.077\,10^{-2}$ (1.42)
.0156	1	$3.745\,10^{-3}$ (1.52)	$4.000\,10^{-3}$ (1.50)	$3.724\,10^{-3}$ (1.53)
.0078	1	$1.281\,10^{-3}$ (1.55)	$1.355\,10^{-3}$ (1.56)	$1.291\,10^{-3}$ (1.53)
.0039	1	$4.480\,10^{-4}$ (1.52)	$4.641\,10^{-4}$ (1.55)	$4.512\,10^{-4}$ (1.52)
.0625	2	$4.523\,10^{-3}$	$4.789\,10^{-3}$	$5.696\,10^{-3}$
.0312	2	$9.545\,10^{-4}$ (2.24)	$1.199\,10^{-3}$ (2.00)	$1.473\,10^{-3}$ (1.95)
.0156	2	$1.908\,10^{-4}$ (2.32)	$2.777\,10^{-4}$ (2.11)	$3.781\,10^{-4}$ (1.96)
.0078	2	$3.543\,10^{-5}$ (2.43)	$5.797\,10^{-5}$ (2.26)	$9.480\,10^{-5}$ (2.00)
.0039	2	$6.415\,10^{-6}$ (2.47)	$1.108\,10^{-5}$ (2.39)	$2.207\,10^{-5}$ (2.10)
.0625	3	$1.035\,10^{-3}$	$1.261\,10^{-3}$	$2.134\,10^{-3}$
.0312	3	$1.403\,10^{-4}$ (2.88)	$1.758\,10^{-4}$ (2.84)	$2.500\,10^{-4}$ (3.09)
.0156	3	$1.554\,10^{-5}$ (3.18)	$2.158\,10^{-5}$ (3.03)	$2.830\,10^{-5}$ (3.14)
.0078	3	$1.452\,10^{-6}$ (3.42)	$1.836\,10^{-6}$ (3.56)	$1.928\,10^{-6}$ (3.88)
.0039	3	$1.298\,10^{-7}$ (3.48)	$1.477\,10^{-7}$ (3.64)	$1.501\,10^{-7}$ (3.68)

Table 2: Convergence rates primal $\mathcal{H}^{1,L}$-norm for SUPG, RD-LDA and BUBBLE scheme on the linear advection problem.

h	p	$\|\| z - z_h \|\|^S_{\mathcal{L}^2}$ (rates)	$\|\| z - z_h \|\|^R_{\mathcal{L}^2}$ (rates)	$\|\| z - z_h \|\|^B_{\mathcal{L}^2}$ (rates)
.0625	1	$1.6886\,10^{-2}$	$3.3145\,10^{-2}$	$1.5433\,10^{-2}$
.0312	1	$5.3287\,10^{-3}$ (1.66)	$1.0118\,10^{-2}$ (1.71)	$5.0281\,10^{-3}$ (1.62)
.0156	1	$1.8670\,10^{-3}$ (1.51)	$3.3506\,10^{-3}$ (1.59)	$1.5697\,10^{-3}$ (1.68)
.0078	1	$3.6211\,10^{-4}$ (2.37)	$9.7188\,10^{-4}$ (1.79)	$3.9956\,10^{-4}$ (1.97)
.0039	1	$6.2967\,10^{-5}$ (2.52)	$2.8564\,10^{-4}$ (1.77)	$1.1083\,10^{-4}$ (1.85)
.0625	2	$3.9929\,10^{-3}$	$1.1265\,10^{-2}$	$9.7326\,10^{-3}$
.0312	2	$7.2804\,10^{-4}$ (2.46)	$5.8484\,10^{-3}$ (0.95)	$6.7172\,10^{-3}$ (0.53)
.0156	2	$9.9610\,10^{-5}$ (2.87)	$3.1556\,10^{-3}$ (0.89)	$4.3264\,10^{-3}$ (0.63)
.0078	2	$1.0449\,10^{-5}$ (3.25)	$1.7036\,10^{-3}$ (0.89)	$2.5741\,10^{-3}$ (0.75)
.0039	2	$1.0541\,10^{-6}$ (3.31)	$9.1565\,10^{-4}$ (0.90)	$1.4398\,10^{-3}$ (0.84)
.0625	3	$1.0860\,10^{-3}$	$7.2879\,10^{-3}$	$5.6921\,10^{-3}$
.0312	3	$1.2489\,10^{-4}$ (3.12)	$3.2017\,10^{-3}$ (1.19)	$2.5824\,10^{-3}$ (1.14)
.0156	3	$9.0597\,10^{-6}$ (3.79)	$1.4914\,10^{-3}$ (1.10)	$1.1564\,10^{-3}$ (1.16)
.0078	3	$4.3139\,10^{-7}$ (4.39)	$7.1721\,10^{-4}$ (1.06)	$5.2457\,10^{-4}$ (1.14)
.0039	3	$1.8827\,10^{-8}$ (4.52)	$3.5072\,10^{-4}$ (1.03)	$2.4215\,10^{-4}$ (1.12)

Table 3: Convergence rates adjoint $\mathcal{L}^2$-norm for SUPG, RD-LDA and BUBBLE scheme on the linear advection problem.

h	p	$\|\| z - z_h \|\|^S_{\mathcal{H}^L}$ (rates)	$\|\| z - z_h \|\|^R_{\mathcal{H}^L}$ (rates)	$\|\| z - z_h \|\|^B_{\mathcal{H}^L}$ (rates)
.0625	1	$5.584\,10^{-2}$	$8.500\,10^{-2}$	$5.451\,10^{-2}$
.0312	1	$1.899\,10^{-2}$ (1.56)	$4.132\,10^{-2}$ (1.04)	$2.227\,10^{-2}$ (1.29)
.0156	1	$6.911\,10^{-3}$ (1.46)	$2.079\,10^{-2}$ (0.99)	$9.553\,10^{-3}$ (1.22)
.0078	1	$2.351\,10^{-3}$ (1.56)	$1.039\,10^{-2}$ (1.00)	$4.197\,10^{-3}$ (1.19)
.0039	1	$8.075\,10^{-4}$ (1.54)	$5.193\,10^{-3}$ (1.00)	$1.941\,10^{-3}$ (1.11)
.0625	2	$1.384\,10^{-2}$	$7.765\,10^{-2}$	$9.175\,10^{-2}$
.0312	2	$4.081\,10^{-3}$ (1.76)	$5.238\,10^{-2}$ (0.57)	$8.237\,10^{-2}$ (0.16)
.0156	2	$1.100\,10^{-3}$ (1.89)	$3.711\,10^{-2}$ (0.50)	$7.094\,10^{-2}$ (0.22)
.0078	2	$2.242\,10^{-4}$ (2.29)	$2.697\,10^{-2}$ (0.46)	$5.774\,10^{-2}$ (0.30)
.0039	2	$4.113\,10^{-5}$ (2.45)	$1.968\,10^{-2}$ (0.45)	$4.471\,10^{-2}$ (0.37)
.0625	3	$6.150\,10^{-3}$	$1.538\,10^{-1}$	$1.097\,10^{-1}$
.0312	3	$9.672\,10^{-4}$ (2.67)	$9.178\,10^{-2}$ (0.74)	$7.250\,10^{-2}$ (0.60)
.0156	3	$1.204\,10^{-4}$ (3.01)	$5.813\,10^{-2}$ (0.66)	$4.901\,10^{-2}$ (0.56)
.0078	3	$1.050\,10^{-5}$ (3.52)	$3.863\,10^{-2}$ (0.59)	$3.380\,10^{-2}$ (0.54)
.0039	3	$8.145\,10^{-7}$ (3.69)	$2.652\,10^{-2}$ (0.54)	$2.361\,10^{-2}$ (0.52)

Table 4: Convergence rates adjoint $\mathcal{H}^{1,L}$-norm for SUPG, RD-LDA and BUBBLE scheme on the linear advection problem.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert^S$ (rates)	$\lvert\mathcal{J}-\mathcal{J}_h\rvert^R$ (rates)	$\lvert\mathcal{J}-\mathcal{J}_h\rvert^B$ (rates)
.0625	1	$4.043\,10^{-4}$	$7.810\,10^{-4}$	$3.952\,10^{-4}$
.0312	1	$5.860\,10^{-5}$ (2.79)	$2.215\,10^{-4}$ (1.82)	$9.611\,10^{-5}$ (2.04)
.0156	1	$8.999\,10^{-6}$ (2.70)	$5.767\,10^{-5}$ (1.94)	$2.572\,10^{-5}$ (1.90)
.0078	1	$1.175\,10^{-6}$ (2.94)	$1.514\,10^{-5}$ (1.93)	$6.717\,10^{-6}$ (1.94)
.0039	1	$1.470\,10^{-7}$ (3.00)	$3.867\,10^{-6}$ (1.97)	$1.711\,10^{-6}$ (1.97)
.0625	2	$6.607\,10^{-6}$	$1.549\,10^{-5}$	$2.296\,10^{-5}$
.0312	2	$5.680\,10^{-7}$ (3.54)	$2.147\,10^{-6}$ (2.85)	$4.210\,10^{-6}$ (2.45)
.0156	2	$2.202\,10^{-8}$ (4.69)	$2.755\,10^{-7}$ (2.96)	$6.578\,10^{-7}$ (2.68)
.0078	2	$7.481\,10^{-10}$ (4.88)	$3.770\,10^{-8}$ (2.87)	$8.938\,10^{-8}$ (2.88)
.0039	2	$2.363\,10^{-11}$ (4.98)	$4.770\,10^{-9}$ (2.98)	$1.159\,10^{-8}$ (2.95)
.0625	3	$7.046\,10^{-7}$	$2.165\,10^{-7}$	$9.227\,10^{-8}$
.0312	3	$7.712\,10^{-9}$ (6.51)	$5.824\,10^{-8}$ (1.89)	$5.541\,10^{-8}$ (0.74)
.0156	3	$4.181\,10^{-11}$ (7.52)	$5.982\,10^{-10}$ (6.61)	$4.096\,10^{-9}$ (3.76)
.0078	3	$6.675\,10^{-15}$ (12.6)	$1.509\,10^{-10}$ (1.99)	$2.548\,10^{-10}$ (4.01)
.0039	3	$2.828\,10^{-13}$ (−5.4)	$1.286\,10^{-11}$ (3.55)	$1.605\,10^{-11}$ (3.99)

Table 5: Convergence rates target quantity for SUPG, RD-LDA, BUBBLE scheme on the linear advection problem.

of $\mathcal{B}$ and $\mathcal{J}$, an exact *error representation formula* results from the following steps

$$\begin{aligned}
\mathcal{J}(u)-\mathcal{J}(u_h) &= \mathcal{J}(u-u_h) && \text{(linearity } \mathcal{J}\text{)}\\
&= \mathcal{B}^*(z,u-u_h) && \text{(adjoint problem)}\\
&= \mathcal{B}(u-u_h,z) && \text{(compatibility condition)}\\
&= \mathcal{B}(u-u_h,z-\tilde{z}_h) && \text{(orthogonality)}\\
&= \mathcal{B}(u,z-\tilde{z}_h)-\mathcal{B}(u_h,z-\tilde{z}_h) && \text{(linearity } \mathcal{B}\text{)}\\
&= \ell(z-\tilde{z}_h)-\mathcal{B}(u_h,z-\tilde{z}_h) && \text{(primal problem)}
\end{aligned}$$

so recapitulating

$$\mathcal{J}(u)-\mathcal{J}(u_h)=\mathcal{R}_\Omega(u_h,z-\tilde{z}_h)\equiv\sum_{\kappa\in\mathcal{K}_h}\eta_\kappa, \tag{30}$$

where $\mathcal{R}_\Omega(u_h,z-\tilde{z}_h)=\ell(z-\tilde{z}_h)-\mathcal{B}(u_h,z-\tilde{z}_h)$ and η_κ the *local adjoint-based indicator* originating from the element κ given by

$$\eta_\kappa=\int_\kappa(z-\tilde{z}_h)\cdot R(u_h)\,d\boldsymbol{x}+\int_{\partial\kappa\cap\Gamma}(z-\tilde{z}_h)\cdot r(u_h)\,ds, \tag{31}$$

with the local residuals $R(u)=f-Lu$ and $r(u)=g-Bu$. Thereby

$$\eta_\kappa\equiv\mathcal{R}_\kappa=\ell|_\kappa(z-\tilde{z}_h)-\mathcal{B}|_\kappa(u_h,z-\tilde{z}_h).$$

5.1 *A posteriori* error bound

Unfortunately, the error representation formula written in the global abstract form (30) does not indicate which elements in the mesh should be refined to reduce the measured

error in the functional. To do this, an error localisation procedure has to be developed to point out a local contribution of each element to the global functional error. By applying the *triangle inequality*, indeed, we have

$$\begin{aligned} \big|\mathcal{J}(u) - \mathcal{J}(u_h)\big| &= \big|\mathcal{R}_\Omega(u_h, z - \tilde{z}_h)\big| && \text{(error representation)} \\ &= \Big|\sum_{\kappa\in\mathcal{K}} \mathcal{R}_\kappa(u_h, z - \tilde{z}_h)\Big| && \text{(element assembly)} \\ &\leq \sum_{\kappa\in\mathcal{K}} \big|\mathcal{R}_\kappa(u_h, z - \tilde{z}_h)\big|. && \text{(triangle inequality)} \end{aligned} \tag{32}$$

We define $\mathcal{R}_{|\Omega|} \equiv \sum_{\kappa\in\mathcal{K}} \big|\mathcal{R}_\kappa(u_h, z - \tilde{z}_h)\big|$ then, the following *a posteriori* error bound araises naturally

$$\big|\mathcal{J}(u) - \mathcal{J}(u_h)\big| \leq \mathcal{R}_{|\Omega|} \equiv \sum_{\kappa\in\mathcal{K}_h} |\eta_\kappa|. \tag{33}$$

Thereby, the local error indicator η_κ will select which elements to refine and coarsen through a given adaptive mesh procedure.
Let us suppose a given tolerance $\mathsf{TOL} > 0$ and we consider the design of an adaptive algorithm with the stopping criterion as follows

$$|\mathcal{J}(u) - \mathcal{J}(u_h)| \leq \mathsf{TOL}.$$

From (30), this condition is equivalent to imposing

$$|\mathcal{R}_\Omega(u_h, z - \tilde{z}_h)| \leq \mathsf{TOL}.$$

Unfortunately, this estimation is not computable because of the unknown analytical solutions, u and z. Thus, in order to make these error estimates computable, both u and z must be replaced by suitable approximations which do not affect negatively the quality of the error bound. Thereby, the analytical solution z in (31) must be numerically approximated on a sequence of suitable adjoint finite element space $\widetilde{\mathcal{V}}_h^{\bar{p}}$, based on a adjoint partition $\overline{\mathcal{K}}_h$ or an adjoint polynomial $\bar{p}$. For sake of clearness, in the following this space will be simply named as $\overline{\mathcal{V}}_h$. So, let $\bar{z}_h$ be the approximation to the analytical adjoint solution z from the new finite element space $\overline{\mathcal{V}}_h$, while $\tilde{z}_h$ denotes the numerical adjoint solution solved on the discrete space $\widetilde{\mathcal{V}}_h$ over the subdivision $\mathcal{K}_h$. However, we notice that, the adjoint discrete space $\overline{\mathcal{V}}_h$ has to be richer than the primal one $\widetilde{\mathcal{V}}_h$. Indeed, if $\overline{\mathcal{V}}_h \equiv \widetilde{\mathcal{V}}_h$ it follows that $\bar{z}_h = \tilde{z}_h$ and then $\mathcal{R}_\Omega = 0$. There are usually three main approaches to guarantee it. The first approach is to compute $\bar{z}_h$, over the same mesh $\mathcal{K}_h$ but using a polynomial degree $\bar{p}$ higher than the one of u_h, i.e. $\bar{p} > p$. A variant of this, is to keep the degree p but to compute $\bar{z}_h$ over a different and finer mesh $\overline{\mathcal{K}}_h > \mathcal{K}_h$. And finally, the third option is to compute the discrete adjoint solution using the same polynomial degree p and over the same mesh $\mathcal{K}_h$ and then to take a global or patchwise higher order recovery, such that $\bar{z}_h = R_p^{\bar{p}} z_h$ and $\bar{z}_h \in \overline{\mathcal{V}}_h > \widetilde{\mathcal{V}}_h$.

