# Higher order and adaptive DG methods for compressible flows

Ralf Hartmann and Tobias Leicht *
DLR (German Aerospace Center)

December 11, 2013

## Contents



*The authors acknowledge the financial support by the President's Initiative and Networking Fund of the Helmholtz Association of German Research Centres, and the financial support of the European Union, under the EU projects ADIGMA (Kroll et al., 2010) and IDIHOM (Kroll, 2012).

# 1 Introduction

In recent years higher order discretization methods are of increasing importance in computational fluid dynamics. In particular, for compressible flows as considered in aerodynamic flow simulations the development of high order accurate, stable and efficient discretization methods is a hot topic. The European projects ADIGMA (Kroll et al., 2010) and IDIHOM (Kroll, 2012) concentrate and focus the European research effort on the development of these methods towards aerodynamic applications in industry.

## 1.1 Higher order discretization methods

The discretization error of higher order discretization methods decreases with a higher order in the mesh size $h$ than that of low order schemes. A discretization method is of order $n$ if the discretization error behaves like $\mathcal{O}(h^n)$. When halving the mesh size $h$ by performing one global mesh refinement step the discretization error decreases by e.g. a factor of 16 for a forth order scheme in comparison to a factor of only 4 for a second order scheme. As a consequence a required accuracy in the solution can be obtained on coarser meshes and in general with less degrees of freedom and potentially less computing resources required than for second order schemes.

The advantages of higher order methods over second order methods are particularly important in aerodynamic flow simulations:

- Higher order methods allow a significantly improved resolution of flow features like vortices in comparison to second order methods. This is particularly important for the simulation of *vortex creation* and *blade-vortex interaction* at helicopter rotor blades as well as for the simulation of *wake-vortices* behind transport aircraft. Current *second order* based flow solvers tend to be too dissipative leading to strong damping of flow features and a premature dissipation of vortices in numerical simulation although being still present in reality. In contrast to that, the vortices can be well resolved and accurately tracked for a significantly longer time/distance by higher order methods (see e.g. Figure 16 in Section 6.8.3). This is particularly important for improving the shape and control of helicopter rotor blades which is required for reducing helicopter noise. It is even more important for optimizing aircraft shapes in order to reduce wake-vortices and to cause wake-vortices to interact and vanish earlier, which is required for reducing the minimum distance of aircraft at take-off or landing at airports, eventually increasing the transport capacity of airports.

- Higher order methods allow accurate solutions on relatively coarse meshes. For example, it has been shown for a laminar flow over a flat plate (cf. Figure 13 and Table 1 in Section 6.8.1), that a $4^{\text{th}}$ order discretization requires 3 elements in the boundary layer to give the same accuracy as a $2^{\text{nd}}$ order discretization with 36 elements in the boundary layer. This promises a significant reduction of mesh sizes potentially allowing for larger-scale applications with the same computing resources than with current flow solver technologies.

The maximum order one encounters when applying a higher order discretization method to a particular problem depends on the smoothness of the solution. Whereas for (arbitrarily) smooth solutions a method of order $n$ shows in fact a discretization error of order $\mathcal{O}(h^n)$, the order of convergence is reduced for non-smooth solutions.

In general, solutions are not smooth in the whole computational domain; in fact, they might exhibit some irregularities like shocks or singularities in some parts of the domain but are perfectly smooth in other parts. In order to fully exploit the regularity of the solution the order of the discretization should be adapted to the smoothness of the solution. Here, the general idea is to employ discretization methods of higher order in smooth parts of the solution and of low order in irregular parts of the solution ($p$-refinement). Together with local mesh refinement ($h$-refinement) this leads to the so-called $hp$-refinement.

## 1.2 Discontinuous Galerkin discretizations

Over the last about fifteen years the development of discontinuous Galerkin (DG) methods for compressible flows has attracted more and more research groups from all over the world, significantly increasing the pace of the development of these methods (Bassi and Rebay, 1997a,b; van der Vegt and van der Ven, 2002; Hartmann and Houston, 2002b, 2008; Bassi et al., 2005; Fidkowski et al., 2005; Fidkowski and Darmofal, 2007; Bassi et al., 2009; Hartmann et al., 2011; Wallraff et al., 2013; Hartmann, 2013b). In fact, it can be observed that to an increasing extent discontinuous Galerkin methods are now applied to problems which traditionally where solved using finite volume methods. The reason for this trend can be identified in several advantages of the discontinuous Galerkin methods over finite volume methods. Second order finite volume methods are achieved by employing a second order accurate reconstruction. The extension of a second order finite volume scheme to a (theoretically) third order scheme requires a third order accurate reconstruction which is very cumbersome on unstructured meshes and shows a deterioration of order in practise. On unstructured meshes finite volume methods of even higher order are virtually impossible. These difficulties bound the order of numerical computations in industrial applications to second order. In contrast to this, the order of discontinuous Galerkin methods, applied to problems with regular solutions, depends only on the degree of the approximating polynomials which can easily be increased, dramatically simplifying the use of higher order methods on unstructured meshes. Furthermore, the stencil of most discontinuous Galerkin schemes is minimal in the sense that each element communicates only with its direct neighbors. In contrast to the increasing number of elements or mesh points communicating for increasing accuracy of finite volume methods, the inter-element communication of discontinuous Galerkin methods is the same for any order. The compactness of the discontinuous Galerkin method has clear advantages in parallelization, which does not require additional element layers at partition boundaries. Also due to simple communication at element interfaces, elements with so-called ‘hanging nodes’ can be treated just as easily as elements without hanging nodes, a fact that simplifies local mesh refinement ($h$-refinement). In addition to this, the communication at element interfaces is identical for any order of the method which simplifies the use of methods of differing orders in adjacent elements. This allows for the variation of the order of the numerical scheme over the computational domain, which in combination with $h$-refinement leads to the $hp$-refinement algorithms.

## 1.3 Numerical analysis of finite element methods

Discontinuous Galerkin methods are a special type of *finite element methods*. Thus, there are many powerful tools of finite element analysis available which – with some DG specific modifications – can be applied to the numerical analysis of discontinuous Galerkin discretizations.

Consider, for simplicity, a linear partial differential equation of the form

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

with $f \in L^2(\Omega)$ and $g \in L^2(\Gamma)$, where $L$ denotes a linear differential operator on the domain $\Omega$, and $B$ denotes a linear differential operator on the boundary $\Gamma = \partial\Omega$. Furthermore, consider the following finite element discretization: find $u_h \in V_h$ such that

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h. \tag{2}$$

Here, $V_h$ is a discrete function space and $L_h : V \times V \to \mathbb{R}$ is a bilinear form, where $V$ is an appropriately chosen function space such that $V_h \subset V$ and $u \in V$, where $u$ is the exact, i.e., analytical, solution to (1). Then, some of the most important topics in the numerical analysis of this discretization are the following:

- **Consistency:** Does relation (2) still hold when we replace $u_h$ by the exact solution $u$ to the differential equation (1)? I.e., do we have

  $$L_h(u, v) = F_h(v) \quad \forall v \in V. \tag{3}$$

  This answers the question: Do we solve the right equations?

  If the discretization is consistent, we can subtract (2) from (3) for $v_h \in V_h \subset V$ which immediately gives us the so-called *Galerkin orthogonality*:

  $$L_h(u - u_h, v_h) = 0 \quad \forall v_h \in V_h, \tag{4}$$

  which means that the discretization error $e = u - u_h$ is orthogonal (with respect to the bilinear form $L_h$) to the discrete test space $V_h$. This is a basic property of all Galerkin finite element methods, among them e.g. the standard Galerkin (or continuous) finite element method as well as the discontinuous Galerkin finite element method.

- **Coercivity & Stability**: Is there a constant $\gamma > 0$, such that

  $$L_h(v_h, v_h) \geq \gamma |||v_h|||^2 \quad \forall v_h \in V_h, \tag{5}$$

  where $|||v|||$ is a norm (or seminorm) on $V$. Furthermore, we assume that $F_h$ in (2) is continuous, i.e., there is a $C_F > 0$ such that

  $$F_h(v_h) \leq C_F |||v_h||| \quad \forall v_h \in V_h. \tag{6}$$

  Then, for the solution $u_h \in V_h$ to the discrete problem (2) we obtain

  $$\gamma |||u_h|||^2 \leq L_h(u_h, u_h) = F_h(u_h) \leq C_F |||u_h|||, \tag{7}$$

  and thus $|||u_h||| \leq \frac{C_F}{\gamma}$, i.e., we have control over all terms occurring in $|||u_h|||$. If $|||\cdot|||$ is a norm (and not only a semi-norm) on the space in which weak solutions to (1) are to be searched then the discretization (2) is stable.

- **Convergence (Order of convergence):** Does the discrete solution $u_h$ converge to the exact solution $u$? What is the order of convergence, i.e., given a solution $u$ with $\|u\|_{**} < \infty$, what is (the maximum) $r$ such that

$$\|u - u_h\|_* \leq ch^r \|u\|_{**}, \tag{8}$$

  where $\|\cdot\|_*$ is an appropriate (global) norm to measure the error in, e.g. $\|\cdot\|_* = \|\cdot\|_{L^2}$, and $\|\cdot\|_{**}$ is a norm on (possibly a subset of) $V$.

- **Convergence in specific target quantities $J(\cdot)$:** Instead of measuring the error in terms of (global) norms, one might be interested in the error measured in terms of some physically relevant quantity. Let $J_h : V_h \to \mathbb{R}$ be a discretization of a functional $J : V \to \mathbb{R}$, like e.g. a (weighted) mean value of the solution on $\Omega$ or on parts of the boundary $\Gamma$. Then we are interested in the order of convergence with respect to $J(\cdot)$, i.e., given a $u$ with $\|u\|_{**} < \infty$, what is (the maximum) $s$ such that

$$|J(u) - J_h(u_h)| \leq ch^s \|u\|_{**}. \tag{9}$$

  We note, that in aerodynamics the functional $J(\cdot)$ might represent important quantities like aerodynamic force coefficients (drag, lift or moment coefficients).

Some error estimates like the $L^2$-estimate in the case of Poisson's equation and the error estimates with respect to target functionals $J(\cdot)$ require the use of duality arguments including the solutions to appropriately defined dual or adjoint problems. Therefore, we continue the above list as follows:

- **Adjoint consistency:** Given the primal problem (1) and a discretization $J_h(\cdot)$ of the target functional

$$J(u) = \int_\Omega j_\Omega \, u \,\mathrm{d}\mathbf{x} + \int_\Gamma j_\Gamma \, u \,\mathrm{d}s, \tag{10}$$

  with $j_\Omega \in L^2(\Omega)$ and $j_\Gamma \in L^2(\Gamma)$, we define the adjoint problem

$$L^* z = j_\Omega \quad \text{in } \Omega, \qquad\qquad B^* z = j_\Gamma \quad \text{on } \Gamma. \tag{11}$$

  where $L^*$ and $B^*$ denote the adjoint operators to $L$ and $B$, respectively. Then we say that the discretization (2) together with $J_h(\cdot)$ is *adjoint consistent* if the exact solution $z$ to the adjoint problem (11) satisfies:

$$L_h(w, z) = J_h(w) \quad \forall w \in V. \tag{12}$$

Depending on the discretization being adjoint consistent or not the corresponding discretization errors measured in $J(\cdot)$ (or in $L^2$) are optimal or not. In fact, there are discontinuous Galerkin discretizations which are adjoint inconsistent, e.g. the non-symmetric interior penalty (NIPG) method for the discretization of Poisson's equation, and which show a reduced order of convergence as compared to adjoint consistent discretizations like the symmetric interior penalty (SIPG) method or the method of Bassi and Rebay (BR2). Whereas consistency can be considered as basic requirement of a discretization to be reasonable at all (without consistency the discrete solutions might even not converge to the exact solution) the adjoint consistency property represents an additional, and very desirable, quality of the discretization.

There are further topics of high interest in the numerical analysis of finite element methods, among them *a posteriori* error estimates and indicators for local h-refinement which will also be covered in this lecture:

- **A priori and a posteriori error estimates:** We distinguish between *a priori* error estimates and *a posteriori* error estimates.
  - *A priori* error estimates involve norms of the exact solution $u$. As $u$ is unknown (otherwise we would not need to solve the problem numerically) an *a priori* error estimate gives no quantitative size of the error of the numerical solution. It gives, however, the order the error converges under mesh refinement, $h \to 0$; see e.g. estimates (8) and (9).
  - *A posteriori* error estimates do not include the exact solution $u$ but only computable values which depend on e.g. the numerical solution $u_h$ like in

$$J(u) - J_h(u_h) \approx E(u_h). \tag{13}$$

- **Indicators for local refinement:** In most cases global refinement of the computational mesh or global enrichment of the polynomial degree is a very inefficient way of improving the accuracy of a numerical solution. In practice, usually only local mesh refinement is affordable. For deciding which elements to refine local error indicators $\eta_\kappa$ are needed. Here, a variety of different indicators exist, many of which are purely heuristic, some are designed to reduce the error in specific global norms and some to reduce the error in specific target quantities $J(\cdot)$. In this work, we consider:
  - Adjoint-based (goal-oriented) refinement indicators which are tailored to the accurate and efficient approximation of target quantities like aerodynamic force coefficients. These indicators are based on primal residuals multiplied by the solution to an adjoint problem connected to the target quantity.
  - Residual-based refinement indicators which are tailored to resolving the overall flow field. These indicators are based on the primal residual but do not depend on an adjoint solution. They are targeted at resolving all flow features irrespective of any target quantity.

Further topics of interest are the derivation of indicators for anisotropic mesh refinement and of indicators for *hp*-refinement. Both topics are not covered by this lecture. Instead we refer to e.g. (Hartmann and Houston, 2010; Leicht and Hartmann, 2010) for an extension of the presented refinement strategies to anisotropic mesh refinement. Furthermore, we refer to e.g. (Hartmann and Houston, 2010; Leicht and Hartmann, 2011) for an extension to *hp*-refinement.

## 1.4 Novelties in this work

To our knowledge, following theoretical results included in this work have not been published before:

- **Extension of the adjoint consistency analysis of Poisson's equation.** Adjoint consistency has been shown in (Arnold et al., 2002) for interior fluxes for a wide range of DG discretizations of Poisson's equation with homogeneous boundary conditions. Furthermore, (Hartmann, 2007a, 2008b) considered the adjoint consistency analysis of the discretization of interior fluxes, boundary fluxes and target quantities for Poisson's equation with *in*homogeneous Dirichlet and Neumann boundary conditions with numerical fluxes specialized for the interior penality DG discretization.

  In the current work (cf. Theorem 4.9), we provide the adjoint consistency analysis for the discretization of interior fluxes, of boundary fluxes and of target quantities for Poisson's equation with inhomogeneous Dirichlet and Neumann boundary conditions for any consistent and conservative numerical flux function, i.e., for a wide range of DG discretizations of diffusive terms.

- **Extension of the adjoint consistency analysis of the compressible Euler equations.** In previous work (Lu, 2005; Hartmann, 2007a,b) a discretization of the pressure induced force coefficients was proposed which is adjoint consistent in combination with a DG discretization for the special case of a normal wall boundary flux.

  In the current work (cf. Theorem 5.13), we generalize this to any consistent discretization of convective boundary fluxes, i.e., for any consistent discretization of boundary fluxes, including the use of numerical flux functions at the wall boundary, we give an associated discretization of the force coefficients which makes the discretization adjoint consistent.

- **Extension of the adjoint consistency analysis of the compressible Navier-Stokes equations.** In previous work (Lu, 2005; Hartmann, 2007a) a discretization of the total (i.e., pressure induced plus viscous) force coefficients was proposed which is adjoint consistent in combination with a DG discretization for the special case of normal convective and diffusive wall boundary fluxes.

  In the current work (cf. Theorem 6.9), we generalize this to any consistent discretization of convective and diffusive boundary fluxes, i.e., for any discretization of boundary fluxes, including the use of numerical flux functions at the wall boundary, we give an associated discretization of the force coefficients which makes the discretization adjoint consistent.

## 1.5 Outline

In Section 2 we begin by introducing the consistency and adjoint consistency analysis of discontinuous Galerkin discretizations. In particular, we follow (Hartmann, 2007a) and give a general framework for analyzing the consistency and adjoint consistency of DG discretizations for *linear* problems with inhomogeneous boundary conditions in Section 2.1. This includes the derivation of continuous adjoint problems associated to specific target quantities $J(\cdot)$, the derivation of primal and adjoint residual forms of the discretizations and the discussion whether the discretizations of the primal equations in combination with the discretizations of the target quantities, $J_h(\cdot)$, are adjoint consistent or not. Following (Hartmann, 2007a), this analysis is then extended to *nonlinear* problems in Section 2.2.

In subsequent sections, we consider DG discretizations of *scalar model problems*,

- the linear advection(-reaction) equation (cf. Section 3), and
- Poisson's equation (cf. Section 4),

and of *compressible flow problems* governed by

- the compressible Euler equations (cf. Section 5),
- the compressible Navier-Stokes equations (cf. Section 6), and
- the Reynolds-averaged Navier-Stokes (RANS) equations and the $k$-$\omega$ turbulence model equations (cf. Section 7).

Each of these discretizations will be analyzed with respect to consistency, adjoint consistency, and global and local conservation properties. Furthermore, for the DG discretizations of the scalar model problems we discuss stability properties and give *a priori* estimates of the discretization error measured in terms of related DG-norms, in terms of the $L^2$-norm as well as in terms of target quantities $J(\cdot)$. Here, Sections 3 and 4 are supported by the accompanying appendix sections A.3 and A.4 which provide more detailed notations and proofs. Furthermore, some of the theoretical results, in particular the *a priori* error estimates in $L^2$ and $J(\cdot)$, are supported by experimental measurements of the order of convergence in the numerical results sections 4.9 and 4.11. For the compressible flow problems, for which no *a priori* error estimates are available, we provide numerical results and experimental measurements of the order of convergence for inviscid and viscous laminar flows in Sections 5.8 and 6.8, respectively.

Then, Section 8 is devoted to the derivation of adjoint-based *a posteriori* estimates of the discretization error in computed target quantity values $J_h(u_h)$ like aerodynamic force coefficients. These estimates are then decomposed into sums of local error indicators (the so-called *goal-oriented* or *adjoint-based* indicators) which can be used in an adaptive mesh refinement algorithm tailored to the accurate and efficient approximation of the target quantity. After introduction of this adjoint-based error estimation and mesh refinement approach for *single* target quantities in Section 8.1, it is extended to the treatment of *multiple* target quantities in Section 8.2. Section 8.4 finalizes the adaptivity section with the derivation of *residual-based* indicators. Not depending on an adjoint problem these indicators target at resolving the overall flow field.

Finally, in Section 9 the performance of the adjoint-based error estimation, the adjoint-based mesh refinement and the residual-based mesh refinement will be demonstrated for a number of aerodynamic test cases of increasing compexity.

# 2 Consistency and adjoint consistency

One of the most important properties of a discretization is its consistency with the differential equations to be discretized. In fact, consistency ensures that the “right” equations are solved. In finite element methods consistency directly implies the well-known Galerkin orthogonality. Provided the discretization is stable and using standard interpolation/approximation estimates this gives optimal *a priori* order estimates in the so-called energy norm, like e.g. the $H^1$-norm for Poisson's equation.

Furthermore, in finite element methods usually a duality argument is applied for deriving error estimates in the $L^2$-norm. This approach introduces an appropriate adjoint (dual) problem, which is then used to represent the $L^2$-norm of the discretization error $e = u - u_h$ in terms of the discretization and the adjoint solution $z$. Again by Galerkin orthogonality and by using smoothness properties of the adjoint solution the $L^2$-error estimates are derived (cf. e.g. Lemma A.23 for the proof of the $L^2$-error estimate of the discontinuous Galerkin discretization of Poisson's equation).

Optimal order $L^2$-error estimates depend on the applicability of the duality argument as well as on the smoothness of the adjoint solution. Both, however, are connected to the so-called *adjoint consistency* of the discretization. As we will see in Lemma A.23 adjoint consistency of the SIPG discretization results in optimal error estimates in the $L^2$-norm whereas the lack of adjoint consistency in the case of the NIPG discretization results in a suboptimal order of convergence in the $L^2$-norm.

In the following, we introduce the consistency and adjoint consistency analysis following (Hartmann, 2007a). In particular, in Section 2.1 we provide a definition of consistency and adjoint consistency for the case of linear problems with inhomogeneous boundary conditions. Furthermore, we give a general framework for analyzing consistency and adjoint consistency. In Section 2.2 the definition of consistency and adjoint consistency as well as the framework for analyzing them is then extended to the case of nonlinear problems. Finally, we note, that this framework will be applied to the DG discretizations of the linear advection equation, of Poissons's equations, and of the compressible Euler and Navier-Stokes equations in subsequent sections.

## 2.1 The analysis for linear problems

We introduce the consistency and adjoint consistency analysis first for linear problems.

### 2.1.1 Definition of consistency and adjoint consistency

Let $\Omega \subset \mathbb{R}^d$ be a bounded open domain with boundary $\Gamma = \partial\Omega$. Given the linear problem

$$Lu = f \quad \text{in } \Omega, \qquad\qquad Bu = g \quad \text{on } \Gamma, \tag{14}$$

where $f \in L^2(\Omega)$, $g \in L^2(\Gamma)$, $L$ denotes a linear differential operator on $\Omega$, and $B$ denotes a linear boundary operator on $\Gamma$. Let $J(\cdot)$ be a linear target functional (also referred to as target quantity) given by

$$J(u) = (j_\Omega, u)_\Omega + (j_\Gamma, Cu)_\Gamma \equiv \int_\Omega j_\Omega \, u \, \mathrm{d}\mathbf{x} + \int_\Gamma j_\Gamma \, Cu \, \mathrm{d}s, \tag{15}$$

where $j_\Omega \in L^2(\Omega)$, $j_\Gamma \in L^2(\Gamma)$, $C$ is an operator on $\Gamma$ which may be differential, and $(\cdot,\cdot)_\Omega$ and $(\cdot,\cdot)_\Gamma$ denote the $L^2(\Omega)$ and $L^2(\Gamma)$ scalar products, respectively. We assume that the target quantity (15) is *compatible* with the primal problem (14), i.e., we assume that there are linear operators $L^*$, $B^*$ and $C^*$ such that following *compatibility condition* holds:

$$(Lu, z)_\Omega + (Bu, C^*z)_\Gamma = (u, L^*z)_\Omega + (Cu, B^*z)_\Gamma. \tag{16}$$

Then, $L^*$, $B^*$ and $C^*$ are the so-called *adjoint operators* to $L$, $B$ and $C$, respectively. We note that for given operators $L$ and $B$ associated with the primal problem (14) only some target quantities (15) with operators $C$ are compatible whereas others are not. However, *assuming* that (16) holds the adjoint problem associated to (14) and (15) is given by

$$L^*z = j_\Omega \quad \text{in } \Omega, \qquad\qquad B^*z = j_\Gamma \quad \text{on } \Gamma. \tag{17}$$

Let $\Omega$ be subdivided into a shape-regular mesh $\mathcal{T}_h = \{\kappa\}$ consisting of elements $\kappa$ and let $V_h$ be a discrete function space on $\mathcal{T}_h$. Furthermore, let problem (14) be discretized as follows: find $u_h \in V_h$ such that

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h, \tag{18}$$

where $L_h(\cdot,\cdot)$ is a bilinear form and $F_h(\cdot)$ a linear form including the prescribed primal force and boundary data functions $f$ and $g$. Then the discretization (18) is said to be *consistent* if the exact solution $u \in V$ to the primal problem (14) satisfies:

$$L_h(u, v) = F_h(v) \quad \forall v \in V, \tag{19}$$

where $V$ is a suitably chosen function space such that $u \in V$ and $V_h \subset V$. Similarly, a discretization $J_h(\cdot)$ of the target quantity $J(\cdot)$ in (15) is said to be *consistent* if the exact solution $u \in V$ to the primal problem (14) satisfies:

$$J_h(u) = J(u). \tag{20}$$

Given the discretizations $L_h(\cdot,\cdot)$ and $J_h(\cdot)$ of, respectively, the primal problem (14) and the target quantity (15), the discretization is said to be *adjoint consistent* if the exact solution $z \in V$ to the adjoint problem (17) satisfies:

$$L_h(w, z) = J(w) \quad \forall w \in V. \tag{21}$$

In other words, a discretization is adjoint consistent if the discrete adjoint problem is a consistent discretization of the continuous adjoint problem.

### 2.1.2 The consistency and adjoint consistency analysis

Based on the definition of consistency and adjoint consistency in the previous subsection we continue following (Hartmann, 2007a) and outline a framework for analyzing consistency and adjoint consistency of discontinuous Galerkin discretizations. This framework can also be used to find specific terms due to which some DG discretizations may not be adjoint consistent. In these cases the analysis gives some insight into how an adjoint inconsistent DG discretization together with a specific target functional could be modified to recover an adjoint consistent discretization.

Given a discretization of the primal problem and a target quantity, the adjoint consistency analysis consists of the following steps:

- **Derivation of the continuous adjoint problem:** Let the primal problem be given by (14). Furthermore, assume that $J(\cdot)$ is a linear target functional as in (15) which is compatible with (14). Then, we derive the continuous adjoint problem with continuous adjoint boundary conditions as given in (17).

- **Consistency analysis of the discrete primal problem:** We rewrite the discontinuous Galerkin discretization (18) of problem (14) in following element-based primal residual form: find $u_h \in V_h$ such that

$$\int_\Omega R(u_h) v_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} r(u_h) v_h \, \mathrm{d}s + \int_\Gamma r_\Gamma(u_h) v_h \, \mathrm{d}s = 0 \quad \forall v_h \in V_h, \tag{22}$$

where $R(u_h)$ denotes the element residual, $r(u_h)$ denotes the interior face residual, and $r_\Gamma(u_h)$ denotes the boundary residual. According to (19) the discretization (18) is consistent if the exact solution $u$ to (14) satisfies

$$\int_\Omega R(u) v \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} r(u) v \, \mathrm{d}s + \int_\Gamma r_\Gamma(u) v \, \mathrm{d}s = 0 \quad \forall v \in V, \tag{23}$$

which holds provided $u$ satisfies

$$R(u) = 0 \quad \text{in } \kappa, \kappa \in \mathcal{T}_h, \qquad r(u) = 0 \quad \text{on } \partial\kappa \setminus \Gamma, \kappa \in \mathcal{T}_h, \qquad r_\Gamma(u) = 0 \quad \text{on } \Gamma. \tag{24}$$

- **Adjoint consistency of element, interior face and boundary terms:** Given the discretization (18) of the primal problem (14) and a discretization $J_h(\cdot)$ of the target quantity (15), we rewrite the discrete adjoint problem: find $z_h \in V_h$ such that

$$L_h(w_h, z_h) = J_h(w_h) \quad \forall w_h \in V_h, \tag{25}$$

in following element-based adjoint residual form: find $z_h \in V_h$ such that

$$\int_\Omega w_h \, R^*(z_h) \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} w_h \, r^*(z_h) \, \mathrm{d}s + \int_\Gamma w_h \, r_\Gamma^*(z_h) \, \mathrm{d}s = 0 \quad \forall w_h \in V_h, \tag{26}$$

where $R^*(z_h)$, $r^*(z_h)$ and $r_\Gamma^*(z_h)$ denote the element, interior face and boundary adjoint residuals, respectively. According to (21) the discretization (18) is adjoint consistent if the exact solution $z \in V$ to (17) satisfies

$$\int_\Omega w \, R^*(z) \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} w \, r^*(z) \, \mathrm{d}s + \int_\Gamma w \, r_\Gamma^*(z) \, \mathrm{d}s = 0 \quad \forall w \in V, \tag{27}$$

which holds provided $z$ satisfies

$$R^*(z) = 0 \quad \text{in } \kappa, \kappa \in \mathcal{T}_h, \qquad r^*(z) = 0 \quad \text{on } \partial\kappa \setminus \Gamma, \kappa \in \mathcal{T}_h, \qquad r_\Gamma^*(z) = 0 \quad \text{on } \Gamma. \tag{28}$$

**Remark 2.1** *We note that the adjoint problem and consequently the adjoint consistency of a discretization depends on the specific target functional $J(\cdot)$ under consideration. Given a target functional of the form* (15)*, we see that $R^*(z)$ depends on $j_\Omega(\cdot)$, and $r^*_\Gamma(z)$ depends on $j_\Gamma(\cdot)$. Among all consistent discretizations $J_h(\cdot)$ of $J(\cdot)$ according to* (20) *we want to find a discretization $J_h(\cdot)$ which gives an adjoint consistent discretization according to* (21)*. We will see in subsequent sections that ingredients like e.g. numerical fluxes which are incorporated in the discretization of the primal problem also need to be incorporated in the discretization of the target quantity.*

## 2.2 The analysis for nonlinear problems

In the previous Section we introduced the consistency and adjoint consistency analysis for linear problems. In this section, we now extend this analysis to nonlinear problems.