5.2 Numerical example

The linear advection problem given in Barth (2002) and proposed in §4.4 is again considered.

$$\begin{aligned} \boldsymbol{b}\cdot\nabla u = 0 \quad &\text{in } \Omega, \\ u = g \quad &\text{on } \Gamma_-. \end{aligned}$$

with circular advection field $\boldsymbol{b} = (-y, x)$ and boundary conditions that lead to following exact solution

$$u_{\text{exact}}(x, y) = g(r),$$

with $r = \sqrt{x^2 + y^2}$ and $g(x)$ defined in (29). The target quantity is the weighted outflow flux functional

$$\mathcal{J}(u) = \int_0^1 (\boldsymbol{b} \cdot \boldsymbol{n}) \psi_{\text{outflow}}(y)\, u(x, y)\, dy.$$

Hence, according to (7), the exact adjoint solution will be as follows

$$z_{\text{exact}}(x, y) = \psi_{\text{outflow}}(r).$$

As we have already seen, both primal and adjoint solutions are sufficiently smooth functions and their error convergence rate follows the theoretical orders. Now, by the same example, we want to highlight the accuracy of the error representation formula (30) and the *a posteriori* error bound (33) when the Petrov-Galerkin numerical discretisation is applied.
We compute first the error representation formula, $|\mathcal{R}_\Omega|$, and the error bound, $\mathcal{R}_{|\Omega|}$, for the three different numerical schemes, by using the exact adjoint solution, z, when computing the estimates. So Tables 5.2, 5.2 and 5.2 show the reliability of the discretisation. Here, BUBBLE and RD-LDA schemes show analogous behaviours, with estimates always very near to the exact error but error bounds less strict than for the corresponding SUPG results while the second order solution slightly increases the overestimation by reducing the mesh size. However, for all the schemes and orders, the efficiency index $\theta_1 = |\mathcal{R}_\Omega/(\mathcal{J} - \mathcal{J}_h)|$ is always close to one, even on the first coarse mesh. Moreover, the second index $\theta_2 = |\mathcal{R}_{|\Omega|}/(\mathcal{J} - \mathcal{J}_h)|$ bounds the true error by a consistent and relatively small factor showing the validity of the error localisation procedure.
Looking at Tables 5.2, 5.2 and 5.2, we tabulate the approximated error estimates $|\overline{\mathcal{R}}_\Omega|$ and $\overline{\mathcal{R}}_{|\Omega|}$ when the analytical adjoint solution is replaced by the numerical $\bar{z}_h \in \overline{\mathcal{V}}_h$ with $\bar{p} = p + 1$ over the same mesh $\mathcal{K}_h$. This approximation does not harm the accuracy of the estimation and the error representation formula keeps its accuracy extremely well for both θ_1 and θ_2. Only the P2 case for BUBBLE scheme seems to underestimate the real error by half, likely because in this case the term $\mathcal{R}(u_h, z - \bar{z}_h)$ is not trivial, showing the need for a better approximation. In addition, but as expected, these discrete representations deteriorate over coarse meshes where the discrete approximation differs more and then catches up when the degrees of freedom increase. In addition, a smaller bound is observed for the $\overline{\mathcal{R}}_{|\Omega|}$ compared to $\mathcal{R}_{|\Omega|}$. This could be simply explained by considering the exclusion of a positive contribution, which is supposed to be small indeed, see (34).
So, apart from some trivial differences and especially when we use primal P1 and adjoint P2 case, the numerical error representation formula and the corresponding error bounds keep consistent to the *a posteriori* error estimate theory for continuous solutions. For this reason, we can later apply the discrete approach with enough safety for problems where the analytical solution is not accessible and only its numerical expression is available. Therefore, according to this analysis, we expect to provide always consistent and suitable estimates close to the exact error and where the the effect of the approximation keeps negligible.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$4.043\,10^{-4}$	$4.033\,10^{-4}$	(1.00)	$1.204\,10^{-3}$	(2.98)
.0312	1	$5.860\,10^{-5}$	$5.953\,10^{-5}$	(1.02)	$1.985\,10^{-4}$	(3.39)
.0156	1	$8.999\,10^{-6}$	$9.008\,10^{-6}$	(1.00)	$3.115\,10^{-5}$	(3.46)
.0078	1	$1.175\,10^{-6}$	$1.173\,10^{-6}$	(1.00)	$3.380\,10^{-6}$	(2.88)
.0039	1	$1.470\,10^{-7}$	$1.468\,10^{-7}$	(1.00)	$3.642\,10^{-7}$	(2.48)
.0625	2	$6.607\,10^{-6}$	$7.400\,10^{-6}$	(1.12)	$2.269\,10^{-5}$	(3.43)
.0312	2	$5.680\,10^{-7}$	$5.653\,10^{-7}$	(1.00)	$1.286\,10^{-6}$	(2.26)
.0156	2	$2.202\,10^{-8}$	$2.184\,10^{-8}$	(0.99)	$4.816\,10^{-8}$	(2.19)
.0078	2	$7.481\,10^{-10}$	$7.458\,10^{-10}$	(1.00)	$1.447\,10^{-9}$	(1.93)
.0039	2	$2.363\,10^{-11}$	$2.387\,10^{-11}$	(1.01)	$4.385\,10^{-11}$	(1.86)

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$	(θ_{eff_1})	$\overline{\mathcal{R}}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$4.043\,10^{-4}$	$3.917\,10^{-4}$	(0.97)	$1.168\,10^{-3}$	(2.89)
.0312	1	$5.860\,10^{-5}$	$5.888\,10^{-5}$	(1.00)	$1.928\,10^{-4}$	(3.29)
.0156	1	$8.999\,10^{-6}$	$9.028\,10^{-6}$	(1.00)	$3.082\,10^{-5}$	(3.42)
.0078	1	$1.175\,10^{-6}$	$1.174\,10^{-6}$	(1.00)	$3.368\,10^{-6}$	(2.87)
.0039	1	$1.470\,10^{-7}$	$1.468\,10^{-7}$	(1.00)	$3.640\,10^{-7}$	(2.48)
.0625	2	$6.607\,10^{-6}$	$6.821\,10^{-6}$	(1.03)	$1.896\,10^{-5}$	(2.87)
.0312	2	$5.680\,10^{-7}$	$5.444\,10^{-7}$	(0.96)	$1.241\,10^{-6}$	(2.18)
.0156	2	$2.202\,10^{-8}$	$2.158\,10^{-8}$	(0.98)	$4.669\,10^{-8}$	(2.12)
.0078	2	$7.481\,10^{-10}$	$7.439\,10^{-10}$	(0.99)	$1.439\,10^{-9}$	(1.92)
.0039	2	$2.363\,10^{-11}$	$2.386\,10^{-11}$	(1.01)	$4.380\,10^{-11}$	(1.85)

Table 6: Exact (a) and approximated (b) efficiency rates of error estimates for SUPG scheme on the linear advection problem.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$7.810\,10^{-4}$	$8.025\,10^{-4}$	(1.03)	$2.505\,10^{-3}$	(3.21)
.0312	1	$2.215\,10^{-4}$	$2.219\,10^{-4}$	(1.00)	$5.572\,10^{-4}$	(2.52)
.0156	1	$5.767\,10^{-5}$	$5.778\,10^{-5}$	(1.00)	$1.336\,10^{-4}$	(2.32)
.0078	1	$1.514\,10^{-5}$	$1.514\,10^{-5}$	(1.00)	$3.197\,10^{-5}$	(2.11)
.0039	1	$3.867\,10^{-6}$	$3.868\,10^{-6}$	(1.00)	$7.858\,10^{-6}$	(2.03)
.0625	2	$1.549\,10^{-5}$	$1.556\,10^{-5}$	(1.00)	$1.029\,10^{-4}$	(6.64)
.0312	2	$2.147\,10^{-6}$	$2.166\,10^{-6}$	(1.01)	$1.972\,10^{-5}$	(9.18)
.0156	2	$2.755\,10^{-7}$	$2.752\,10^{-7}$	(1.00)	$3.325\,10^{-6}$	(12.0)
.0078	2	$3.770\,10^{-8}$	$3.763\,10^{-8}$	(1.00)	$5.051\,10^{-7}$	(13.4)
.0039	2	$4.770\,10^{-9}$	$4.766\,10^{-9}$	(1.00)	$7.218\,10^{-8}$	(15.1)

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$	(θ_{eff_1})	$\overline{\mathcal{R}}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$7.810\,10^{-4}$	$7.360\,10^{-4}$	(0.94)	$2.540\,10^{-3}$	(3.25)
.0312	1	$2.215\,10^{-4}$	$2.094\,10^{-4}$	(0.95)	$5.520\,10^{-4}$	(2.49)
.0156	1	$5.767\,10^{-5}$	$5.503\,10^{-5}$	(0.95)	$1.320\,10^{-4}$	(2.29)
.0078	1	$1.514\,10^{-5}$	$1.451\,10^{-5}$	(0.96)	$3.146\,10^{-5}$	(2.08)
.0039	1	$3.867\,10^{-6}$	$3.720\,10^{-6}$	(0.96)	$7.683\,10^{-6}$	(1.99)
.0625	2	$1.549\,10^{-5}$	$1.554\,10^{-5}$	(1.00)	$8.778\,10^{-5}$	(5.67)
.0312	2	$2.147\,10^{-6}$	$2.030\,10^{-6}$	(0.95)	$1.613\,10^{-5}$	(7.51)
.0156	2	$2.755\,10^{-7}$	$2.571\,10^{-7}$	(0.93)	$2.672\,10^{-6}$	(9.70)
.0078	2	$3.770\,10^{-8}$	$3.524\,10^{-8}$	(0.93)	$4.070\,10^{-7}$	(10.8)
.0039	2	$4.770\,10^{-9}$	$4.509\,10^{-9}$	(0.95)	$5.891\,10^{-8}$	(12.3)

Table 7: Exact (a) and approximated (b) efficiency rates of error estimates for RD-LDA scheme on the linear advection problem.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$3.952\,10^{-4}$	$3.983\,10^{-4}$	(1.01)	$1.223\,10^{-3}$	(3.10)
.0312	1	$9.611\,10^{-5}$	$9.575\,10^{-5}$	(1.00)	$2.707\,10^{-4}$	(2.82)
.0156	1	$2.572\,10^{-5}$	$2.570\,10^{-5}$	(1.00)	$6.223\,10^{-5}$	(2.42)
.0078	1	$6.717\,10^{-6}$	$6.715\,10^{-6}$	(1.00)	$1.440\,10^{-5}$	(2.14)
.0039	1	$1.711\,10^{-6}$	$1.711\,10^{-6}$	(1.00)	$3.519\,10^{-6}$	(2.06)
.0625	2	$2.296\,10^{-5}$	$2.320\,10^{-5}$	(1.01)	$7.619\,10^{-5}$	(3.32)
.0312	2	$4.210\,10^{-6}$	$4.192\,10^{-6}$	(1.00)	$1.954\,10^{-5}$	(4.64)
.0156	2	$6.578\,10^{-7}$	$6.580\,10^{-7}$	(1.00)	$4.555\,10^{-6}$	(6.92)
.0078	2	$8.938\,10^{-8}$	$8.941\,10^{-8}$	(1.00)	$9.539\,10^{-7}$	(10.6)
.0039	2	$1.159\,10^{-8}$	$1.160\,10^{-8}$	(1.00)	$1.672\,10^{-7}$	(14.4)

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$	(θ_{eff_1})	$\overline{\mathcal{R}}_{\lvert\Omega\rvert}$	(θ_{eff_2})
.0625	1	$3.952\,10^{-4}$	$3.452\,10^{-4}$	(0.87)	$1.239\,10^{-3}$	(3.13)
.0312	1	$9.611\,10^{-5}$	$7.928\,10^{-5}$	(0.82)	$2.675\,10^{-4}$	(2.78)
.0156	1	$2.572\,10^{-5}$	$2.163\,10^{-5}$	(0.84)	$6.120\,10^{-5}$	(2.38)
.0078	1	$6.717\,10^{-6}$	$5.752\,10^{-6}$	(0.86)	$1.394\,10^{-5}$	(2.08)
.0039	1	$1.711\,10^{-6}$	$1.484\,10^{-6}$	(0.87)	$3.340\,10^{-6}$	(1.95)
.0625	2	$2.296\,10^{-5}$	$1.103\,10^{-5}$	(0.48)	$7.301\,10^{-5}$	(3.18)
.0312	2	$4.210\,10^{-6}$	$2.054\,10^{-6}$	(0.49)	$1.892\,10^{-5}$	(4.49)
.0156	2	$6.578\,10^{-7}$	$3.347\,10^{-7}$	(0.51)	$4.337\,10^{-6}$	(6.59)
.0078	2	$8.938\,10^{-8}$	$4.567\,10^{-8}$	(0.51)	$9.115\,10^{-7}$	(10.2)
.0039	2	$1.159\,10^{-8}$	$5.936\,10^{-9}$	(0.51)	$1.603\,10^{-7}$	(13.8)

Table 8: Exact (a) and approximated (b) efficiency rates of error estimates for BUBBLE scheme on the linear advection problem.

5.3 Mesh adaptation

Let us suppose a given tolerance $\mathsf{TOL} > 0$ and we consider the design of an adaptive algorithm with the stopping criterion as follows

$$|\mathcal{J}(u) - \mathcal{J}(u_h)| \leq \mathsf{TOL}.$$

From (30), this condition is equivalent to imposing

$$|\mathcal{R}_\Omega(u_h, z - \tilde{z}_h)| \leq \mathsf{TOL}.$$

Let us now decompose the error representation formula into two terms, one computable and the other not,

$$\mathcal{R}_\Omega(u_h, z - \tilde{z}_h) = \mathcal{R}_\Omega(u_h, \bar{z}_h - \tilde{z}_h) + \mathcal{R}_\Omega(u_h, z - \bar{z}_h), \tag{34}$$

then it follows that

$$\begin{aligned} |\mathcal{J}(u) - \mathcal{J}(u_h)| &\leq \overline{\mathcal{R}}_{|\Omega|} + |\mathcal{R}_{\overline{\Omega}}| \\ &\equiv \sum_{\kappa \in \mathcal{K}_h} |\bar{\eta}_\kappa| + |\mathcal{R}_\Omega(u_h, z - \bar{z}_h)|, \end{aligned}$$

where $\bar{\eta}_h$ is defined by (31) with $\bar{z}_h$ replacing z. Therefore, the stopping criterion becomes

$$\overline{\mathcal{R}}_{|\Omega|} + |\mathcal{R}_{\overline{\Omega}}| \leq \mathsf{TOL}.$$

As it has been shown in Becker and Rannacher (2001) and we will prove through a numerical example as well, with a suitable choice of the adjoint discrete space $\overline{\mathcal{V}}_h$, the estimate term $|\mathcal{R}_{\overline{\Omega}}|$ is typically negligible with respect to $\overline{\mathcal{R}}_{|\Omega|}$. So finally, despite of all these approximations, the accuracy of the error representation formula (30) and the error bound (33) are not contaminated such that $\overline{\mathcal{R}}_\Omega$ and $\overline{\mathcal{R}}_{|\Omega|}$ keep on approaching to the true error in the target functional $\mathcal{J}(\cdot)$. So, after that, the stopping criterion can be safely set as

$$\overline{\mathcal{R}}_{|\Omega|} \leq \mathsf{TOL}.$$

This final condition is defined by

$$|\mathcal{J}(u) - \mathcal{J}(u_h)| \leq \overline{\mathcal{R}}_{|\Omega|} \equiv \sum_{\kappa \in \mathcal{K}_h} |\bar{\eta}_\kappa|.$$

It involves the numerical solution $\bar{z}_h$ of the adjoint problem and is known also as *Type I a posteriori* error bound (Süli and Houston (2001) and Süli and Houston (2002)). However, as already noticed in §5, in order to have an accurate Type I error bound, it is necessary that $|\mathcal{R}_{\overline{\Omega}}| \ll \overline{\mathcal{R}}_{|\Omega|}$ and then, the adjoint discrete space $\overline{\mathcal{V}}_h$ has to be richer than the primal one $\widetilde{\mathcal{V}}_h$.
An alternative of this error bound is to avoid the computation of the adjoint solution. Such estimation is called *Type II a posteriori* error bound. It can be achieved by the use of the Cauchy-Schwarz inequality on η_κ, based on (31), such that

$$|\eta_\kappa| \leq \| R(u_h) \|_\kappa \, \| z - \tilde{z}_h \|_\kappa.$$

Now, since $\tilde{z}_h$ is a finite element interpolant of the exact solution z from the function space $\widetilde{\mathcal{V}}_h$, let us apply the interpolant error estimate (27) in terms of powers of h and Sobolev semi-norms of z,

$$\| z - \tilde{z}_h \| \leq C h^{\tilde{q}} | z |$$

Finally, we employ for these Sobolev semi-norms a *strong stability* estimation (Eriksson et al. (1995), Süli and Houston (2002)) to end up with Sobolev norms of the data and so, avoiding the involvement of the adjoint solution, e.g.

$$\| z \|_{\mathcal{L}^2(\Omega)} \leq C_{\text{stab}} \quad \text{and} \quad |\eta_\kappa| \leq C_{\text{stab}} \| u \|_\kappa .$$

However, two direct drawbacks are found for this estimate. Firstly, the proof of the strong stability estimation and the study of its constants is depending on the particular problem and requires a relevant amount of analytical work. Secondly, bounds deriving from this type estimate give typically a pessimistic over-estimation of the error (Houston et al. (1999), Houston et al. (2000)).
Once the error bound is defined, given a tolerance TOL, a simple mesh adaptation strategy can be outlined as follows:

1. Construct an initial mesh $\mathcal{K}_h$.
2. Compute the numerical approximation $u_h \in \mathcal{V}_h$ on the current mesh $\mathcal{K}_h$.
3. Compute the numerical approximation $\tilde{z}_h \in \widetilde{\mathcal{V}}_h$
4. Compute the numerical approximation $\bar{z}_h \in \overline{\mathcal{V}}_h$
   (a) on the same mesh $\mathcal{K}_h$ and $\bar{p} > p$, or
   (b) on the mesh $\overline{\mathcal{K}}_h > \mathcal{K}_h$ and $\bar{p} = p$, or
   (c) on the same mesh $\mathcal{K}_h$ and $\bar{p} = p$ and apply a reconstruction strategy
5. Evaluate the error indicators, $\bar{\eta}_\kappa$, for all elements $\kappa \in \mathcal{K}_h$ and sum them all up.
6. If $\sum_{\kappa \in \mathcal{K}_h} |\bar{\eta}_\kappa| \leq$ TOL then STOP, otherwise, refine and coarsen a specified fraction of the total number of elements according to a specified mesh refinement criteria based on the size of $|\eta_\kappa|$, generate a new mesh $\mathcal{K}_h$ and GOTO 2.

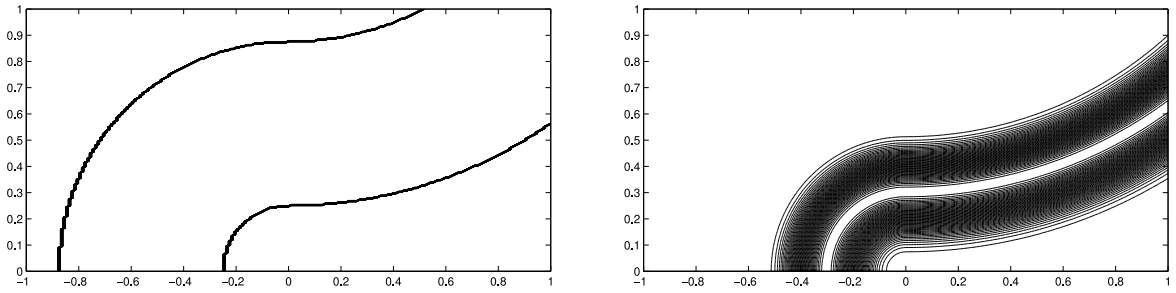
5.4 Numerical example

Here, by using a simple scalar example, we are able to demonstrate the advantages of an adaptive algorithm based on the *Type I a posteriori* error indicators $|\eta_\kappa|$ based on (31)

$$\eta_\kappa = \int_\kappa (z - \tilde{z}_h) \cdot R(u_h)\, dx + \int_{\partial\kappa \cap \Gamma} (z - \tilde{z}_h) \cdot r(u_h)\, ds,$$

with respect to the traditional refinement strategies, *Type II* error bound, which ignore the information coming from the adjoint solution, such as

$$\eta_\kappa^{\text{std}} = \| h\, R(u_h) \|_{\mathcal{L}^2(\kappa)} + \| h^{1/2}\, r(u_h) \|_{\mathcal{L}^2(\partial\kappa \cap \Gamma)}.$$



a: b:

Figure 6: Linear advection problem R. Hartmann (2002). (a) Exact primal and (b) exact adjoint solution.

In this example, we consider the linear hyperbolic problem used in R. Hartmann (2002),

$$\begin{aligned} \boldsymbol{b}\cdot\nabla u &= 0 \qquad \text{in } \Omega \\ u &= g \qquad \text{on } \Gamma, \end{aligned}$$

with $\Omega = [-1,1]\times[0,1] \subset \mathbb{R}^2$ and advection field governed by $\boldsymbol{b} = \frac{\hat{\boldsymbol{b}}}{|\hat{\boldsymbol{b}}|}$ where

$$\hat{\boldsymbol{b}}(x,y) = \begin{cases} (y,x) & \text{if } x<0 \\ (2-y,-x) & \text{otherwise} \end{cases}.$$

The boundary function g is defined as follows

$$g(x,y) = \begin{cases} 1 & \text{for } (x,y)\in[-\frac{7}{8},-\frac{1}{4}]\times\{0\} \\ 0 & \text{otherwise.} \end{cases}$$

Let us suppose we are interested in the outflow solution on a section of the right boundary, i.e. $(x,y)\in\{1\}\times[\frac{1}{4},1]$. Thereby, we set the target functional as

$$\mathcal{J}(u) = \int_0^1 \psi(x,y)\,u(x,y)\,dy,$$

with the weighted function

$$\psi(x,y) = \begin{cases} \exp\left[(\frac{3}{8})^{-2} - \left((y-\frac{5}{8})^2 - \frac{3}{8}\right)^{-2}\right] & \text{for } (x,y)\in\{1\}\times[\frac{1}{4},1] \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the primal solution consists of two discontinuities from the jumps of the boundary function g that are carried over the domain through the advection field $\boldsymbol{b}$ as in Figure a. On the other hand, the adjoint solution is transported backward from the right outlet to the inlet boundary, Figure b.
The primal solution is approximated by first order polynomials, i.e. $u_h \in \mathcal{V}_h^1$ and the adjoint solutions described by a first and second order approximation, respectively, $\tilde{z}_h \in \widetilde{\mathcal{V}}_h^1$ and $\bar{z}_h \in \widetilde{\mathcal{V}}_h^2$. SUPG is the numerical scheme used. Finally, the refinement strategy applied is a special pointwise fixed fraction strategy (see D'Angelo (2014)) with 5% of flag

refinement fraction and 0% of derefinement, combined with a remeshing algorithm by the MMG[12] mesh generator.

Figure 7 shows the final meshes for the adjoint-based and residual adaptive methods, re-

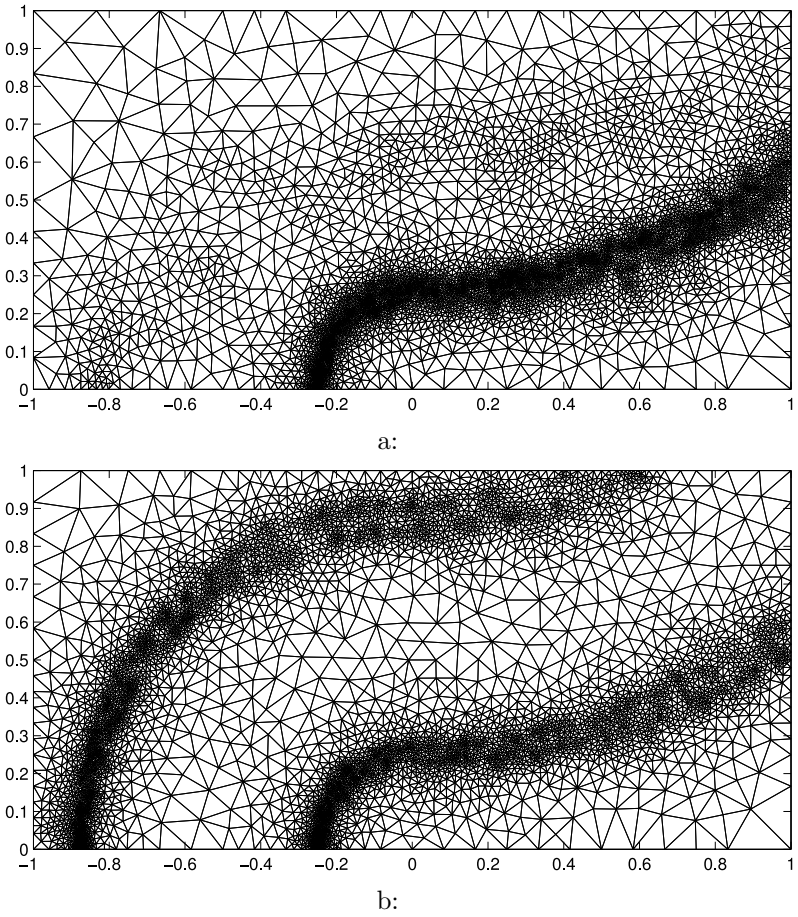


Figure 7: Linear advection problem R. Hartmann (2002). Final grids obtained for (a) goal oriented adjoint-based with 10522 triangles and $\mathcal{J}(e) = 4.924\,10^{-6}$ and (b) residual-based adaptation with 11658 triangles and $\mathcal{J}(e) = 5.520\,10^{-5}$.

spectively. The first polarizes the refinement simply along the right jump passed through the adjoint solution, Figure b. As likely the minimum element size has been soaked, some further spurious refinements are noted somewhere else due more to the adjoint bubble upper tail on the boundary than the left primal jump. On the other hand, in Figure a, both jumps are refined uniformly and these are also the only two features over the domain interested in the refinement.

Table 9 resumes the convergence results for the adaptive algorithm by using the adjoint-based indicators. The number of degrees of freedom, the exact target error $|\mathcal{J} - \mathcal{J}_h|$ and the numerical error representation $|\overline{\mathcal{R}}_\Omega|$ and $\overline{\mathcal{R}}_{|\Omega|}$ with their corresponding efficiencies are shown. The exact error reduces ten times every two iterations till a value close to 10^{-6}. The corresponding estimate follows the error with efficiencies tightly close to

[1]http://www.math.u-bordeaux1.fr/ cdobrzyn/logiciels/mmg3d.php

[2]http://hal.inria.fr/IMB/hal-00681813

one. The error bound $\overline{\mathcal{R}}_{|\Omega|}$ is also bounded and does not increase to very higher order of magnitude. Hence, the inter-triangle cancellations, lost because of the triangle inequality (32), does not play a significant role to reproduce a correct error estimate. Finally, the error evolution is highlighted in Figure 8 where the adjoint adaptive error is compared to the standard one. The latter seems to converge linearly with a smaller rate, while as we already mentioned, the adjoint error decreases faster and with a rate more than linear.

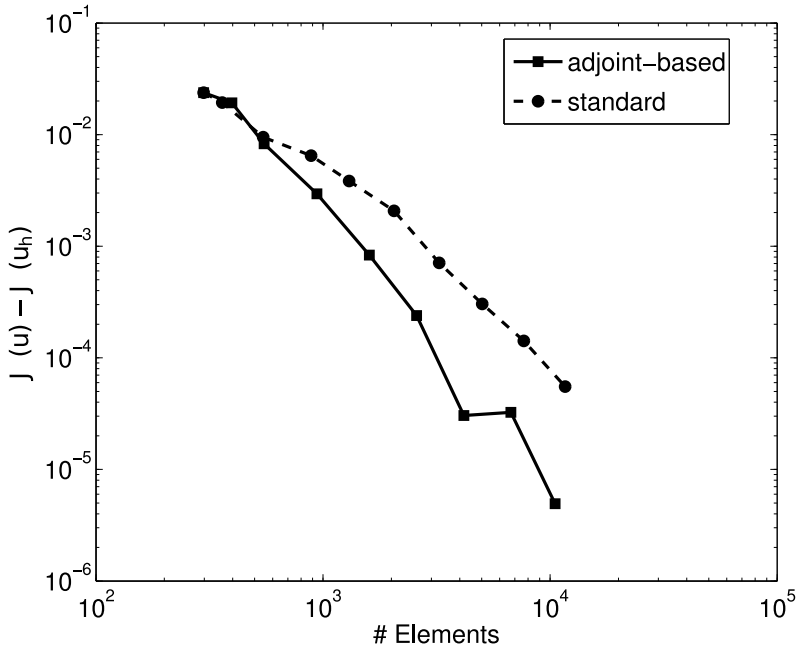


Figure 8: Linear advection problem R. Hartmann (2002). Comparison of target error, $|\mathcal{J}(u) - \mathcal{J}(u_h)|$, computed with standard and adjoint-based adaptivity.

#DoF	$\lvert\mathcal{J} - \mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_{\Omega}\rvert$	(θ_{eff_1})	$\overline{\mathcal{R}}_{\lvert\Omega\rvert}$	(θ_{eff_2})
168	$2.375 \cdot 10^{-2}$	$2.054 \cdot 10^{-2}$	(0.86)	$4.613 \cdot 10^{-2}$	(1.94)
221	$1.928 \cdot 10^{-2}$	$1.869 \cdot 10^{-2}$	(0.97)	$4.966 \cdot 10^{-2}$	(2.58)
303	$8.273 \cdot 10^{-3}$	$7.766 \cdot 10^{-3}$	(0.94)	$2.388 \cdot 10^{-2}$	(2.89)
502	$2.951 \cdot 10^{-3}$	$2.959 \cdot 10^{-3}$	(1.00)	$9.861 \cdot 10^{-3}$	(3.34)
839	$8.334 \cdot 10^{-4}$	$8.725 \cdot 10^{-4}$	(1.05)	$4.026 \cdot 10^{-3}$	(4.83)
1334	$2.393 \cdot 10^{-4}$	$2.440 \cdot 10^{-4}$	(1.02)	$2.083 \cdot 10^{-3}$	(8.70)
2137	$3.042 \cdot 10^{-5}$	$2.957 \cdot 10^{-5}$	(0.97)	$8.084 \cdot 10^{-4}$	(26.5)
3414	$3.247 \cdot 10^{-5}$	$3.265 \cdot 10^{-5}$	(1.01)	$3.048 \cdot 10^{-4}$	(9.39)
5330	$4.924 \cdot 10^{-6}$	$4.868 \cdot 10^{-6}$	(0.99)	$1.318 \cdot 10^{-4}$	(26.7)

Table 9: Linear advection problem R. Hartmann (2002). Efficiency of adjoint-based *a posteriori* error estimation.