### 2.2.1 Definition of consistency and adjoint consistency

We consider a nonlinear problem of the form

$$Nu = 0 \quad \text{in } \Omega, \qquad\qquad Bu = 0 \quad \text{on } \Gamma, \tag{29}$$

where $N$ is a nonlinear differential (and Fréchet-differentiable) operator and $B$ is a (possibly nonlinear and Fréchet-differentiable) boundary operator. Let $J(\cdot)$ be a nonlinear target quantity

$$J(u) = \int_\Omega j_\Omega(u)\,\mathrm{d}\mathbf{x} + \int_\Gamma j_\Gamma(Cu)\,\mathrm{d}s, \tag{30}$$

with Fréchet derivative

$$J'[u](w) = \int_\Omega j'_\Omega[u]\,w\,\mathrm{d}\mathbf{x} + \int_\Gamma j'_\Gamma[Cu]\,C'[u]w\,\mathrm{d}s, \tag{31}$$

where $j_\Omega(\cdot)$ and $j_\Gamma(\cdot)$ may be nonlinear with derivatives $j'_\Omega$ and $j'_\Gamma$, respectively, and $C$ is a differential boundary operator on $\Gamma$ and may be nonlinear with derivative $C'$. Here, $'$ denotes the (total) Fréchet derivative and the square bracket $[\cdot]$ denotes the state about which linearization is performed. Again, we say that the target functional (30) is *compatible* with (29) provided the following compatibility condition holds

$$(N'[u]w, z)_\Omega + (B'[u]w, (C'[u])^*z)_\Gamma = (w, (N'[u])^*z)_\Omega + (C'[u]w, (B'[u])^*z)_\Gamma, \tag{32}$$

where $(N'[u])^*$, $(B'[u])^*$ and $(C'[u])^*$ denote the adjoint operators to $N'[u]$, $B'[u]$ and $C'[u]$. This condition is analogous to (16), with $L$, $B$ and $C$ replaced by $N'[u]$, $B'[u]$ and $C'[u]$, respectively. *Assuming* that (32) holds the continuous adjoint problem associated to (29) and (31) is:

$$(N'[u])^*z = j'_\Omega[u] \quad \text{in } \Omega, \qquad\qquad (B'[u])^*z = j'_\Gamma[Cu]C'[u] \quad \text{on } \Gamma. \tag{33}$$

Let $N_h(\cdot,\cdot)$ be a semi-linear form, nonlinear in its first and linear in its second argument, such that the nonlinear problem (29) is discretized as follows: find $u_h \in V_h$ such that

$$N_h(u_h, v_h) = 0 \quad \forall v_h \in V_h. \tag{34}$$

Then, the discretization (34) is said to be *consistent* if the exact solution $u \in V$ to the primal problem (29) satisfies the following equation:

$$N_h(u, v) = 0 \quad \forall v \in V. \tag{35}$$

Similarly, a discretization $J_h(\cdot)$ of the target quantity $J(\cdot)$ in (30) is said to be *consistent* if the exact solution $u \in V$ to the primal problem (29) satisfies:

$$J_h(u) = J(u). \tag{36}$$

Given the discretizations $N_h(\cdot, \cdot)$ and $J_h(\cdot)$ of, respectively, the primal problem (29) and the target quantity (30), the discretization is said to be *adjoint consistent* if the exact solution $z \in V$ to the adjoint problem (33) satisfies

$$N_h'[u](w, z) = J_h'[u](w) \quad \forall w \in V, \tag{37}$$

where $N_h'[u]$ and $J_h'[u]$ denote the Fréchet derivatives of $N_h(u, v)$ and $J_h(u)$ with respect to $u$. In other words, a discretization is adjoint consistent if the discrete adjoint problem is a consistent discretization of the continuous adjoint problem. Finally, we note that in case of a linear problem and target functional the definition of adjoint consistency in (37) reduces to the definition of linear adjoint consistency given in Section 2.1. The definition of adjoint consistency for nonlinear problems as given in (37) was introduced by (Lu, 2005). Furthermore, we note that (Lu, 2005) also gives a definition of asymptotically adjoint consistent methods.

### 2.2.2 The consistency and adjoint consistency analysis

Based on the definition of consistency and adjoint consistency in the previous subsection we now follow (Hartmann, 2007a) and generalize the framework for analyzing consistency and adjoint consistency of discontinuous Galerkin discretizations for linear problems as given in Section 2.1 to the case of nonlinear problems. We recall that this framework can also be used to find specific terms due to which some DG discretizations may not be adjoint consistent. In these cases the analysis gives some insight into how an adjoint inconsistent DG discretization together with a specific target functional could be modified to recover an adjoint consistent discretization.

Given a discretization of the primal problem and a target quantity, the adjoint consistency analysis consists of the following steps:

- **Derivation of the continuous adjoint problem:** Let the primal problem be given by (29). Furthermore, assume that $J(\cdot)$ is a nonlinear target quantity of the form (30) which is compatible with the primal problem (29). Then we derive the continuous adjoint problem (33) including adjoint boundary conditions.

  **Remark 2.2** *We note that the derivation of the adjoint operator $(N'[u])^*$ for nonlinear systems is a considerably more complicated task than deriving $L^*$ for scalar linear problems. Still more involved is the derivation of the adjoint boundary operators $(B'[u])^*$. In the framework of optimal design, (Giles and Pierce, 1997) gives a general approach of deriving $(B'[u])^*$ and $(C'[u])^*$ assumed to be connect to $B$,*

*$C$, $N$ and $(N'[u])^*$ through (32). This approach is based on a matrix representation of boundary operators which for systems of equations leads to lengthy and error prone derivations. In contrast to optimization where both $(B'[u])^*$ and $(C'[u])^*$ are required, in the following analysis we require only the adjoint operator $(B'[u])^*$. Due to this we can circumvent the approach described in (Giles and Pierce, 1997) and use a simpler way of deriving the adjoint operators $(B'[u])^*$.*

- **Consistency analysis of the discrete primal problem:** We rewrite the discontinuous Galerkin discretization (34) of problem (29) in the following element-based primal residual form: find $u_h \in V_h$ such that

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa R(u_h)v_h\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} r(u_h)v_h\,\mathrm{d}s + \int_\Gamma r_\Gamma(u_h)v_h\,\mathrm{d}s = 0 \quad \forall v_h\in V_h, \qquad (38)$$

where $R(u_h)$, $r(u_h)$ and $r_\Gamma(u_h)$ denote the element, interior face and boundary residuals, respectively. According to (35), the discretization (34) is consistent if the exact solution $u$ to (29) satisfies

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa R(u)v\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} r(u)v\,\mathrm{d}s + \int_\Gamma r_\Gamma(u)v\,\mathrm{d}s = 0 \quad \forall v\in V, \qquad (39)$$

which holds provided $u$ satisfies

$$R(u)=0 \;\text{ in } \kappa,\kappa\in\mathcal{T}_h, \quad r(u)=0 \;\text{ on } \partial\kappa\setminus\Gamma,\kappa\in\mathcal{T}_h, \quad r_\Gamma(u)=0 \;\text{ on } \Gamma. \qquad (40)$$

- **Derivation of the discrete adjoint problem** Given the discretization (34) of the primal problem (29) and a discretization $J_h(\cdot)$ of the target quantity (30), we derive the discrete adjoint problem: find $z_h\in V_h$ such that

$$N_h'[u_h](w_h,z_h) = J_h'[u_h](w_h) \quad \forall w_h\in V_h. \qquad (41)$$

$N_h'[u_h]$ is called the Jacobian of the numerical scheme and is required also for implicit solution methods like e.g. Newton or Backward-Euler methods. The discrete adjoint problem (41) includes the transpose of the Jacobian and is required in *a posteriori* error estimation and goal-oriented adaptation as well as in optimization.

- **Adjoint consistency of element, interior face and boundary terms** We rewrite the discrete adjoint problem (41) in element-based adjoint residual form: find $z_h\in V_h$ such that

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa w_h\,R^*[u_h](z_h)\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} w_h\,r^*[u_h](z_h)\,\mathrm{d}s + \int_\Gamma w_h\,r_\Gamma^*[u_h](z_h)\,\mathrm{d}s = 0, \qquad (42)$$

for all $w_h\in V_h$, where $R^*[u_h](z_h)$, $r^*[u_h](z_h)$ and $r_\Gamma^*[u_h](z_h)$ denote the element, interior face and boundary adjoint residuals, respectively. According to (37), the discretization (34) is adjoint consistent if the exact solutions $u$ and $z$ satisfy

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa w\,R^*[u](z)\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} w\,r^*[u](z)\,\mathrm{d}s + \int_\Gamma w\,r_\Gamma^*[u](z)\,\mathrm{d}s = 0 \quad \forall w\in V, \qquad (43)$$

which holds provided $u$ and $z$ satisfy

$$R^*[u](z)=0 \;\text{ in } \kappa, \quad r^*[u](z)=0 \;\text{ on } \partial\kappa\setminus\Gamma,\kappa\in\mathcal{T}_h, \quad r_\Gamma^*[u](z)=0 \;\text{ on } \Gamma. \qquad (44)$$

# 3 DG discretization of the linear advection equation

In this section we derive and analyze DG discretizations of the linear advection-reaction equation.

## 3.1 The linear advection-reaction equation

Let $\Omega \subset \mathbb{R}^2$ be a bounded open domain with boundary $\Gamma = \partial\Omega$. We consider the linear advection-reaction equation

$$\begin{aligned} \nabla \cdot (\mathbf{b}u) + cu &= f \quad \text{in } \Omega, \\ u &= g \quad \text{on } \Gamma_-, \end{aligned} \tag{45}$$

with $f \in L^2(\Omega)$, $\mathbf{b} \in [C^1(\Omega)]^2$, $c \in L^\infty(\Omega)$, and with $g \in L^2(\Gamma_-)$ on the inflow part of the boundary,

$$\Gamma_- = \{\mathbf{x} \in \Gamma, \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\}. \tag{46}$$

Furthermore, we assume that there exists a $c_0 \in L^\infty(\Omega)$ and a number $\gamma_0 > 0$ such that

$$c(\mathbf{x}) + \frac{1}{2}\nabla \cdot \mathbf{b}(\mathbf{x}) = c_0^2(\mathbf{x}) \geq \gamma_0 > 0. \tag{47}$$

## 3.2 The continuous adjoint equation

In order to derive the continuous adjoint equation, we multiply the left hand side of (45) by $z$, integrate over the domain $\Omega$ and integrate by parts. Thereby, we obtain

$$(\nabla \cdot (\mathbf{b}u) + cu, z)_\Omega + (u, -\mathbf{b} \cdot \mathbf{n}\, z)_{\Gamma_-} = (u, -\mathbf{b} \cdot \nabla z + cz)_\Omega + (u, \mathbf{b} \cdot \mathbf{n}\, z)_{\Gamma_+}. \tag{48}$$

Comparing with the compatibility condition (16),

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

we see that for $Lu = \nabla \cdot (\mathbf{b}u) + cu$ in $\Omega$ and

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

the adjoint operators are given by $L^* z = -\mathbf{b} \cdot \nabla z + cz$ in $\Omega$ and

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

In particular, for

$$J(u) = \int_\Omega j_\Omega\, u \,\mathrm{d}\mathbf{x} + \int_\Gamma j_\Gamma\, Cu \,\mathrm{d}s = \int_\Omega j_\Omega\, u \,\mathrm{d}\mathbf{x} + \int_{\Gamma_+} j_\Gamma\, u \,\mathrm{d}s, \tag{49}$$

with $j_\Omega \in L^2(\Omega)$ and $j_\Gamma \in L^2(\Gamma_+)$, the continuous adjoint problem is given by

$$\begin{aligned} -\mathbf{b} \cdot \nabla z + cz &= j_\Omega \quad \text{in } \Omega, \\ \mathbf{b} \cdot \mathbf{n}\, z &= j_\Gamma \quad \text{on } \Gamma_+. \end{aligned} \tag{50}$$

## 3.3 Derivation of the DG discretization

Suppose that $\mathcal{T}_h$ is a subdivision of $\Omega$ into disjoint open element domains $\kappa$ such that $\bar{\Omega} = \cup_{\kappa\in\mathcal{T}_h}\bar{\kappa}$. Let us assume that each $\kappa \in \mathcal{T}_h$ is a smooth bijective image of a fixed reference element $\hat{\kappa}$, that is, $\kappa = F_\kappa(\hat{\kappa})$ for each $\kappa \in \mathcal{T}_h$. On the reference element $\hat{\kappa}$ we define the tensor-product polynomial space $\mathcal{Q}_p$ and the complete polynomial space $\mathcal{P}_p$ of degree $p \geq 0$ as follows:

$$\mathcal{Q}_p = \operatorname{span}\left\{\hat{\mathbf{x}}^\alpha : 0 \leq \alpha_i \leq p, i = 1, 2\right\}, \qquad \mathcal{P}_p = \operatorname{span}\left\{\hat{\mathbf{x}}^\alpha : 0 \leq |\alpha| \leq p\right\}. \tag{51}$$

We now introduce the finite element function space $V_h^p$ consisting of discontinuous piecewise polynomial functions of degree $p \geq 0$, defined by

$$\begin{aligned} V_h^p = \{v_h \in L^2(\Omega) : v_h|_\kappa \circ F_\kappa \in Q_p(\hat{\kappa}) &\text{ if } \hat{\kappa} \text{ is the unit square, and} \\ v_h|_\kappa \circ F_\kappa \in P_p(\hat{\kappa}) &\text{ if } \hat{\kappa} \text{ is the unit triangle}, \kappa \in \mathcal{T}_h\}. \end{aligned} \tag{52}$$

Given an element $\kappa \in \mathcal{T}_h$, we multiply (45) by a test function $v$, integrate over $\kappa$,

$$\int_\kappa (\nabla \cdot (\mathbf{b}u) + cu)\, v \,\mathrm{d}\mathbf{x} = \int_\kappa f v \,\mathrm{d}\mathbf{x},$$

and integrate by parts

$$-\int_\kappa (\mathbf{b}u) \cdot \nabla v \,\mathrm{d}\mathbf{x} + \int_\kappa cuv \,\mathrm{d}\mathbf{x} + \int_{\partial\kappa} \mathbf{b}\cdot\mathbf{n}\, uv \,\mathrm{d}s = \int_\kappa f v \,\mathrm{d}\mathbf{x}. \tag{53}$$

Summing over all elements $\kappa \in \mathcal{T}_h$ and replacing $u$ by $g$ on $\Gamma_-$ we obtain following weak formulation of the linear advection equation (45): find $u \in V$

$$\int_\Omega (-\mathbf{b}u)\cdot\nabla v + cuv \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \mathbf{b}\cdot\mathbf{n}\, uv \,\mathrm{d}s + \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, uv \,\mathrm{d}s + \int_{\Gamma_-} \mathbf{b}\cdot\mathbf{n}\, gv \,\mathrm{d}s = \int_\kappa f v \,\mathrm{d}\mathbf{x}$$

for all $v \in V$. Then we replace the analytical solution $u$ by the discrete function $u_h$ and the test function $v$ by $v_h$, where $u_h$ and $v_h$ both belong to the finite element space $V_h^p$. In addition, since the numerical solution $u_h$ may be discontinuous at interfaces $\partial\kappa \cap \partial\kappa'$ between elements $\kappa, \kappa' \in \mathcal{T}_h$, $\kappa' \neq \kappa$, we replace the flux $\mathbf{b}\cdot\mathbf{n}\, u$ by a *numerical flux function* $\hat{h}_h = \hat{h}(u_h, \mathbf{n}) = \hat{h}(u_h^+, u_h^-, \mathbf{n})$, which depends on both the interior and exterior traces, $u_h^+$ and $u_h^-$, respectively, of $u_h$ on $\partial\kappa \setminus \Gamma$, $\kappa \in \mathcal{T}_h$, and the unit outward normal $\mathbf{n}$ to $\partial\kappa$. Furthermore, we define the numerical flux function at the boundary

$$\hat{h}_h|_\Gamma = h_{\Gamma,h} = h_\Gamma(u_h^+, \mathbf{n}) = \begin{cases} \mathbf{b}\cdot\mathbf{n}\, g & \text{on } \Gamma_-, \\ \mathbf{b}\cdot\mathbf{n}\, u_h^+ & \text{on } \Gamma_+. \end{cases} \tag{54}$$

Then, the discontinuous Galerkin discretization of (45) is given as follows: find $u_h \in V_h^p$ such that

$$\int_\Omega (-\mathbf{b}u_h)\cdot\nabla_h v_h + cu_h v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{h}_h\, v_h^+ \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} \qquad \forall v_h \in V_h^p, \tag{55}$$

where $\nabla_h$ is the broken gradient operator defined in Definition A.4 (cf. Appendix A.1). Examples of numerical flux function $\hat{h}_h = \hat{h}(u_h^+, u_h^-, \mathbf{n})$ will be given later.

## 3.4 Consistency

Integrating (55) back by parts on each element $\kappa$ we obtain following equivalent form of the discretization: find $u_h \in V_h^p$ such that

$$\int_\Omega (\nabla_h \cdot (\mathbf{b}u_h) + cu_h)\, v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \left(\hat{h}_h - \mathbf{b}\cdot\mathbf{n}\, u_h^+\right) v_h^+ \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} \quad \forall v_h \in V_h^p.$$

Thus, we obtain the primal residual form: find $u_h \in V_h^p$ such that

$$\sum_{\kappa\in\mathcal{T}_h} \int_\kappa R(u_h) v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} r(u_h) v_h \,\mathrm{d}s + \int_\Gamma r_\Gamma(u_h) v_h \,\mathrm{d}s = 0 \quad \forall v_h \in V_h^p, \tag{56}$$

where $R(u_h)$, $r(u_h)$ and $r_\Gamma(u_h)$ denote the element, interior face and boundary residuals, respectively, given by

$$\begin{aligned}
R(u_h) &= f - \nabla_h \cdot (\mathbf{b}u_h) - cu_h && \text{in } \kappa, \kappa \in \mathcal{T}_h,\\
r(u_h) &= \mathbf{b}\cdot\mathbf{n}\, u_h^+ - \hat{h}(u_h^+, u_h^-, \mathbf{n}) && \text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h,\\
r_\Gamma(u_h) &= \mathbf{b}\cdot\mathbf{n}\, u_h^+ - h_\Gamma(u_h^+, \mathbf{n}) = \mathbf{b}\cdot\mathbf{n}\,(u_h^+ - g) && \text{on } \Gamma_-,\\
r_\Gamma(u_h) &= \mathbf{b}\cdot\mathbf{n}\, u_h^+ - h_\Gamma(u_h^+, \mathbf{n}) = 0 && \text{on } \Gamma_+.
\end{aligned}$$

Furthermore, we see that the exact solution $u$ to (45) satisfies (23) with $R(u) = 0$ and $r_\Gamma(u) = 0$. Furthermore, we have $r(u) = 0$ if (and only if) $\hat{h}(u, u, \mathbf{n}) = \mathbf{b}\cdot\mathbf{n}\, u$. Motivated by this, we arrive at following definition and statement.

**Definition 3.1** *A numerical flux function $\hat{h}$ is said to be* consistent *if*

$$\hat{h}(v, v, \mathbf{n}) = \mathbf{b}\cdot\mathbf{n}\, v, \tag{57}$$

*whenever $v$ is a smooth function.*

**Lemma 3.2 (Consistency)** *The discretization* (55) *of the linear advection equation* (45) *is* consistent *if and only if the numerical flux function $\hat{h}$ is* consistent.

## 3.5 Global and local conservation property

Let $S$ be the union of any collection of elements $\kappa \in \mathcal{T}_h$. Furthermore, let $\mathcal{T}_{h,S}$ be the corresponding part of the mesh $\mathcal{T}_h$, and let $\Gamma_{\mathcal{I}_S}$ denote the union of all interior faces of $\mathcal{T}_{h,S}$. Choosing $v_h \in V_h^p$ with $v_h = 1$ on $S$ and $v_h = 0$ on $\Omega\setminus S$ in (55) for $c = 0$, we obtain

$$\sum_{\kappa\subset S} \int_{\partial\kappa\setminus\partial S} \hat{h}(u_h^+, u_h^-, \mathbf{n}) \,\mathrm{d}s + \int_{\partial S_-} \hat{h}_h \,\mathrm{d}s + \int_{\partial S_+} \hat{h}_h \,\mathrm{d}s = \int_S f \,\mathrm{d}\mathbf{x}$$

Noting, that any interior edge $e \in \Gamma_{\mathcal{I}_S}$ occurs twice in the sum $\sum_{\kappa\in S}\int_{\partial\kappa\setminus\partial S}$ with opposite normals and states $u_h^+$ and $u_h^-$, we obtain

$$\sum_{e\in\Gamma_{\mathcal{I}_S}} \int_e \hat{h}(u_h^+, u_h^-, \mathbf{n}) + \hat{h}(u_h^-, u_h^+, -\mathbf{n}) \,\mathrm{d}s + \int_{\partial S_-} \hat{h}_h \,\mathrm{d}s + \int_{\partial S_+} \hat{h}_h \,\mathrm{d}s = \int_S f \,\mathrm{d}\mathbf{x}.$$

Motivated by this, we arrive at following definition and statement.

**Definition 3.3** *A numerical flux function* $\hat{h}(u^+, u^-, \mathbf{n})$ *is said to be* conservative *if*

$$\hat{h}(u^+, u^-, \mathbf{n}) = -\hat{h}(u^-, u^+, -\mathbf{n}).$$

**Lemma 3.4 (Conservation property)** *Let linear advection equation* (45) *with* $c \equiv 0$ *be discretized based on* (55)*. Then, the discretization is* conservative*, i.e., for any union* $S$ *of elements,*

$$\int_{\partial S_-} \hat{h}_h \, ds + \int_{\partial S_+} \hat{h}_h \, ds = \int_\Omega f \, d\mathbf{x}, \tag{58}$$

*if and only if the numerical flux function* $\hat{h}(u^+, u^-, \mathbf{n})$ *is* conservative.

*In particular, we then have the* global conservation property,

$$\int_{\Gamma_-} \mathbf{b}\cdot\mathbf{n}\, g \, ds + \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, u_h^+ \, ds = \int_\Omega f \, d\mathbf{x}, \tag{59}$$

*and the* local conservation property *for any* $\kappa \in \mathcal{T}_h$,

$$\int_{\partial\kappa_-} \hat{h}_h \, ds + \int_{\partial\kappa_+} \hat{h}_h \, ds = \int_\kappa f \, d\mathbf{x}. \tag{60}$$

## 3.6 Numerical flux functions

We have seen that for the discretization (55) to be consistent and conservative, the numerical flux $\hat{h}(\cdot,\cdot,\cdot)$ must be consistent and conservative. In the following, we introduce two different numerical flux functions which are both consistent and conservative.

**The mean value flux** First we define the *mean value flux* as follows,

$$h_{\mathrm{mv}}(u_h^+, u_h^-, \mathbf{n}) = \mathbf{b}\cdot\mathbf{n}\, \{\!\{u_h\}\!\}, \tag{61}$$

where

$$\{\!\{u_h\}\!\} = \frac{1}{2}\left(u_h^+ + u_h^-\right) \tag{62}$$

denotes the mean value of $u_h^+$ and $u_h^-$. This seems to be the most natural choice of a numerical flux function approximating $\mathbf{b}\cdot\mathbf{n}\, u$ based on $u^+$ and $u^-$. In fact, this flux is consistent and conservative. However, as we will see later, this flux leads to an *unstable* discontinuous Galerkin discretization.

**The upwind flux** We now define the *upwind flux* as follows,

$$h_{\mathrm{uw}}(u_h^+, u_h^-, \mathbf{n}) = \begin{cases} \mathbf{b}\cdot\mathbf{n}\, u_h^-, & \text{for } (\mathbf{b}\cdot\mathbf{n})(\mathbf{x}) < 0, \text{ i.e., } \mathbf{x} \in \partial\kappa_-, \\ \mathbf{b}\cdot\mathbf{n}\, u_h^+, & \text{for } (\mathbf{b}\cdot\mathbf{n})(\mathbf{x}) \geq 0, \text{ i.e., } \mathbf{x} \in \partial\kappa_+, \end{cases}, \tag{63}$$

where $\partial\kappa_-$ and $\partial\kappa_+$ are the inflow and outflow boundaries of element $\kappa$ defined by

$$\begin{aligned} \partial\kappa_- &= \{\mathbf{x} \in \partial\kappa, \mathbf{b}(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) < 0\}, \\ \partial\kappa_+ &= \{\mathbf{x} \in \partial\kappa, \mathbf{b}(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) \geq 0\} = \partial\kappa \setminus \partial\kappa_-. \end{aligned} \tag{64}$$

This flux always takes the value from upstream (upwind) direction. This numerical flux is consistent and conservative. Additionally, as we will see later, a discretization based on this flux is *stable*.

**Generic flux** The mean value flux and the upwind flux can be written in unified form as follows

$$h_{b_0}(u_h^+, u_h^-, \mathbf{n}) = \mathbf{b}\cdot\mathbf{n}\,\{\!\{u_h\}\!\} + b_0\,[u_h], \tag{65}$$

where

$$[u_h] = u_h^+ - u_h^- \tag{66}$$

denotes the (simple) jump of $u_h$. By setting $b_0 = 0$ the generic flux (65) reduces to the mean value flux (61) and by setting $b_0 = \frac{1}{2}|\mathbf{b}\cdot\mathbf{n}|$ it reduces to the upwind flux (63).

Replacing the numerical flux $\hat{h}$ in (55) by the generic flux $h_{b_0}$ we obtain the discontinuous Galerkin discretization: find $u_h \in V_h^p$ such that

$$L_h(u_h, v_h) \equiv \int_\Omega (\mathbf{b}u_h)\cdot\nabla_h v_h + c u_h v_h\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} \left(\mathbf{b}\cdot\mathbf{n}\,\{\!\{u_h\}\!\} + b_0\,[u_h]\right)\,v_h^+\,\mathrm{d}s$$
$$+ \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\,u_h v_h\,\mathrm{d}s = \int_\Omega f v_h\,\mathrm{d}\mathbf{x} - \int_{\Gamma_-}\mathbf{b}\cdot\mathbf{n}\,g v_h\,\mathrm{d}s \qquad \forall v_h \in V_h^p. \tag{67}$$

## 3.7 Stability and *a priori* error estimates

For the discontinuous Galerkin discretization (67) following properties can be shown (cf. Appendix A.3 for more details):

- Coercivity:

$$L_h(v,v) = \|c_0 v\|^2 + \sum_{e\in\Gamma_\mathcal{I}}\int_e b_0\,[v]^2\,\mathrm{d}s + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\,v^2\,\mathrm{d}s. \tag{68}$$

- Stability:

$$\|c_0 v\|^2 + \sum_{e\in\Gamma_\mathcal{I}}\int_e b_0\,[v]^2\,\mathrm{d}s + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\,v^2\,\mathrm{d}s \le C^2, \tag{69}$$

  where $C$ depends on the data $f$ and $g$ in (45).

We note, that the discontinuous Galerkin discretization based on the upwind flux, i.e., $b_0 = \frac{1}{2}|\mathbf{b}\cdot\mathbf{n}|$, has an improved stability as compared to the discretization based on the mean value flux where $b_0 = 0$. In fact, from (69) we see that we have control over $\sum_{e\in\Gamma_\mathcal{I}}\int_e [v]^2\,\mathrm{d}s$ for $b_0 \neq 0$ which we do not have for $b_0 = 0$. In fact, the DG discretization based on the upwind flux turns out to be stable whereas the DG discretization based on the mean value flux is unstable.

**Theorem 3.5 (*A priori* error estimate, (Brezzi et al., 2004))** *Let $u \in H^{p+1}(\Omega)$ be the exact solution to the linear advection equation* (45)*. Furthermore, let $u_h \in \tilde{V}_h^p := \{v_h \in L^2(\Omega) : v_h|_\kappa \in P_p(\kappa), \kappa \in \mathcal{T}_h\}$ be the solution to*

$$L_h(u_h, v_h) = \int_\Omega f v_h\,d\mathbf{x} - \int_{\Gamma_-}\mathbf{b}\cdot\mathbf{n}\,g v_h\,ds \qquad \forall v_h \in \tilde{V}_h^p,$$

*where $L_h(\cdot,\cdot)$ is the bilinear form given in* (67).

*Then, for $b_0 = c_e|\mathbf{b}\cdot\mathbf{n}|$ with $c_e > 0$ (which for $c_e = 1/2$ gives the* upwind flux*) we have*

$$|||u - u_h|||_{b_0} \le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}, \tag{70}$$

*and for $b_0 = 0$, i.e., when using the* mean value flux (61)*, we have*

$$|||u - u_h|||_{b_0} \le Ch^{p}|u|_{H^{p+1}(\Omega)}, \tag{71}$$

*where $|||\cdot|||_{b_0}^2$ is the DG-norm defined by*

$$|||v|||_{b_0}^2 = \|c_0 v\|^2 + \sum_{e\in\Gamma_\mathcal{I}} \int_e b_0\,[v]^2\,ds + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\,v^2\,ds. \tag{72}$$

**Proof:** See (Brezzi et al., 2004) or Appendix A.3. □

From the proof one can see that the order of convergence is connected to the stability of the discretization, in particular of the interior face terms. For sufficiently smooth solutions, $u \in H^{p+1}(\Omega)$, the discretization based on the upwind flux is of order $\mathcal{O}(h^{p+1/2})$ and the discretization based on the mean value flux is of the order $\mathcal{O}(h^p)$. In contrast to that the order of convergence is reduced for solutions with a lower smoothness. In fact, for $u \in H^{s+1}(\Omega)$ with $s < p$ the estimates (70) and (71) are replaced by the estimates

$$|||u - u_h|||_{b_0} \le Ch^{s+1/2}|u|_{H^{s+1}(\Omega)}, \tag{73}$$

for the upwind flux, and by

$$|||u - u_h|||_{b_0} \le Ch^{s}|u|_{H^{s+1}(\Omega)}, \tag{74}$$

for the mean value flux, respectively. In particular, for $u \in H^1(\Omega)$ we see that the discontinuous Galerkin solution based on the mean value flux does not converge under $h \to 0$. In fact, for $u \in H^1(\Omega)$, the discretization based on the mean value flux is unstable, whereas the upwind flux yields a stable discretization. Furthermore, we see that already the use of *some* upwinding, i.e., when $b_0 = c_e|\mathbf{b}\cdot\mathbf{n}|$ with $c_e > 0$, gives a stable discretization.

Finally, we note that estimate (70) is suboptimal by $h^{1/2}$ as compared to the $\mathcal{O}(h^{p+1})$ approximation order of $V_h^p$. However, (Peterson, 1991) confirmed by considering so–called Peterson meshes, that $O(h^{p+1/2})$ is actually a sharp estimate.

As the DG discretization based on the mean value flux $h_{\text{mv}}$ is unstable, in the following we concentrate on the discretization based on the upwind flux $h_{\text{uw}}$ which is stable. By substituting the upwind flux as defined in (63), and the boundary flux as defined in (54) into the discretization (55), we obtain following discontinuous Galerkin discretization of the linear advection equation (45): find $u_h \in V_h^p$ such that

$$\int_\Omega (-\mathbf{b}u_h)\cdot\nabla_h v_h + c u_h v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa_+} \mathbf{b}\cdot\mathbf{n}\,u_h^+\,v_h^+\,\mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa_-\setminus\Gamma} \mathbf{b}\cdot\mathbf{n}\,u_h^-\,v_h^+\,\mathrm{d}s$$
$$= \int_\Omega f v_h\,\mathrm{d}\mathbf{x} - \int_{\Gamma_-} \mathbf{b}\cdot\mathbf{n}\,g\,v_h^+\,\mathrm{d}s \qquad \forall v_h \in V_h^p. \tag{75}$$

## 3.8 Adjoint consistency and error estimates in $J(\cdot)$

Let the target quantity (49) be discretized as follows

$$J_h(u_h) = J(u_h) = \int_\Omega j_\Omega\, u_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_+} j_\Gamma\, u_h \,\mathrm{d}s. \tag{76}$$

In the following, we want to show that the discontinuous Galerkin discretization (75) of the linear advection equation (45) in combination with the discretization (76) of the target quantity is adjoint consistent. To this end, we substitute

$$\sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa_-\setminus\Gamma} \mathbf{b}\cdot\mathbf{n}\, u_h^- v_h^+ \,\mathrm{d}s = -\sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa_+\setminus\Gamma} \mathbf{b}\cdot\mathbf{n}\, u_h^+ v_h^- \,\mathrm{d}s$$

in (75), and find that the discrete adjoint problem (25) to the discretization (55) is given by: find $z_h \in V_h$ such that

$$L_h(w_h, z_h) \equiv \int_\Omega w_h\left(-\mathbf{b}\cdot\nabla_h z_h + c z_h\right)\mathrm{d}\mathbf{x} + \sum_\kappa \int_{\partial\kappa_+\setminus\Gamma} w_h^+\, \mathbf{b}{\cdot}\mathbf{n}[z_h]\,\mathrm{d}s + \int_{\Gamma_+} w_h^+\, \mathbf{b}{\cdot}\mathbf{n}\, z_h^+ = J_h(w_h),$$

for all $w_h \in V_h$. Hence, for the discrete target functional $J_h(\cdot)$ given by (76), we have (26) with

$$\begin{aligned}
R^*(z_h) &= j_\Omega + \mathbf{b}\cdot\nabla_h z_h - c z_h && \text{in } \kappa, \kappa\in\mathcal{T}_h,\\
r^*(z_h) &= -\mathbf{b}\cdot\mathbf{n}\,[z_h] && \text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h,\\
r^*(z_h) &= j_\Gamma - \mathbf{b}\cdot\mathbf{n}\, z_h^+ && \text{on } \Gamma_+,
\end{aligned}$$

and $r^*(z_h) \equiv 0$ on $\Gamma_-$. As (27) with (28) holds for the exact (and assumingly smooth) solution $z$ to (50), we conclude, that discretization (75) in combination with (76) is adjoint consistent.

**Corollary 3.6 (*A priori* error estimate in $J(\cdot)$)** *Let $u_h \in V_h^p$ be the solution to (75) and $J(\cdot)$ be given by (76). Furthermore, we assume that the solutions $u$ to the advection equation (45) is smooth, $u \in H^{p+1}(\Omega)$, and the solution $z$ to the adjoint equation (50) is smooth, $z \in H^{p+1}(\Omega)$. Then, there is a constant $C > 0$ such that*

$$|J(u) - J_h(u_h)| \le C h^{2p+1} |u|_{H^{p+1}(\Omega)} |z|_{H^{p+1}(\Omega)} \qquad \forall u \in H^{p+1}(\Omega). \tag{77}$$

**Proof:** See (Houston and Süli, 2001; Harriman et al., 2003). □

Here, we see that the order of convergence $\mathcal{O}(h^{2p+1})$ in $J(\cdot)$ is twice the order of convergence $\mathcal{O}(h^{p+1/2})$ in the DG-norm $|||\cdot|||_{b_0}$ (cf. Theorem 3.5). Note, that for this *order doubling* an essential ingredient is the adjoint consistency of the discretization.

Furthermore note, that the order of convergence in $J(\cdot)$ is $\mathcal{O}(h^{2p+1})$ provided both primal and adjoint solutions are smooth, $u \in H^{p+1}(\Omega)$ and $z \in H^{p+1}(\Omega)$. If, however, $u$ or $z$ are less regular we obtain an estimate with a correspondingly reduced order of convergence in $J(\cdot)$:

**Corollary 3.7 (*A priori* error estimate in $J(\cdot)$ with reduced regularity)** *Let $u_h \in V_h^p$ be the solution to* (67) *and $J(\cdot)$ be given by* (76). *Furthermore, we assume that $u \in H^{s+1}(\Omega)$ and $z \in H^{\tilde{s}+1}(\Omega)$ hold for the exact solutions $u$ and $z$ to the primal and adjoint problems* (45) *and* (50)*, respectively. Then, there is a constant $C > 0$ such that*

$$|J(u) - J_h(u_h)| \leq Ch^{t+\tilde{t}+1}|u|_{H^{t+1}(\Omega)}|z|_{H^{\tilde{t}+1}(\Omega)} \qquad \forall u \in H^{s+1}(\Omega), \tag{78}$$

*where $t = \min(s, p)$ and $\tilde{t} = \min(\tilde{s}, p)$.*

**Proof:** See (Houston and Süli, 2001; Harriman et al., 2003). □

# 4 DG discretization of Poisson's equation

In this section we derive and analyze DG discretizations of Poisson's equation.

## 4.1 The Poisson's equation

Let $\Omega \subset \mathbb{R}^2$ again be a bounded open domain with boundary $\Gamma = \partial\Omega$. We consider the elliptic model problem

$$-\Delta u = f \quad \text{in } \Omega, \qquad u = g_D \quad \text{on } \Gamma_D, \qquad \mathbf{n}\cdot\nabla u = g_N \quad \text{on } \Gamma_N, \tag{79}$$

where $f \in L^2(\Omega)$, $g_D \in L^2(\Gamma_D)$ and $g_N \in L^2(\Gamma_N)$ are given functions. We assume that $\Gamma_D$ and $\Gamma_N$ are disjoint subsets with union $\Gamma$, that is $\Gamma_D \cup \Gamma_N = \Gamma$ and $\Gamma_D \cap \Gamma_N = \emptyset$. Furthermore, we assume that $\Gamma_D \neq \emptyset$. Problem (79) represents the general Dirichlet-Neumann problem of Poisson's equation.

## 4.2 The continuous adjoint equation

In order to derive the continuous adjoint equation, we multiply the left hand side of (79) by $z$ and integrate twice by parts over the domain $\Omega$. Thereby, we obtain

$$(-\Delta u, z)_\Omega = (\nabla u, \nabla z)_\Omega - (\mathbf{n}\cdot\nabla u, z)_\Gamma = (u, -\Delta z)_\Omega + (u, \mathbf{n}\cdot\nabla z)_\Gamma - (\mathbf{n}\cdot\nabla u, z)_\Gamma.$$

Splitting the boundary terms according to $\Gamma = \Gamma_D \cup \Gamma_N$ and shuffling terms we arrive at

$$(-\Delta u, z)_\Omega + (u, -\mathbf{n}\cdot\nabla z)_{\Gamma_D} + (\mathbf{n}\cdot\nabla u, z)_{\Gamma_N} = (u, -\Delta z)_\Omega + (\mathbf{n}\cdot\nabla u, -z)_{\Gamma_D} + (u, \mathbf{n}\cdot\nabla z)_{\Gamma_N}.$$

Comparing with the compatibility condition (16),

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

we see that for $Lu = -\Delta u$ in $\Omega$ and

$$\begin{aligned} &Bu = u, && Cu = \mathbf{n}\cdot\nabla u && \text{on } \Gamma_D,\\ &Bu = \mathbf{n}\cdot\nabla u, && Cu = u && \text{on } \Gamma_N, \end{aligned}$$

the adjoint operators are given by $L^*z = -\Delta z$ on $\Omega$ and

$$\begin{aligned} &B^*z = -z, && C^*z = -\mathbf{n}\cdot\nabla z && \text{on } \Gamma_D,\\ &B^*z = \mathbf{n}\cdot\nabla z, && C^*z = z && \text{on } \Gamma_N. \end{aligned}$$

In particular, for

$$\begin{aligned} J(u) &= \int_\Omega j_\Omega\, u \,\mathrm{d}\mathbf{x} + \int_\Gamma j_\Gamma\, Cu \,\mathrm{d}s \\ &= \int_\Omega j_\Omega\, u \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\, \mathbf{n}\cdot\nabla u \,\mathrm{d}s + \int_{\Gamma_N} j_N\, u \,\mathrm{d}s, \end{aligned} \tag{80}$$

the continuous adjoint problem is given by

$$\begin{aligned} -\Delta z &= j_\Omega \quad \text{in } \Omega,\\ -z &= j_D \quad \text{on } \Gamma_D,\\ \mathbf{n}\cdot\nabla z &= j_N \quad \text{on } \Gamma_N. \end{aligned} \tag{81}$$

**Remark 4.1** *The differential operator $L^*$ given by $L^*z = -\Delta z$ is the adjoint operator to the differential operator $L$ given by $Lu = -\Delta u$ of the primal problem. As the Laplace operator is self-adjoint, the adjoint problem to Poisson's equation is again Poisson's equation.*

**Remark 4.2** *The right hand side $j_\Omega$ in (81) may be any arbitrary (but fixed) function in $L^2(\Omega)$. Depending on the "purpose" of the adjoint problem the function $j_\Omega$ may be chosen appropriately. For example, when deriving* a priori *error estimates in $L^2(\Omega)$ one can choose $j_\Omega = e = u - u_h$ which gives $\|e\|^2 = \hat{L}_h(e, z)$. An adjoint problem like in (81) will also be required when deriving error estimates with respect to target quantities of the form $J(v) = \int_\Omega j_\Omega v \, d\mathbf{x}$ which gives $J(e) = \hat{L}_h(e, z)$.*

## 4.3 Derivation of the DG discretization

Like in Section 3.3 we assume that the computational domain $\Omega \in \mathbb{R}^2$ is subdivided into a shape regular mesh $\mathcal{T}_h = \{\kappa\}$ consisting of elements $\kappa$. In addition to the finite element space $V_h^p$ of *scalar* discrete functions defined in (52) we define the space $\boldsymbol{\Sigma}_h^p := [V_h^p]^2$ of vector-valued discontinuous piecewise polynomial functions of degree $p \geq 0$.

For deriving a discontinuous Galerkin discretization of (79), we begin by rewriting (79) as a system of first-order equations as follows

$$\boldsymbol{\sigma} = \nabla u, \qquad -\nabla \cdot \boldsymbol{\sigma} = f \quad \text{in } \Omega, \qquad u = g_D \quad \text{on } \Gamma_D, \qquad \mathbf{n} \cdot \nabla u = g_N \quad \text{on } \Gamma_N. \tag{82}$$

We multiply the first and second equation by test functions $\boldsymbol{\tau}$ and $v$, respectively, integrate over an element $\kappa \in \mathcal{T}_h$, and integrate by parts. Thus

$$\begin{aligned} \int_\kappa \boldsymbol{\sigma} \cdot \boldsymbol{\tau} \, \mathrm{d}\mathbf{x} &= -\int_\kappa u \, \nabla \cdot \boldsymbol{\tau} \, \mathrm{d}\mathbf{x} + \int_{\partial\kappa} u \, \mathbf{n} \cdot \boldsymbol{\tau} \, \mathrm{d}s, \\ \int_\kappa \boldsymbol{\sigma} \cdot \nabla v \, \mathrm{d}\mathbf{x} &= \int_\kappa f v \, \mathrm{d}\mathbf{x} + \int_{\partial\kappa} \boldsymbol{\sigma} \cdot \mathbf{n} \, v \, \mathrm{d}s, \end{aligned} \tag{83}$$

where $\mathbf{n}$ is the unit outward normal vector to $\partial\kappa$.

Then, we sum (83) over all elements $\kappa \in \mathcal{T}_h$, and replace the functions $u$ and $\boldsymbol{\sigma}$ by discrete functions $u_h \in V_h^p$ and $\boldsymbol{\sigma}_h \in \boldsymbol{\Sigma}_h^p$ and the test functions $v$ and $\boldsymbol{\tau}$ by discrete test functions $v_h \in V_h^p$ and $\boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h^p$. In addition, since the discrete functions $u_h$ and $\boldsymbol{\sigma}_h$ may be discontinuous at interfaces $\partial\kappa \cap \partial\kappa'$ between elements $\kappa, \kappa' \in \mathcal{T}_h$, $\kappa' \neq \kappa$, we replace them by numerical flux functions $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ which are approximations to $u$ and $\boldsymbol{\sigma} = \nabla u$, respectively. Thus we obtain the following discretization in *system flux formulation*: find $u_h \in V_h^p$ and $\boldsymbol{\sigma}_h \in \boldsymbol{\Sigma}_h^p$ such that

$$\int_\Omega \boldsymbol{\sigma}_h \cdot \boldsymbol{\tau}_h \, \mathrm{d}\mathbf{x} = -\int_\Omega u_h \nabla_h \cdot \boldsymbol{\tau}_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa} \hat{u}_h \, \mathbf{n} \cdot \boldsymbol{\tau}_h \, \mathrm{d}s \quad \forall \boldsymbol{\tau}_h \in \boldsymbol{\Sigma}_h^p, \tag{84}$$

$$\int_\Omega \boldsymbol{\sigma}_h \cdot \nabla_h v_h \, \mathrm{d}\mathbf{x} = \int_\Omega f v_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n} \, v_h \, \mathrm{d}s \qquad \forall v_h \in V_h^p. \tag{85}$$

Here, $\hat{u}_h = \hat{u}(u_h) = \hat{u}(u_h^+, u_h^-)$ is a scalar numerical flux function, and $\hat{\boldsymbol{\sigma}}_h = \hat{\boldsymbol{\sigma}}(u_h, \nabla u_h) = \hat{\boldsymbol{\sigma}}(u_h^+, u_h^-, \nabla u_h^+, \nabla u_h^-)$ is a vector-valued numerical flux function. Depending on the particular choice of the numerical flux functions $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ several different DG discretizations can be derived, each with specific properties with respect to stability and accuracy. Examples of numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ will be given later.

Equations (84) and (85) represent a first order system in $u_h$ and $\boldsymbol{\sigma}_h$ with three times as many unknowns as the original (scalar) problem in $u_h$. In order to reduce the problem size, the auxiliary variable $\boldsymbol{\sigma}_h$ is usually eliminated to gain a so-called *primal formulation* involving only the primal variable $u_h$. To this end, we perform a second integration by parts on each element $\kappa$ in (84) and set $\boldsymbol{\tau}_h = \nabla_h v_h$ which gives

$$\int_\Omega \boldsymbol{\sigma}_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} = \int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{u}_h - u_h)\mathbf{n}\cdot\nabla_h v_h \,\mathrm{d}s. \tag{86}$$

Substituting (86) into (85) yields the element-based *primal formulation*: find $u_h \in V_h^p$ such that

$$\int_\Omega \nabla_h u_h\cdot\nabla_h v_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\, v_h \,\mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{u}_h - u_h)\mathbf{n}\cdot\nabla_h v_h \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} \quad \forall v_h \in V_h^p. \tag{87}$$

We note that (87) is in *element-based* form, i.e., it is given in terms of $\sum_\kappa \int_{\partial\kappa}$. This means that each interior face $e\in\Gamma_\mathcal{I}$ occurs twice in the sum over all elements $\kappa$ (once in $\int_{\partial\kappa}$ and once in $\int_{\partial\kappa'}$ for $\kappa'\neq\kappa$ and $e = \partial\kappa\cap\partial\kappa'\neq 0$). Defining following mean-value and jump operators on interior faces $e\in\Gamma_\mathcal{I}$ and on boundary faces $e\in\Gamma$, for scalar-valued functions $q_h\in V_h^p$,

$$\begin{aligned} \{\!\{q_h\}\!\} &= \tfrac{1}{2}(q_h^+ + q_h^-) \quad &&\text{on } \Gamma_\mathcal{I}, \qquad & \{\!\{q_h\}\!\} &= q_h^+ \quad &&\text{on } \Gamma,\\ [\![q_h]\!] &= q_h^+\mathbf{n}^+ + q_h^-\mathbf{n}^- \quad &&\text{on } \Gamma_\mathcal{I}, \qquad & [\![q_h]\!] &= q_h^+\mathbf{n}^+, \quad &&\text{on } \Gamma, \end{aligned} \tag{88}$$

and for vector-valued functions $\boldsymbol{\phi}_h \in \boldsymbol{\Sigma}_h^p$,

$$\begin{aligned} \{\!\{\boldsymbol{\phi}_h\}\!\} &= \tfrac{1}{2}(\boldsymbol{\phi}_h^+ + \boldsymbol{\phi}_h^-) \quad &&\text{on } \Gamma_\mathcal{I}, \qquad & \{\!\{\boldsymbol{\phi}_h\}\!\} &= \boldsymbol{\phi}_h^+ \quad &&\text{on } \Gamma,\\ [\![\boldsymbol{\phi}_h]\!] &= \boldsymbol{\phi}_h^+\cdot\mathbf{n}_h^+ + \boldsymbol{\phi}_h^-\cdot\mathbf{n}^- \quad &&\text{on } \Gamma_\mathcal{I} \qquad & [\![\boldsymbol{\phi}_h]\!] &= \boldsymbol{\phi}_h^+\cdot\mathbf{n}^+ \quad &&\text{on } \Gamma. \end{aligned} \tag{89}$$

the element-based primal formulation (87) can equivalently be written in following *face-based primal formulation* (cf. Appendix A.4 for the derivation): find $u_h\in V_h^p$ such that

$$\begin{aligned} \int_\Omega \nabla_h u_h\cdot\nabla_h v_h \,\mathrm{d}\mathbf{x} &+ \int_{\Gamma_\mathcal{I}\cup\Gamma} \left([\![\hat{u}_h - u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\} - \{\!\{\hat{\boldsymbol{\sigma}}_h\}\!\}\cdot[\![v_h]\!]\right) \mathrm{d}s \\ &+ \int_{\Gamma_\mathcal{I}} \left(\{\!\{\hat{u}_h - u_h\}\!\}\,[\![\nabla_h v_h]\!] - [\![\hat{\boldsymbol{\sigma}}_h]\!]\,\{\!\{v_h\}\!\}\right) \mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} \quad \forall v_h\in V_h^p. \end{aligned} \tag{90}$$

## 4.4 Consistency

Integrating (87) back by parts on each element $\kappa$ we obtain following equivalent form of the discretization: find $u_h\in V_h^p$ such that

$$-\int_\Omega \Delta_h u_h\, v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\nabla_h u_h - \hat{\boldsymbol{\sigma}}_h)\cdot\mathbf{n}\, v_h \,\mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{u}_h - u_h)\mathbf{n}\cdot\nabla_h v_h \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x}$$

for all $v_h \in V_h^p$. This can expressed in the following primal residual form,

$$\int_\Omega R(u_h)v_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} r(u_h)v_h + \boldsymbol{\rho}(u_h)\cdot\nabla_h v_h \,\mathrm{d}s \\ + \int_\Gamma r_\Gamma(u_h)v_h + \boldsymbol{\rho}_\Gamma(u_h)\cdot\nabla_h v_h \,\mathrm{d}s = 0 \quad \forall v_h\in V_h,$$

where the residuals are given by $R(u_h) = f + \Delta_h u_h$ on $\kappa\in\mathcal{T}_h$, and

$$\begin{aligned} r(u_h) &= (\hat{\boldsymbol{\sigma}}_h - \nabla u_h)\cdot\mathbf{n}, & \boldsymbol{\rho}(u_h) &= (u_h - \hat{u}_h)\mathbf{n} & &\text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h,\\ r_\Gamma(u_h) &= (\hat{\boldsymbol{\sigma}}_h - \nabla u_h)\cdot\mathbf{n}, & \boldsymbol{\rho}_\Gamma(u_h) &= (u_h - \hat{u}_h)\mathbf{n} & &\text{on } \Gamma = \Gamma_D\cup\Gamma_N. \end{aligned} \tag{91}$$

Note, that this is an extension of the primal residual form in (22) to include face and boundary residuals $\boldsymbol{\rho}(u_h)$ and $\boldsymbol{\rho}_\Gamma(u_h)$ multiplied by $\nabla_h v_h$. For the discretization to be consistent, all residuals in (91) must vanish if evaluated for the exact solution $u$ to (79),

$$\begin{aligned} r(u) &= (\hat{\boldsymbol{\sigma}}(u,\nabla u) - \nabla u)\cdot\mathbf{n} = 0, & \boldsymbol{\rho}(u) &= (u - \hat{u}(u))\mathbf{n} = 0 & &\text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h,\\ r_\Gamma(u) &= (\hat{\boldsymbol{\sigma}}(u,\nabla u) - \nabla u)\cdot\mathbf{n} = 0, & \boldsymbol{\rho}_\Gamma(u) &= (g_D - \hat{u}(u))\mathbf{n} = 0 & &\text{on } \Gamma_D,\\ r_\Gamma(u) &= \hat{\boldsymbol{\sigma}}(u,\nabla u)\cdot\mathbf{n} - g_N = 0, & \boldsymbol{\rho}_\Gamma(u) &= (u - \hat{u}(u))\mathbf{n} = 0 & &\text{on } \Gamma_N. \end{aligned} \tag{92}$$

Motivated by this we arrive at following definition and statement:

**Definition 4.3** *The numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are said to be* consistent *if*

$$\begin{aligned} &\hat{u}(v) = v, & &\hat{\boldsymbol{\sigma}}(v,\nabla v) = \nabla v, & &\text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h,\\ &\hat{u}(v) = g_D, & &\hat{\boldsymbol{\sigma}}(v,\nabla v) = \nabla v, & &\text{on } \Gamma_D,\\ &\hat{u}(v) = v, & &\hat{\boldsymbol{\sigma}}(v,\nabla v)\cdot\mathbf{n} = g_N, & &\text{on } \Gamma_N, \end{aligned}$$

*whenever $v$ is a smooth function satisfying the Dirichlet and Neumann boundary conditions in* (79)*.*

**Lemma 4.4 (Consistency)** *The discretization* (87) *of Poisson's equation* (79) *is* consistent *if and only if the numerical flux functions $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are* consistent*.*

In order to simplify the notation in the proceeding analysis let us define $\hat{u}_\theta$ on $\Gamma_D$ by

$$\hat{u}_\theta(u_h) = u_h + \theta(u_h - g_D) = \begin{cases} g_D & \text{for } \theta = -1,\\ 2u_h - g_D & \text{for } \theta = 1. \end{cases} \tag{93}$$

**Remark 4.5** *This notation covers a symmetric discretization of boundary terms for $\theta = -1$ like for the SIPG or BR2 and a non-symmetric discretization for $\theta = 1$ like for NIPG. Thus* (93) *covers all discretizations that will be introduced in Section 4.7 and is thus sufficient for this work. We avoid a more general notation based on e.g. $\hat{u}_u u_h + \hat{u}_g g_D$ in order to keep notation simple.*

We note that $\hat{u}_\theta(u_h)$ is consistent (cf. Definition 4.3), as it reduces to $\hat{u}_\theta(v) = g_D$ for any function $v$ satisfying the Dirichlet boundary condition in (79).

Let us now set $\hat{u}(u_h) = u_\theta(u_h)$ on $\Gamma_D$ and $\hat{\boldsymbol{\sigma}}(u_h, \nabla u_h) \cdot \mathbf{n} = g_N$ on $\Gamma_N$. Then we can rewrite (87) as follows: find $u_h \in V_h^p$ such that

$$\begin{aligned}
\hat{L}_h(u_h, v_h) &= F_h(v_h) \qquad \forall v_h \in V_h^p, \\
\hat{L}_h(u_h, v_h) &= \int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma_N} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\, v_h \,\mathrm{d}s \\
&\quad + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma_D} (\hat{u}_h - u_h)\mathbf{n} \cdot \nabla_h v_h \,\mathrm{d}s + \int_{\Gamma_D} \theta u_h\, \mathbf{n} \cdot \nabla_h v_h \,\mathrm{d}s, \\
F_h(v_h) &= \int_\Omega f v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} \theta g_D\, \mathbf{n} \cdot \nabla_h v_h \,\mathrm{d}s + \int_{\Gamma_N} g_N v_h \,\mathrm{d}s.
\end{aligned} \tag{94}$$

The bilinear form in (94) is denoted by $\hat{L}_h$ (and not $L_h$) as it includes the (still unspecified) numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$.

## 4.5 Global and local conservation property

Let $S$ be the union of any collection of elements, $\mathcal{T}_{h,S}$ the corresponding part of the mesh, and $\Gamma_{\mathcal{I}_S}$ the collection of interior faces of $\mathcal{T}_{h,S}$ like in Section 3.5. Choosing $v_h \in V_h^p$ with $v_h = 1$ on $S$ and $v_h = 0$ on $\Omega \setminus S$ in (87), we obtain

$$\sum_{\kappa\subset S} \int_{\partial\kappa\setminus\partial S} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n} \,\mathrm{d}s + \int_{\partial S} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n} \,\mathrm{d}s + \int_S f \,\mathrm{d}\mathbf{x} = 0. \tag{95}$$

Motivated by this we arrive at following definition and statement:

**Definition 4.6** *A numerical flux function $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}(u^+, u^-, \nabla u^+, \nabla u^-)$ is said to be* conservative *if*

$$\hat{\boldsymbol{\sigma}}(u^+, u^-, \nabla u^+, \nabla u^-) = \hat{\boldsymbol{\sigma}}(u^-, u^+, \nabla u^-, \nabla u^+).$$

*Note, that this property is also referred to as: "$\hat{\boldsymbol{\sigma}}$ is* single-valued*" (cf. (Arnold et al., 2002)).*

Furthermore, note that $\hat{\boldsymbol{\sigma}}$ is single-valued if and only if

$$\begin{aligned}
[\![\hat{\boldsymbol{\sigma}}]\!] &= \hat{\boldsymbol{\sigma}}(u^+, u^-, \nabla u^+, \nabla u^-) \cdot \mathbf{n} + \hat{\boldsymbol{\sigma}}(u^-, u^+, \nabla u^-, \nabla u^+) \cdot (-\mathbf{n}) \\
&= \big(\hat{\boldsymbol{\sigma}}(u^+, u^-, \nabla u^+, \nabla u^-) - \hat{\boldsymbol{\sigma}}(u^-, u^+, \nabla u^-, \nabla u^+)\big) \cdot \mathbf{n} = 0.
\end{aligned} \tag{96}$$

**Lemma 4.7 (Conservation property)** *Let Poisson's equation* (79) *be discretized based on* (87)*. Then, the discretization is* conservative*, i.e., for any union $S$ of elements,*

$$\int_{\partial S} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\, ds + \int_S f\, d\mathbf{x} = 0,$$

*if and only if the numerical flux function $\hat{\boldsymbol{\sigma}}_h$ is* conservative*.*

*In particular, we then have the* global conservation property*,*

$$\int_\Gamma \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\, ds + \int_\Omega f\, d\mathbf{x} = 0,$$

*and the* local conservation property *for any $\kappa \in \mathcal{T}_h$,*

$$\int_{\partial\kappa} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\, ds + \int_\kappa f\, d\mathbf{x} = 0.$$

**Proof:** Note, that in (95) the term

$$\sum_{\kappa\subset S}\int_{\partial\kappa\setminus\partial S}\hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\,\mathrm{d}s=\sum_{e\in\Gamma_{\mathcal{I}_S}}\int_e\big(\hat{\boldsymbol{\sigma}}(u_h^+,u_h^-,\nabla_h u_h^+,\nabla_h u_h^-)\cdot\mathbf{n}+\hat{\boldsymbol{\sigma}}(u_h^-,u_h^+,\nabla_h u_h^-,\nabla_h u_h^+)\cdot(-\mathbf{n})\big)\,\mathrm{d}s$$

vanishes if and only if $\hat{\boldsymbol{\sigma}}_h$ is conservative. □

We end this section by defining analogously the following.