6 Hyperbolic conservation laws

6.1 Variational formulation for conservation laws

Now, let us consider a steady conservation law, such as

$$\nabla \cdot \mathcal{F}(u) = 0 \quad \text{in} \quad \Omega, \tag{35}$$

where $\mathcal{F}(u)$ is a nonlinear flux of a conservative quantity u, under appropriate boundary conditions on Γ such that $B(u, \boldsymbol{n}) = \partial_u \mathcal{F}(u) \cdot \boldsymbol{n}$ has real eigenvalues for all vectors $\boldsymbol{n} \in \partial\Omega$. In flux form this condition is replaced by

$$I^-(u, \boldsymbol{n}) \left[\mathcal{F}(u) \cdot \boldsymbol{n} - \mathcal{F}(g) \cdot \boldsymbol{n} \right] = 0, \tag{36}$$

where $I^{\pm} = R I_{\Lambda^{\pm}} L$, with R and L are the right and left eigenvectors of $B(u, \boldsymbol{n})$, $I_{\Lambda^{\pm}} = \text{diag}(\lambda^{\pm}/|\lambda|)$ and satisfying $I^- + I^+ = I$, the unit matrix.
Following the same procedure applied in §2.3, we end up with the corresponding nonlinear operator

$$\mathcal{N}(u, v) = \int_{\Omega} v \, \nabla \cdot \mathcal{F}(u) \, d\boldsymbol{x} + \int_{\Gamma} v^+ \, \mathcal{H}(u^+, u_{\Gamma}(u^+), \boldsymbol{n}) \, ds, \tag{37}$$

where a boundary flux function

$$\mathcal{H}(u^+, u^-, \boldsymbol{n}) = I^- \left(\mathcal{F}(u^-) - \mathcal{F}(u^+) \right) \cdot \boldsymbol{n}, \tag{38}$$

has been defined with u^+ and u^- the inner and outer-trace of the solution u on the boundary Γ, and $\boldsymbol{n}$ the unit outward normal Γ. The boundary integral corresponds to the weak imposition of the boundary data only for the incoming characteristics. Therefore, let us define a weak formulation of (35) given by: find $u \in \mathcal{V}$ such that

$$\mathcal{N}(u, v) = 0 \qquad \forall v \in \widetilde{\mathcal{V}}, \tag{39}$$

where $\mathcal{N}(\cdot, \cdot)$ is a semi-linear form, nonlinear in the first argument and linear in its second argument. Further, let us construct the mean value linearisation of the target functional given by

$$\overline{\mathcal{J}}(u, u_h; u - u_h) = \mathcal{J}(u) - \mathcal{J}(u_h) = \int_0^1 \mathcal{J}'[su + (1-s)u_h](u - u_h) ds,$$

where $\mathcal{J}'[w](\cdot)$ means the functional (Fréchet) derivative of $\mathcal{J}(\cdot)$ evaluated at some $w \in \mathcal{V}$ and again, the dependence of $\overline{\mathcal{J}}$ on the exact solution u will be ignored in the notation. By the compatibility condition and using infinite-dimensional trial and test spaces $\mathcal{V}$, we replace the corresponding adjoint operator $\overline{\mathcal{N}}^*(\cdot, \cdot)$ by the current $\overline{\mathcal{N}}(\cdot, \cdot)$ and define the adjoint problem as follows

$$\overline{\mathcal{N}}(w, z) = \overline{\mathcal{N}}^*(z, w) = \overline{\mathcal{J}}(w) \qquad \forall w \in \mathcal{V}. \tag{40}$$

6.2 Numerical discretization

Through a suitable Petrov-Galerkin numerical discretisation of (39) with test space $\widetilde{\mathcal{V}}_h$, we recover the semi-linear form, $\mathcal{N}_h(\cdot,\cdot)$, such that the nonlinear problem in discrete form becomes: find $u_h \in \mathcal{V}_h$ such that

$$\mathcal{N}_h(u_h, \tilde{v}_h) = 0 \quad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h. \tag{41}$$

This discretisation is called *consistent* if after replacing u_h by the exact solution u for continuum test functions v, it still holds

$$\mathcal{N}_h(u, v) = 0 \quad \forall v \in \widetilde{\mathcal{V}}. \tag{42}$$

This is indeed satisfied if we choose $\mathcal{N}_h(u,v) := \mathcal{N}(u,v)$ as is done in this work.

Galerkin orthogonality We show Galerkin orthogonality and derive the error representation for the nonlinear case. To this end, we first introduce $\overline{\mathcal{N}}(u, u_h; \cdot, \cdot)$ to denote the mean-value linearisation given by

$$\begin{aligned} \overline{\mathcal{N}}(u, u_h; u - u_h, v) &= \mathcal{N}(u,v) - \mathcal{N}(u_h, v) \\ &= \int_0^1 \mathcal{N}'[su + (1-s)u_h](u - u_h, v) ds, \end{aligned} \tag{43}$$

for all $v \in \mathcal{V}$ and where $\mathcal{N}'[u](w,v)$ denotes the Fréchet derivative of $\mathcal{N}(u,v)$ with respect to u, for a fixed $v \in \mathcal{V}$, at some direction $w \in \mathcal{V}$. As R. Hartmann (2002) claims, the linearisation of the semilinear form is only a formal calculation and the derivative might not in general exist. However, in the following, we assume that the $\mathcal{N}'[u](\cdot,\cdot)$ is well-defined and for sake of shortness, the dependence of $\overline{\mathcal{N}}$ on the integration path $u \to u_h$ will be suppressed in the notation. Now, if the continuum function space $\mathcal{V}$ satisfies $\widetilde{\mathcal{V}}_h \subset \mathcal{V}$, subtracting (41) from (42) we obtain the Galerkin orthogonality

$$\mathcal{N}_h(u, \tilde{v}_h) - \mathcal{N}_h(u_h, \tilde{v}_h) = 0 \quad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h.$$

which combined with (43) gives

$$\overline{\mathcal{N}}(u, u_h; u - u_h, \tilde{v}_h) = 0 \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h,$$

hence the error is orthogonal to the space $\widetilde{\mathcal{V}}_h$.

Discrete primal problem After constructing a basis for $\mathcal{V}_h$ and $\widetilde{\mathcal{V}}_h$, given respectively by $\{\phi_i, i = 1, \ldots, N_h\}$ and $\{\tilde{\phi}_j, j = 1, \ldots, N_h\}$, the discrete nonlinear system of algebraic equations for the solution $\{u_i, i = 1, \ldots, N_h\}$ is given by

$$\mathcal{N}_h\big(\sum_i \phi_i u_i, \tilde{\phi}_j\big) = 0 \qquad j = 1, \ldots, N_h. \tag{44}$$

This can be solved by employing a nonlinear iteration scheme, like the *Newton iteration*. The latter generates a sequence of iterands u_h^k by the following method. Given an iterative solution u_h^k,

$$u_h^{k+1} = u_h^k + \omega_N \Delta u_h^k,$$

with Δu_h^k the solution of the linear system

$$\mathcal{N}_h'[u_h^k](\Delta u_h^k, \tilde{v}_h) = -\mathcal{N}_h(u_h^k, \tilde{v}_h) \qquad \forall \tilde{v}_h \in \tilde{\mathcal{V}}_h^p,$$

Here ω_N indicates a damping parameter and $\mathcal{N}_h'[w](\cdot, \tilde{v})$ is the Jacobian of the nonlinear operator $\mathcal{N}_h$, i.e. the functional Fréchet derivative $u \to \mathcal{N}_h(u, \tilde{v})$, for the component $\tilde{v}$ fixed, at some direction w in $\mathcal{V}$. So, $w \to \mathcal{H}'_{u^+}(w^+, w^-, \boldsymbol{n})$ and $w \to \mathcal{H}'_{u^-}(w^+, w^-, \boldsymbol{n})$ denote the derivative of the boundary flux function $\mathcal{H}(\cdot, \cdot, \cdot)$ with respect to its first and second arguments and by the same reasoning, $w \to \mathcal{F}'_u(w)$ is the derivative of the flux $\mathcal{F}(\cdot)$ with respect to its argument. Hence the discrete Fréchet derivative of $\mathcal{N}_h$ with respect to the first argument is defined as

$$\begin{aligned} \mathcal{N}_h'[u_h](w_h, \tilde{v}_h) = \int_\Omega \tilde{v}_h \nabla \cdot \left(\mathcal{F}'_{u_h} w_h\right) d\boldsymbol{x} + \int_\Omega \tilde{v}'_{u_h} w_h \nabla \cdot \mathcal{F}_h \, d\boldsymbol{x} \\ + \int_\Gamma \tilde{v}_h^+ \left(\mathcal{H}'_{u_h^+} + \mathcal{H}'_{u_h^-} u'_\Gamma\right) w_h \, ds, \qquad \forall \tilde{v}_h \in \widetilde{\mathcal{V}}_h \end{aligned} \tag{45}$$

where all the fluxes and their derivatives have been discretised. According to (19) and applying it for the current problem (35), the test function of $\widetilde{\mathcal{V}}_h$ space is constructed as follows

$$\tilde{v}_h = \tilde{v}_h(\tau, \mathcal{F}_u[u_h] \cdot \nabla v_h).$$

Thereby, due to the flux Jacobian dependence, the function $\tilde{v}_h$ is nonlinear with respect to the solution u_h, therefore a $\tilde{v}'_{u_h}$ term appears in (45) given by

$$\tilde{v}'_{u_h}(\tau, \mathcal{F}_{uu}[u_h] \cdot \nabla v),$$

where $\mathcal{F}_{uu}[u_h]$ is a Hessian tensor describing the derivative of $\mathcal{F}_u$ with respect to the conservative quantity. Besides, an approximation on τ has been applied as we ignore its dependence on u_h so τ is considered as a constant.

Discrete adjoint problem A linearised compatibility condition and the numerical approximation of the mean value linearisation (43) induce the identity in Fréchet operator form as

$$\mathcal{N}'[u](w, \tilde{v}) = \mathcal{N}'^*[u](\tilde{v}, w),$$

So, given a differentiable target functional $\mathcal{J}(\cdot)$ and its linearization $\mathcal{J}'[u](\cdot)$, the discrete adjoint problem is given by: find $\tilde{z}_h \in \widetilde{\mathcal{V}}_h$ such that

$$\mathcal{N}_h'[u_h](w_h, \tilde{z}_h) = \mathcal{J}_h'[u_h](w_h) \quad \forall w_h \in \mathcal{V}_h. \tag{46}$$

Therefore, the adjoint problem (40) must be solved numerically and since the formal Fréchet derivative might not exist, it is replaced by a suitable approximation, R. Hartmann (2002), in order to ensure the well-posedness of the adjoint problem; i.e. we replace $\mathcal{N}'[w](\cdot, \cdot)$ by $\mathcal{N}_h'[w](\cdot, \cdot)$ and thus, we define the approximate adjoint problem as: find $z \in \widetilde{\mathcal{V}}$ such that

$$\overline{\mathcal{N}}_h(u, u_h; w, z) = \overline{\mathcal{J}}(w) \qquad \forall w \in \mathcal{V}. \tag{47}$$

However, the estimation is still not computable because of the unknown analytical solutions, u and z. Thus, in order to make these error estimates computable, both u and z

must be replaced by suitable approximations which do not affect negatively the quality of the error bound.
Concerning u, the proof of the identity (30) implies the dependence on the exact primal solution through the mean-value linearisation of the semilinear form $\mathcal{N}(\cdot,\cdot)$ and the non-linear target functional $\mathcal{J}(\cdot)$. In order to overcome this dependence, we will approximate these linearisations simply at the numerical solution u_h rather than at the convex combination of u and u_h; i.e. let us use $\overline{\mathcal{N}}(u_h, u_h\cdot,\cdot)$ and $\overline{\mathcal{J}}(u_h, u_h;\cdot)$ instead of $\overline{\mathcal{N}}(u, u_h;\cdot,\cdot)$ and $\overline{\mathcal{J}}(u, u_h;\cdot)$, see Becker and Rannacher (2001).

6.3 Consistency and adjoint consistency analysis

Similar to Hartmann (2008), let us derive the corresponding adjoint problem for steady hyperbolic problems in conservation law form. To do that, we multiply both sides of (35) by z, integrate by parts and linearise around u

$$\big(\nabla\cdot(\mathcal{F}_u[u]w), z\big)_\Omega = -\big(\mathcal{F}_u[u]w, \nabla z\big)_\Omega + \big(\boldsymbol{n}\cdot\mathcal{F}_u[u]w, z\big)_\Gamma \quad \forall z\in\widetilde{\mathcal{V}},$$

Thereby, according to (36) and some algebra identity, we find the following compatibility condition

$$\begin{aligned}\big(\nabla\cdot(\mathcal{F}_u[u]w), z\big)_\Omega + \big(-I^-(\boldsymbol{n}\cdot\mathcal{F}_u[u]w), z\big)_\Gamma = \\ \big(w, \nabla z\,(-\mathcal{F}_u[u])\big)_\Omega + \big(w, z\,I^+(\boldsymbol{n}\cdot\mathcal{F}_u[u])\big)_\Gamma \quad \forall z\in\widetilde{\mathcal{V}}.\end{aligned}$$

Following the same approach as of the linear case, we define each entry for the linearised primal problem

$$\begin{aligned} N'[u]w &= \nabla\cdot(\mathcal{F}_u[u]w), && \text{in} \quad \Omega \\ B'[u]w &= -I^-(\boldsymbol{n}\cdot\mathcal{F}_u[u]), \qquad C'[u]w = w, && \text{on} \quad \Gamma \end{aligned}$$

while the corresponding adjoint operators are

$$\begin{aligned} N'[u]^*z &= \nabla z\,(-\mathcal{F}_u[u]), && \text{in} \quad \Omega \\ B'[u]^*z &= z\,I^+(\boldsymbol{n}\cdot\mathcal{F}_u[u]), \qquad C'[u]^*z = z, && \text{on} \quad \Gamma \end{aligned}$$

The continuum adjoint problem is then defined as follows

$$\nabla z\,(-\mathcal{F}_u[u]) = j'_\Omega[u] \quad \text{in} \quad \Omega, \qquad z\,I^+(\boldsymbol{n}\cdot\mathcal{F}_u[u]) = j'_\Gamma[Cu]C'[u] \quad \text{on} \quad \Gamma, \tag{48}$$

where $j'_\Omega[u]$ and $j'_\Gamma[Cu]C'[u]$ are the linearised weight functions of a target functional $\mathcal{J}(u)$.
Therefore, by writing the discrete problem (41) in term of residuals as follows

$$\int_\Omega \tilde{v}_h\,R(u_h)\,d\boldsymbol{x} + \int_\Gamma \tilde{v}_h^+\,r(u_h)\,ds = 0 \qquad \forall\tilde{v}_h\in\widetilde{\mathcal{V}}_h,$$

where $R(u_h)$ and $r(u_h)$ are the volume and boundary residual, respectively, given by

$$\begin{aligned} R(u_h) &= -\nabla\cdot\mathcal{F}(u_h) && \text{in} \quad \Omega, \\ r(u_h) &= -I^-\big[\mathcal{F}(u_\Gamma(u_h))\cdot\boldsymbol{n} - \mathcal{F}(u_h)\cdot\boldsymbol{n}\big] && \text{on} \quad \Gamma, \end{aligned}$$

as long as $u_\Gamma(u) = u$, it is easy to show that the discrete problem automatically verifies to be consistent with respect to the equations (35) and its boundary conditions (36).
As seen in §3.4, the numerical discretisation is also said to be *adjoint consistent* if the discrete problem is a consistent discretisation of the continuum adjoint problem. Now we rewrite the discrete adjoint problem (46) in the adjoint residual form by integrating by parts the volume integral and we obtain: find $\tilde{z}_h \in \widetilde{\mathcal{V}}_h$ such that

$$\int_\Omega w_h R^*[u_h](\tilde{z}_h) d\boldsymbol{x} + \int_\Gamma w_h r^*[u_h](\tilde{z}_h) ds = 0,$$

for all $w_h \in \mathcal{V}_h$, where by algebraic identities

$$\begin{aligned} R^*[u_h](\tilde{z}_h) &= j'_\Omega[u_h] + \nabla \tilde{z}_h \left(\mathcal{F}_u[u_h]\right) - \tilde{z}'_{u_h}(\nabla \cdot \mathcal{F}_h(u_h)) && \text{in} \quad \Omega, \\ r^*[u_h](\tilde{z}_h) &= j'_\Gamma[Cu_h]C'[u_h] - \tilde{z}_h \left(\mathcal{F}_u[u_h] \cdot \boldsymbol{n} + \mathcal{H}'_{u^+} + \mathcal{H}'_{u^-} u'_\Gamma[u_h]\right) && \text{on} \quad \Gamma. \end{aligned}$$

The product of the numerical derivative of the adjoint solution $\tilde{\boldsymbol{z}}'_{\boldsymbol{u}_h}$ and the flux divergence $\nabla \cdot \mathcal{F}(\boldsymbol{u}_h)$ disappears once we replace the exact solutions in the discrete operator. So the numerical adjoint residual vanishes $R^*[u_h](z) = 0$ and the adjoint consistency holds for the inner part. On the other hand, according to (38) and reminding $I = I^+ + I^-$, the discrete adjoint boundary condition gives

$$\tilde{z}_h \left(I^+ \mathcal{F}_u[u_h] \cdot \boldsymbol{n} + \mathcal{H}'_{u^-}\, u'_\Gamma[u_h]\right) = j'_\Gamma[Cu_h]C'[u_h]. \tag{49}$$