**Definition 4.8** *A numerical flux function $\hat{u}=\hat{u}(u^+,u^-)$ is said to be* single-valued *if*

$$\hat{u}(u^+,u^-)=\hat{u}(u^-,u^+). \tag{97}$$

Note that $\hat{u}$ is single-valued if and only if

$$[\![\hat{u}]\!]=\hat{u}(u^+,u^-)\mathbf{n}+\hat{u}(u^-,u^+)\,(-\mathbf{n})=\big(\hat{u}(u^+,u^-)-\hat{u}(u^-,u^+)\big)\,\mathbf{n}=0. \tag{98}$$

## 4.6 Adjoint consistency

Let $\hat{u}_h=\hat{u}(u_h)$ and $\hat{\boldsymbol{\sigma}}_h=\hat{\boldsymbol{\sigma}}(u_h,\nabla u_h)$ be consistent numerical flux functions (cf. Definition 4.3). Then we discretize the target quantity (80) as follows

$$J_h(u_h)=\int_\Omega j_\Omega\,u_h\,\mathrm{d}\mathbf{x}+\int_{\Gamma_D}j_D\,\hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\,\mathrm{d}s+\int_{\Gamma_N}j_N\,\hat{u}_h\,\mathrm{d}s, \tag{99}$$

Note, that this $J_h(u_h)$ is a *consistent* discretization of the target quantity (80), i.e., the exact solution $u$ to Poisson's equation (79) satisfies $J_h(u)=J(u)$. To see this, recall that due to consistency of $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ the exact solution $u$ satisfies $\hat{u}(u)=u$ on $\Gamma_N$ and $\hat{\boldsymbol{\sigma}}(u)=\nabla u$ on $\Gamma_D$ (cf. Definition 4.3).

In the following, we want to show that the discretization (94) of Poisson's equation (79) in combination with the discretization (99) of the target quantity (80) is *adjoint consistent*, i.e., the solution $z$ to the adjoint equation (81) satisfies

$$\hat{L}_h(w,z)=J_h(w)\quad w\in V. \tag{100}$$

To show this, we perform an integration by parts on (94) to obtain

$$\begin{aligned}\hat{L}_h(u_h,v_h)=-\int_\Omega u_h\cdot\Delta_h v_h\,\mathrm{d}\mathbf{x}-\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma_N}\hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\,v_h\,\mathrm{d}s+\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\hat{u}_h\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s\\+\int_{\Gamma_N}\hat{u}_h\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s+\int_{\Gamma_D}(\theta+1)u_h\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s.\end{aligned}$$

Now we replace $u_h$ by a test function $w$, and $v_h$ by a smooth function $z$ and obtain

$$\begin{aligned}\hat{L}_h(w,z)=-\int_\Omega w\,\Delta_h z\,\mathrm{d}\mathbf{x}-\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma_N}\hat{\boldsymbol{\sigma}}(w,\nabla w)\cdot\mathbf{n}\,z\,\mathrm{d}s+\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\hat{u}(w)\,\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s\\+\int_{\Gamma_N}\hat{u}(w)\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s+\int_{\Gamma_D}(\theta+1)w\,\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s.\end{aligned} \tag{101}$$

Using (81) and rewriting (101) in terms of interior faces $e \in \Gamma_{\mathcal{I}}$ we arrive at

$$\begin{aligned}\hat{L}_h(w,z) = &\int_\Omega w\,\Delta_h z\,\mathrm{d}\mathbf{x} - \sum_{e\in\Gamma_{\mathcal{I}}}\int_e [\![\hat{\boldsymbol{\sigma}}]\!]\,z\,\mathrm{d}s + \sum_{e\in\Gamma_{\mathcal{I}}}\int_e [\![\hat{u}]\!]\cdot\nabla_h z\,\mathrm{d}s \\ &- \int_{\Gamma_D}\hat{\boldsymbol{\sigma}}(w)\cdot\mathbf{n}\,z\,\mathrm{d}s + \int_{\Gamma_N}\hat{u}(w)\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s + \int_{\Gamma_D}(\theta+1)w\,\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s. \quad (102)\end{aligned}$$

which, due to (96) and (98), reduces to

$$\hat{L}_h(w,z) = \int_\Omega w\,\Delta_h z\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D}\hat{\boldsymbol{\sigma}}(w)\cdot\mathbf{n}\,(-z)\,\mathrm{d}s + \int_{\Gamma_N}\hat{u}(w)\mathbf{n}\cdot\nabla_h z\,\mathrm{d}s \quad (103)$$

if and only if the numerical fluxes $\hat{\boldsymbol{\sigma}}$ and $\hat{u}$ are single-valued (cf. Definitions 4.6 and 4.8) and $\theta = -1$. Thus we arrive at following statement:

**Theorem 4.9** *Let Poisson's equation* (79) *be discretized with $\hat{L}_h(\cdot,\cdot)$ as given in* (94). *Furthermore, let the target quantity* (80) *be discretized with $J_h(\cdot)$ as given in* (99). *Then the discretization is* adjoint consistent, *i.e., the exact solution $z$ to the adjoint equation* (81) *satisfies*

$$\hat{L}_h(w,z) = \int_\Omega w\,j_\Omega\,d\mathbf{x} + \int_{\Gamma_D}\hat{\boldsymbol{\sigma}}(w)\cdot\mathbf{n}\,j_D\,ds + \int_{\Gamma_N}\hat{u}(w)\,j_N\,ds = J_h(w), \quad (104)$$

*if and only if the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are* conservative *(*single-valued*) and $\theta = -1$.*

**Remark 4.10** *Note, that we can also show adjoint consistency by deriving the adjoint residual form* (26) *and showing that the adjoint residuals vanish for the adjoint solution $z$. In fact, this analysis has been performed in (Hartmann, 2007a, 2008b) for numerical fluxes $\hat{\boldsymbol{\sigma}}_h$ being specialized for the interior penalty (IP) discretization (cf. Section 4.7). However, for generality, here we want to show adjoint consistency independent of the specific flux employed. Derivation of the adjoint residual form* (26) *would require the numerical fluxes $\hat{u}(w) \equiv \hat{u}(w^+, w^-)$ and $\boldsymbol{\sigma}(w, \nabla w) \equiv \hat{\boldsymbol{\sigma}}(w^+, w^-, \nabla w^+, \nabla w^-)$ which here are linear in their arguments to be rewritten as follows*

$$\begin{aligned}\hat{u}(w) &= \hat{u}_u w \equiv \hat{u}_{u^+}w^+ + \hat{u}_{u^-}w^-,\\ \hat{\boldsymbol{\sigma}}(w,\nabla w) &= \hat{\boldsymbol{\sigma}}_u w + \hat{\boldsymbol{\sigma}}_{\nabla u}\nabla w \equiv \hat{\boldsymbol{\sigma}}_{u^+}w^+ + \hat{\boldsymbol{\sigma}}_{u^-}w^- + \hat{\boldsymbol{\sigma}}_{\nabla u^+}\nabla w^+ + \hat{\boldsymbol{\sigma}}_{\nabla u^-}\nabla w^-,\end{aligned}$$

*with a special notation required on the boundary like indicated in Remark 4.5. Due to the relative complex notation we obmit this derivation which would yield nothing else then Theorem 4.9.*

*Note, however, that this kind of analysis is performed for the compressible Navier-Stokes equation in Section 6, as there, due to the nonlinearity of the problem, a linearization of the numerical fluxes (and thus the complex notation) is required anyway.*

## 4.7 Numerical flux functions

We recall the DG discretization (90) in face-based form: find $u_h \in V_h$ such that

$$\begin{aligned}&\int_\Omega \nabla_h u_h\cdot\nabla_h v_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}}\cup\Gamma}\left([\![\hat{u}_h - u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\} - \{\!\{\hat{\boldsymbol{\sigma}}_h\}\!\}\cdot[\![v_h]\!]\right)\mathrm{d}s \\ &\quad + \int_{\Gamma_{\mathcal{I}}}\left(\{\!\{\hat{u}_h - u_h\}\!\}\,[\![\nabla_h v_h]\!] - [\![\hat{\boldsymbol{\sigma}}_h]\!]\{\!\{v_h\}\!\}\right)\mathrm{d}s = \int_\Omega f v_h\,\mathrm{d}\mathbf{x} \quad \forall v_h\in V_h, \quad (105)\end{aligned}$$

where $\hat{u}_h := \hat{u}(u_h)$ and $\hat{\boldsymbol{\sigma}}_h := \hat{\boldsymbol{\sigma}}(u_h, \nabla u_h)$. Here the numerical flux functions $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are still unspecified. According to Lemma 4.4 the discretization (105) of Poisson's equation (79) is consistent provided the fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are consistent. Furthermore, according to Lemma 4.7 the discretization is conservative provided the numerical flux $\hat{\boldsymbol{\sigma}}$ is conservative. Finally, according to Theorem 4.9, the discretization in combination with the discrete target quantity (99) is adjoint consistent, provided the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are conservative. Depending on the specific choice of $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ several different DG methods can be derived, each with specific properties with respect to accuracy and stability. Before continuing with the derivation of specific DG discretizations we first collect some elementary relations of the mean value and jump operators:

**Lemma 4.11** *Let $\{\!\{\cdot\}\!\}$ and $[\![\cdot]\!]$ be the mean value and jump operators on $\Gamma_{\mathcal{I}}$ as defined in* (88) *and* (89). *Then for scalar valued functions $q$ and vector-valued functions $\boldsymbol{\phi}$ we have*

$$\{\!\{\{\!\{q\}\!\}\}\!\} = \{\!\{q\}\!\}, \qquad \{\!\{[\![q]\!]\}\!\} = [\![q]\!], \qquad [\![[\![q]\!]]\!] = 0, \qquad [\![\{\!\{q\}\!\}]\!] = 0, \tag{106}$$

$$\{\!\{\{\!\{\boldsymbol{\phi}\}\!\}\}\!\} = \{\!\{\boldsymbol{\phi}\}\!\}, \qquad \{\!\{[\![\boldsymbol{\phi}]\!]\}\!\} = [\![\boldsymbol{\phi}]\!], \qquad [\![[\![\boldsymbol{\phi}]\!]]\!] = 0, \qquad [\![\{\!\{\boldsymbol{\phi}\}\!\}]\!] = 0. \tag{107}$$

### 4.7.1 The symmetric interior penalty Galerkin (SIPG) discretization

Let the fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ in (105) be given by

$$\begin{aligned}
&\hat{u}_h = \{\!\{u_h\}\!\}, && \hat{\boldsymbol{\sigma}}_h = \{\!\{\nabla_h u_h\}\!\} - \boldsymbol{\delta}^{\mathrm{ip}}(u_h) && \text{on } \Gamma_{\mathcal{I}},\\
&\hat{u}_h = \hat{u}_{\theta=-1}(u_h) = g_D, && \hat{\boldsymbol{\sigma}}_h = \nabla_h u_h - \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h) && \text{on } \Gamma_D,\\
&\hat{u}_h = u_h, && \hat{\boldsymbol{\sigma}}_h = g_N \mathbf{n} && \text{on } \Gamma_N,
\end{aligned} \tag{108}$$

where

$$\boldsymbol{\delta}^{\mathrm{ip}}(u_h) = \delta[\![u_h]\!] = C_{\mathrm{IP}}\frac{p^2}{h}[\![u_h]\!] \qquad \text{on } \Gamma_{\mathcal{I}}, \tag{109}$$

$$\boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h) = \delta(u_h - g_D)\mathbf{n} = C_{\mathrm{IP}}\frac{p^2}{h}(u_h - g_D)\mathbf{n} \qquad \text{on } \Gamma_D. \tag{110}$$

Using Lemma 4.11 we obtain $[\![\hat{u}_h]\!] = [\![\{\!\{u_h\}\!\}]\!] = 0$, $\{\!\{\hat{u}_h\}\!\} = \{\!\{\{\!\{u_h\}\!\}\}\!\} = \{\!\{u_h\}\!\}$, $\{\!\{\hat{\boldsymbol{\sigma}}_h\}\!\} = \{\!\{\{\!\{\nabla_h u_h\}\!\}\}\!\} - \{\!\{\boldsymbol{\delta}^{\mathrm{ip}}(u_h)\}\!\} = \{\!\{\nabla_h u_h\}\!\} - \boldsymbol{\delta}^{\mathrm{ip}}(u_h)$, and $[\![\hat{\boldsymbol{\sigma}}_h]\!] = [\![\{\!\{\nabla_h u_h\}\!\}]\!] - [\![\boldsymbol{\delta}^{\mathrm{ip}}(u_h)]\!] = 0$ on $\Gamma_{\mathcal{I}}$, and (105) reduces to the symmetric IP discretization given by: find $u_h \in V_h$ such that

$$\begin{aligned}
\int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} \left(-[\![u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\} - \{\!\{\nabla_h u_h\}\!\}\cdot[\![v_h]\!]\right)\mathrm{d}s + \int_{\Gamma_{\mathcal{I}}} \boldsymbol{\delta}^{\mathrm{ip}}(u_h)\cdot[\![v_h]\!]\,\mathrm{d}s\\
+ \int_{\Gamma_D} \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h)\cdot\mathbf{n}\, v_h\,\mathrm{d}s = \int_\Omega f v_h\,\mathrm{d}\mathbf{x} - \int_{\Gamma_D} g_D\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s + \int_{\Gamma_N} g_N v_h\,\mathrm{d}s
\end{aligned} \tag{111}$$

for all $v_h \in V_h$. The numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ are consistent and conservative/single-valued. Thereby, the SIPG discretization (111) is *consistent* and *conservative*. Furthermore, in combination with the discretization (99) of the target quantity (80), i.e.,

$$\begin{aligned}
J_h(u_h) &= \int_\Omega j_\Omega\, u_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\,\hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma_N} j_N\,\hat{u}_h\,\mathrm{d}s,\\
&= \int_\Omega j_\Omega\, u_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\,\left(\nabla_h u_h - \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h)\right)\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma_N} j_N\, u_h\,\mathrm{d}s,
\end{aligned} \tag{112}$$

the SIPG discretization (111) is also *adjoint consistent.* Note, however, that a *direct discretization* of (80) based on

$$J(u_h) = \int_\Omega j_\Omega\, u_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\, \nabla_h u_h \cdot \mathbf{n} \,\mathrm{d}s + \int_{\Gamma_N} j_N\, u_h \,\mathrm{d}s, \tag{113}$$

would result in an adjoint *in*consistent discretization. In fact, the difference of both discretizations, i.e., the term

$$J_h(u_h) - J(u_h) = -\int_{\Gamma_D} j_D\, \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h) \cdot \mathbf{n} \,\mathrm{d}s = -\int_{\Gamma_D} j_D\, \delta\,(u_h - g_D)\,\mathrm{d}s, \tag{114}$$

was found in (Harriman et al., 2004; Hartmann, 2007a) to be necessary for the adjoint consistency of the discretization and for the smoothness of discrete adjoint solutions. While the need of this term (called *IP modification* of the target quantity in Hartmann (2007a)) might be a surprising outcome of the analysis in (Harriman et al., 2004; Hartmann, 2007a), it seems quite natural in view of (99) and (112). Note, that (99) is based on the numerical flux function and does not require any additional modification. Furthermore, the formulation in (99) is also valid for discretizations other than SIPG.

Finally, note that the discretization (111) is stable provided the $C_{\mathrm{IP}}$ constant in (109) is larger than a constant $C^0_{\mathrm{IP}} > 0$ (cf. Appendix A.4.3 for details). For long time the lower bound $C^0_{\mathrm{IP}}$ was in general not known. Note, that (Shahbazi, 2005) provided an expression for $C^0_{\mathrm{IP}}$ on simplical elements. Very recently, (Hillewaert, 2013) provided an expression for $C^0_{\mathrm{IP}}$ on hybrid meshes, i.e., on quadrilaterals and triangles in two dimensions and on hexahedra, tetrahedra, prisms and pyramids in three dimensions.

### 4.7.2 The non-symmetric interior penalty Galerkin (NIPG) discretization

Let the fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ in (105) be given by

$$\begin{aligned}
&\hat{u}_h = \{\!\{u_h\}\!\} + \mathbf{n}^+ \cdot [\![u_h]\!], && \hat{\boldsymbol{\sigma}}_h = \{\!\{\nabla_h u_h\}\!\} - \boldsymbol{\delta}^{\mathrm{ip}}(u_h) && \text{on } \Gamma_{\mathcal{I}},\\
&\hat{u}_h = \hat{u}_{\theta=1}(u_h) = 2u_h - g_D, && \hat{\boldsymbol{\sigma}}_h = \nabla_h u_h - \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h) && \text{on } \Gamma_D,\\
&\hat{u}_h = u_h, && \hat{\boldsymbol{\sigma}}_h = g_N \mathbf{n} && \text{on } \Gamma_N.
\end{aligned} \tag{115}$$

We use $\mathbf{n}^+ \cdot [\![u_h]\!] = \mathbf{n}^+ \cdot (u_h^+\mathbf{n}^+ + u_h^-\mathbf{n}^-) = u_h^+ - u_h^-$ and $[\![\{\!\{u_h\}\!\}]\!] = 0$, and obtain

$$\begin{aligned}
[\![\hat{u}_h]\!] &= [\![u_h^+ - u_h^-]\!] = (u_h^+ - u_h^-)\mathbf{n}^+ + (u_h^- - u_h^+)\mathbf{n}^- = 2(u_h^+\mathbf{n}^+ + u_h^-\mathbf{n}^-) = 2[\![u_h]\!],\\
\{\!\{\hat{u}_h\}\!\} &= \{\!\{\{\!\{u_h\}\!\}\}\!\} + \{\!\{\mathbf{n}\cdot[\![u_h]\!]\}\!\} = \{\!\{u_h\}\!\} + \tfrac{1}{2}(u_h^+ - u_h^- + u_h^- - u_h^+)) = \{\!\{u_h\}\!\}.
\end{aligned}$$

Then, (105) reduces to the non-symmetric IP discretization: find $u_h \in V_h$ such that

$$\begin{aligned}
\int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} \left([\![u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\} - \{\!\{\nabla_h u_h\}\!\}\cdot[\![v_h]\!]\right)\mathrm{d}s + \int_{\Gamma_{\mathcal{I}}} \boldsymbol{\delta}^{\mathrm{ip}}(u_h)\cdot[\![v_h]\!]\,\mathrm{d}s\\
+ \int_{\Gamma_D} \boldsymbol{\delta}^{\mathrm{ip}}_\Gamma(u_h)\cdot\mathbf{n}\, v_h \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} g_D\, \mathbf{n}\cdot\nabla_h v_h \,\mathrm{d}s + \int_{\Gamma_N} g_N v_h \,\mathrm{d}s
\end{aligned} \tag{116}$$

for all $v_h \in V_h$. We note, that the only difference of this discretization to the SIPG discretization in (111) is the sign of the $\int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} [\![u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\}\,\mathrm{d}s$ and $\int_{\Gamma_D} g_D\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s$ term.

The numerical flux $\hat{\boldsymbol{\sigma}}_h$ is consistent and conservative, as is the identical flux for SIPG (cf. Section 4.7.1). Furthermore, the numerical flux $\hat{u}_h$ is consistent. Thereby, the NIPG discretization (116) is *consistent* and *conservative*. However, the numerical flux $\hat{u}_h$ is *not* conservative/single-valued (recall that $[\![\hat{u}_h]\!] = 2[\![u_h]\!] \neq 0$). Thereby, the NIPG discretization is *adjoint inconsistent*. As we will see later (cf. Sections 4.8 and 4.10) missing adjoint consistency has a degrading effect on the order of convergence measured in the $L^2$-norm and in target quantities $J(\cdot)$. Nevertheless, the NIPG discretization was quite popular in the first years of DG research, as the non-symmetric discretization of face terms made the numerical analysis of NIPG significantly simpler than that of the symmetric one in SIPG. Furthermore, the discretization (116) is stable for any $C_{\mathrm{IP}} > C_{\mathrm{IP}}^0 = 0$ (cf. Appendix A.4).

### 4.7.3 The method of Baumann-Oden (BO)

Let the fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ in (105) be given by

$$\begin{aligned} \hat{u}_h &= \{\!\{u_h\}\!\} + \mathbf{n}^+ \cdot [\![u_h]\!], & \hat{\boldsymbol{\sigma}}_h &= \{\!\{\nabla_h u_h\}\!\} & &\text{on } \Gamma_{\mathcal{I}}, \\ \hat{u}_h &= 2u_h - g_D, & \hat{\boldsymbol{\sigma}}_h &= \nabla_h u_h & &\text{on } \Gamma_D, \\ \hat{u}_h &= u_h, & \hat{\boldsymbol{\sigma}}_h &= g_N \mathbf{n} & &\text{on } \Gamma_N. \end{aligned} \tag{117}$$

Then we obtain the method by Baumann and Oden: find $u_h \in V_h$ such that

$$\begin{aligned} \int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}} \cup \Gamma_D} \left([\![u_h]\!] \cdot \{\!\{\nabla_h v_h\}\!\} - \{\!\{\nabla_h u_h\}\!\} \cdot [\![v_h]\!]\right) \mathrm{d}s& \\ = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} g_D\, \mathbf{n} \cdot \nabla v_h \,\mathrm{d}s + \int_{\Gamma_N} g_N v_h \,\mathrm{d}s& \end{aligned} \tag{118}$$

for all $v_h \in V_h$. We note, that this discretization can be obtained from the NIPG discretization in (116) simply by obmitting the interior penalty term $\boldsymbol{\delta}^{\mathrm{ip}}(u_h) = \delta[\![u_h]\!]$. Like the NIPG discretization this method is *consistent* and *conservative*, but *adjoint inconsistent*. Furthermore, the method of Baumann-Oden is *unstable* whereas the NIPG discretization is stable due to the stabilizing effect of the interior penalty term (cf. Appendix A.4).

### 4.7.4 Unified description for SIPG, NIPG and Baumann-Oden

The SIPG and NIPG discretizations and the method of Baumann-Oden can be written in unified form as follows: find $u_h \in V_h$ such that

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h, \tag{119}$$

where

$$\begin{aligned} L_h(u,v) &= \int_\Omega \nabla_h u \cdot \nabla_h v \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}} \cup \Gamma_D} \left(\theta[\![u]\!] \cdot \{\!\{\nabla_h v\}\!\} - \{\!\{\nabla_h u\}\!\} \cdot [\![v]\!]\right) \mathrm{d}s + \int_{\Gamma_{\mathcal{I}} \cup \Gamma_D} \delta[\![u]\!] \cdot [\![v]\!] \,\mathrm{d}s \\ F_h(v) &= \int_\Omega f v \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} g_D\, \theta \mathbf{n} \cdot \nabla v \,\mathrm{d}s + \int_{\Gamma_D} g_D\, \delta v \,\mathrm{d}s + \int_{\Gamma_N} g_N v \,\mathrm{d}s, \end{aligned} \tag{120}$$

and the constants $\theta$ and $\delta$ are given by

$$\begin{aligned}
&\text{SIPG}: && \theta = -1, && \delta > 0,\\
&\text{NIPG}: && \theta = 1, && \delta > 0,\\
&\text{Baumann-Oden}: && \theta = 1, && \delta = 0.
\end{aligned} \tag{121}$$

### 4.7.5 The original DG discretization of Bassi and Rebay (BR1)

Let the fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ in (105) be given by

$$\begin{aligned}
&\hat{u}_h = \{\!\{u_h\}\!\}, && \hat{\boldsymbol{\sigma}}_h = \{\!\{\nabla_h u_h\}\!\} - \boldsymbol{\delta}^{\mathrm{br1}}(u_h) && \text{on } \Gamma_{\mathcal{I}},\\
&\hat{u}_h = g_D, && \hat{\boldsymbol{\sigma}}_h = \nabla_h u_h - \boldsymbol{\delta}_\Gamma^{\mathrm{br1}}(u_h) && \text{on } \Gamma_D,\\
&\hat{u}_h = u_h, && \hat{\boldsymbol{\sigma}}_h = g_N \mathbf{n} && \text{on } \Gamma_N,
\end{aligned} \tag{122}$$

with

$$\boldsymbol{\delta}^{\mathrm{br1}}(u_h) = \boldsymbol{\delta}_\Gamma^{\mathrm{br1}}(u_h) = \{\!\{\mathbf{L}_{g_D}(u_h)\}\!\}, \tag{123}$$

where the so-called *global lifting operator* including Dirichlet boundary values is a vector-valued affine operator defined by: For a scalar-valued function $w$ let $\mathbf{L}_{g_D}(w) \in \boldsymbol{\Sigma}_h^p$ be the solution to

$$\int_\Omega \mathbf{L}_{g_D}(w) \cdot \boldsymbol{\tau} \,\mathrm{d}\mathbf{x} = \int_{\Gamma_{\mathcal{I}}} [\![w]\!] \cdot \{\!\{\boldsymbol{\tau}\}\!\} \,\mathrm{d}s + \int_{\Gamma_D} (w - g_D)\mathbf{n} \cdot \boldsymbol{\tau} \,\mathrm{d}s \quad \forall \boldsymbol{\tau} \in \boldsymbol{\Sigma}_h^p. \tag{124}$$

Furthermore, we consider the *global lifting operator* $\mathbf{L}_0$ [1] with homogeneous Dirichlet boundary values, which is the vector-valued linear operator given by: Let $\mathbf{L}_0(w) \in \boldsymbol{\Sigma}_h^p$ be the solution to

$$\int_\Omega \mathbf{L}_0(w) \cdot \boldsymbol{\tau} \,\mathrm{d}\mathbf{x} = \int_{\Gamma_{\mathcal{I}} \cup \Gamma} [\![w]\!] \cdot \{\!\{\boldsymbol{\tau}\}\!\} \,\mathrm{d}s \quad \forall \boldsymbol{\tau} \in \boldsymbol{\Sigma}_h^p. \tag{125}$$

In view of (124) and (125), we have

$$\int_\Omega \mathbf{L}_{g_D}(w) \cdot \boldsymbol{\tau} \,\mathrm{d}\mathbf{x} = \int_\Omega \mathbf{L}_0(w) \cdot \boldsymbol{\tau} \,\mathrm{d}\mathbf{x} - \int_{\Gamma_D} g_D \,\mathbf{n} \cdot \boldsymbol{\tau} \,\mathrm{d}s. \tag{126}$$

Using the numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ as given in (122) the DG discretization (105) reduces to: find $u_h \in V_h$ such that

$$\begin{aligned}
\int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}} \cup \Gamma_D} (-[\![u_h]\!] \cdot \{\!\{\nabla_h v_h\}\!\} - \{\!\{\nabla_h u_h\}\!\} \cdot [\![v_h]\!]) \,\mathrm{d}s + \int_{\Gamma_{\mathcal{I}} \cup \Gamma_D} \{\!\{\mathbf{L}_{g_D}(u_h)\}\!\} \cdot [\![v_h]\!] \,\mathrm{d}s \\
= \int_\Omega f v_h \,\mathrm{d}\mathbf{x} - \int_{\Gamma_D} g_D \,\mathbf{n} \cdot \nabla_h v_h \,\mathrm{d}s + \int_{\Gamma_N} g_N v_h \,\mathrm{d}s \qquad \forall v_h \in V_h.
\end{aligned} \tag{127}$$

[1] We note that in some publications, the global lifting operator is defined for vector-valued functions $\boldsymbol{\phi}$ as

$$\int_\Omega \mathbf{l}_0(\boldsymbol{\phi}) \cdot \boldsymbol{\tau} \,\mathrm{d}\mathbf{x} = -\int_{\Gamma_{\mathcal{I}} \cup \Gamma} \boldsymbol{\phi} \cdot \{\!\{\boldsymbol{\tau}\}\!\} \,\mathrm{d}s \quad \forall \boldsymbol{\tau} \in \boldsymbol{\Sigma}_h^p,$$

for which we then have $\mathbf{L}_0(w) = -\mathbf{l}_0([\![w]\!])$.

Using the definition (124) for $\mathbf{L}_{g_D}$ we can rewrite

$$\int_\Omega \mathbf{L}_{g_D}(u_h)\cdot\nabla_h v_h\,\mathrm{d}\mathbf{x} = \int_{\Gamma_\mathcal{I}\cup\Gamma} [\![u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\}\,\mathrm{d}s - \int_{\Gamma_D} g_D\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s. \tag{128}$$

Furthermore, using the relation (125) of $\mathbf{L}_0$ we can rewrite

$$\int_\Omega \nabla_h u_h\cdot\mathbf{L}_0(v_h)\,\mathrm{d}\mathbf{x} = \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\nabla_h u_h\}\!\}\cdot[\![v_h]\!]\,\mathrm{d}s, \tag{129}$$

$$\int_\Omega \mathbf{L}_{g_D}(u_h)\cdot\mathbf{L}_0(v_h)\,\mathrm{d}\mathbf{x} = \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\mathbf{L}_{g_D}(u_h)\}\!\}\cdot[\![v_h]\!]\,\mathrm{d}s. \tag{130}$$

Substituting these relations into (127) we obtain the discretization: find $u_h\in V_h$ such that

$$\int_\Omega (\nabla_h u_h - \mathbf{L}_{g_D}(u_h))\cdot(\nabla_h v_h - \mathbf{L}_0(v_h))\,\mathrm{d}\mathbf{x} = \int_\Omega f v_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N v_h\,\mathrm{d}s \qquad \forall v_h\in V_h, \tag{131}$$

or equivalently, using (126): find $u_h\in V_h$ such that

$$L_h(u_h,v_h) = F_h(v_h) \qquad \forall v_h\in V_h, \tag{132}$$

where

$$\begin{aligned} L_h(u,v) &= \int_\Omega (\nabla_h u - \mathbf{L}_0(u))\cdot(\nabla_h v - \mathbf{L}_0(v))\,\mathrm{d}\mathbf{x},\\ F_h(v) &= \int_\Omega f v\,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N v\,\mathrm{d}s - \int_{\Gamma_D} g_D\,\mathbf{n}\cdot(\nabla_h v - \mathbf{L}_0(v))\,\mathrm{d}s. \end{aligned} \tag{133}$$

This is the original method of Bassi and Rebay introduced in (Bassi and Rebay, 1997a). Here, the numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ are consistent and conservative. Thereby, the BR1 discretization (127) is *consistent* and *conservative*. However, this discretization has several disadvantages: In contrast to most other DG discretizations where an element communicates with its direct neighboring elements only, the stencil of the BR1 discretization is considerably larger as it includes also neighbors of neighbors. Furthermore, this discretization is unstable. In fact, we obtain

$$L_h(v,v) = \|\nabla_h v - \mathbf{L}_0(v)\|^2_{L^2(\Omega)},$$

which vanishes on the set

$$Z := \{v\in V_h : \nabla_h v - \mathbf{L}_0(v) = 0\},$$

where $Z\setminus\{0\}$ can, in general, be nonempty. This discretization is called the *BR1 discretization* in order to distinguish it from the modification of Bassi and Rebay, the so-called *BR2 discretization*, which we will introduce in the following subsection.

### 4.7.6 The modified DG discretization of Bassi and Rebay (BR2)

Let the fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ in (105) be given by

$$
\begin{aligned}
\hat{u}_h &= \{\!\{u_h\}\!\}, & \hat{\boldsymbol{\sigma}}_h &= \{\!\{\nabla_h u_h\}\!\} - \boldsymbol{\delta}^{\mathrm{br2}}(u_h) & &\text{on } \Gamma_{\mathcal{I}},\\
\hat{u}_h &= g_D, & \hat{\boldsymbol{\sigma}}_h &= \nabla_h u_h - \boldsymbol{\delta}_\Gamma^{\mathrm{br2}}(u_h) & &\text{on } \Gamma_D,\\
\hat{u}_h &= u_h, & \hat{\boldsymbol{\sigma}}_h &= g_N \mathbf{n} & &\text{on } \Gamma_N,
\end{aligned}
\tag{134}
$$

with

$$
\boldsymbol{\delta}^{\mathrm{br2}}(u_h) = \boldsymbol{\delta}_\Gamma^{\mathrm{br2}}(u_h) = C_{\mathrm{BR2}}\{\!\{\mathbf{L}^e_{g_D}(u_h)\}\!\} \quad \text{for } e \subset \Gamma_{\mathcal{I}} \cup \Gamma_D, \tag{135}
$$

where the so-called *local lifting operator* including Dirichlet boundary conditions is a vector-valued affine operator defined by: find $\mathbf{L}^e_{g_D}(w) \in \boldsymbol{\Sigma}_h^p$ such that

$$
\begin{aligned}
\int_\Omega \mathbf{L}^e_{g_D}(w)\cdot\boldsymbol{\tau}\,\mathrm{d}\mathbf{x} &= \int_e (w - g_D)\,\mathbf{n}\cdot\boldsymbol{\tau}\,\mathrm{d}s \quad \forall \boldsymbol{\tau}\in\boldsymbol{\Sigma}_h^p, & &\text{for } e\subset\Gamma_D\\
\int_\Omega \mathbf{L}^e_{g_D}(w)\cdot\boldsymbol{\tau}\,\mathrm{d}\mathbf{x} &= \int_e [\![w]\!]\cdot\{\!\{\boldsymbol{\tau}\}\!\}\,\mathrm{d}s \quad \forall \boldsymbol{\tau}\in\boldsymbol{\Sigma}_h^p, & &\text{on } e\subset\Gamma_{\mathcal{I}},
\end{aligned}
\tag{136}
$$

and $\mathbf{L}^e_{g_D}(w)$ is defined to be zero for $e \subset \Gamma_N$. The local lifting operator with homogeneous Dirichlet boundary conditions $\mathbf{L}^e_0$ is defined accordingly. In particular, for $e \subset \Gamma_D$ we have

$$
\int_\Omega \mathbf{L}^e_{g_D}(w)\cdot\boldsymbol{\tau}\,\mathrm{d}\mathbf{x} = \int_\Omega \mathbf{L}^e_0(w)\cdot\boldsymbol{\tau}\,\mathrm{d}\mathbf{x} + \int_e g_D\,\mathbf{n}\cdot\boldsymbol{\tau}\,\mathrm{d}s \quad \forall\boldsymbol{\tau}\in\boldsymbol{\Sigma}_h^p. \tag{137}
$$

We note, that $\mathbf{L}^e_{g_D}(w)$ has support (i.e., is non-equal zero) only on the (one or two) elements sharing the edge $e$. Furthermore, $\mathbf{L}^e_{g_D}(w)$ does not depend on $g_D$ on interior edges $e \subset \Gamma_{\mathcal{I}}$. Using the numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ as given in (134) the DG discretization (105) reduces to: find $u_h \in V_h$ such that

$$
\begin{aligned}
\int_\Omega \nabla_h u_h\cdot\nabla_h v_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} (-[\![u_h]\!]\cdot\{\!\{\nabla_h v_h\}\!\} - \{\!\{\nabla_h u_h\}\!\}\cdot[\![v_h]\!])\,\mathrm{d}s + \int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\{\!\{\mathbf{L}^e_{g_D}(u_h)\}\!\}\cdot[\![v_h]\!]\,\mathrm{d}s\\
= \int_\Omega f v_h\,\mathrm{d}\mathbf{x} - \int_{\Gamma_D} g_D\,\mathbf{n}\cdot\nabla_h v_h\,\mathrm{d}s + \int_{\Gamma_N} g_N v_h\,\mathrm{d}s \qquad \forall v_h\in V_h.
\end{aligned}
\tag{138}
$$

Using the definition (136) of $\mathbf{L}^e_{g_D}$ we can rewrite

$$
\sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\int_\Omega \mathbf{L}^e_{g_D}(u_h)\cdot\mathbf{L}^e_0(v_h)\,\mathrm{d}\mathbf{x} = \int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\{\!\{\mathbf{L}^e_{g_D}(u_h)\}\!\}\cdot[\![v_h]\!]\,\mathrm{d}s. \tag{139}
$$

Substituting this, (128) and (129) into (138) yields: find $u_h \in V_h$ such that

$$
\begin{aligned}
&\int_\Omega (\nabla_h u_h\cdot\nabla_h v_h - \mathbf{L}_{g_D}(u_h)\cdot\nabla_h v_h - \nabla_h u_h\cdot\mathbf{L}_0(v_h))\,\mathrm{d}\mathbf{x}\\
&+ \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\int_\Omega \mathbf{L}^e_{g_D}(u_h)\cdot\mathbf{L}^e_0(v_h)\,\mathrm{d}\mathbf{x} = \int_\Omega f v_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N v_h\,\mathrm{d}s \qquad \forall v_h\in V_h,
\end{aligned}
\tag{140}
$$

or equivalently, using (126) and (137): find $u_h \in V_h$ such that

$$L_h(u_h, v_h) = F_h(v_h) \qquad \forall v_h \in V_h, \tag{141}$$

where

$$\begin{aligned} L_h(u,v) &= \int_\Omega (\nabla_h u \cdot \nabla_h v - \mathbf{L}_0(u) \cdot \nabla_h v - \nabla_h u \cdot \mathbf{L}_0(v))\,\mathrm{d}\mathbf{x} + \sum_{e \subset \Gamma_\mathcal{I} \cup \Gamma_D} C_{\mathrm{BR2}} \int_\Omega \mathbf{L}_0^e(u) \cdot \mathbf{L}_0^e(v)\,\mathrm{d}\mathbf{x}, \\ F_h(v) &= \int_\Omega f v\,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N v\,\mathrm{d}s - \int_{\Gamma_D} g_D\,\mathbf{n} \cdot (\nabla_h v - C_{\mathrm{BR2}} \mathbf{L}_0^e(v))\,\mathrm{d}s. \end{aligned} \tag{142}$$

We note, that (140) can be obtained also by replacing $\int_\Omega \mathbf{L}_{g_D}(u_h)\mathbf{L}_0(v_h)\,\mathrm{d}\mathbf{x}$ in (131) by (139). The BR2 discretization has several advantages over the BR1 scheme. The stencil of BR2 scheme includes only first neighbors instead of additional second neighbors as does the BR1 scheme. Furthermore, the BR2 discretization is stable, provided $C_{\mathrm{BR2}}$ is larger than the number of neighboring elements, i.e., $C_{\mathrm{BR2}} > 3$ for triangular elements and $C_{\mathrm{BR2}} > 4$ for quadrilateral elements (cf. Appendix A.4).

Note, that the BR2 discretization differs from the SIPG discretization (cf. Section 4.7.1) only in the definition of the stabilization/penalization term $\boldsymbol{\delta}^{\mathrm{br2}}(u_h)$ versus $\boldsymbol{\delta}^{\mathrm{ip}}(u_h)$. In particular, for BR2, like for IP, the numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ are consistent and conservative. Thereby, the BR2 discretization (111) is *consistent* and *conservative*. Furthermore, in combination with the discretization (99) of the target quantity (80), i.e.,

$$\begin{aligned} J_h(u_h) &= \int_\Omega j_\Omega\, u_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\, \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\,\mathrm{d}s + \int_{\Gamma_N} j_N\, \hat{u}_h\,\mathrm{d}s, \\ &= \int_\Omega j_\Omega\, u_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D\, \left(\nabla_h u_h - \boldsymbol{\delta}_\Gamma^{\mathrm{br2}}(u_h)\right) \cdot \mathbf{n}\,\mathrm{d}s + \int_{\Gamma_N} j_N\, u_h\,\mathrm{d}s, \end{aligned} \tag{143}$$

the BR2 discretization (138) is also *adjoint consistent*. However, due to the use of lifting operators the BR2 discretization is in general more complicated and more computing time expensive than the SIPG discretization.

## 4.8 *A priori* error estimates in DG- and $L^2$-norm

In this section we provide *a priori* error estimates for the SIPG and NIPG discretization.

**Lemma 4.12 (*A priori* error estimates for SIPG and NIPG)** *Let $u \in H^{p+1}(\Omega)$ be the exact solution to Poisson's equation* (79). *Furthermore, let $u_h \in V_h^p$ be the solution to*

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h^p,$$

*where $L_h$ is as given in* (120) *with $\theta = 1$ (NIPG) or $\theta = -1$ (SIPG) and $\delta = C_{IP}\frac{p^2}{h}$, $C_{IP} > C_{IP}^0$. Then, for both, SIPG and NIPG:*

$$|||u - u_h|||_\delta \le C h^p |u|_{H^{p+1}(\Omega)}, \tag{144}$$

*where $|||\cdot|||_\delta$ is a DG-norm given by*

$$|||v|||_\delta^2 = \|\nabla_h v\|_{L^2(\Omega)}^2 + \int_{\Gamma_\mathcal{I} \cup \Gamma_D} \delta^{-1} \left(\mathbf{n} \cdot \{\!\{\nabla v\}\!\}\right)^2\,ds + \int_{\Gamma_\mathcal{I} \cup \Gamma_D} \delta [v]^2\,ds. \tag{145}$$

*Furthermore, for NIPG:*

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^p |u|_{H^{p+1}(\Omega)}, \tag{146}$$

*and for SIPG:*

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^{p+1} |u|_{H^{p+1}(\Omega)}. \tag{147}$$

**Proof:** See Appendix A.4.4. □

Here, we see that the order of convergence in the $L^2$-norm of the SIPG discretization is one order higher than that of the NIPG discretization. Note, that this difference is due to the fact that SIPG is adjoint consistent whereas NIPG is not (cf. the proof in Appendix A.4.4).

## 4.9 Numerical results: Order of convergence in $L^2$

In the following we investigate the experimental order of convergence in the $H^1$- and the $L^2$-norm of the SIPG ($\theta = -1$) and the NIPG ($\theta = 1$) discretizations, see Section 4.7.4. According to Section 4.7.1 the penalization parameter is given by $\delta = C_{\mathrm{IP}} \frac{p^2}{h}$. In this example we choose $C_{\mathrm{IP}} = 4$.

Let us consider the following *model problem*: Let $\Omega = (0,1)^2$ and consider Poisson's equation (79) with forcing function $f$ which is chosen such that the analytical solution to (79) is given by

$$u(\mathbf{x}) = \sin(\tfrac{1}{2}\pi x_1)\sin(\tfrac{1}{2}\pi x_2). \tag{148}$$

We impose Dirichlet boundary conditions where the boundary value function $g_D$ on $\Gamma_D = \Gamma = \partial\Omega$ is prescribed based on the analytical solution $u$.

Figure 1 plots the error in the $H^1(\Omega)$-seminorm, $|u - u_h|_{H^1(\Omega)}$, against the number of elements. We see that for a given polynomial degree $p$ the discretization errors of the SIPG and the NIPG discretization almost coincide. Furthermore, we see that for the discretizations with polynomial degree $p = 1, \ldots, 5$, the discretization error in the $H^1$-seminorm is of order $\mathcal{O}(h^p)$ which is in agreement with the theoretical order of convergence, see Lemma 4.12.

Figure 2a) shows that the error in the $L^2(\Omega)$-norm of the SIPG discretization for the polynomial degrees $p = 1, \ldots, 5$, is of order $\mathcal{O}(h^{p+1})$ which again is in perfect agreement with the theoretical result, see Lemma 4.12. In comparison to that, Figure 2b) shows the $L^2(\Omega)$-error of the NIPG discretization. Here, we see that the discretization behaves like $\mathcal{O}(h^{p+1})$ for odd $p$ and like $\mathcal{O}(h^p)$ for even $p$. This sub-optimal convergence of the NIPG method is attributed to the lack of adjoint consistency and the resulting lack of smoothness of the adjoint solution, see the proof of Lemma 4.12. We note that similar results for a different test case have been obtained in (Harriman et al., 2003).

Figures 1 and 2 show that there is a significant advantage of using higher order discretizations over using low order discretization methods. In fact, in Figure 1 we see that the discretization error in the $H^1(\Omega)$-seminorm for $p = 3$ on the coarsest mesh is of similar size as the error for $p = 1$ on the finest mesh. Similarly, in Figure 2a) the discretization error in the $L^2(\Omega)$-norm for $p = 4$ on the coarsest mesh is comparable to the error for $p = 1$ on the finest mesh. We emphasize that here the solutions are of similar accuracy although the finest mesh has by a factor of 16384 more elements than the coarsest mesh. Clearly, a discretization method of higher order requires more degrees of freedom (DoFs)

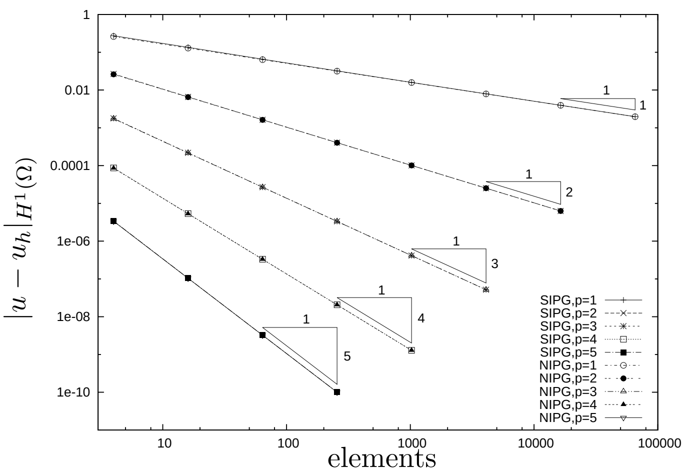
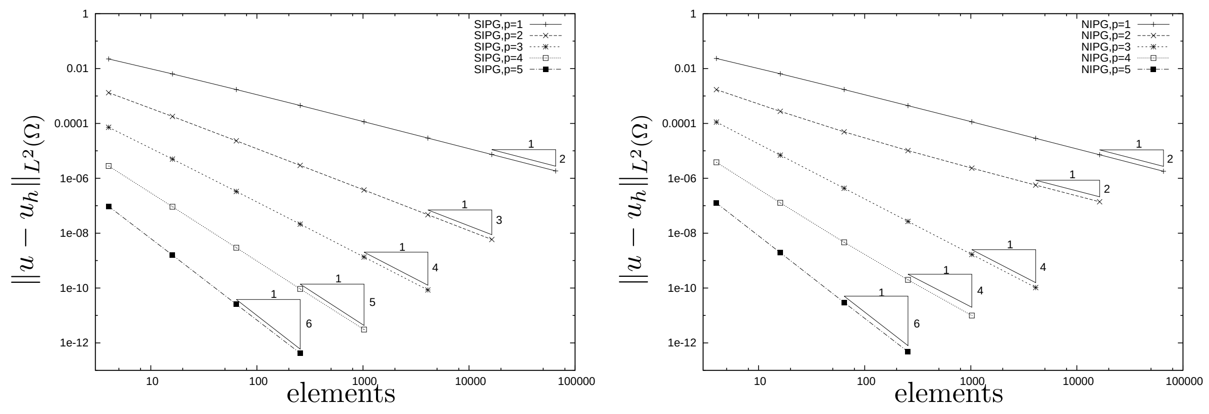
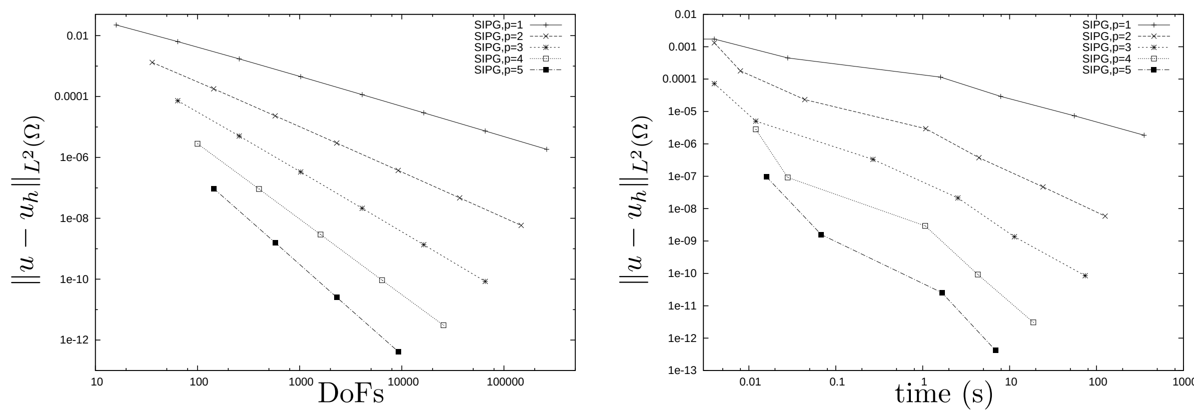



Figure 1: Model problem: The discretization error $|u - u_h|_{H^1(\Omega)}$ of the SIPG and NIPG methods with $p = 1, \ldots, 5$ is of order $\mathcal{O}(h^p)$.




(a) (b)

Figure 2: Model problem: Convergence of the discretization error $\|u - u_h\|_{L^2(\Omega)}$ for a) the SIPG and b) the NIPG methods with global mesh refinement.




(a) (b)

Figure 3: Model problem: The discretization error $\|u - u_h\|_{L^2(\Omega)}$ of the SIPG method plotted a) against the number of degrees of freedom (DoFs) and b) against the computing time in seconds.

per element, 25 DoFs/element for $p = 4$ in comparison to 4 DoFs/element for $p = 1$ in this case, but still the $p = 1$ discretization requires a factor of more than 2600 as many DoFs for the same accuracy as the $p = 4$ discretization. In more detail this is seen in Figure 3a) which plots the $L^2$-error against the number of DoFs. The large factor in the number of DoFs for the specific accuracy translates into a large factor in the computing time required. In fact, in Figure 3b) we see that the discretization with $p = 1$ on the finest mesh requires a by a factor of several thousands larger computing time for the same accuracy as the discretization with $p = 4$ on the coarsest mesh.

Admittedly, the model problem considered here is ideal in the sense that the geometry (unit square) and the governing equations (Poisson's equation) are particularly simple, also the solution is perfectly smooth. However, also for more complicated problems like aerodynamic flows, see e.g. Section 6.8, a significant gain of higher order methods over low order methods can be expected.

## 4.10 *A priori* error estimates in target quantities $J(\cdot)$

In this section we derive *a priori* error estimates with respect to target functionals $J(\cdot)$ for adjoint consistent and adjoint inconsistent discontinuous Galerkin discretizations of Poisson's equation. In particular, we will see that analogous to the (sub-)optimal order of convergence in the $L^2$-norm for adjoint (in-)consistent discretizations also the order of convergence in $J(\cdot)$ is (sub-)optimal.

We assume that Poisson's equation (79) is discretized as follows: find $u_h \in V_h^p$ such that

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h^p, \tag{149}$$

where the bilinear form $L_h(\cdot,\cdot)$ is continuous on $V$ with respect to a specific $|||\cdot|||$-norm, i.e.,

$$L_h(w, v) \leq C_B |||w|||\,|||v||| \quad \forall w, v \in V.$$

Here, $V$ is suitably chosen function space that includes discrete solutions $u_h \in V_h^p$ as well as the exact solution $u$ to Poisson's equation (79).

**Remark 4.13** *Refer to Appendix A.4.2 for the proofs that the bilinear forms of the SIPG, NIPG, Baumann-Oden and BR2 discretization are continuous.*

Also, we assume that the discretization (149) is consistent, i.e., the exact solution $u$ to (79) satisfies

$$L_h(u, v) = F_h(v) \quad \forall v \in V, \tag{150}$$

which implies the Galerkin orthogonality

$$L_h(u - u_h, v_h) = 0 \quad \forall v_h \in V_h^p. \tag{151}$$

Furthermore, we assume that following *a priori* error estimate in the $|||\cdot|||$-norm holds: There is a constant $C > 0$ such that

$$|||u - u_h||| \leq Ch^p |u|_{H^{p+1}(\Omega)} \quad \forall u \in H^{p+1}(\Omega). \tag{152}$$

**Remark 4.14** *Refer to Lemma 4.12 in case of the SIPG and NIPG discretizations.*

Finally, we assume that the local projection operator $P_h^p$ (cf. Section A.2) satisfies following approximation estimate in the $|||\cdot|||$-norm: There is a constant $C > 0$ such that

$$|||v - P_h^p v||| \leq Ch^p |v|_{H^{p+1}(\Omega)} \qquad \forall v \in H^{p+1}(\Omega). \tag{153}$$

In the following we want to measure the discretization error $e = u - u_h$ not in some global norm like the $|||\cdot|||$-norm but with respect to target functionals $J(\cdot)$ of the form (80),

$$J(u) = \int_\Omega j_\Omega \, u \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D \,\mathbf{n}\cdot\nabla u \,\mathrm{d}s + \int_{\Gamma_N} j_N \, u \,\mathrm{d}s. \tag{154}$$

where $j_\Omega \in L^2(\Omega)$, $j_D \in L^2(\Gamma_D)$ and $j_N \in L^2(\Gamma_N)$. According to (99) this target quantity is discretized as follows,

$$J_h(u_h) = \int_\Omega j_\Omega \, u_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} j_D \,\hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n} \,\mathrm{d}s + \int_{\Gamma_N} j_N \, \hat{u}_h \,\mathrm{d}s.$$

Finally, we recall that a discretization together with a target functional is called adjoint consistent, if the exact solution $z$ to the adjoint problem (81) satisfies:

$$L_h(w, z) = J_h(w) \quad \forall w \in V. \tag{155}$$

**Theorem 4.15 (*A priori* error estimates in $J(\cdot)$)** *Let the situation be as described above. Furthermore, assume that the solution $z$ to the adjoint problem* (81) *is smooth, $z \in H^{p+1}(\Omega)$. Then, we have following estimates:*

*a) If the discretization* (149) *together with the target functional $J_h(\cdot)$ is* adjoint consistent, *then there is a constant $C > 0$ such that*

$$|J(u) - J_h(u_h)| \leq Ch^{2p} |u|_{H^{p+1}(\Omega)} |z|_{H^{p+1}(\Omega)} \qquad \forall u \in H^{p+1}(\Omega). \tag{156}$$

*b) If, however, the discretization is* adjoint inconsistent *we only have:*

$$|J(u) - J_h(u_h)| \leq Ch^p |u|_{H^{p+1}(\Omega)} \qquad \forall u \in H^{p+1}(\Omega). \tag{157}$$

**Proof:** a) For an adjoint consistent discretization we set $w := e = u - u_h$ in (155), obtain

$$\begin{aligned}|J(u) - J_h(u_h)| &= |J_h(e)| = |L_h(e, z)| = |L_h(u - u_h, z - P_h z)| \leq C|||u - u_h|||\,|||z - P_h z|||\\ &\leq Ch^p |u|_{H^{p+1}(\Omega)} Ch^p |z|_{H^{p+1}(\Omega)},\end{aligned}$$

and thus (156), where here and in the following we use $P_h$ as a short notation for $P_h^p$.

b) For an adjoint inconsistent discretization we do not have (155). Thereby, in order to represent the error $J(u) - J_h(u_h)$ we define following mesh-dependent adjoint problem: find $\psi \in V$ such that

$$L_h(w, \psi) = J_h(w) \quad \forall w \in V. \tag{158}$$

We note, that for an adjoint consistent discretization the solution $\psi$ to (158) coincides with the solution $z$ to the continuous adjoint solution (81) and thus is smooth. For an adjoint inconsistent discretization, however, we cannot expect $\psi$ to be smooth. In that case $\psi$ is mesh-dependent and in general discontinuous across interior faces. We then proceed as follows

$$\begin{aligned}|J(u) - J_h(u_h)| &= |J_h(e)| = |L_h(e, \psi)| = |L_h(u - u_h, \psi - P_h \psi)| \leq C|||u - u_h|||\,|||\psi - P_h \psi|||\\ &\leq Ch^p |u|_{H^{p+1}(\Omega)},\end{aligned}$$

where due to the lack of smoothness of $\psi$ here we do not gain additional orders of $h$ from $|||\psi - P_h \psi|||$. □

**Remark 4.16** *Note, that the SIPG discretization together with the target quantity* (154) *is adjoint consistent but NIPG is not. Thereby, the error measured in* $J(\cdot)$ *converges with* $\mathcal{O}(h^{2p})$ *for SIPG whereas it converges with* $\mathcal{O}(h^p)$*, only, for NIPG.*

We see, that the order of convergence in $J(\cdot)$ is $\mathcal{O}(h^{2p})$ for an adjoint consistent and $\mathcal{O}(h^p)$ for an adjoint inconsistent discretization provided both primal and adjoint solutions are smooth, $u \in H^{p+1}(\Omega)$ and $z \in H^{p+1}(\Omega)$. If, however, $u$ or $z$ are less regular we obtain an estimate with a correspondingly reduced order of convergence in $J(\cdot)$:

**Corollary 4.17 (*A priori* error estimate in $J(\cdot)$ with reduced regularity)** *Let the situation be as described above. Furthermore, assume that* $u \in H^{s+1}(\Omega)$ *and* $z \in H^{\tilde{s}+1}(\Omega)$ *hold for the exact solutions* $u$ *and* $z$ *to the primal and adjoint problems* (79) *and* (81)*, respectively. Then, we have following estimates for an adjoint consistent discretization (e.g. SIPG):*

$$|J(u) - J_h(u_h)| \leq Ch^{t+\tilde{t}} |u|_{H^{t+1}(\Omega)} |z|_{H^{\tilde{t}+1}(\Omega)} \qquad \forall u \in H^{s+1}(\Omega), \tag{159}$$

*and for an adjoint inconsistent discretization (e.g. NIPG):*

$$|J(u) - J_h(u_h)| \leq Ch^t |u|_{H^{t+1}(\Omega)} \qquad \forall u \in H^{s+1}(\Omega), \tag{160}$$

*where* $t = \min(s, p)$ *and* $\tilde{t} = \min(\tilde{s}, p)$.

## 4.11 Numerical results: Order of convergence in $J(\cdot)$

In this section we investigate the experimental order of convergence of the SIPG and NIPG discretizations when measuring the error in terms of specific target quantities $J(\cdot)$.