According to (38), in order to incorporate boundary conditions in (44), $\mathcal{H}_h$ depends on $u_\Gamma(u_h)$ thereby $\mathcal{H}'_{u^-} \neq 0$. Therefore we require $u'_\Gamma[u_h] \equiv 0$ on Γ where $j_\Gamma \neq 0$, as otherwise the left hand side of (49) consists of two non-null terms which differs from the continuous adjoint boundary condition (48). Unfortunately, for typical target quantities this condition does not hold naturally, see D'Angelo (2014) and Hartmann (2008) and we must use a modification of the target functional $\widehat{\mathcal{J}}(\cdot)$ to recover the adjoint consistent discretisation, as follows

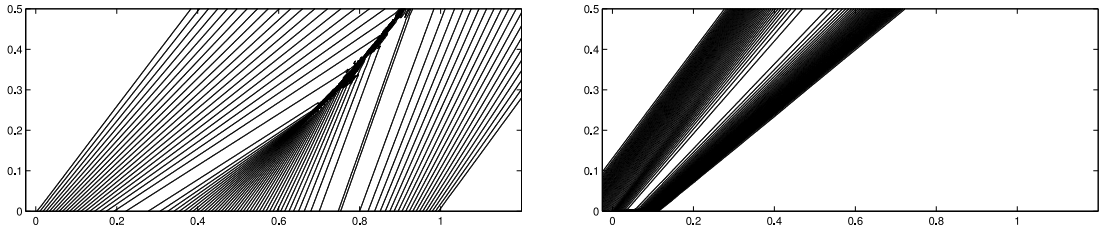
$$\widehat{\mathcal{J}}(u_h) = \mathcal{J}(i(u_h)) + \int_\Gamma r_{\mathcal{J}}(u_h)\, ds, \tag{50}$$

where $i(\cdot)$ and $r_{\mathcal{J}}(\cdot)$ must be specified and where $\widehat{\mathcal{J}}(u) := \mathcal{J}(u)$. Thereby, to be consistent $i(u) = u$ and $r_{\mathcal{J}}(u) = 0$. So, this change does not modify the exact value of the target but the discrete target $\widehat{\mathcal{J}}(u_h)$ will be different and moreover, for nonlinear cases, also $\widehat{\mathcal{J}}'[u_h]$ will differ from $\mathcal{J}'[u_h]$. Thus, wisely defining these two new terms, we are able to recover an adjoint consistent discretisation.

6.4 Numerical example

Let now make a numerical comparison first among the current schemes and then between the two refinement algorithms. Thereby, we follow the problem proposed by Ricchiuto (2005), a simple nonlinear scalar hyperbolic problem with an exponential flux $\mathcal{F}(u) = (e^{au}, u)$, such as

$$a\, e^{au} \partial_x u + \partial_t u = 0,$$



a: b:

Figure 9: Exponential flux problem. (a) Reference primal and (b) adjoint solution.

with $a = 0.75$ and where the unsteady dimension, t, is considered as a second coordinate of a stationary 2D problem on the space-time plane (x, t) with a domain $\Omega = [-0.025, 1.2] \times [0, 0.5]$. Therefore, the boundary conditions are given by

$$\begin{aligned} u(x,0) &= \begin{cases} \sin(2\pi x) & \text{for} \quad 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \\ u(-0.025,t) &= 0 \end{aligned}$$

Figure a shows the level contours of a reference solution over the domain. The decreasing slope of the boundary wave generates a convergent fan collapsing in a shock.
As a target quantity, we are interested in the boundary solution value between $[0.25; 0.75]$ weighted by a function $\psi(r_0, r)$, see Süli and Houston (2002), i.e. a target functional given by

$$\mathcal{J}(u) = \int_\Gamma \psi(0.25, |x - 0.5|)\, u(x, 0.5)\, dx,$$

The reference value computed is $\mathcal{J}(u) = 0.079289840610707$. Figure b plots the current adjoint solution coming back that from the outlet boundary towards the inlet of the domain. The present shock does not affect the adjoint solution that keeps smooth over the whole domain.

Let then run the three numerical schemes over the domain by applying the adjoint refinement procedure. For this computation, we use five nested meshes and we consider $u_h \in \mathcal{V}_h^1$, $\tilde{z}_h \in \mathcal{V}_h^1$ and $\bar{z}_h \in \mathcal{V}_h^2$, i.e. we opt for the richer space solution strategy. Table 10, 11 and 12 resume the corresponding results for RD-LDA, SUPG and BUBBLE scheme, respectively. The third column lists the numerical error estimate based on (30) while column four shows the sum of the error indices based on the localisation (33). The corresponding index θ_{eff} stands by the effectivity index between the present error estimate and the real error on the second column. Hence all the three schems seem to well estimate the real error with effectivity indices strictly close to unity. The best estimate comes from the RD-LDA scheme even if the SUPG achieve smaller error at the same conditions. The BUBBLE scheme behaves in between the other two. The localisation process does not affect the estimate and in fact the corresponding index keeps small and bounded.
Finally, using the RD-LDA scheme, we provide a comparison between the adjoint and residual based refinement. Figure 10 shows the corresponding final meshes, using a pointwise fixed fraction strategy (see D'Angelo (2014)) with 5% of flag refinement fraction. The residual based procedure focus the refinement just along the shock and discards the rest of

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
0.1000	1	$2.573\,10^{-3}$	$2.588\,10^{-3}$	(1.01)	$1.325\,10^{-2}$	(5.15)
0.0500	1	$1.518\,10^{-3}$	$1.775\,10^{-3}$	(1.17)	$3.057\,10^{-3}$	(2.01)
0.0250	1	$3.983\,10^{-4}$	$3.925\,10^{-4}$	(0.99)	$6.422\,10^{-4}$	(1.61)
0.0125	1	$1.096\,10^{-4}$	$1.087\,10^{-4}$	(0.99)	$1.765\,10^{-4}$	(1.61)
0.1000	2	$2.482\,10^{-4}$	$2.153\,10^{-3}$	(8.68)	$3.063\,10^{-3}$	(12.3)
0.0500	2	$4.597\,10^{-5}$	$9.634\,10^{-5}$	(2.10)	$2.473\,10^{-4}$	(5.38)
0.0250	2	$8.972\,10^{-6}$	$8.684\,10^{-6}$	(0.97)	$4.227\,10^{-5}$	(4.71)
0.0125	2	$2.126\,10^{-6}$	$3.938\,10^{-6}$	(1.85)	$1.285\,10^{-5}$	(6.04)

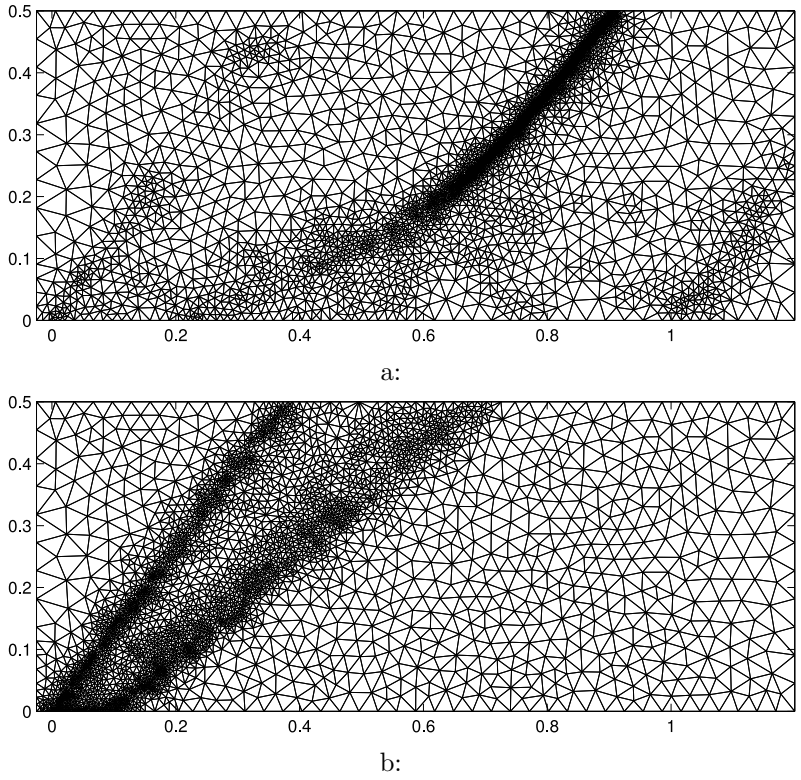
Table 10: Exponential flux problem. RD-LDA scheme, effectivity of adjoint-based *a posteriori* error estimation.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
0.1000	1	$1.066\,10^{-2}$	$1.023\,10^{-2}$	(0.96)	$1.778\,10^{-2}$	(1.67)
0.0500	1	$8.518\,10^{-4}$	$8.544\,10^{-4}$	(1.00)	$3.225\,10^{-3}$	(3.79)
0.0250	1	$1.399\,10^{-4}$	$1.538\,10^{-4}$	(1.10)	$3.926\,10^{-4}$	(2.81)
0.0125	1	$1.389\,10^{-5}$	$1.674\,10^{-5}$	(1.21)	$6.250\,10^{-5}$	(4.50)
0.1000	2	$2.596\,10^{-4}$	$3.830\,10^{-4}$	(1.48)	$1.117\,10^{-3}$	(4.30)
0.0500	2	$6.791\,10^{-6}$	$6.275\,10^{-6}$	(0.92)	$6.421\,10^{-5}$	(9.46)
0.0250	2	$3.401\,10^{-6}$	$2.338\,10^{-6}$	(0.69)	$3.973\,10^{-6}$	(1.17)
0.0125	2	$6.371\,10^{-7}$	$3.627\,10^{-7}$	(0.57)	$4.708\,10^{-7}$	(0.74)

Table 11: Exponential flux problem. SUPG scheme, effectivity of adjoint-based *a posteriori* error estimation.

h	p	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\mathcal{R}_\Omega\rvert$	(θ_{eff_1})	$\mathcal{R}_{\lvert\Omega\rvert}$	(θ_{eff_2})
0.1000	1	$8.518\,10^{-3}$	$8.408\,10^{-3}$	(0.99)	$1.688\,10^{-2}$	(1.98)
0.0500	1	$9.066\,10^{-4}$	$8.873\,10^{-4}$	(0.98)	$3.071\,10^{-3}$	(3.39)
0.0250	1	$1.508\,10^{-4}$	$1.659\,10^{-4}$	(1.10)	$3.745\,10^{-4}$	(2.48)
0.0125	1	$2.178\,10^{-5}$	$2.418\,10^{-5}$	(1.11)	$6.946\,10^{-5}$	(3.19)
0.1000	2	$8.710\,10^{-4}$	$9.012\,10^{-4}$	(1.03)	$1.949\,10^{-3}$	(2.24)
0.0500	2	$1.484\,10^{-4}$	$1.342\,10^{-4}$	(0.90)	$2.706\,10^{-4}$	(1.82)
0.0250	2	$3.764\,10^{-5}$	$3.757\,10^{-5}$	(1.00)	$5.969\,10^{-5}$	(1.59)
0.0125	2	$9.576\,10^{-6}$	$1.045\,10^{-5}$	(1.09)	$1.691\,10^{-5}$	(1.77)

Table 12: Exponential flux problem. BUBBLE scheme, effectivity of adjoint-based *a posteriori* error estimation.



a:

b:

Figure 10: Exponential flux problem. RD-LDA scheme, final grids obtained for (a) residual-based adaptation with 28874 triangles and $\mathcal{J}(e) = 3.546\,10^{-4}$ and (b) goal oriented adjoint-based with 7081 triangles and $\mathcal{J}(e) = 2.448\,10^{-5}$.

the domain while the adjoint based strategy refines along the characteristics that bound the adjoint solution. A simple comparison between the corresponding errors highlights the efficiency of the adjoint refinement with respect to the residual based. In fact, for the former $\mathcal{J}(e) = 2.448 \cdot 10^{-5}$ with 7081 elements while for the latter $\mathcal{J}(e) = 3.546 \cdot 10^{-4}$ with 28874 elements, i.e. an error ten times smaller with one quarter of triangles.

7 Euler equations

As it is well known, the Euler system is an archetype of this class of nonlinear partial differential equations. Indeed, the compressible Euler equations are a homogeneous nonlinear hyperbolic system of conservative equations for mass, momentum and energy, describing inviscid compressible flow. Velocity and pressure of the fluid appear as secondary variables by the so-called *equation of state* and the characteristic controlling dimensionless parameter is the Mach number $M = |\boldsymbol{v}|/c$ with $\boldsymbol{v}$ the typical velocity and c the typical speed of sound. The Euler equations for a perfect gas are of interest for a number of reasons. In fact, even though a solution of the Euler equations is only an approximation to a real fluids problem, for some problems especially external aerodynamic flows over streamlined bodies, it provides a good model of reality and makes this problem an interesting test for the numerical analysis.

The stationary case of the 2D compressible Euler problem is then given by

$$\nabla \cdot \mathcal{F}(\boldsymbol{u}) = 0 \quad \text{in} \quad \Omega,$$

where in a two-dimensional space, the flux vector $\mathcal{F}(\boldsymbol{u}) = (f_1(\boldsymbol{u}), f_2(\boldsymbol{u}))^T$ and the state vector, $\boldsymbol{u}$, in conservative variables, are defined as

$$\boldsymbol{u} = \begin{bmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ \rho E \end{bmatrix}, \quad f_1(\boldsymbol{u}) = \begin{bmatrix} \rho v_1 \\ \rho v_1^2 + p \\ \rho v_1 v_2 \\ \rho H v_1 \end{bmatrix} \quad \text{and} \quad f_2(\boldsymbol{u}) = \begin{bmatrix} \rho v_2 \\ \rho v_1 v_2 \\ \rho v_2^2 + p \\ \rho H v_2 \end{bmatrix},$$

with ρ the density of the fluid, $\boldsymbol{v} = (v_1, v_2)$ the flow speed and E the total energy per unit mass, and where H, the specific total enthalpy, is given by

$$H = E + \frac{p}{\rho} = e + \frac{1}{2}v^2 + \frac{p}{\rho},$$

with $v^2 = v_1^2 + v_2^2$ and where the pressure p is determined by the state equation of an ideal gas as

$$p = (\gamma - 1)\rho e,$$

with e the specific internal energy and $\gamma = c_p/c_v$ the ratio of specific heat capacities at constant pressure, c_p, and constant volume, c_v. Thereby, let us define the Jacobian matrix in the direction $\boldsymbol{n}$ as

$$B(\boldsymbol{u}, \boldsymbol{n}) = \sum_{i=1}^{d} n_i A_i(\boldsymbol{u}) = \partial_{\boldsymbol{u}}\big(\mathcal{F}(\boldsymbol{u}) \cdot \boldsymbol{n}\big), \tag{51}$$

with $A_i(\boldsymbol{u}) = \frac{\partial f_i(\boldsymbol{u})}{\partial \boldsymbol{u}}$ in conservative variables given by

$$A_1(\boldsymbol{u}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -v_1^2 + \frac{1}{2}(\gamma-1)v^2 & (3-\gamma)v_1 & -(\gamma-1)v_2 & \gamma - 1 \\ -v_1 v_2 & v_2 & v_1 & 0 \\ v_1(\frac{1}{2}(\gamma-1)v^2 - H) & H - (\gamma-1)v_1^2 & -(\gamma-1)v_1 v_2 & \gamma v_1 \end{pmatrix},$$

$$A_2(\boldsymbol{u}) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ -v_1 v_2 & v_2 & v_1 & 0 \\ -v_2^2 + \frac{1}{2}(\gamma-1)v^2 & -(\gamma-1)v_1 & (3-\gamma)v_2 & \gamma - 1 \\ v_2(\frac{1}{2}(\gamma-1)v^2 - H) & -(\gamma-1)v_1 v_2 & H - (\gamma-1)v_2^2 & \gamma v_2 \end{pmatrix}.$$

and the corresponding eigenvalues of $B(\boldsymbol{u}, \boldsymbol{n})$ defined by

$$\lambda_1 = \boldsymbol{v} \cdot \boldsymbol{n} - c, \qquad \lambda_2 = \lambda_3 = \boldsymbol{v} \cdot \boldsymbol{n}, \qquad \lambda_4 = \boldsymbol{v} \cdot \boldsymbol{n} + c, \tag{52}$$

with $c = \sqrt{\gamma p/\rho}$ the speed of sound.

7.1 Boundary conditions

According to (37), for elements with faces along the boundary, $\partial\kappa \cap \Gamma$, a boundary contribution appears on the flux $\mathcal{H}_h$ through a boundary function $\boldsymbol{u}_\Gamma(\boldsymbol{u})$, which depends on the flow field solution. This function is defined differently according to the type of boundary. In hyperbolic problems, this depends on the number of characteristics entering the domain each one imposing a physical quantity. This number is equal to the number of negative eigenvalues of the Jacobian matrix (51) in the direction of the outward normal. So, depending on the sign of the local eigenvalues λ_i, $i = 1, \ldots, 4$ defined in (52), we can distinguish four different *flowfield boundary conditions*:

Supersonic inflow : when $\lambda_i < 0$, $i = 1, \ldots, 4$ and corresponding to imposing a Dirichlet boundary condition for all flow variables, we prescribe

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \boldsymbol{g}.$$

Supersonic outflow : when $\lambda_i > 0$, $i = 1, \ldots, 4$ and corresponding to imposing on the boundary the solution fully taken from the flow field, with

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \boldsymbol{u}.$$

Subsonic inflow : when $\lambda_i < 0$, $i = 1, \ldots, 3$, $\lambda_4 > 0$. In this case the pressure is taken from the flow field while the other variables are imposed by the inlet state, $\boldsymbol{u}_\infty$

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \left(u_{1,\infty}, u_{2,\infty}, u_{3,\infty}, \frac{p(\boldsymbol{u})}{\gamma - 1} + \frac{u_{2,\infty}^2 + u_{3,\infty}^2}{2u_{1,\infty}}\right)^T.$$

Subsonic outflow : when $\lambda_1 < 0$, $\lambda_i > 0$, $i = 2, \ldots, 4$. In this case the pressure is based on the outflow state and the other variables are imposed by the local flow field solution

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \left(u_1, u_2, u_3, \frac{p(\boldsymbol{u}_\infty)}{\gamma - 1} + \frac{u_2^2 + u_3^2}{2u_1}\right)^T.$$

Finally, let us also define

Farfield (or freestream) : corresponds to imposing Dirichlet condition based on freestream conditions for all characteristics, no matter if they are ingoing or outgoing.

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \boldsymbol{u}_\infty.$$

and the *wall boundary* condition. For the latter, we remember that because Euler equations neglect the viscous effects, the flow is allowed to slip at a solid surface. However, any flow penetration is forbidden which leads to ensure the condition $\boldsymbol{v} \cdot \boldsymbol{n} = 0$, i.e. suppressing the wall normal component. So on the wall boundary we impose

Slip wall : the goal boundary state, $\boldsymbol{u}_\Gamma$, is computed as the difference of the current state and the wall normal velocity, $\boldsymbol{v_n} = (\boldsymbol{v} \cdot \boldsymbol{n})\boldsymbol{n}$. Besides, the energy contribution

has to be changed to keep constant only the internal energy while modifying the kinetic contribution consistent with the velocity moidification. Therefore,

$$\rho_\Gamma = \rho, \quad \boldsymbol{v}_\Gamma = \boldsymbol{v} - \boldsymbol{v_n} \qquad \text{and} \qquad \rho E_\Gamma = \rho E - \frac{1}{2}\rho \boldsymbol{v}^2 + \frac{1}{2}\rho \boldsymbol{v}_\Gamma^2,$$

so,

$$\boldsymbol{u}_\Gamma(\boldsymbol{u}) = \Big(u_{\Gamma_1}, u_{\Gamma_2}, u_{\Gamma_3}, u_{\Gamma_4}\Big),$$

with

$$\begin{aligned} u_{\Gamma_1} &= u_1 \\ u_{\Gamma_2} &= (1 - n_1^2)u_2 + (-n_1 n_2)u_3 \\ u_{\Gamma_3} &= (-n_1 n_2)u_2 + (1 - n_2^2)u_3 \\ u_{\Gamma_3} &= u_4 - \frac{u_2^2 + u_3^2}{2u_1} + \frac{u_{\Gamma_2}^2 + u_{\Gamma_3}^2}{2u_1}. \end{aligned}$$

7.2 Shock capturing

When the analytical solution presents sharp features (as maybe the case for the Euler system) and the numerical discretisation implies a non-positive scheme (der Weide (1998), Paillere (1995) and Ricchiuto (2005)), spurious oscillations may arise in this numerical solution, especially when high order discretisations are used. So the numerical discretisation (37) must be enhanced by the addition of some form of nonlinear dissipation mechanism which does not affect harmfully the formal order of accuracy of the scheme and the consistency of the discrete problem.

We emphasise that the present *a posteriori* error analysis is based on the Galerkin orthogonality property of the finite element method, see (30). For this reason, any stabilization technique used to enhance the numerical discretisation performances should not violate this property and therefore cures bases on local projections or slope limiters Cockburn and Shu (1998) must be discarded. A consistent stabilization approach is instead the use of an artificial viscosity term, which depends on both the mesh size h and on the local discrete residual, $\nabla \cdot \mathcal{F}(\boldsymbol{u}_h)$. So, the semi-linear form of (37) is augmented as follows by adding a Laplace type diffusion term

$$\begin{aligned} \mathcal{N}(\boldsymbol{u}, \tilde{\boldsymbol{v}}) = \int_\Omega \nabla \cdot \mathcal{F}(\boldsymbol{u})\, \tilde{\boldsymbol{v}}\, d\boldsymbol{x} + \int_\Gamma \mathcal{H}(\boldsymbol{u}, \boldsymbol{u}_\Gamma(\boldsymbol{u}), \boldsymbol{n})\, \boldsymbol{v}\, ds \\ + \int_\Omega \varepsilon(\boldsymbol{u}) \nabla \boldsymbol{u} \cdot \nabla \boldsymbol{v}\, d\boldsymbol{x}, \end{aligned} \tag{53}$$

with

$$\varepsilon(\boldsymbol{u}) = C_\epsilon h^{2-\beta} \big| \mathcal{F}'_{\boldsymbol{u}}[\boldsymbol{u}] \cdot \nabla \boldsymbol{u} \big| \geq 0, \tag{54}$$

and where we applied a conservative linearisation of the divergence of the flux. Here C_ϵ is a positive constant and $0 < \beta < 1/2$, (R. Hartmann (2002), Jaffre et al. (1995)). Based on (54) and (53), the corresponding term in the Jacobian of the nonlinear operator $\mathcal{N}'_h$ appears as follows

$$\int_\Omega \big(\varepsilon(\boldsymbol{u}) \nabla \boldsymbol{w} + \varepsilon'[\boldsymbol{u}](\boldsymbol{w}) \nabla \boldsymbol{u} \big) \cdot \nabla \boldsymbol{v}\, d\boldsymbol{x},$$

with

$$\varepsilon'[\boldsymbol{u}](\boldsymbol{w}) = C_\epsilon h^{2-\beta}\mathrm{sgn}(\varepsilon(\boldsymbol{u}))\big(\mathcal{F}'_{\boldsymbol{u}}(\boldsymbol{u})\cdot\nabla\boldsymbol{w} + \mathcal{F}'_{\boldsymbol{uu}}(\boldsymbol{u})\boldsymbol{w}\nabla\boldsymbol{u}\big),$$

coming from the Fréchet derivative of the functional $\boldsymbol{u} \to \varepsilon(\boldsymbol{u})$.

7.3 Numerical examples

In the following, we finally consider some different examples where we apply the compressible Euler equations and which include smooth solutions and solutions with shocks. The primal solution will be approximated by first order polynomials, i.e. $\boldsymbol{u}_h \in \boldsymbol{\mathcal{V}}_h^1$, the lower adjoint solution is computed as $\tilde{\boldsymbol{z}}_h \in \widetilde{\boldsymbol{\mathcal{V}}}_h^1$ while the higher solution $\bar{\boldsymbol{z}}_h$ is sought from $\widetilde{\boldsymbol{\mathcal{V}}}_h^2$. The nonlinear residual is reduced over 6 orders of magnitude on each mesh, in order to be sure that the resulting primal solutions are enough converged and that the iterative solver error contributions are negligible compared to the discrete approximation. Finally, the refinement strategy used for all the computations is the pointwise fixed fraction strategy with 5% of flag refinement fraction and no derefinement, combined with a remeshing algorithm by the MMG mesh generator.

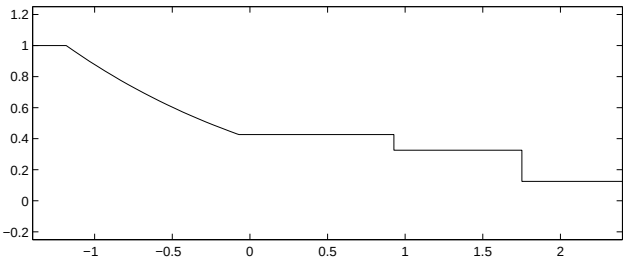
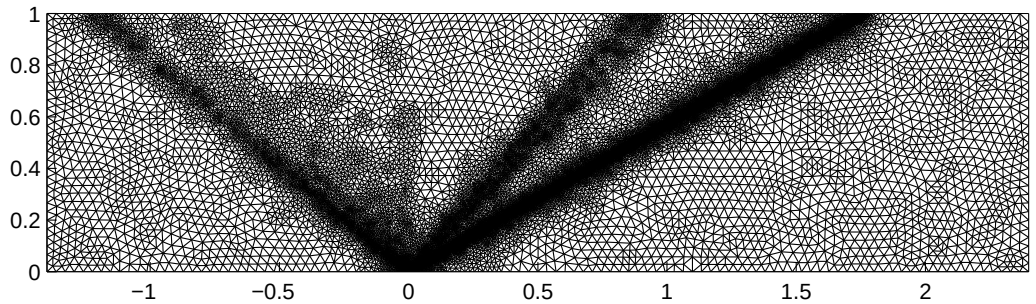


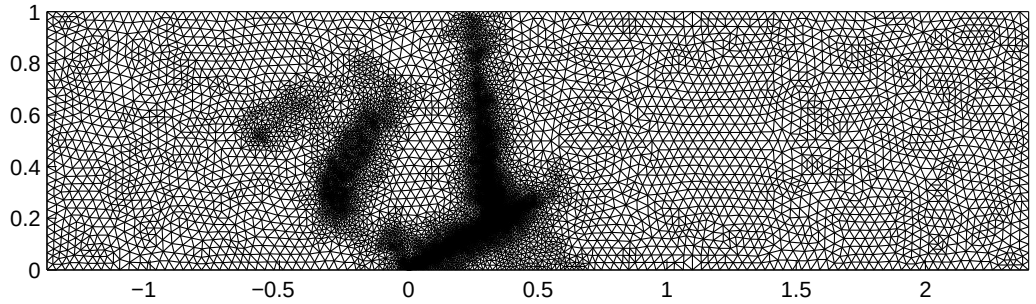
Figure 11: Unsteady 1D Euler problem. Final solution $\rho(x, 1)$.

7.3.1 Unsteady 1D Euler problem

Let start by the one dimensional time-dependent Euler equations. This unsteady problem is seen as a steady two-dimensional case where the coordinates are $x_1 \equiv x$ and $x_2 \equiv t$. The state vector in conservative variables is given by $\boldsymbol{u} = (\rho, \rho v, \rho E)^T = (u_1, u_2, u_3)^T$, where ρ, v and E stand for the density, the velocity and the specific total energy, respectively. Then the governing equations can be found i.e. in D'Angelo (2014).
The boundary condition at $t = 0$ imposes a discontinuity at $x = 0$, while on the left and right, freestream boundary conditions are imposed, and at the top boundary no condition is allowed. The current problem is also called *Sod's problem* and its analytical solution can be found to R. Hartmann (2002). Then we apply on this problem an adaptive meshing procedure for a given functional, by assuming to be interested in the value of the density on the upper boundary $t = 1$, at $x = 0.25$, which is located in the area of the constant intermediate state between the rarefaction tail and the contact discontinuity. Figure 11 plots the cross section of the final solution along the outflow boundary located at $t = 1$.



a:



b:

Figure 12: Unsteady 1D Euler problem. SUPG scheme, (a) residual-based adaptation with 129106 triangles and $|\mathcal{J} - \mathcal{J}_h| = 2.538\,10^{-5}$ and (b) goal oriented adjoint-based with 96915 triangles and $|\mathcal{J} - \mathcal{J}_h| = 3.801\,10^{-6}$.

Hence, the target functional is given by

$$\mathcal{J}(\boldsymbol{u}) = \rho(0.25, 1.0).$$

Since discontinuities appear for this test case, SUPG has been prefered as numerical scheme because it can garantee better results of convergence and accuracy for the primal solution. Figure 12 compares the final adaptive meshes when using residual and adjoint-based indicators, respectively. Here, the two final meshes appear again completely different. The residual indicators drive the refinement along the two jumps in the solution, shock and contact discontinuity and partially over the head and the tail of the rarefaction wave where discontinuities in the solution gradients occur. On the contrary, the adjoint refinement follows the characteristics backward from the target point until the crossing with the three main features of the primal solution. Furthermore, comparing the corresponding target errors, $|\mathcal{J}(\boldsymbol{u}) - \mathcal{J}(\boldsymbol{u}_h)|$, for the adjoint-based procedure ($|\mathcal{J} - \mathcal{J}_h| = 3.801 \cdot 10^{-6}$) is one order of magnitude smaller with almost 70% of the number of elements used on the final mesh for residual based adaptation ($|\mathcal{J} - \mathcal{J}_h| = 2.538 \cdot 10^{-5}$).
Finally, we tabulate the iterative results of this mesh adaptation in order to analyze the accuracy of error estimation, Table 13. In particular, the effectivity index of the error representation, θ_{eff_1}, keeps bounded even not strictly close to the unit value. This is due to the higher complexity of the problem with respect to a scalar equation, but mainly resulting from the fact that the adjoint solution crosses through discontinuities of the primal solution which affects the error estimation, as we deduce from the adjoint-based final mesh. In fact, a discontinuous solution u corresponds to a singularity on

#DoF	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$	(θ_{eff_1})
11085	$1.122\cdot10^{-3}$	$7.314\cdot10^{-4}$	(0.65)
14064	$7.357\cdot10^{-4}$	$3.138\cdot10^{-3}$	(4.27)
18843	$8.786\cdot10^{-5}$	$5.132\cdot10^{-4}$	(5.84)
25353	$1.162\cdot10^{-4}$	$5.884\cdot10^{-4}$	(5.06)
35277	$5.195\cdot10^{-5}$	$1.087\cdot10^{-4}$	(2.09)
49809	$1.091\cdot10^{-4}$	$1.303\cdot10^{-4}$	(1.19)
72477	$2.538\cdot10^{-5}$	$5.453\cdot10^{-5}$	(2.15)
105273	$9.167\cdot10^{-6}$	$2.444\cdot10^{-5}$	(2.67)
145926	$3.801\cdot10^{-6}$	$8.351\cdot10^{-6}$	(2.20)

Table 13: Unsteady 1D Euler problem. SUPG scheme, efficiency of adjoint-based *a posteriori* error estimation.

the discrete linearization; therefore, it is unavoidable that under these conditions, an $\mathcal{O}[1]$ error is produced in the approximated adjoint solution at that point and consequently, the estimate is also compromised. The same happens when the adjoint spike passes through a pure shock without any artificial viscosity inserted to smooth the discontinuity. So the shock capturing term is applied not only for improving the accuracy of the discrete primal solution but also for improving the error estimate.