**Example 1** To this end, we revisit the experimental *model problem* introduced in Section 4.9. This problem is based on Poisson's equations with an inhomogeneous Dirichlet boundary value function $g_D$ and a forcing function $f$ chosen so that the analytical solution $u$ is given by Eqn. (148). In this first example, we choose the target quantity to represent the (weighted) mean value of $u$ over $\Omega$, i.e.,

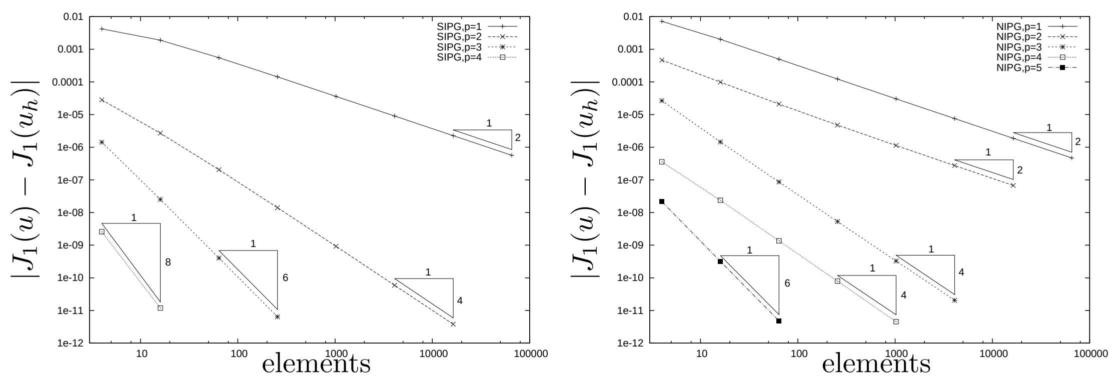
$$J_1(u_h) = \int_\Omega j_\Omega \, u_h \, \mathrm{d}\mathbf{x}; \tag{161}$$

here, we define the weight function $j_\Omega$ by

$$j_\Omega(\mathbf{x}) = \sin(\pi x_1)\sin(\pi x_2).$$

Thereby the true value of the target quantity is given by $J_1(u) = 0.1801265486975$. We note that the target quantity (161) is compatible with Poisson's equation (79). In fact, it is a special case of the target quantity given in (80) with $\Gamma_N = \emptyset$ and $j_D = 0$ on $\Gamma_D = \Gamma$.

Figure 4a) shows the error of the SIPG discretization measured in terms of the target quantity $J_1(\cdot)$ given by (161). We see that under global mesh refinement the error $|J_1(u) - J_1(u_h)|$ behaves like $O(h^{2p})$ which is in perfect agreement with the theoretical order of convergence, see estimate (156). Figure 4b) shows the respective plot for the NIPG discretization. Here, we see that the error $|J_1(u) - J_1(u_h)|$ behaves like $\mathcal{O}(h^{p+1})$ for odd $p$ and like $\mathcal{O}(h^p)$ for even $p$. This convergence behavior is similar to the convergence behavior in the $L^2(\Omega)$-norm encountered for the NIPG scheme in Section 4.9. Again, due to the lack of adjoint consistency the order of convergence in $J_1(\cdot)$ for the NIPG scheme is lower than in the case of the adjoint consistent SIPG scheme. We note that similar results for a different test case have been obtained in (Harriman et al., 2003).




(a) (b)

Figure 4: Example 1: Convergence of the error $|J_1(u) - J_1(u_h)|$ for a) the SIPG and b) the NIPG discretizations with global mesh refinement.

**Example 2** We consider the same *model problem* as in the previous example. However, instead of the mean value quantity (161), here we now choose the target quantity to represent the mean value of the normal derivative of $u$ over the boundary $\Gamma = \partial\Omega$, i.e.,

$$J_2(u_h) = \int_\Gamma j_D \, \mathbf{n} \cdot \nabla_h u_h \, \mathrm{d}s, \tag{162}$$

with $j_D \equiv 1$ on $\Gamma_D = \Gamma$. Thereby the true value of the target quantity is given by $J_2(u) = -2$. We note that this target quantity is compatible with Poisson's equation (79). In fact, it is a special case of the target quantity given in (80) with $\Gamma_N = \emptyset$, and $j_\Omega = 0$ on $\Omega$. Furthermore, we note that the solution $z$ to the corresponding continuous adjoint problem

$$-\Delta z = 0 \quad \text{in } \Omega, \qquad\qquad -z = j_D \quad \text{on } \Gamma_D \tag{163}$$

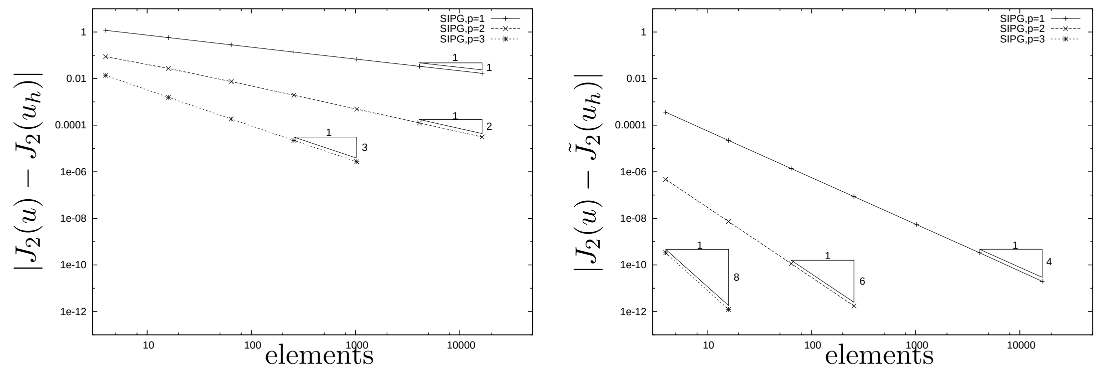
is given by $z \equiv -1$ on $\Omega$. Figure 5a) shows that the error $|J_2(u) - J_2(u_h)|$ behaves like $O(h^p)$ for the SIPG discretizations with $p = 1, 2, 3$. Following the discussion in Section 4.7.1 we recognize that the SIPG discretization in combination with the target functional $J_2(\cdot)$ in (162) is adjoint *in*consistent. Thus the order of convergence $O(h^p)$ encountered in Figure 5a) is, in fact, the expected order of convergence for this adjoint inconsistent discretization (cf. estimate (157)).

However, we recall from Section 4.7.1 that the following discretization

$$J_{2,h}(u_h) = \int_{\Gamma_D} j_D \, \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n} \, \mathrm{d}s = J_2(u_h) - \int_\Gamma j_D \, \delta(u_h - g_D) \, \mathrm{d}s \tag{164}$$

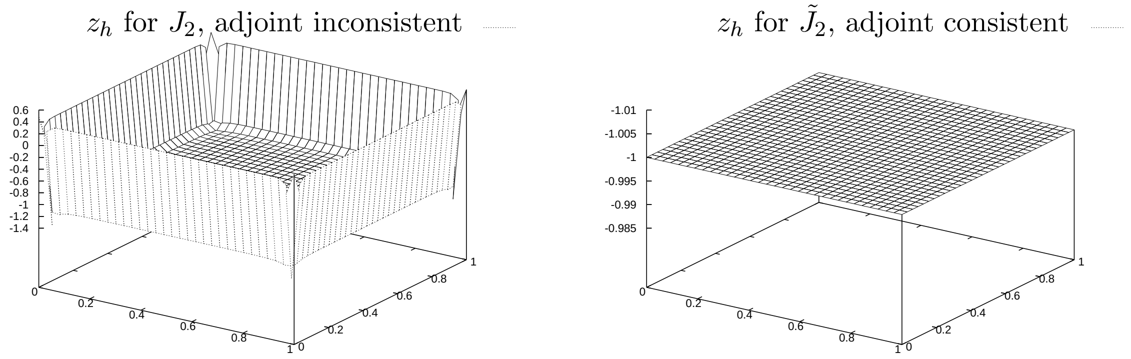
of the target functional $J_2(\cdot)$ leads to an adjoint consistent discretization. Here, $\delta$ is the penalization parameter of the IP discretization and $g_D$ is the boundary value function of the *model problem* considered. Note, that $J_{2,h}(u_h)$ in (164) is a consistent discretization of $J_2(\cdot)$ as the true value of the target quantity is unchanged: $J_{2,h}(u) = J_2(u) = -2$ holds for the exact solution $u$.

Figure 5b) shows the error of the SIPG discretization measured in terms of the target quantity discretization $J_{2,h}(\cdot)$ given in (164). We see that the adjoint consistent discretization of the target functional leads to a significant increase in the accuracy and the order




(a) (b)

Figure 5: Example 2: Convergence of a) the error $|J_2(u) - J_2(u_h)|$ and b) the error $|J_2(u) - J_{2,h}(u_h)|$ for the SIPG discretization with global mesh refinement.




(a) (b)

Figure 6: Example 2: Discrete adjoint solution $z_h$ corresponding to the SIPG discretization combined with a) the adjoint inconsistent $J_2(u_h)$ and b) the adjoint consistent $J_{2,h}(u_h)$.

of convergence of the discretization. In fact, we see that under global mesh refinement the error $|J_2(u) - J_{2,h}(u_h)|$ behaves like $O(h^{2(p+1)})$ which is even larger than the expected order $O(h^{2p})$ (cf. estimate (156)) of an adjoint consistent discretization.

In the following we want to highlight the connection between adjoint consistency and the smoothness of the adjoint solution. To this end, Figure 6a) shows the discrete adjoint solution $z_h$ connected to the (original) target quantity $J_2(\cdot)$; i.e., $z_h$ is the solution to the discrete adjoint problem (155) with right hand side $J_2(\cdot)$. In Figure 6a) we see that $z_h$ is irregular in the neighborhood of the boundary. We note that this irregularity does not vanish under mesh refinement. Thereby, the discrete adjoint solution does not converge to the exact solution, $z \equiv -1$, of the continuous adjoint problem (163). This behavior corresponds to the fact that the SIPG discretization in combination with the target quantity $J_2(\cdot)$ is adjoint *in*consistent.

In comparison to that, Figure 6b) shows the discrete adjoint solution $z_h$ connected to the discretization $J_{2,h}(\cdot)$ (cf. (164)) of the target quantity $J_2(\cdot)$. Here, we see that $z_h$ is perfectly smooth. Furthermore, we note that $z_h$ converges to the exact adjoint solution $z \equiv -1$. In fact, we have $z_h \equiv -1$ from the second coarsest mesh onwards. That is, the discrete adjoint solution is a consistent discretization of the continuous adjoint solution. In other words: the SIPG discretization in combination with the discrete target quantity $J_{2,h}(\cdot)$ is adjoint consistent.

Finally, we recall that the experimental order of convergence of the error $|J_2(u) - J_{2,h}(u_h)|$ of the adjoint consistent SIPG discretization behaves like $O(h^{2(p+1)})$ which is two powers of $h$ larger than the theoretically expected order $O(h^{2p})$ (cf. estimate (156)). A possible reason for this might be a too simple *model problem* in combination with a particularly simple target quantity which results in the constant continuous adjoint solution $z \equiv -1$.

**Example 3** In order to demonstrate that the estimate (156) is sharp we consider the following problem: Let $\Omega = (0,1) \times (0.1,1)$ and consider Poisson's equation (79) with forcing function $f$ which is chosen so that the analytical solution to (79) is given by

$$u(\mathbf{x}) = \tfrac{1}{4}(1+x_1)^2 \sin(2\pi x_1 x_2). \tag{165}$$

We note that this is a modification of the problem considered in (Harriman et al., 2003). Again, we impose Dirichlet boundary conditions where the boundary value function $g_D$ on $\Gamma_D = \Gamma$ is prescribed based on the solution $u$. We consider two discretizations of the target quantity $J_3(\cdot)$, the direct one, $J_3(u_h)$, and the adjoint consistent one, $J_{3,h}(u_h)$, given as follows

$$J_3(u_h) = \int_{\Gamma_D} j_D\, \mathbf{n}\cdot\nabla_h u_h \,\mathrm{d}s, \tag{166}$$

$$J_{3,h}(u_h) = \int_{\Gamma_D} j_D\, \hat{\boldsymbol{\sigma}}_h\cdot\mathbf{n}\,\mathrm{d}s = J_3(u_h) - \int_\Gamma \delta(u_h - g_D) j_D \,\mathrm{d}s. \tag{167}$$

and choose $j_D \in L^2(\Gamma)$ to be given by

$$j_D(\mathbf{x}) = \begin{cases} \exp\left(4 - \frac{1}{16}((x_1-\frac14)^2-\frac18)^{-2}\right) & \text{for } \mathbf{x}\in(0,\frac14)\times(0.1,1),\\ \exp\left(4 - \frac{1}{16}((x_1-\frac34)^2-\frac18)^{-2}\right) & \text{for } \mathbf{x}\in(\frac34,1)\times(0.1,1),\\ 1 & \text{for } \mathbf{x}\in(\frac14,\frac34)\times(0.1,1),\\ 0 & \text{elsewhere on } \Gamma.\end{cases}$$

Thereby, the true value of the target quantity is $J_3(u) = -1.2825165799606$.

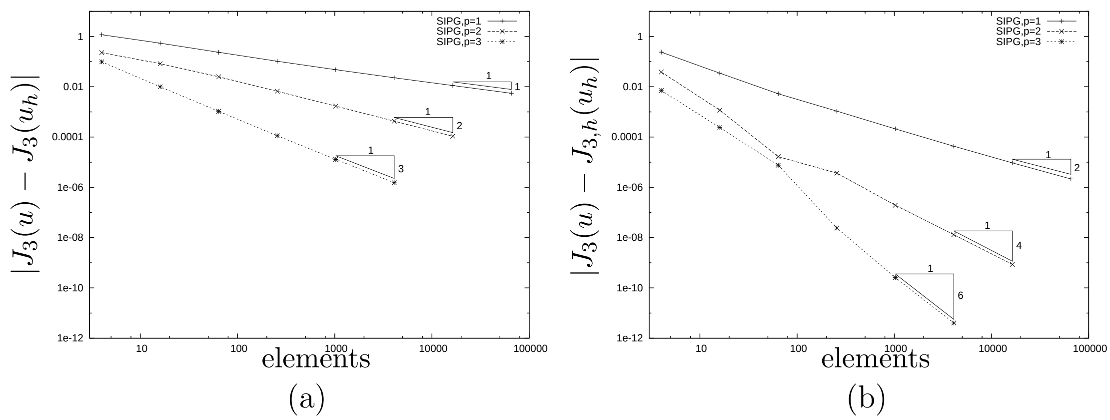
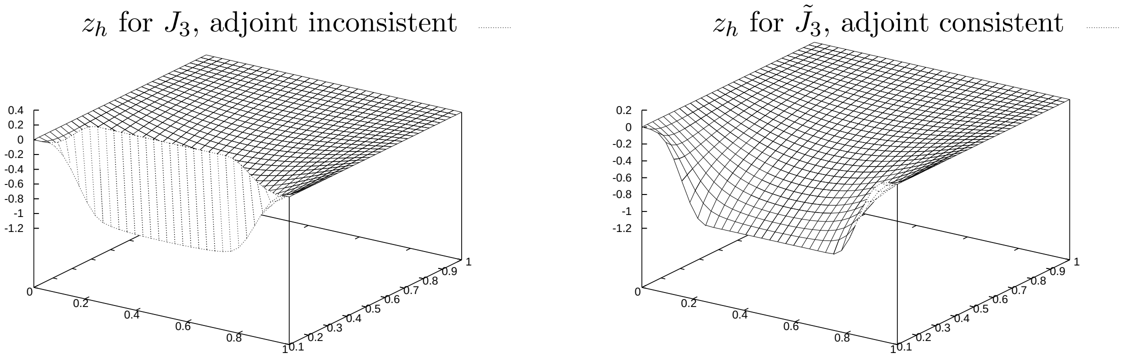



Figure 7: Example 3: Convergence of a) the error $|J_3(u) - J_3(u_h)|$ and b) the error $|J_3(u) - J_{3,h}(u_h)|$ for the SIPG discretization with global mesh refinement.

Figure 7a) shows that the convergence behavior of the error $|J_3(u) - J_3(u_h)|$ behaves like $O(h^p)$ which is in perfect agreement with the estimate (157) for an adjoint *in*consistent discretization. Furthermore, Figure 7b) shows that the convergence behavior of the error $|J_3(u) - J_{3,h}(u_h)|$ behaves like $O(h^{2p})$ which is as expected (cf. estimate (156)), for an adjoint consistent discretization. Finally, Figure 8a) shows the discrete adjoint solution $z_h$ connected to the (original) target quantity $J_3(\cdot)$. We see that in the neighborhood of the bottom boundary $[0,1]\times\{0.1\}\subset\Gamma$ the discrete adjoint solution is irregular which corresponds to the fact that the SIPG discretization in combination with the target quantity $J_3(\cdot)$ is adjoint inconsistent. In contrast to that the corresponding Figure 8b) shows that the discrete adjoint solution $z_h$ connected to the adjoint consistent target quantity $J_{3,h}(u_h)$ is entirely smooth which corresponds to the fact that the SIPG discretization in combination with the discrete target quantity $J_{3,h}(u_h)$ is adjoint consistent.




(a) (b)

Figure 8: Example 3: Discrete adjoint solution $z_h$ corresponding to the SIPG discretization combined with a) the adjoint inconsistent $J_3(u_h)$ and b) the adjoint consistent $J_{3,h}(u_h)$.

# 5 DG discretization of the compressible Euler equations

In this section we consider the discontinuous Galerkin discretization of the compressible Euler equations. Including conservation equations for mass, momentum and energy, these equations describe inviscid compressible flows and are frequently used as a simple model for gas flows.

## 5.1 The compressible Euler equations

We consider the two-dimensional steady state compressible Euler equations given by

$$\nabla \cdot \mathcal{F}^c(\mathbf{u}) = 0 \quad \text{in } \Omega \subset \mathbb{R}^2, \qquad\qquad B\mathbf{u} = 0 \quad \text{on } \Gamma, \tag{168}$$

where the nonlinear boundary operator $B = B(\mathbf{u}, \mathbf{n})$ on the boundary $\Gamma = \partial\Omega$ with unit outwards normal vector $\mathbf{n}$ will be specified at the end of this section. Here, the vector of conservative variables $\mathbf{u}$ and the convective flux $\mathcal{F}^c(\mathbf{u}) = (\mathbf{f}_1^c(\mathbf{u}), \mathbf{f}_2^c(\mathbf{u}))^\top$ are defined by

$$\mathbf{u} = \begin{bmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ \rho E \end{bmatrix}, \quad \mathbf{f}_1^c(\mathbf{u}) = \begin{bmatrix} \rho v_1 \\ \rho v_1^2 + p \\ \rho v_1 v_2 \\ \rho H v_1 \end{bmatrix} \quad \text{and} \quad \mathbf{f}_2^c(\mathbf{u}) = \begin{bmatrix} \rho v_2 \\ \rho v_1 v_2 \\ \rho v_2^2 + p \\ \rho H v_2 \end{bmatrix}, \tag{169}$$

where $\rho$, $\mathbf{v} = (v_1, v_2)^\top$, $p$ and $E$ denote the density, velocity vector, pressure and specific total energy, respectively. Additionally, $H$ is the total enthalpy given by

$$H = E + \frac{p}{\rho} = e + \tfrac{1}{2}\mathbf{v}^2 + \frac{p}{\rho}, \tag{170}$$

where $e$ is the specific static internal energy, and the pressure is determined by the equation of state of an ideal gas

$$p = (\gamma - 1)\rho e, \tag{171}$$

where $\gamma = c_p/c_v$ is the ratio of specific heat capacities at constant pressure, $c_p$, and constant volume, $c_v$; for dry air, $\gamma = 1.4$. The flux Jacobians $A_i(\mathbf{u}) := \partial_{\mathbf{u}} \mathbf{f}_i^c(\mathbf{u})$, $i = 1, 2$, are given by

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

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

and the eigenvalues $\lambda_i$, $i = 1, \dots, 4$, of the normal flux Jacobian $A(\mathbf{u}, \mathbf{n}) := \sum_{i=1}^2 n_i A_i(\mathbf{u})$ are

$$\mathbf{v}\cdot\mathbf{n} - c,\ \mathbf{v}\cdot\mathbf{n},\ \mathbf{v}\cdot\mathbf{n},\ \mathbf{v}\cdot\mathbf{n} + c, \tag{172}$$

where $c = \sqrt{\gamma p/\rho}$ denotes the speed of sound. The matrix $A(\mathbf{u}, \mathbf{n})$ can be diagonalized as follows

$$A(\mathbf{u}, \mathbf{n}) = P \Lambda P^{-1}, \tag{173}$$

where $\Lambda = \mathrm{diag}(\lambda_i)$ denotes the diagonal matrix consisting of the eigenvalues (172), and $P = [\mathbf{r}_1, \dots, \mathbf{r}_4]$ denotes the $4 \times 4$ matrix of right eigenvectors of $A(\mathbf{u}, \mathbf{n})$ with $A\mathbf{r}_i = \lambda_i \mathbf{r}_i$, $i = 1, \dots, 4$. Finally, the positive and negative parts of $A(\mathbf{u}, \mathbf{n})$ are given by

$$A^{\pm}(\mathbf{u}, \mathbf{n}) = P \Lambda^{\pm} P^{-1}, \tag{174}$$

where $\Lambda^+ = \mathrm{diag}(\max(\lambda_i, 0))$ and $\Lambda^- = \mathrm{diag}(\min(\lambda_i, 0))$ denote the $4 \times 4$ diagonal matrices of the, respectively, positive and negative eigenvalues of $A(\mathbf{u}, \mathbf{n})$.

Finally, we define the boundary operator $B$ introduced in (168) and a corresponding *boundary value function* $\mathbf{u}_\Gamma(\mathbf{u})$ on $\Gamma$. On the wall boundary $\Gamma_W$ they are given by

$$B\mathbf{u} = n_1 u_2 + n_2 u_3, \qquad \mathbf{u}_\Gamma(\mathbf{u}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - n_1^2 & -n_1 n_2 & 0 \\ 0 & -n_1 n_2 & 1 - n_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{u} \quad \text{on} \quad \Gamma_{\mathrm{W}}.$$

Note, that $\mathbf{u}_\Gamma(\mathbf{u})$ originates from $\mathbf{u}$ by removing the normal velocity component of $\mathbf{u}$, i.e., $\mathbf{v} = (v_1, v_2)$ is replaced by $\mathbf{v}_\Gamma = \mathbf{v} - (\mathbf{n} \cdot \mathbf{v})\mathbf{n}$. This choice ensures a vanishing normal velocity, $B\mathbf{u}_\Gamma(\mathbf{u}) = n_1 u_{\Gamma,2} + n_2 u_{\Gamma,3} = \rho\, \mathbf{n} \cdot \mathbf{v}_\Gamma = 0$, on $\Gamma_W$.

For defining the boundary operator $B$ on the farfield boundary $\Gamma \setminus \Gamma_W$, we consider the signs of the eigenvalues $\lambda_i$, $i = 1, \dots, 4$, in (172), ordered such that $\lambda_i < \lambda_j$ for $i < j$. Each eigenvalue smaller than zero corresponds to an inflow characteristic. The number of variables to be prescribed on the farfield boundary depend on the number of inflow characteristics. Thereby, we distinguish four cases of farfield boundary conditions and define the corresponding boundary operator $B$ and corresponding *boundary value functions* $\mathbf{u}_\Gamma(\mathbf{u})$ as follows:

- The supersonic inflow boundary condition on the part of the farfield boundary where $\lambda_i < 0$, $i = 1, \dots, 4$, takes all variables from the freestream state $\mathbf{u}_\infty$, corresponding to Dirichlet boundary conditions, i.e.,

$$B\mathbf{u} = \mathbf{u} - \mathbf{u}_\infty, \qquad \mathbf{u}_\Gamma(\mathbf{u}) = \mathbf{g}_D = \mathbf{u}_\infty.$$

- The supersonic outflow boundary condition on the part of farfield boundary where $\lambda_i > 0$, $i = 1, \dots, 4$, takes all variables from the flow field, corresponding to Neumann boundary conditions, i.e.,

$$B\mathbf{u} \equiv 0, \qquad \mathbf{u}_\Gamma(\mathbf{u}) = \mathbf{u}.$$

- The subsonic inflow boundary condition on the part of the farfield boundary where $\lambda_i < 0$, $i = 1, 2, 3$, $\lambda_4 > 0$, takes the pressure from the flow field and imposes all other variables based on freestream conditions $\mathbf{u}_\infty$, i.e.,

$$B\mathbf{u} = \begin{pmatrix} u_1 - u_{\infty,1} \\ u_2 - u_{\infty,2} \\ u_3 - u_{\infty,3} \end{pmatrix}, \qquad \mathbf{u}_\Gamma(\mathbf{u}) = \left( u_{\infty,1}, u_{\infty,2}, u_{\infty,3}, \frac{p(\mathbf{u})}{\gamma - 1} + \frac{u_{\infty,2}^2 + u_{\infty,3}^2}{2u_{\infty,1}} \right)^\top.$$

Here, $p(\mathbf{u})$ denotes the pressure evaluated at the state $\mathbf{u}$ using the equation of state (171).

- The subsonic outflow boundary condition on the part of the farfield boundary where $\lambda_1 < 0$, $\lambda_i > 0$, $i = 2,3,4$, imposes an outflow pressure $p_{\text{out}}$ and takes all other variables from the flow field, i.e.,

$$B\mathbf{u} = p(\mathbf{u}) - p_{\text{out}}, \qquad \mathbf{u}_\Gamma(\mathbf{u}) = \left(u_1, u_2, u_3, \frac{p_{\text{out}}}{\gamma - 1} + \frac{u_2^2 + u_3^2}{2u_1}\right)^\top.$$

Alternatively, we can apply characteristic boundary conditions on the farfield boundary with

$$B\mathbf{u} = A^-(\mathbf{u} - \mathbf{u}_\infty) = 0 \quad \text{on } \Gamma \setminus \Gamma_W, \tag{175}$$

where $A^- = A^-(\mathbf{u}, \mathbf{n})$ denotes the negative part of the normal flux Jacobian (174).

## 5.2 The continuous adjoint equations

The most important (and compatible) target quantities in inviscid compressible flows are the pressure induced drag and lift coefficients, $C_{\text{dp}}$ and $C_{\text{lp}}$, defined by

$$J(\mathbf{u}) = \int_\Gamma j(\mathbf{u})\,\mathrm{d}s = \int_{\Gamma_W} p\,\mathbf{n}\cdot\boldsymbol{\psi}\,\mathrm{d}s, \tag{176}$$

where $j(\mathbf{u}) = p(\mathbf{u})\,\mathbf{n}\cdot\boldsymbol{\psi}$ on $\Gamma_W$ and $j(\mathbf{u}) \equiv 0$ on $\Gamma \setminus \Gamma_W$. Here, $\boldsymbol{\psi}$ is given by $\boldsymbol{\psi}_d = \frac{1}{C_\infty}(\cos(\alpha), \sin(\alpha))^\top$ or $\boldsymbol{\psi}_l = \frac{1}{C_\infty}(-\sin(\alpha), \cos(\alpha))^\top$ for the drag and lift coefficient, respectively, where $\alpha$ is the angle of attack. Furthermore, $C_\infty = \frac{1}{2}\gamma p_\infty M_\infty^2 \bar{l} = \frac{1}{2}\gamma \frac{|\mathbf{v}_\infty|^2}{c_\infty^2} p_\infty \bar{l} = \frac{1}{2}\rho_\infty |\mathbf{v}_\infty|^2 \bar{l} = q_\infty \bar{l}$, where $M$ denotes the Mach number, $q = \frac{1}{2}\rho|\mathbf{v}|^2$ denotes the dynamic pressure and $\bar{l}$ denotes a reference length. Subscripts $_\infty$ indicate freestream quantities.

In order to derive the continuous adjoint problem, we multiply the left hand side of (168) by $\mathbf{z}$, integrate by parts and linearize about $\mathbf{u}$ to obtain: find $\mathbf{z} \in V$ such that

$$\left(\nabla\cdot\left(\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}](\mathbf{w})\right), \mathbf{z}\right)_\Omega = -\left(\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}](\mathbf{w}), \nabla\mathbf{z}\right)_\Omega + \left(\mathbf{n}\cdot\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}](\mathbf{w}), \mathbf{z}\right)_\Gamma \quad \forall \mathbf{w} \in \mathbf{V}_A, \tag{177}$$

where $\mathbf{V}_A := \{\mathbf{w} \in V : B'[\mathbf{u}]\mathbf{w} = 0\}$ is the space of admissible variations which do not alter the boundary values, and $B'[\mathbf{u}]$ is the derivative with respect to $\mathbf{u}$ of the boundary operator $B$ introduced in (168). Here, $\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}] := (\mathcal{F}^c)'[\mathbf{u}]$ denotes the Fréchet derivative of $\mathcal{F}^c$ with respect to $\mathbf{u}$. Here, we already use the subscript $_{\mathbf{u}}$ notation, which we require in Section 6 to distinguish from subscript $_{\nabla\mathbf{u}}$ denoting the derivative with respect to $\nabla\mathbf{u}$. Thereby, the variational formulation of the continuous adjoint problem is given by: find $\mathbf{z}$ such that

$$-\left(\mathbf{w}, \left(\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \nabla\mathbf{z}\right)_\Omega + \left(\mathbf{w}, \left(\mathbf{n}\cdot\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \mathbf{z}\right)_\Gamma = J'[\mathbf{u}](\mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V}_A, \tag{178}$$

and the continuous adjoint problem is given by

$$-\left(\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \nabla\mathbf{z} = 0 \quad \text{in } \Omega, \qquad \left(\mathbf{n}\cdot\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \mathbf{z} = j'[\mathbf{u}] \quad \text{on } \Gamma_W. \tag{179}$$

Using $\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}) = p(0,n_1,n_2,0)^\top$ on $\Gamma_W$, and the definition of $j$ in (176) we obtain

$$p'[\mathbf{u}](0,n_1,n_2,0)^\top\cdot\mathbf{z} = p'[\mathbf{u}]\mathbf{n}\cdot\boldsymbol{\psi} \quad \text{on } \Gamma_W,$$

which reduces to the wall boundary condition of the adjoint compressible Euler equations,

$$(B'[\mathbf{u}])^*\mathbf{z} = n_1 z_2 + n_2 z_3 = \mathbf{n}\cdot\boldsymbol{\psi} \quad \text{on } \Gamma_W. \tag{180}$$

Finally, we deduce from (178) following adjoint condition on the farfield boundary $\Gamma\setminus\Gamma_W$,

$$\left(\mathbf{w}, \left(\mathbf{n}\cdot\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \mathbf{z}\right)_{\Gamma\setminus\Gamma_W} = 0 \quad \forall\mathbf{w}\in\mathbf{V}_A. \tag{181}$$

## 5.3 Derivation of the DG discretization

We begin by introducing the vector-valued counterpart of the discrete function space $V_h^p$ defined in (52). Let $\mathbf{V}_h^p$ be the finite element space consisting of discontinuous vector-valued polynomial functions of degree $p\geq 0$, defined by

$$\begin{aligned}\mathbf{V}_h^p = \{\mathbf{v}_h \in \left[L^2(\Omega)\right]^4 : \mathbf{v}_h|_\kappa\circ F_\kappa \in [Q_p(\hat\kappa)]^4 \text{ if } \hat\kappa \text{ is the unit square, and}\\ \mathbf{v}_h|_\kappa\circ F_\kappa \in [P_p(\hat\kappa)]^4 \text{ if } \hat\kappa \text{ is the unit triangle}, \kappa\in\mathcal{T}_h\},\end{aligned} \tag{182}$$

where here, like for $V_h^p$ in (52), we use the tensor-product polynomial space $Q_p$ on the unit square, and the polynomial space $P_p$ on the unit triangle.

For deriving the discontinuous Galerkin discretization of the compressible Euler equations we proceed similarly to the case of the linear advection equation in Section 3. In order to introduce a weak formulation of (168) we multiply it by an arbitrary smooth vector-valued function $\mathbf{v}$ and integrate by parts over an element $\kappa$ in the mesh $\mathcal{T}_h$; thereby, we obtain

$$-\int_\kappa \mathcal{F}^c(\mathbf{u}) : \nabla\mathbf{v}\,\mathrm{d}\mathbf{x} + \int_{\partial\kappa}(\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}))\cdot\mathbf{v}\,\mathrm{d}s = 0, \tag{183}$$

where we use the standard notation $\underline{\sigma}:\underline{\tau} = \sum_{k=1}^m\sum_{l=1}^n \sigma_{kl}\tau_{kl}$ for matrices $\underline{\sigma},\underline{\tau}\in\mathbb{R}^{m\times n}$.

After summation over all elements $\kappa\in\mathcal{T}_h$, we replace the analytical solution $\mathbf{u}$ by the Galerkin finite element approximation $\mathbf{u}_h$ and the test function $\mathbf{v}$ by $\mathbf{v}_h$, where $\mathbf{u}_h$ and $\mathbf{v}_h$ both belong to the finite element space $\mathbf{V}_h^p$. In addition, since the numerical solution $\mathbf{u}_h$ may be discontinuous at interfaces $\partial\kappa\cap\partial\kappa'$ between elements $\kappa,\kappa'\in\mathcal{T}_h$, $\kappa'\neq\kappa$, we replace the normal flux $\mathcal{F}^c(\mathbf{u})\cdot\mathbf{n}$ by a *numerical flux function*

$$\hat{\mathbf{h}}_h = \hat{\mathbf{h}}(\mathbf{u}_h,\mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_h^-,\mathbf{n}), \tag{184}$$

which depends on both the interior and exterior traces, $\mathbf{u}_h^+$ and $\mathbf{u}_h^-$, of $\mathbf{u}_h$ on $\partial\kappa\setminus\Gamma$, $\kappa\in\mathcal{T}_h$, and the unit outward normal $\mathbf{n}$ to $\partial\kappa$. On the boundary $\Gamma$ we replace the normal flux $\mathcal{F}^c(\mathbf{u})\cdot\mathbf{n}$ by a *numerical boundary flux function*,

$$\hat{\mathbf{h}}_h|_\Gamma = \hat{\mathbf{h}}_{\Gamma,h} = \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}). \tag{185}$$

Then, the DG discretization of (168) is given by: find $\mathbf{u}_h\in\mathbf{V}_h^p$ such that

$$\begin{aligned}\hat{N}_h(\mathbf{u}_h,\mathbf{v}_h) \equiv -\int_\Omega \mathcal{F}^c(\mathbf{u}_h):\nabla_h\mathbf{v}_h\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_h^-,\mathbf{n})\cdot\mathbf{v}_h^+\,\mathrm{d}s\\ + \int_\Gamma \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n})\cdot\mathbf{v}_h^+\,\mathrm{d}s = 0 \qquad \forall\mathbf{v}_h\in\mathbf{V}_h^p.\end{aligned} \tag{186}$$

Using the notation $\hat{\mathbf{h}}_h$ and $\hat{\mathbf{h}}_h|_\Gamma = \hat{\mathbf{h}}_{\Gamma,h}$ introduced in (184) and (185) this can be rewritten in following short notation: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$-\int_\Omega \mathcal{F}^c(\mathbf{u}_h) : \nabla_h \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\mathbf{h}}_h \cdot \mathbf{v}_h^+ \,\mathrm{d}s = 0 \qquad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \tag{187}$$

## 5.4 Consistency and convervation property

Integrating (187) back by parts on each element $\kappa$ we obtain following equivalent form of the discretization: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega (\nabla\cdot\mathcal{F}^c(\mathbf{u}_h))\cdot\mathbf{v}_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \left(\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_h) - \hat{\mathbf{h}}_h\right)\cdot \mathbf{v}_h^+ \,\mathrm{d}s = 0 \qquad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \tag{188}$$

Thus, we obtain the primal residual form: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega \mathbf{R}(\mathbf{u}_h)\cdot\mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \mathbf{r}(\mathbf{u}_h)\cdot\mathbf{v}_h^+ \,\mathrm{d}s + \int_\Gamma \mathbf{r}_\Gamma(\mathbf{u}_h)\cdot\mathbf{v}_h^+ \,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p, \tag{189}$$

where the primal residuals are given by

$$\begin{aligned} \mathbf{R}(\mathbf{u}_h) &= -\nabla\cdot\mathcal{F}^c(\mathbf{u}_h) && \text{in } \kappa, \kappa\in\mathcal{T}_h, \\ \mathbf{r}(\mathbf{u}_h) &= \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_h^+) - \hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_h^-,\mathbf{n}) && \text{on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h, \\ \mathbf{r}_\Gamma(\mathbf{u}_h) &= \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_h^+) - \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) && \text{on } \Gamma. \end{aligned} \tag{190}$$

Motivated by this we arrive at following two definitions:

**Definition 5.1** *A numerical flux function* $\hat{\mathbf{h}}$ *is said to be* consistent *if*

$$\hat{\mathbf{h}}(\mathbf{v},\mathbf{v},\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{v}) \tag{191}$$

*whenever* $\mathbf{v}$ *is a smooth function.*

Note, that this is an extension of Definition (3.1) to a general normal flux $\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u})$.

**Definition 5.2** *A numerical flux function* $\hat{\mathbf{h}}_\Gamma$ *at the boundary is said to be* consistent *if the exact solution* $\mathbf{u}$ *to* (168) *satisfies*

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u},\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}). \tag{192}$$

**Lemma 5.3 (Consistency)** *Let the numerical fluxes* $\hat{\mathbf{h}}$ *and* $\hat{\mathbf{h}}_\Gamma$ *in* (186) *be* consistent*. Then* (186) *is a* consistent *discretization of the primal equations* (168)*.*

**Proof:** As shown before, the discretization (186) is equivalent to the discretization in (189). (189) is a consistent discretization of (168) because the exact solution $\mathbf{u}$ to (168) satisfies

$$\int_\Omega \mathbf{R}(\mathbf{u})\cdot\mathbf{v} \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \mathbf{r}(\mathbf{u})\cdot\mathbf{v}^+ \,\mathrm{d}s + \int_\Gamma \mathbf{r}_\Gamma(\mathbf{u})\cdot\mathbf{v}^+ \,\mathrm{d}s = 0 \quad \forall \mathbf{v}\in\mathbf{V}, \tag{193}$$

where $\mathbf{R}(\mathbf{u}) = 0$ due to (186), $\mathbf{r}(\mathbf{u}) = 0$ due to the consistency of $\hat{\mathbf{h}}$, and $\mathbf{r}_\Gamma(\mathbf{u}) = 0$ due to the consistency of $\hat{\mathbf{h}}_\Gamma$. □

Like in Section 3.5 we define the following.

**Definition 5.4** *A numerical flux function* $\hat{\mathbf{h}}(\mathbf{u}^+, \mathbf{u}^-, \mathbf{n})$ *is said to be* conservative *if*

$$\hat{\mathbf{h}}(\mathbf{u}^+, \mathbf{u}^-, \mathbf{n}) = -\hat{\mathbf{h}}(\mathbf{u}^-, \mathbf{u}^+, -\mathbf{n}).$$

**Lemma 5.5** *Let the numerical flux* $\hat{\mathbf{h}}$ *be conservative. Then the discretization* (186) *is conservative, i.e., for any union* $S$ *of elements we have*

$$\int_{\partial S} \hat{\mathbf{h}}_h \, ds = 0. \tag{194}$$

*In particular, we have the* global *and* local conservation property *for any* $\kappa \in \mathcal{T}_h$*,*

$$\int_{\Gamma} \hat{\mathbf{h}}_h \, ds = 0, \qquad \int_{\partial\kappa} \hat{\mathbf{h}}_h \, ds = 0. \tag{195}$$

**Proof:** Starting from (187) and proceeding along the lines of Section 3.5. □

## 5.5 Numerical flux functions

As seen in the last section, a consistent and conservative discretization requires the use of consistent and conservative numerical flux functions. There are many numerical flux functions satisfying these conditions, such as the Godunov, Engquist–Osher, Lax–Friedrichs, Roe or the Vijayasundaram flux. As examples, in the following we give details for three different numerical flux functions:

- The (local) Lax–Friedrichs flux $\mathbf{h}_{\mathrm{LF}}(\cdot,\cdot,\cdot)$ is defined by

  $$\mathbf{h}_{\mathrm{LF}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})|_{\partial\kappa} = \tfrac{1}{2}\left(\mathcal{F}^c(\mathbf{u}_h^+)\cdot\mathbf{n} + \mathcal{F}^c(\mathbf{u}_h^-)\cdot\mathbf{n} + \alpha\left(\mathbf{u}_h^+ - \mathbf{u}_h^-\right)\right),$$

  for $\kappa \in \mathcal{T}_h$, where $\alpha$ is the maximum over $\mathbf{u}_h^+$ and $\mathbf{u}_h^-$,

  $$\alpha = \max_{\mathbf{v}=\mathbf{u}_h^+,\mathbf{u}_h^-} \{|\lambda(B(\mathbf{v},\mathbf{n}))|\},$$

  of the largest (in absolute value) of the eigenvalues (cf. (172)) of the matrix $B(\mathbf{v},\mathbf{n}) = \sum_{i=0}^d n_i A_i(\mathbf{u})$.

- The Vijayasundaram flux $\mathbf{h}_{\mathrm{Vija}}(\cdot,\cdot,\cdot)$ is defined by

  $$\mathbf{h}_{\mathrm{Vija}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})|_{\partial\kappa} = A^+(\{\!\{\mathbf{u}_h\}\!\},\mathbf{n})\mathbf{u}_h^+ + A^-(\{\!\{\mathbf{u}_h\}\!\},\mathbf{n})\mathbf{u}^- \quad \text{for } \kappa \in \mathcal{T}_h, \tag{196}$$

  where $A^+(\cdot,\mathbf{n})$ and $A^-(\cdot,\mathbf{n})$ denote the positive and negative parts (cf. (174)) of the matrix $A(\cdot,\mathbf{n})$ and are evaluated at the mean value $\{\!\{\mathbf{u}_h\}\!\} = \frac{1}{2}(\mathbf{u}_h^+ + \mathbf{u}_h^-)$.

- `roe`: The Roe flux with entropy fix, $\mathbf{h}_{\mathrm{Roe}}(\cdot,\cdot,\cdot)$, is defined by

  $$\mathbf{h}_{\mathrm{Roe}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})|_{\partial\kappa} = \tfrac{1}{2}\left(\mathcal{F}^c(\mathbf{u}_h^+)\cdot\mathbf{n} + \mathcal{F}^c(\mathbf{u}_h^-)\cdot\mathbf{n} + A_{\mathrm{ef}}(\bar{\mathbf{u}}_h,\mathbf{n})(\mathbf{u}_h^+ - \mathbf{u}_h^-)\right), \tag{197}$$

  Here, $A_{\mathrm{ef}} = P\Lambda_{\mathrm{ef}}P^{-1}$ with $\Lambda_{\mathrm{ef}} = \mathrm{diag}(\alpha_i)$ is based on the diagonalization of $A(\bar{\mathbf{u}},\mathbf{n}) = P\Lambda P^{-1}$ (cf. (173)) and $\bar{\mathbf{u}}_h$ is the Roe mean value given by $\bar{\rho} = \sqrt{\rho^+\rho^-}$. Furthermore, $\bar{v}_1$, $\bar{v}_2$, $\bar{H}$ are computed according to

  $$\bar{u} := \frac{u^+\sqrt{\rho^+} + u^-\sqrt{\rho^-}}{\sqrt{\rho^+} + \sqrt{\rho^-}},$$

for $u = v_1, v_2, H$, respectively. Without entropy fix the $\alpha_i$ values are given by the absolute eigenvalues $\alpha_i = |\lambda_i|$ of $A(\bar{\mathbf{u}}, \mathbf{n})$. With entropy fix the $\alpha_i$ values are given by

$$\alpha_i = \begin{array}{ll} \frac{|\lambda_i|^2+\delta^2}{2\delta} & \text{if } |\lambda_i| < \delta \\ |\lambda_i| & \text{else.} \end{array} \qquad \text{with } \delta = \delta_{\text{ef}}\lambda_{\max}, \tag{198}$$

where $\lambda_{\max} = \max_i\{|\lambda_i|\}$, and $0 < \delta_{\text{ef}} \leq 1$ is a parameter usually taken in the range of $[0.05, 0.2]$. It it inspired by Harten (Harten and Hyman, 1983) and has the advantage of being differentiable at the point $|\lambda_i| = \delta$.

**Remark 5.6** *We note that for the linear advection equation* (45) *we have* $\mathcal{F}^c(u) = \mathbf{b}u$. *For that model problem most numerical fluxes, in particular the numerical fluxes introduced above, reduce to the upwind flux given in* (63)*:*

$$\mathbf{h}_{uw}(u_h^+, u_h^-, \mathbf{n}) = \begin{cases} \mathbf{b}\cdot\mathbf{n}\, u_h^-, & \textit{for } (\mathbf{b}\cdot\mathbf{n})(\mathbf{x}) < 0, \ \textit{i.e.}, \ \mathbf{x} \in \partial\kappa_-, \\ \mathbf{b}\cdot\mathbf{n}\, u_h^+, & \textit{for } (\mathbf{b}\cdot\mathbf{n})(\mathbf{x}) \geq 0, \ \textit{i.e.}, \ \mathbf{x} \in \partial\kappa_+. \end{cases}$$

## 5.6 Numerical flux functions at the boundary

In this section we introduce possible choices of the numerical flux function $\hat{\mathbf{h}}_\Gamma$ at the boundary $\Gamma$. First, we consider $\hat{\mathbf{h}}_\Gamma$ at the farfield boundary $\Gamma \setminus \Gamma_W$ (cf. Section 5.6.1) and then $\hat{\mathbf{h}}_\Gamma$ at the wall boundary $\Gamma_W$ (cf. Section 5.6.2).

### 5.6.1 Farfield boundary

One of the possible choices of numerical flux functions at the farfield boundary is given by

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)), \tag{199}$$

where $\mathbf{u}_\Gamma(\cdot)$ denotes the *boundary value function* as defined in Section 5.1 for a variety of farfield boundary conditions. The boundary value function $\mathbf{u}_\Gamma(\cdot)$ might depend on boundary values, like the freestream state $\mathbf{u}_\infty$ on the farfield boundary, but it might also depend on the interior state $\mathbf{u}_h^+$, thus $\mathbf{u}_\Gamma = \mathbf{u}_\Gamma(\mathbf{u}_h^+)$.

**Definition 5.7** *A boundary value function* $\mathbf{u}_\Gamma(\cdot)$ *is said to be consistent if the exact solution* $\mathbf{u}$ *to the flow problem* (168) *satisfies* $\mathbf{u}_\Gamma(\mathbf{u}) = \mathbf{u}$.

**Lemma 5.8** *Let* $\mathbf{u}_\Gamma(\cdot)$ *be a consistent boundary value function. Then the numerical boundary flux function given by*

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) \tag{200}$$

*is consistent.*

**Proof:** Note, that the exact solution to $\mathbf{u}$ to the flow problem (168) satisfies

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}, \mathbf{n}) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u})) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}),$$

due to consistency of $\mathbf{u}_\Gamma(\cdot)$. □

**Remark 5.9** *All boundary value functions* $\mathbf{u}_\Gamma(\cdot)$ *introduced Section 5.1 are consistent. Thus, the numerical boundary flux function* $\hat{\mathbf{h}}_\Gamma$ *in* (199) *is consistent for any of those boundary value functions.*

Furthermore, we consider characteristic farfield boundary conditions. While imposing Dirichlet boundary conditions on characteristic inflow variables based on freestream conditions $\mathbf{u}_\infty$, it imposes no boundary conditions on characteristic outflow variables,

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = A^+(\mathbf{u}_h^+,\mathbf{n})\mathbf{u}_h^+ + A^-(\mathbf{u}_h^+,\mathbf{n})\mathbf{u}_\infty, \tag{201}$$

where $A^\pm(\mathbf{u},\mathbf{n})$ are the positive and negative parts of the normal flux Jacobian $A(\mathbf{u},\mathbf{n})$ defined in (174). This corresponds to using the Vijayasundaram flux (196) on the farfield boundary with the normal flux Jacobian being evaluated at the interior state $\mathbf{u}_h^+$. As an alternative, a characteristic farfield boundary condition could be based on the Roe flux (197). Due to consistency of the Vijayasundaram (or Roe) flux also the corresponding numerical boundary flux $\hat{\mathbf{h}}_\Gamma$ is consistent.

### 5.6.2 Wall boundary

We recall that on interior faces, $\partial\kappa\backslash\Gamma$, the numerical flux $\hat{\mathbf{h}}$ in (186) is evaluated connecting the interior state $\mathbf{u}_h^+$ with the exterior state $\mathbf{u}_h^-$. On interior faces, i.e., faces with an neighboring element $\kappa'$, the exterior state $\mathbf{u}_h^-$ with respect to $\kappa$ is just the interior state of the neighboring element $\kappa'$. On boundary faces $\partial\kappa\cap\Gamma$ there is no neighboring element and thus no exterior state $\mathbf{u}_h^-$. As a replacement one could define a *boundary exterior state* $\mathbf{u}_\Gamma^-$ which depends on the *boundary state* $\mathbf{u}_\Gamma(\mathbf{u}_h^+)$ but which might also depend on the interior state $\mathbf{u}_h^+$, thus $\mathbf{u}_\Gamma^- = \mathbf{u}_\Gamma^-(\mathbf{u}_h^+) \equiv \mathbf{u}_\Gamma^-(\mathbf{u}_h^+, \mathbf{u}_\Gamma(\mathbf{u}_h^+))$. Using the same numerical flux $\hat{\mathbf{h}}$ on the boundary like on interior faces, one would obtain $\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n})$. However, one could define $\hat{\mathbf{h}}_\Gamma$ totally independent of the numerical flux $\hat{\mathbf{h}}$ employed on interior faces, as long as it is consistent (cf. Definition 5.2) as this is required for the consistency of the discretization (186) (cf. Lemma 5.3).

In the following we introduce two possible choises of $\hat{\mathbf{h}}_\Gamma$ resulting in two (of the arbitrary many different possible) discretizations at the boundary. The first one is based on the normal (component of the non-numerical) flux evaluated at the wall boundary state $\mathbf{u}_\Gamma(\mathbf{u}_h^+)$ with zero normal velocity. The second one is based on a numerical flux connecting the interior state $\mathbf{u}^+$ with a mirrored exterior boundary state $\mathbf{u}_\Gamma^-$. In detail we consider the following two choices:

1. The numerical boundary flux function $\hat{\mathbf{h}}_\Gamma$ is based on the normal boundary flux,

   $$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)), \tag{202}$$

   i.e., the normal flux is evaluated at the *boundary state* $\mathbf{u}_\Gamma(\mathbf{u}_h^+)$ where the *boundary value function* $\mathbf{u}_\Gamma(\cdot)$ is given by (5.1).

2. The numerical boundary flux function $\hat{\mathbf{h}}_\Gamma$ is based on the interior numerical flux,

   $$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n}), \tag{203}$$

i.e., the numerical flux function $\hat{\mathbf{h}}$ employed on interior faces is also employed on boundary faces, where the *boundary exterior state* $\mathbf{u}_\Gamma^-(\mathbf{u}_h^+)$ is obtained by mirroring the interior state $\mathbf{u}_h^+$ at the boundary state $\mathbf{u}_\Gamma(\mathbf{u}_h^+)$ (cf. (5.1)) according to

$$\frac{1}{2}\left(\mathbf{u}_h^+ + \mathbf{u}_\Gamma^-(\mathbf{u}_h^+)\right) = \mathbf{u}_\Gamma(\mathbf{u}_h^+). \tag{204}$$

This corresponds to introducing a ghost layer of elements at the wall boundary and evaluating the discretization on the wall boundary like on interior faces.

**Remark 5.10** *First numerical tests indicate that the discretization based on* (202) *seems to be more accurate on coarse grids and low polynomial degree $p$ of $\mathbf{u}_h \in \mathbf{V}_{h,p}$ than that based on* (203) *while the discretization based on* (203) *seems to be more stable than that based on* (202).

**Remark 5.11** *The two different discretizations of boundary terms given in* (202) *and* (203) *can be viewed as extremes of a whole class of discretizations. Consider*

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_\Gamma^+(\mathbf{u}_h^+), \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n}), \tag{205}$$

*with*

$$\mathbf{u}_\Gamma^+(\mathbf{u}_h^+) = \xi\mathbf{u}_h^+ + (1-\xi)\mathbf{u}_\Gamma(\mathbf{u}_h^+), \qquad \mathbf{u}_\Gamma^-(\mathbf{u}_h^+) = 2\mathbf{u}_\Gamma(\mathbf{u}_h^+) - \mathbf{u}_\Gamma^+(\mathbf{u}_h^+). \tag{206}$$

*For $\xi = 0$ this reduces to* (202) *while for $\xi = 1$ it reduces to* (203). *The parameter $\xi \in [0,1]$ would allow to choose a discretization of boundary fluxes as a compromise between accuracy and stability.*

**Remark 5.12** *Note, that the boundary value function $\mathbf{u}_\Gamma(\cdot)$ in* (5.1) *is consistent. Thereby,*

- *the numerical boundary flux in* (202) *is consistent according to Lemma 5.8, and*
- *and the numerical flux functions* (203) *and* (205) *are consistent.*

## 5.7 Adjoint consistency

In this section, we analyze the adjoint consistency of the discretization (186) of the compressible Euler equations (168) in combination with specific discretizations of the force coefficients (176).

While in previous work (Hartmann, 2007a,b) a discretization of the force coefficients (176) was proposed which is adjoint consistent in combination with the DG discretization (186) and the special case of the normal wall boundary flux (202), in the following we generalize this to any *consistent* discretization $\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n})$ of boundary fluxes. By this we mean, that for any discretization $\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n})$ of the boundary fluxes (not only for the examples given in Section 5.6), we give an associated discretization of the force coefficients (176) which makes the discretization adjoint consistent.

To this end, assuming that $\hat{\mathbf{h}}_{\Gamma,h} = \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h, \mathbf{n})$ is consistent, we define the following discretization of the force coefficients,

$$J_h(\mathbf{u}_h) = \int_\Gamma \hat{\mathbf{h}}_{\Gamma,h} \cdot \tilde{\boldsymbol{\psi}}\, \mathrm{d}s, \tag{207}$$

with $\tilde{\boldsymbol{\psi}} = (0, \psi_1, \psi_2, 0)^\top$ on $\Gamma_W$ for $\boldsymbol{\psi} = (\psi_1, \psi_2)^\top$ in (176), and $\tilde{\boldsymbol{\psi}} \equiv 0$ on $\Gamma \setminus \Gamma_W$. Note, that this is a consistent discretization of the target quantity (176), i.e., the values of the target quantity (176) and its discretization (207) evaluated for the exact solution $\mathbf{u}$ to (168) are identical, $J_h(\mathbf{u}) = J(\mathbf{u})$. To see this, recall that due to consistency of $\hat{\mathbf{h}}_\Gamma$, the exact solution $\mathbf{u}$ to (168) satisfies $\hat{\mathbf{h}}_\Gamma(\mathbf{u}, \mathbf{n}) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u})$. Using $\mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}) = p(0, n_1, n_2, 0)^\top$ on $\Gamma_W$ we obtain

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}, \mathbf{n}) \cdot \tilde{\boldsymbol{\psi}} = (\mathbf{n} \cdot \mathcal{F}^c(\mathbf{u})) \cdot \tilde{\boldsymbol{\psi}} = p(\mathbf{u})(0, n_1, n_2, 0)^\top \cdot \tilde{\boldsymbol{\psi}} = p(\mathbf{u})\mathbf{n} \cdot \boldsymbol{\psi} \tag{208}$$

and arrive at $J_h(\mathbf{u}) = J(\mathbf{u})$.

**Theorem 5.13** *Let the numerical flux* $\hat{\mathbf{h}}$ *in* (186) *be consistent and conservative and the numerical boundary flux* $\hat{\mathbf{h}}_\Gamma$ *in* (186) *be consistent. Then, the discretization based on* (186) *and* (207) *is* adjoint consistent.