7.3.2 Ringleb problem

The Ringleb problem, proposed for the first time by Ringleb (1940), consists of a smooth transonic flow in a channel, drawn in Figure a. The left and right boundaries are considered as walls while the bottom and upper boundary are the inlet and outlet, respectively. This problem is one of the few non-trivial examples of the 2D Euler equations where a smooth analytical solution can be deduced, see D'Angelo (2014). Hence, it becomes an interesting test to accurately prove the sharpness of the error representation for the 2D Euler equations.
We decide to solve this problem by using RD-LDA scheme and we also choose two different target quantities. In the first we test the usual pointwise target (proposed in Barth (2002)) while in the second, we focus on the horizontal force applied on the right wall. Therefore, first, we consider the internal energy value, E, at given point, then

$$\mathcal{J}(\boldsymbol{u}) = E(-0.63, 1.70).$$

The reference value is then $\mathcal{J}(\boldsymbol{u}) = 1.83439675995224$ and the present adjoint solution is a singularity driven backward by a spike from that point to the inlet boundary. Figure 13 shows the corresponding final meshes for the residual and adjoint-based refinements. Being mainly a subsonic problem, the standard residual-based algorithm does not favour any particular zone or direction and brings a uniform refinement over the whole domain. On the other hand, the adjoint-based procedure focuses along the adjoint spike by strongly reducing the local element size and weighting more the corresponding local residuals. Besides, we notice a refinement around the right bottom corner because of a supersonic inlet and wall boundary that share this corner generating a spurious singularity on the higher

adjoint solution.

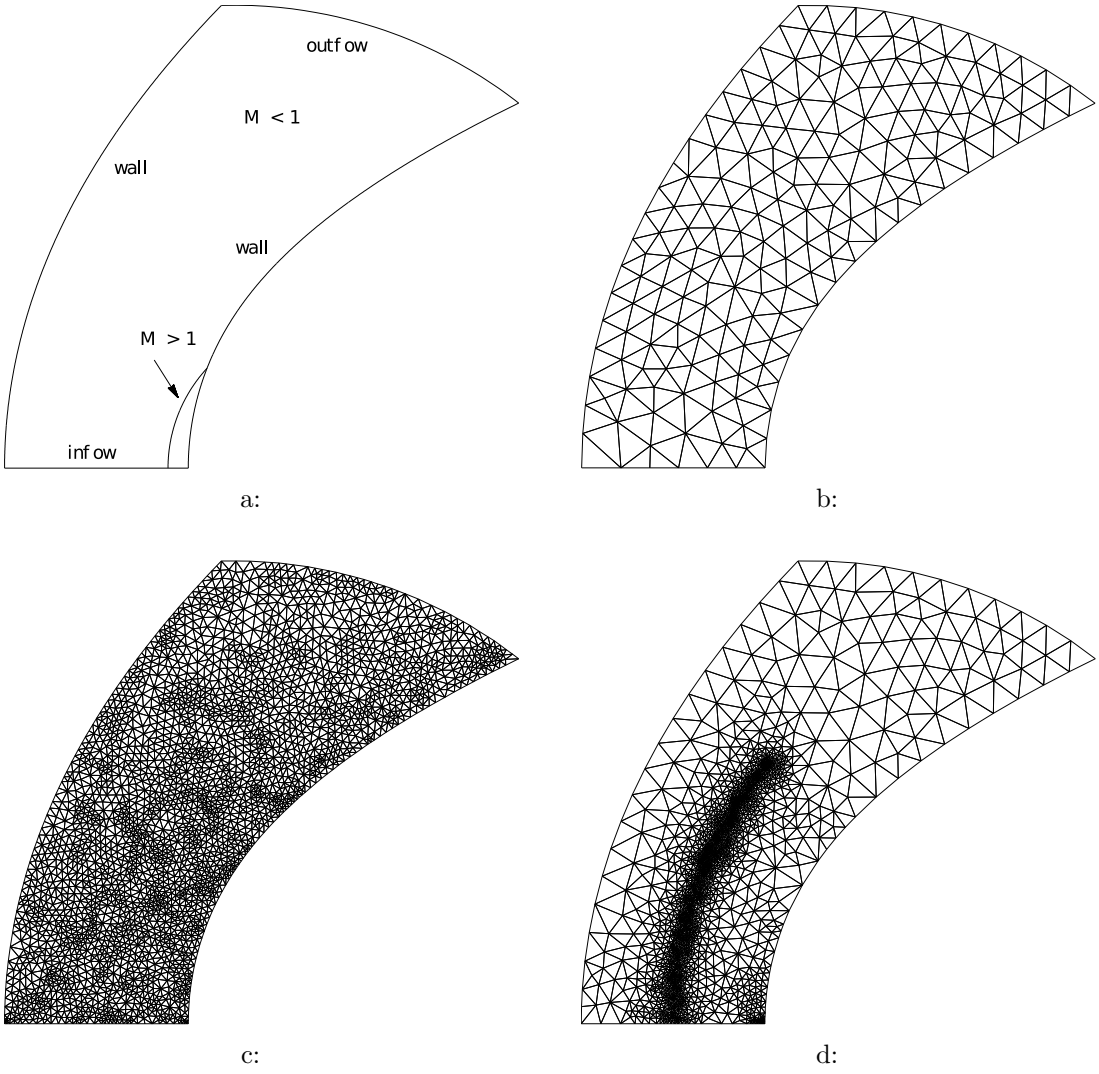


Figure 13: Ringleb problem for pointwise target: RD-LDA scheme on P1 triangles, (a) geometry of the Ringleb problem, (b) initial mesh with 317 triangles, (c) residual-based adaptation with 5829 triangles and $|\mathcal{J} - \mathcal{J}_h| = 2.284\,10^{-4}$, (d) goal oriented adjoint-based with 4069 triangles and $|\mathcal{J} - \mathcal{J}_h| = 2.277\,10^{-5}$ for point target.

This time, the related error converges slowlier and after a dozen of iterations it reduces till $|\mathcal{J} - \mathcal{J}_h| = 2.277 \cdot 10^{-5}$. Indeed, as we start from the exact solution, we reach good results already with coarse meshes and because the problem is mainly subsonic, the convergence keeps slow. Nevertheless all these issues, the residual based refinement is even slowier and it achieves an error $|\mathcal{J} - \mathcal{J}_h| = 2.284 \cdot 10^{-4}$ with more elements 5829 vs 4069 for the adjoint-based. The reason is well explained by the two final meshes. Whilst the residual algorithm refines isotropically and uniformly over the wall domain, (Figure c), the adjoint based refinement focus its attention only along the mass path that, from the inlet boundary goes towards the target point (Figure d).

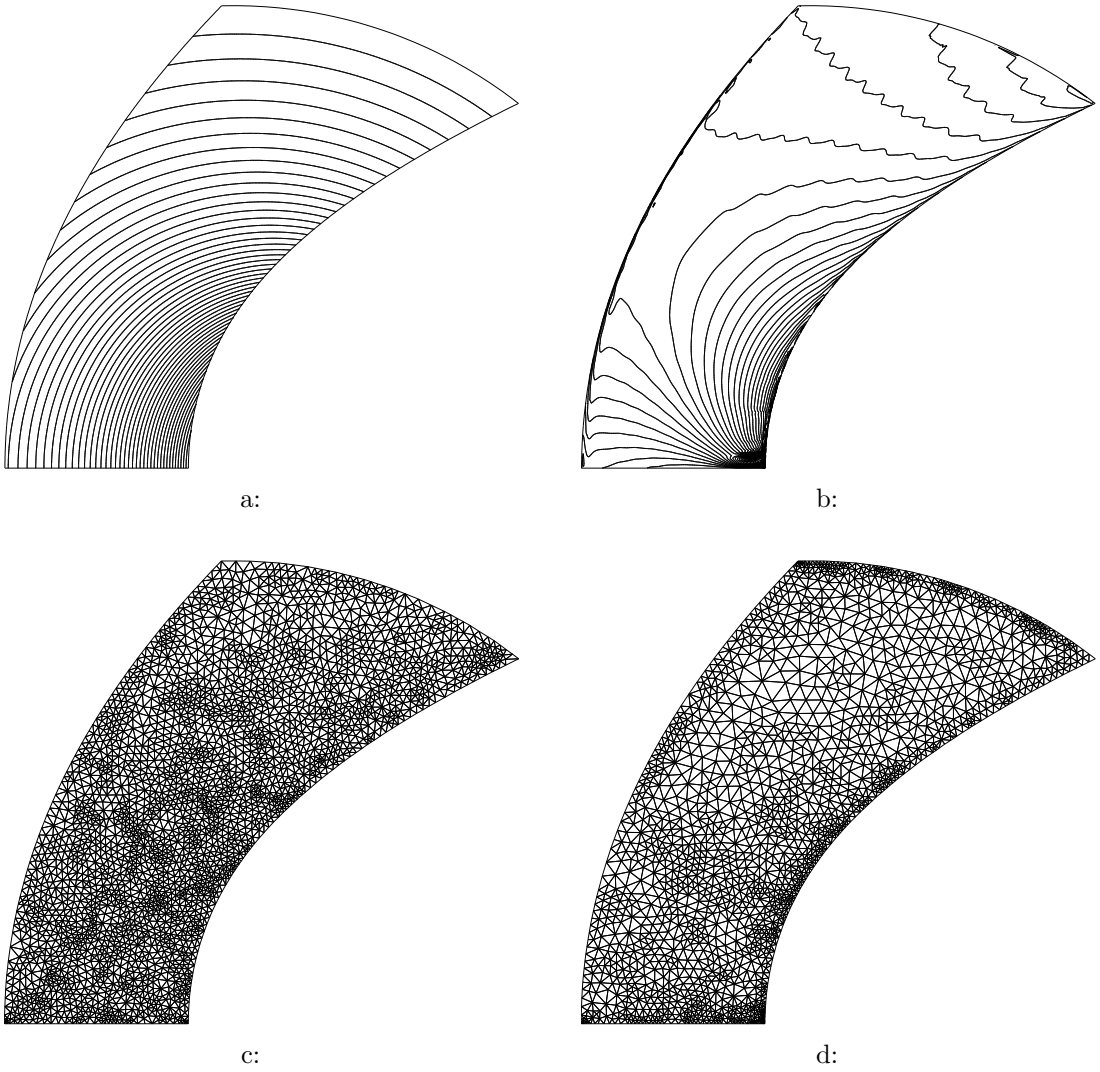


Figure 14: Ringleb problem for force target: RD-LDA scheme on P1 triangles, (a) primal solution, (b) adjoint solution, (c) residual-based adaptation with 5829 triangles and $|\mathcal{J} - \mathcal{J}_h| = 5.854\,10^{-5}$, (d) goal oriented adjoint-based with 2647 triangles and $|\mathcal{J} - \mathcal{J}_h| = 4.213\,10^{-5}$ for force target.

#DoF	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$ (θ_{eff_1})	$\overline{\mathcal{R}}_{\lvert\Omega\rvert}$ (θ_{eff_2})
744	$1.777\cdot10^{-3}$	$3.004\cdot10^{-5}$ (0.02)	$1.231\cdot10^{-2}$ (6.93)
348	$1.126\cdot10^{-3}$	$1.063\cdot10^{-3}$ (0.94)	$5.174\cdot10^{-3}$ (4.60)
1224	$6.234\cdot10^{-4}$	$8.452\cdot10^{-4}$ (1.36)	$2.935\cdot10^{-3}$ (4.71)
1540	$5.803\cdot10^{-5}$	$8.931\cdot10^{-5}$ (1.54)	$2.494\cdot10^{-3}$ (43.0)
2196	$1.859\cdot10^{-4}$	$2.280\cdot10^{-4}$ (1.23)	$1.338\cdot10^{-3}$ (7.20)
3100	$1.975\cdot10^{-4}$	$2.070\cdot10^{-4}$ (1.05)	$8.242\cdot10^{-4}$ (4.17)
4244	$1.036\cdot10^{-4}$	$1.193\cdot10^{-4}$ (1.15)	$6.000\cdot10^{-4}$ (5.79)
5752	$4.213\cdot10^{-5}$	$5.316\cdot10^{-5}$ (1.26)	$4.142\cdot10^{-4}$ (9.83)

Table 14: Ringleb force problem. RD-LDA scheme, efficiency of adjoint-based *a posteriori* error estimation.

In the second test case of the Ringleb problem, we consider a target quantity the horizontal force over the right wall of the channel. Unlike the previous case, this functional is an integral quantity defined as

$$\mathcal{J}(\boldsymbol{u}) = \int_{\Gamma_{\text{w}}} p(\boldsymbol{u}_\Gamma)\,(\boldsymbol{n}\cdot\boldsymbol{n}_h)\,ds,$$

where p is the local pressure on the wall computed by the $\boldsymbol{u}_\Gamma$ values, $\boldsymbol{n}$ the local normal of the current wall and the horizontal force normal $\boldsymbol{n}_h = (1,0)$. For this case, the exact value is $\mathcal{J}(\boldsymbol{u}) = 1.10567714227773$ while primal and adjoint densities are plotted in Figure 14. For the latter, it is important to notice the singularity arising on the right bottom corner and strong gradients along the low part of the right wall.
As we already pointed out, the standard iterative procedure is not affected by the choice of the target quantity and so, it generates the same meshes for any target considered. Hence, by the last mesh of the previous iteration, we obtain a current error equal to $\lvert\mathcal{J}-\mathcal{J}_h\rvert = 5.854\cdot10^{-5}$. On the other hand, by using the adjoint-based indicators and then the goal oriented refinement, we are able to reach a target error $\lvert\mathcal{J}-\mathcal{J}_h\rvert = 4.213\cdot10^{-5}$ with 2647 triangles, Figure d. Here, due to the adjoint sigularity, the refinement is mainly focused around the right bottom corner and only later along the right wall boundary. However, besides the comparison of the exact errors $\mathcal{J}(e)$, for this such a problem, where the target is an intergral functional and not just a pointwise one (way too sensitive to numerical perturbations), we can also focus our attention on the error estimations $\overline{\mathcal{R}}_\Omega$ and $\overline{\mathcal{R}}_{\lvert\Omega\rvert}$ and their effectivity indices. In Table 14 they are presented along the iterative refinement. There, we realize how the estimate is close to the real error (θ_{eff_1} always around the unity) and the remeshing does not overdo in the refinement (θ_{eff_2} bounded under ten).

7.3.3 Supersonic flow

We conclude this overview with a supersonic flow around the asymmetric NACA23012 profile, Figure 15. Farfield conditions are set at Mach 1.2, angle of attack $\alpha = 5°$, density and pressure are respectively given by $\rho = 1$ and $p = 1$. The SUPG scheme has been taken for the discretisation and a shock capturing term is also used by setting $C_\epsilon = 0.0125$ and $\beta = 0.2$. The corresponding solution for this problem envisages a bow shock in front

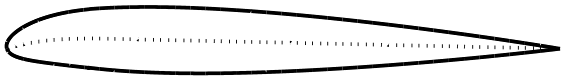
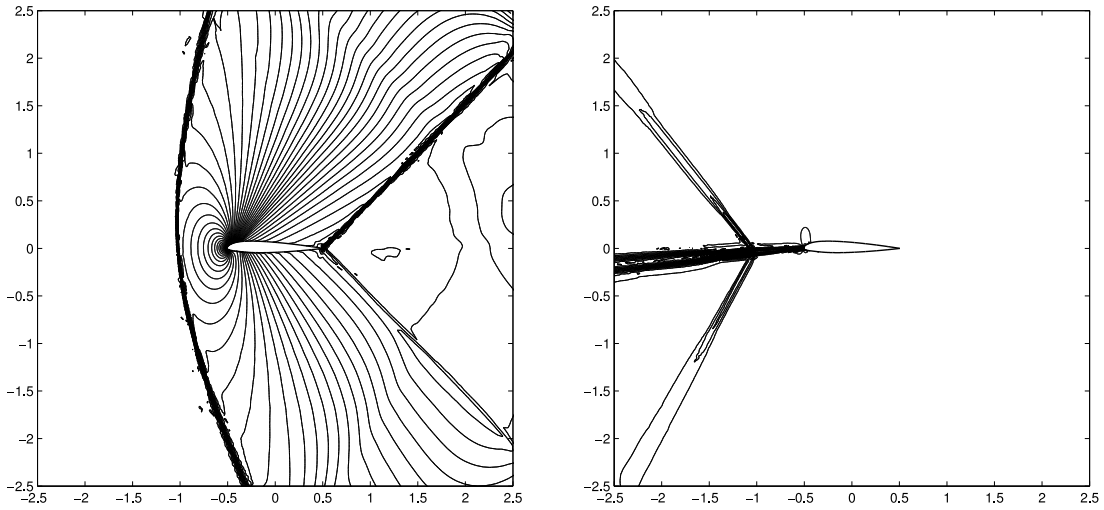


Figure 15: Profile of the NACA23012 airfoil

of the profile, generating a subsonic bubble that wraps by two sonic lines on the surface of the airfoil. So, the leading edge belongs to this subsonic region while outside, the flow is supersonic.
As for similar cases presented in R. Hartmann (2002) and Houston and Hartmann (2002), the quantity of interest for this problem has been chosen as the pressure value at the stagnation point, i.e.

$$\mathcal{J}(\boldsymbol{u}) = p(-l/2, 0),$$

where $l = 1.00893$ is the lenght of the profile since the reference system has been settled at the center of the airfoil. Its exact value has been computed through a fine mesh, providing $\mathcal{J}(\boldsymbol{u}) = 2.24950$.



a: b:

Figure 16: Supersonic flow problem around a NACA23012 profile. Isolines of (a) density, ρ, and (b) adjoint density, z_1.

Figure 16 shows the isolines for the density and the adjoint density around the airfoil. Looking at Figure a, we notice the detached bow shock in front of the profile and two other oblique shocks starting from the trailing edge where the upper and lower supersonic flows join. On the other hand, the adjoint density in Figure b allows to illustrate a comparison about the transport of information between supersonic and subsonic flow. Firstly, we point out that downstream the subsonic region, the adjoint solution is zero as the supersonic region cannot enter and pass any information to the upstream subsonic one. However, where the flow is subsonic the adjoint solution is non-zero because sound waves can reach the leading edge from any point in the subsonic area. Nevertheless, in this case, the adjoint solution is concentrated in a unique spike along the material transport direction going backward from the target stagnation point. When this spike crosses the

#DoF	$\lvert\mathcal{J}-\mathcal{J}_h\rvert$	$\lvert\overline{\mathcal{R}}_\Omega\rvert$	(θ_{eff_1})
372	$4.776 \cdot 10^{-1}$	$1.430 \cdot 10^{-0}$	(2.99)
484	$2.691 \cdot 10^{-1}$	$4.296 \cdot 10^{-1}$	(1.60)
652	$1.161 \cdot 10^{-1}$	$4.737 \cdot 10^{-0}$	(40.8)
848	$2.446 \cdot 10^{-2}$	$7.051 \cdot 10^{-1}$	(28.8)
1176	$6.954 \cdot 10^{-2}$	$5.601 \cdot 10^{-1}$	(8.05)
1644	$2.189 \cdot 10^{-2}$	$1.159 \cdot 10^{-1}$	(5.29)
2184	$5.783 \cdot 10^{-2}$	$1.539 \cdot 10^{-1}$	(2.66)
2988	$6.451 \cdot 10^{-2}$	$1.295 \cdot 10^{-1}$	(2.01)
3924	$6.509 \cdot 10^{-2}$	$1.467 \cdot 10^{-1}$	(2.25)

Table 15: Supersonic flow problem around a NACA23012 profile with target quantity the pressure at stagnation point. SUPG scheme, effectivity of adjoint-based *a posteriori* error estimation.

shock and enters the supersonic region, it splits into three spikes along the characteristic directions, θ and $\theta \pm \mu$, respectively, the material transport and Mach wave directions.
Because of the complexity of the problem and a local point-wise target, the exact error struggles to converge. This is well pointed out on Table 15 where the effectivity is resumed. The target error reduces slowly and it also seems to reach a plateau value, probably due to a high minimum triangle size imposed for the remeshing tool. Regarding the error estimation, we notice the same behaviour observed in the precedent examples; the residual method definitely over estimates the real error and it is not able to reduce it. Instead, apart from the first iterations, the adjoint error representation formula keeps close to the target error and even when shocks are present, its final effectivity is not far from unit.
Finally, in Figure 17, we show the meshes produced by using the two adaptive methods. So, Figure a and c present the refinement driven by the residual based error indicator; there we notice a uniform refinement along the upstream and downstream shock as well as at the leading and trailing edge. On the right hand side, in Figure b and d, adjoint error indicators lead to a remarkable refinement only around the leading edge and the portion of the bow shock where the material transport path crosses the latter; in addition a mild refinement over all the subsonic region is also applied. However, the rest of the shock and the whole domain are completely ignored.

8 Conclusions

In this work, we have developed an optimum adaptive meshing design based on an *a posteriori* error analysis of Petrov-Galerkin finite element methods for systems of nonlinear hyperbolic conservation laws. This procedure, already defined and fully applied in the last decade on discontinuous Galerkin methods (R. Hartmann (2002)) and finite volumes (Barth and Larson (2002)), was here for the first time applied to Petrov-Galerkin methods, with particular attention on stabilized finite element schemes.
Inspired by R. Hartmann (2002) and Süli and Houston (2001), an innovative approach for Petrov-Galerkin discretisation has been completely developed. The essence is to keep the Galerkin structure for the variational operators in order to assure consistency and allow

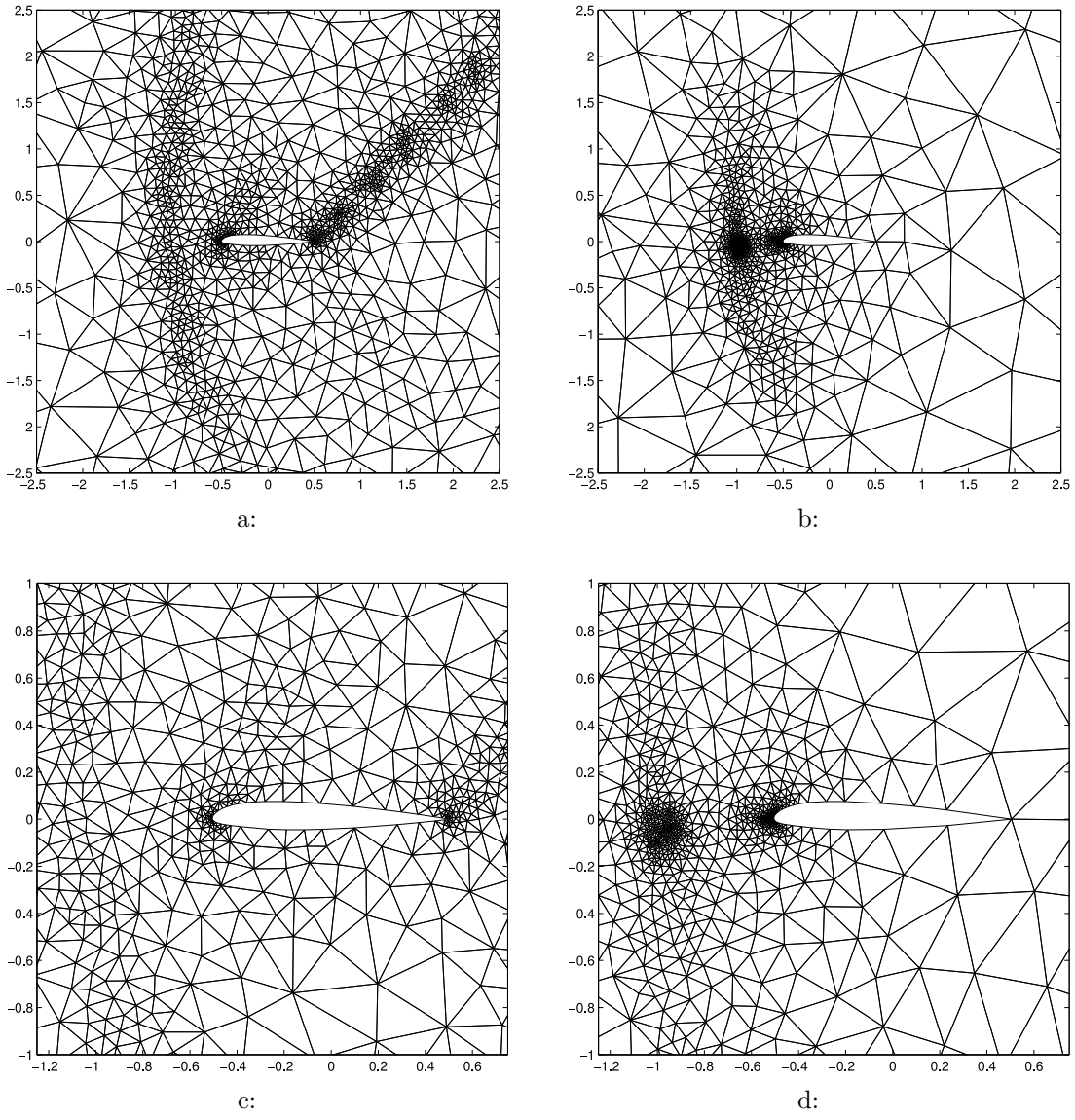


Figure 17: Supersonic flow around a NACA23012 profile with target quantity pressure at stagnation point. SUPG scheme, zoomed and wide view of (a)-(c) residual-based adaptation with 5405 triangles and $|\mathcal{J} - \mathcal{J}_h| = 9.791\,10^{-2}$ and (b)-(d) goal oriented adjoint-based with 1901 triangles and $|\mathcal{J} - \mathcal{J}_h| = 6.509\,10^{-2}$.

a unique numerical discretisation for both primal and adjoint problems, simply swapping trial and test functions. The stabilizing Petrov-Galerkin nature is given by using a different and (usually) discontinuous primal test space which implies an adjoint discontinuous solution space. Throught the analysis of its numerical properties and establishing the convergence rate of its solutions, a well-posed framework of a consistent Petrov-Galerkin discretisations is now available for scalar and systems of conservation equations.
New versions of the some well-known Petrov-Galerkin schemes (SUPG, RD, BUBBLE) have been selected in order to make them more general and applicable for high order elements. Regarding the numerical accuracy, we are usually not interested directly in the solution but in some linear or nonlinear target functional, that represents some physical quantity. Using the new Petrov-Galerkin formulation, we rederived the weighted residual-based error representation formula and the so-called *adjoint*-based *a posteriori* error bounds with respect to these target quantities. This error representation consists then of the element-residual sum multiplied by local weights involving the adjoint solution, which describes how the information is driven over the domain with respect to this functional. So, in order to compute this representation formula, the adjoint problem has to be solved numerically by suitable and consistent approximations. Therefore, a study of adjoint consistency has been provided for scalar and system problems and some commonly target quantities. As the adjoint problem is always linear even if the primal problem were nonlinear, the additional cost of solving a single adjoint problem is negligible.
Finally, local error indicators have been implemented into adaptive mesh finite element algorithms, capable of delivering optimised meshes tailored on the current target quantity and accurate within a given tolerance. Some examples have been proposed to validate the quality of this numerical error estimate and adaptive design comforming the theoretical developments. Besides, we have compared and prove the superiority of this approach over the standard mesh refinement algorithms which employs simple residual-based error estimates, avoiding the adjoint information. On the basis of these computational experiments we could state that the need to perform an additional computation for the adjoint solution is a profitable and valuable price to pay for the availability of a reliable error control based on a rigorous and general mathematical framework.

References

Abgrall, R., Larat, A., Ricchiuto, M., and Tave, C. (2009). A simple construction of very high order non oscillatory compact schemes on unstructured meshes. *Comp. & Fluids*, 38(7):1314–1323.

Abgrall, R. and Roe, P. L. (2003). High order fluctuation schemes on triangular meshes. *Journal of Scientific Computing*, 19(1-3):3–36.

Aziz, A. (1972). *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*. Academic Press, New York.

Barth, T. (2002). A Posteriori Error Estimation and Mesh Adaptivity for Finite Volume and Finite Element Method. Technical report, NASA Ames Research Center, Moffet Field.

Barth, T. and Larson, M. (2002). A Posteriori Error Estimation for Higher Order Godunov Finite Volume Methods on Unstructured Meshes. Technical Report NAS-02-001, NASA Ames Research Center, Moffet Field.

Becker, R. and Rannacher, R. (2001). An optimal control approach to a-posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102.

Bochev, P. (2005). A discourse on variational and geometric aspects of stability of discretizations. In *33rd VKI LS Computational Fluid Dynamics*.

Braess, D. (1997). *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge.

Cockburn, B. and Shu, C. (1998). The Runge-Kutta discountinuous Galerkin finite element method for conservation laws. *Journal of Computational Physics*, 141:199–224.

D'Angelo, S. (2014). *Adjoint-based error estimation for adaptive Petrov-Galerkin finite element methods*. PhD thesis, Universit'e Libre de Bruxelles/VKI.

der Weide, E. V. (1998). *Compressible Flow Simulation on Unstructured Grids usinig Multi-dimensional Upwind Schems*. PhD thesis, Technische Universiteit Delft.

Eriksson, K., Estep, D., Hansbo, P., and Johnson, C. (1995). Introduction to adaptive methods for differential equations. *Acta Numerica*, 4:105–158.

Ern, A. and Guermond, J. (2004). *Theory and Practice of Finite Elements*. Springer.

Giles, M. and Pierce, N. (1997). Adjoint equations in CFD: duality, boundary conditions and solution behaviour. *AIAA*, 97-1850:182–198.

Gresho, P. and Lee, R. (1979). Finite Element Methods for Convection Dominated Flows. In *ASME AMD*, volume 34, pages 37–61.

Hartmann, R. (2007). Adjoint consistency analysis of discountinuous Galerkin discretizations. *SIAM J. Numer. Anal.*, 46:2671–2696.

Hartmann, R. (2008). Numerical Analysis of Higher Order Discountinuos Galerkin Finite Element Methods. In *35th CFD VKI / ADIGMA*.

Houston, P. and Hartmann, R. (2002). Adaptive discountinuous Galerkin finite element methods for Compressible Euler Equations. *SIAM J. Sci. Comput.*, 183:508–532.

Houston, P., Mackenzie, J., Süli, E., and Warnecke, G. (1999). *A posteriori* error analysis for numerical approximations of Friedrichs systems. *Numer. Math.*, 82:433–470.

Houston, P., Rannacher, R., and Süli, E. (2000). *A posteriori* error analysis for stabilised finite element approximations of transport problems. *Comput. Meth. Appl. Mech. Engrg.*, 190:1483–1508.

J. A. Nitsche (1968). Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens. *Numer. Math.*, 11:346–348.

Jaffre, J., Johnson, C., and Szepessy, A. (1995). Convergence of the discountinuous Galerkin finite element method for hyperbolic conservation laws. *Math. Models Methods Appl. Sci.*, 5:367–386.

Lax, P. D. and Wendroff, B. (1960). Systems of conservation laws. *Comm. Pure Appl. Math.*, 13:217–237.

Paillere, H. (1995). *Multidimensional upwind residual distribution schemes for the Euler and Navier-Stokes equations on unstructured grids*. PhD thesis, ULB/VKI.

R. Hartmann (2002). *Adaptive Finite Element Methods for the Compressible Euler Equations*. PhD thesis, Universität Heidelberg.

R. Hartmann and P. Houston (2002). Adaptive Discontinuous Galerkin Finite Element Methods for Nonlinear Hyperbolic Conservation Laws. *SIAM J. Sci. Comp.*, 24:979–1004.

Ricchiuto, M. (2005). *Construction and analysis of compact residual discretizations for conservation laws on unstructured meshes*. PhD thesis, Université Libre de Bruxelles.

Ricchiuto, M. (2010). RD-like PG schemes...or variable β RD. private notes.

Ringleb, F. (1940). Lösungen der Differentialgleichung einer adiabatischen Strömung. *ZAMM*, 20(4):185–198.

Süli, E. and Houston, P. (2001). *hp*-Adaptive discountinuous Galerkin finite element methods for hyperbolic problems. *SIAM J. Sci. Comp.*, 23:1225–1251.

Süli, E. and Houston, P. (2002). *Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics*, chapter Adaptive Finite Element Approximation of Hyperbolic Problems, pages 269–344. Springer.

Villedieu, N. (2009). *High Order Discretisation by Residual Distribution Schemes*. PhD thesis, Université Libre de Bruxelles.

Vymazal, M., L. Koloszár, S. D., N. Villedieu, M. R., and Deconinck, H. (2014). *IDIHOM - Industrialisation of High-Order Methods, A Top Down Approach*, chapter High-order Residual Distribution and error estimation for steady and unsteady compressible flow. Springer.