**Proof:** To show adjoint consistency, we consider the discrete adjoint problem: find $\mathbf{z}_h \in \mathbf{V}_h^p$ such that

$$\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h) = J_h'[\mathbf{u}_h](\mathbf{w}_h) \qquad \forall \mathbf{w}_h \in \mathbf{V}_h^p. \tag{209}$$

Here, $J_h'[\mathbf{u}_h](\mathbf{w}_h)$ is the Fréchet derivative of $J_h(\mathbf{u}_h)$ in (207) at $\mathbf{u}_h$ in the direction of $\mathbf{w}_h$, given by

$$J_h'[\mathbf{u}_h](\mathbf{w}_h) = \int_{\Gamma_W} \hat{\mathbf{h}}_{\Gamma,\mathbf{u}} \mathbf{w}_h \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s, \tag{210}$$

where $\hat{\mathbf{h}}_{\Gamma,\mathbf{u}}$ is a short notation for $\partial_\mathbf{u} \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h, \mathbf{n})$. Furthermore, $\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h)$ is the Fréchet derivative of $\hat{N}_h(\mathbf{u}_h, \mathbf{z}_h)$ in (186) with respect to $\mathbf{u}_h$ in the direction of $\mathbf{w}_h$ given by

$$\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h) = -\int_\Omega \mathcal{F}^c_\mathbf{u}[\mathbf{u}_h]\mathbf{w}_h : \nabla_h \mathbf{z}_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} \left( \hat{\mathbf{h}}_{\mathbf{u}^+} \mathbf{w}_h^+ + \hat{\mathbf{h}}_{\mathbf{u}^-} \mathbf{w}_h^- \right) \mathbf{z}_h^+ \, \mathrm{d}s + \int_\Gamma \hat{\mathbf{h}}_{\Gamma,\mathbf{u}} \mathbf{w}_h \cdot \mathbf{z}_h \, \mathrm{d}s. \tag{211}$$

Here, $\hat{\mathbf{h}}_{\mathbf{u}^+}$ and $\hat{\mathbf{h}}_{\mathbf{u}^-}$ denote the partial derivatives of $\hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})$ with respect to its first and second argument, respectively. As the numerical flux function $\hat{\mathbf{h}}$ is conservative, $\hat{\mathbf{h}}(\mathbf{v}, \mathbf{w}, \mathbf{n}) = -\hat{\mathbf{h}}(\mathbf{w}, \mathbf{v}, -\mathbf{n})$, we obtain $\hat{\mathbf{h}}_{\mathbf{u}^-}(\mathbf{v}, \mathbf{w}, \mathbf{n}) = \partial_\mathbf{w} \hat{\mathbf{h}}(\mathbf{v}, \mathbf{w}, \mathbf{n}) = -\partial_\mathbf{w} \hat{\mathbf{h}}(\mathbf{w}, \mathbf{v}, -\mathbf{n}) = -\hat{\mathbf{h}}_{\mathbf{u}^+}(\mathbf{w}, \mathbf{v}, -\mathbf{n})$. Thus on an interior face $e$ we have

$$\int_e \hat{\mathbf{h}}_{\mathbf{u}^-}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}^+)\mathbf{w}_h^- \mathbf{z}_h^+ \, \mathrm{d}s = -\int_e \hat{\mathbf{h}}_{\mathbf{u}^+}(\mathbf{u}_h^-, \mathbf{u}_h^+, \mathbf{n}^-)\mathbf{w}_h^- \mathbf{z}_h^+ \, \mathrm{d}s = -\int_e \hat{\mathbf{h}}_{\mathbf{u}^+}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}^+)\mathbf{w}_h^+ \mathbf{z}_h^- \, \mathrm{d}s, \tag{212}$$

where in the last step we exchanged notation $^+$ and $^-$ on $e$. Thus (211) reduces to

$$\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h) = -\int_\Omega \mathcal{F}^c_\mathbf{u}[\mathbf{u}_h]\mathbf{w}_h : \nabla_h \mathbf{z}_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} \hat{\mathbf{h}}_{\mathbf{u}^+} \mathbf{w}_h^+ \left( \mathbf{z}_h^+ - \mathbf{z}_h^- \right) \mathrm{d}s + \int_\Gamma \hat{\mathbf{h}}_{\Gamma,\mathbf{u}} \mathbf{w}_h \cdot \mathbf{z}_h \, \mathrm{d}s.$$

Then, the discrete adjoint problem (209) can be rewritten in adjoint residual form (like in (42)) as follows: find $\mathbf{z}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega \mathbf{w}_h \cdot \mathbf{R}^*[\mathbf{u}_h](\mathbf{z}_h) \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} \mathbf{w}_h^+ \cdot \mathbf{r}^*[\mathbf{u}_h](\mathbf{z}_h) \, \mathrm{d}s + \int_\Gamma \mathbf{w}_h \cdot \mathbf{r}_\Gamma^*[\mathbf{u}_h](\mathbf{z}_h) \, \mathrm{d}s = 0, \tag{213}$$

for all $\mathbf{w}_h \in \mathbf{V}_h^p$, where the adjoint residuals are given by

$$\begin{aligned}
\mathbf{R}^*[\mathbf{u}_h](\mathbf{z}_h) &= (\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}_h])^\top \nabla \mathbf{z}_h && \text{in } \kappa, \kappa \in \mathcal{T}_h,\\
\mathbf{r}^*[\mathbf{u}_h](\mathbf{z}_h) &= -\hat{\mathbf{h}}^\top_{\mathbf{u}^+}\left(\mathbf{z}_h^+ - \mathbf{z}_h^-\right) && \text{on } \partial\kappa \setminus \Gamma, \kappa \in \mathcal{T}_h,\\
\mathbf{r}^*_\Gamma[\mathbf{u}_h](\mathbf{z}_h) &= \hat{\mathbf{h}}^\top_{\Gamma,\mathbf{u}}\left(\tilde{\boldsymbol{\psi}} - \mathbf{z}_h\right) && \text{on } \Gamma.
\end{aligned} \tag{214}$$

For proving adjoint consistency, we want to show that the exact solution $\mathbf{u}$ to the flow equations (168) and the exact (and assumingly smooth) solution $\mathbf{z}$ to the adjoint equations (179) satisfy

$$\int_\Omega \mathbf{w}\cdot\mathbf{R}^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}} \int_{\partial\kappa\setminus\Gamma} \mathbf{w}\cdot\mathbf{r}^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}s + \int_\Gamma \mathbf{w}\cdot\mathbf{r}^*_\Gamma[\mathbf{u}](\mathbf{z})\,\mathrm{d}s = 0, \tag{215}$$

for all admissible variations $\mathbf{w} \in \mathbf{V}_A$. To this end, note that on the farfield boundary $\Gamma \setminus \Gamma_W$ we have $\tilde{\boldsymbol{\psi}} = 0$. Furthermore, by using consistency of $\hat{\mathbf{h}}_\Gamma$, i.e., the exact solution $\mathbf{u}$ satisfies $\hat{\mathbf{h}}_\Gamma(\mathbf{u}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u})$, we arrive at

$$\int_{\Gamma\setminus\Gamma_W} \mathbf{w}\cdot\mathbf{r}^*_\Gamma[\mathbf{u}](\mathbf{z})\,\mathrm{d}s = -\int_{\Gamma\setminus\Gamma_W} \mathbf{w}\cdot\hat{\mathbf{h}}^\top_{\Gamma,\mathbf{u}}\,\mathbf{z}\,\mathrm{d}s = -\left(\mathbf{w}, \left(\mathbf{n}\cdot\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}]\right)^\top \mathbf{z}\right)_{\Gamma\setminus\Gamma_W} = 0,$$

for all $\mathbf{w} \in \mathbf{V}_A$ due to (181). Furthermore, note that due to the adjoint equations (179) and due to the smoothness of $\mathbf{z}$ we have

$$\mathbf{R}^*[\mathbf{u}](\mathbf{z}) = 0 \;\text{ in } \kappa, \kappa \in \mathcal{T}_h, \qquad \mathbf{r}^*[\mathbf{u}](\mathbf{z}) = 0 \;\text{ on } \partial\kappa\setminus\Gamma, \kappa\in\mathcal{T}_h.$$

Finally, at the wall boundary $\Gamma_W$ we have, due to the relation (208) and the adjoint boundary condition (180),

$$\mathbf{r}^*_\Gamma[\mathbf{u}](\mathbf{z}) = p'[\mathbf{u}]\,(0, n_1, n_2, 0)^\top \cdot \left(\tilde{\boldsymbol{\psi}} - \mathbf{z}\right) = p'[\mathbf{u}]\,\left(\mathbf{n}\cdot\boldsymbol{\psi} - \mathbf{n}\cdot(z_2, z_3)^\top\right) = 0 \;\text{ on } \Gamma_W.$$

Thus, the discretization (186) in combination with (207) is adjoint consistent. □

In the following we discuss some typical combinations of numerical boundary fluxes and discretizations of the force coefficients (176), i.e.,

$$J(\mathbf{u}) = \int_{\Gamma_W} p(\mathbf{u})\,\mathbf{n}\cdot\boldsymbol{\psi}\,\mathrm{d}s. \tag{216}$$

Before doing so, let us first consider following most “simple” discretization of this target quantity given by

$$J(\mathbf{u}_h) = \int_{\Gamma_W} p(\mathbf{u}_h)\,\mathbf{n}\cdot\boldsymbol{\psi}\,\mathrm{d}s, \tag{217}$$

which we denote as *direct discretization* of the target quantity in the following.

**Example 5.14** *Let the numerical boundary flux in* (186) *be given as in* (202) *by*

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)). \tag{218}$$

*Then the discretization is given by: find* $\mathbf{u}_h \in \mathbf{V}_h^p$ *such that*

$$-\int_\Omega \mathcal{F}^c(\mathbf{u}_h) : \nabla_h \mathbf{v}_h \, d\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) \cdot \mathbf{v}_h^+ \, ds + \int_\Gamma \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) \cdot \mathbf{v}_h^+ \, ds = 0 \qquad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \quad (219)$$

(a) *The discretization* (219) *in combination with the following discretization of* $J(\cdot)$,

$$\begin{aligned} J_h(\mathbf{u}_h) &= \int_{\Gamma_W} \hat{\mathbf{h}}_{\Gamma,h} \cdot \tilde{\boldsymbol{\psi}} \, ds = \int_{\Gamma_W} \left(\mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \, ds \\ &= \int_{\Gamma_W} p\left(\mathbf{u}_\Gamma(\mathbf{u}_h^+)\right) \, \mathbf{n} \cdot \boldsymbol{\psi} \, ds = J(\mathbf{u}_\Gamma(\mathbf{u}_h^+)), \end{aligned} \quad (220)$$

*is* adjoint consistent. *Note, that here we used* (208). *This coincides with the adjoint consistent discretization of the compressible Euler equations discussed in (Hartmann, 2007a,b).*

*Any other discretization of the target quantity does* not *yield an adjoint consistent discretization*

(b) *In particular, the discretization* (219) *in combination with the* direct discretization *of the target quantity as given in* (217) *is adjoint* in*consistent.*

**Example 5.15** *Let the numerical boundary flux in* (186) *be given as in* (203) *by*

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n}). \quad (221)$$

*Then the discretization is given by: find* $\mathbf{u}_h \in \mathbf{V}_h^p$ *such that*

$$-\int_\Omega \mathcal{F}^c(\mathbf{u}_h) : \nabla_h \mathbf{v}_h \, d\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) \cdot \mathbf{v}_h^+ \, ds + \int_\Gamma \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n}) \cdot \mathbf{v}_h^+ \, ds = 0 \qquad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \quad (222)$$

(a) *The discretization* (222) *in combination with the following discretization of* $J(\cdot)$,

$$J_h(\mathbf{u}_h) = \int_{\Gamma_W} \hat{\mathbf{h}}_{\Gamma,h} \cdot \tilde{\boldsymbol{\psi}} \, ds = \int_{\Gamma_W} \hat{\mathbf{h}}(\mathbf{u}_h^+, \mathbf{u}_\Gamma^-(\mathbf{u}_h^+), \mathbf{n}) \cdot \tilde{\boldsymbol{\psi}} \, ds, \quad (223)$$

*is* adjoint consistent.

*Any other discretization of the target quantity does* not *yield an adjoint consistent discretization.*

(b) *In particular, the discretization* (222) *in combination with the* direct discretization *of the target quantity as given in* (217) *is adjoint* in*consistent.*

(c) *Also, the discretization* (222) *in combination with a discretization of the target quantity based on a different numerical flux, like e.g.* (220), *is adjoint* in*consistent.*

## 5.8 Numerical results

**Ringleb flow problem** For discretizations of the 2D steady state compressible Euler equations there are virtually no *a priori* error estimates available. Therefore, in the following we examine the order of convergence of the DG discretization experimentally. In particular, we consider the solution to the 2d compressible Euler equations for the Ringleb flow problem. This is one of the few non-trivial problems of the 2d Euler equations for which a smooth analytical solution is known. For this problem the analytical solution may be obtained be employing the hodograph transformation, see (Chiocchia, 1985) or the appendix of (Hartmann, 2002). This problem represents a transonic flow in a channel, see Figure 9a), with inflow and outflow boundaries given by the lower and upper boundaries of the domain, and reflective (slip wall) boundaries with vanishing normal velocity, $\mathbf{v}\cdot\mathbf{n}=0$, on the left and right boundary. The solution to this flow problem is smooth but it is transonic with a small supersonic region near the lower right corner. The computational domain is subdivided into quadrilateral elements. Figure 9 shows the coarsest three meshes in a sequence of globally refined meshes.

Here we impose characteristic boundary conditions, $A^-(\mathbf{u},\mathbf{n})\,(\mathbf{u}-\mathbf{g})=0$ on the whole boundary $\Gamma$ of the domain, where $\mathbf{g}$ is the boundary value function taken from the exact solution to the Ringleb flow problem. This boundary condition represents an inflow boundary condition for characteristic variables on inflow parts (with respect to the corresponding characteristics) of the boundary. Figure 10, taken from (Hartmann, 2002), plots the $L^2(\Omega)$-error of the DG$(p)$, $0\le p\le 4$, solutions against the number of degrees of freedom (DoFs) on the sequence of globally refined meshes. We observe an experimental order $O(h^{p+1})$ of convergence which is optimal for polynomial trial and test functions of degree $p$.

**Inviscid flow around the NACA0012 airfoil** In the following, we investigate the smoothness of the discrete adjoint solution when employing an adjoint consistent discretization (cf. Example 5.14(a)) in comparison to an adjoint inconsistent discretization (cf. Example 5.15(b)). To this end, we consider an inviscid Mach $M=0.5$ flow at a zero angle of attack, i.e., $\alpha=0°$, around the NACA0012 airfoil. Here, the upper and lower

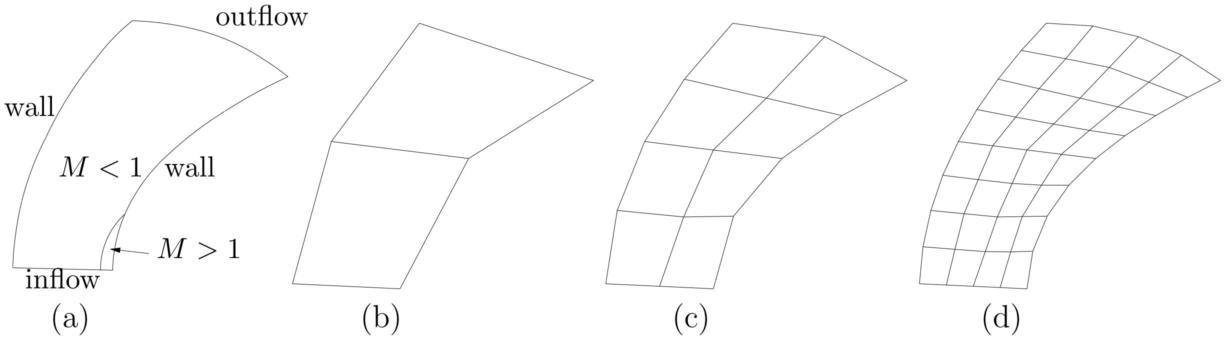



Figure 9: Ringleb flow problem: a) Regions of sub- and supersonic flow denoted by the Mach number $M<1$ and $M>1$; b)-d) Coarse meshes with 2, 8 and 32 elements, respectively.

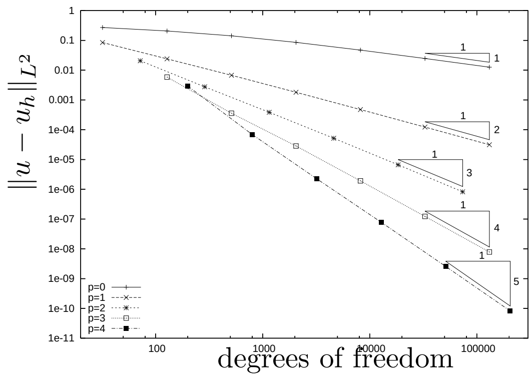



Figure 10: Ringleb flow problem: The $L^2$-error of the DG($p$), $p = 0, \ldots, 4$, discretizations of the compressible Euler equations is of order $\mathcal{O}(h^{p+1})$, (Hartmann, 2002).

surfaces of the airfoil geometry are specified by the function $g^{\pm}$, respectively, where

$$g^{\pm}(s) = \pm 5 \times 0.12 \times (0.2969 s^{1/2} - 0.126 s - 0.3516 s^2 + 0.2843 s^3 - 0.1015 s^4).$$

As the chord length $l$ of the airfoil is $l \approx 1.00893$ we use a rescaling of $g$ in order to yield an airfoil of unit (chord) length. The computational domain $\Omega$ is subdivided into quadrilateral elements. Curved boundaries are approximated by piecewise quadratic polynomials. In Figure 11, taken from (Hartmann, 2007b), we compare the (primal) flow solutions $\mathbf{u}_h \in \mathbf{V}_h^1$ for the adjoint consistent and the adjoint *in*consistent DG discretizations and find no visible difference. However, when comparing the adjoint solutions corresponding to the pressure induced drag coefficient $C_{\mathrm{dp}}$, see Figure 12, we notice that the discrete adjoint solution to the adjoint *in*consistent DG discretization is irregular near and upstream the airfoil. In contrast to that, the adjoint solution to the adjoint consistent discretization is entirely smooth. Furthermore, in (Hartmann, 2007b) it has been shown that for this test case on a sequence of locally refined meshes the error in the $C_{\mathrm{dp}}$ value for the adjoint consistent discretization is by a factor 1.3-2.4 smaller than for the adjoint inconsistent discretization.

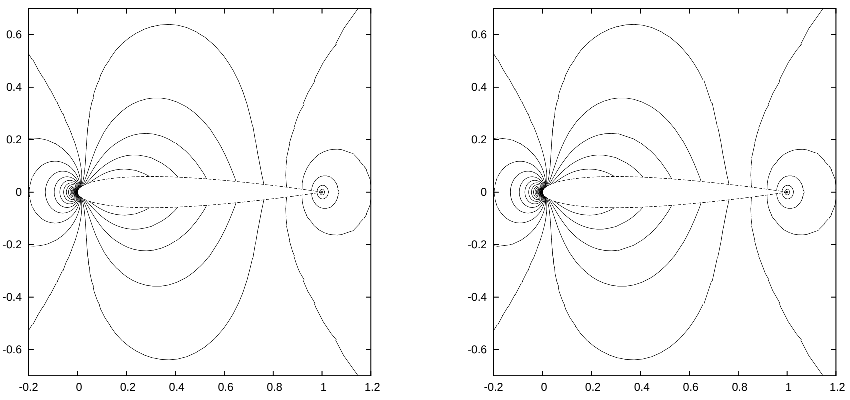



Figure 11: Inviscid flow around the NACA0012 airfoil: Mach isolines of the flow solution $\mathbf{u}_h$ to (left) the adjoint consistent and (right) the adjoint inconsistent DG discretization, (Hartmann, 2007b).

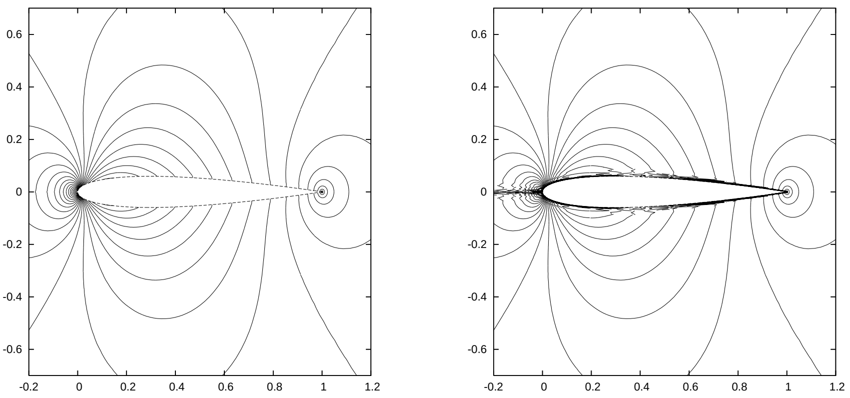



Figure 12: Inviscid flow around the NACA0012 airfoil: $z_1$ isolines of the discrete adjoint solution $\mathbf{z}_h$ to (left) the adjoint consistent and (right) the adjoint inconsistent DG discretization, (Hartmann, 2007b).

# 6 DG discretization of the compressible Navier-Stokes equations

The compressible Euler equations as considered in the last section serve as a simple model for gas flows. In fact, while ignoring all viscous effects they describe an inviscid compressible flow. In the following, we will enrich the physical model by including also diffusive terms. The resulting compressible Navier-Stokes equations serve as a model for laminar viscous compressible flows.

## 6.1 The compressible Navier-Stokes equations

In the following we give a detailed description of the two–dimensional steady state compressible Navier-Stokes equations. Like in Section 5.1, $\rho$, $\mathbf{v} = (v_1, v_2)^\top$, $p$ and $E$ denote the density, velocity vector, pressure and specific total energy, respectively. Furthermore, $T$ denotes the temperature. The equations of motion are given by

$$\nabla \cdot (\mathcal{F}^c(\mathbf{u}) - \mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u})) = 0 \quad \text{in } \Omega \subset \mathbb{R}^2, \qquad B\mathbf{u} = 0, \quad B_1\mathbf{u} = 0 \quad \text{on } \Gamma. \qquad (224)$$

where the nonlinear boundary operator $B = B(\mathbf{u}, \mathbf{n})$ and the nonlinear differential boundary operator $B_1 = B_1(\mathbf{u}, \mathbf{n})$ on $\Gamma$ will be specified at the end of this section, and $\Gamma = \partial\Omega$ is the boundary of the domain $\Omega$ with unit outwards normal vector $\mathbf{n}$. The vector of conservative variables $\mathbf{u}$ and the convective fluxes $\mathbf{f}_k^c$, $k = 1, 2$, are given by (169). Furthermore, the diffusive fluxes $\mathbf{f}_k^v$, $k = 1, 2$, are defined by

$$\mathbf{f}_1^v(\mathbf{u}, \nabla\mathbf{u}) = \begin{bmatrix} 0 \\ \tau_{11} \\ \tau_{21} \\ \tau_{1j}v_j + \mathcal{K}T_{x_1} \end{bmatrix} \quad \text{and} \quad \mathbf{f}_2^v(\mathbf{u}, \nabla\mathbf{u}) = \begin{bmatrix} 0 \\ \tau_{12} \\ \tau_{22} \\ \tau_{2j}v_j + \mathcal{K}T_{x_2} \end{bmatrix},$$

respectively, where $\mathcal{K}$ is the thermal conductivity coefficient. Finally, the viscous stress tensor is defined by

$$\tau = \mu \left( \nabla\mathbf{v} + (\nabla\mathbf{v})^\top - \tfrac{2}{3}(\nabla \cdot \mathbf{v})I \right),$$

where $\mu$ is the dynamic viscosity coefficient, and the temperature $T$ is given by $e = c_v T$; thus

$$\mathcal{K}T = \tfrac{\mu\gamma}{Pr} \left( E - \tfrac{1}{2}\mathbf{v}^2 \right),$$

where $Pr = \frac{\mu c_p}{\mathcal{K}} = 0.72$ is the Prandtl number.

For the purposes of discretization, we rewrite the compressible Navier–Stokes equations (224) in the following (equivalent) form:

$$\frac{\partial}{\partial x_k} \left( \mathbf{f}_k^c(\mathbf{u}) - \mathbf{f}_k^v(\mathbf{u}, \nabla\mathbf{u}) \right) = \frac{\partial}{\partial x_k} \left( \mathbf{f}_k^c(\mathbf{u}) - G_{kl}(\mathbf{u}) \frac{\partial \mathbf{u}}{\partial x_l} \right) = 0 \quad \text{in } \Omega. \qquad (225)$$

Here, the matrices $G_{kl}(\mathbf{u}) = \partial \mathbf{f}_k^v(\mathbf{u}, \nabla\mathbf{u}) / \partial u_{x_l}$, for $k, l = 1, 2$, are the homogeneity tensors defined by $\mathbf{f}_k^v(\mathbf{u},\nabla\mathbf{u}) = G_{kl}(\mathbf{u})\partial\mathbf{u}/\partial x_l$, $k=1,2$, where

$$
\begin{gathered}
G_{11} = \frac{\mu}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ -\frac{4}{3}v_1 & \frac{4}{3} & 0 & 0 \\ -v_2 & 0 & 1 & 0 \\ -\left(\frac{4}{3}v_1^2 + v_2^2 + \frac{\gamma}{Pr}\left(E-\mathbf{v}^2\right)\right) & \left(\frac{4}{3}-\frac{\gamma}{Pr}\right)v_1 & \left(1-\frac{\gamma}{Pr}\right)v_2 & \frac{\gamma}{Pr} \end{pmatrix}, \\
G_{12} = \frac{\mu}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{2}{3}v_2 & 0 & -\frac{2}{3} & 0 \\ -v_1 & 1 & 0 & 0 \\ -\frac{1}{3}v_1v_2 & v_2 & -\frac{2}{3}v_1 & 0 \end{pmatrix}, G_{21} = \frac{\mu}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ -v_2 & 0 & 1 & 0 \\ \frac{2}{3}v_1 & -\frac{2}{3} & 0 & 0 \\ -\frac{1}{3}v_1v_2 & -\frac{2}{3}v_2 & v_1 & 0 \end{pmatrix}, \\
G_{22} = \frac{\mu}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ -v_1 & 1 & 0 & 0 \\ -\frac{4}{3}v_2 & 0 & \frac{4}{3} & 0 \\ -\left(v_1^2 + \frac{4}{3}v_2^2 + \frac{\gamma}{Pr}\left(E-\mathbf{v}^2\right)\right) & \left(1-\frac{\gamma}{Pr}\right)v_1 & \left(\frac{4}{3}-\frac{\gamma}{Pr}\right)v_2 & \frac{\gamma}{Pr} \end{pmatrix}.
\end{gathered} \tag{226}
$$

For viscous flows, we distinguish between *isothermal* and *adiabatic* no-slip wall boundary conditions. To this end, decomposing $\Gamma_W = \Gamma_{\text{iso}} \cup \Gamma_{\text{adia}}$, we set

$$
\mathbf{v} = 0 \;\text{ on }\; \Gamma_W, \qquad T = T_{\text{wall}} \;\text{ on }\; \Gamma_{\text{iso}}, \qquad \mathbf{n}\cdot\nabla T = 0 \;\text{ on }\; \Gamma_{\text{adia}}, \tag{227}
$$

where $T_{\text{wall}}$ is a prescribed wall temperature. Thereby, the boundary operators $B$ and $B_1$ introduced in (224) are given on $\Gamma_W$ as follows

$$
\begin{aligned}
B\mathbf{u} &= (u_2,u_3)^\top = 0, & B_1\mathbf{u} &= \mathbf{n}\cdot\nabla T, & \mathbf{u}_\Gamma(\mathbf{u}) &= (u_1,0,0,u_4)^\top \quad \text{on } \Gamma_{\text{adia}}, \\
B\mathbf{u} &= (u_2,u_3,T(\mathbf{u})-T_{\text{wall}})^\top, & B_1\mathbf{u} &\equiv 0, & \mathbf{u}_\Gamma(\mathbf{u}) &= (u_1,0,0,u_1c_vT_{\text{wall}})^\top \quad \text{on } \Gamma_{\text{iso}},
\end{aligned} \tag{228}
$$

where $T(\mathbf{u})$ is the temperature computed from the flow state $\mathbf{u}$ as follows

$$
T(\mathbf{u}) = \tfrac{1}{c_v}(E - \tfrac{1}{2}v^2) = \tfrac{1}{c_v}E = \tfrac{1}{c_v}\tfrac{u_4}{u_1} \quad \text{on } \Gamma_W, \tag{229}
$$

due to $\mathbf{v}=0$ on the wall boundary. Furthermore, on the farfield boundary $\Gamma\setminus\Gamma_W$, $B$ is given as in Section 5.1 and $B_1 \equiv 0$.

**Remark 6.1** *We recall from Section 5.2 the definition of the space of admissible variations which do not alter the boundary values,*

$$
\mathbf{V}_A := \left\{\mathbf{w}\in V : B'[\mathbf{u}]\mathbf{w} = 0\right\}, \tag{230}
$$

*where $B'[\mathbf{u}]$ is the derivative with respect to $\mathbf{u}$ of boundary operator $B$ introduced in Section 5.1 and Eqn (228) above. This definition was used in Section 5.2 on the derivation of the continuous adjoint problem to the compressible Euler equations and will again be used in the next section in the derivation the continuous adjoint problem to the compressible Navier-Stokes equations. Note, that for following alternative boundary operator*

$$
\tilde{B}\mathbf{u} = \begin{cases} \mathbf{u} - \mathbf{u}_\Gamma(\mathbf{u}) & \text{on } \Gamma\setminus\Gamma_{\text{iso}}, \\ B\mathbf{u} & \text{on } \Gamma_{\text{iso}}, \end{cases} \tag{231}
$$

*we have $\left\{\mathbf{w}\in V : \tilde{B}'[\mathbf{u}]\mathbf{w} = 0\right\} = \mathbf{V}_A$. Here, we keep the original operator $B$ on $\Gamma_{\text{iso}}$ in order to make sure that the temperature $T$ (and not $u_4 = u_1c_vT$) shall be constant on $\Gamma_{\text{iso}}$.*

## 6.2 The continuous adjoint equations

The most important target quantities in viscous compressible flows are the total (i.e., the pressure induced plus viscous) drag and lift coefficients, $C_{\mathrm{d}}$ and $C_{\mathrm{l}}$, defined by

$$J(\mathbf{u}) = \int_\Gamma j(\mathbf{u})\,\mathrm{d}s = \int_{\Gamma_W} (p\,\mathbf{n} - \underline{\tau}\,\mathbf{n})\cdot\boldsymbol{\psi}\,\mathrm{d}s = \int_{\Gamma_W} (p\,n_i - \tau_{ij}n_j)\,\psi_i\,\mathrm{d}s, \tag{232}$$

where $j(\mathbf{u}) = (p\,\mathbf{n} - \underline{\tau}\,\mathbf{n})\cdot\boldsymbol{\psi}$ on $\Gamma_W$, $j(\mathbf{u}) \equiv 0$ on $\Gamma\setminus\Gamma_W$, and $\boldsymbol{\psi}$ is as in (176). In order to derive the adjoint problem, we multiply the left hand side of (224) by $\mathbf{z}$, integrate by parts and linearize about $\mathbf{u}$ to obtain: find $\mathbf{z}\in V$ such that

$$\begin{aligned}&(\nabla\cdot(\mathcal{F}^c_{\mathbf{u}}\mathbf{w} - \mathcal{F}^v_{\mathbf{u}}\mathbf{w} - \mathcal{F}^v_{\nabla\mathbf{u}}\nabla\mathbf{w}),\mathbf{z})_\Omega\\ &\qquad = -((\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}})\,\mathbf{w} - \mathcal{F}^v_{\nabla\mathbf{u}}\nabla\mathbf{w},\nabla\mathbf{z})_\Omega + (\mathbf{n}\cdot(\mathcal{F}^c_{\mathbf{u}}\mathbf{w} - \mathcal{F}^v_{\mathbf{u}}\mathbf{w} - \mathcal{F}^v_{\nabla\mathbf{u}}\nabla\mathbf{w}),\mathbf{z})_\Gamma\end{aligned} \tag{233}$$

for all $\mathbf{w}\in\mathbf{V}_A$, where $\mathcal{F}^v_{\mathbf{u}} := \partial_{\mathbf{u}}\mathcal{F}^v(\mathbf{u},\nabla\mathbf{u}) = G'[\mathbf{u}]\nabla\mathbf{u}$ and $\mathcal{F}^v_{\nabla\mathbf{u}} := \partial_{\nabla\mathbf{u}}\mathcal{F}^v(\mathbf{u},\nabla\mathbf{u}) = G(\mathbf{u})$ denote the derivatives of $\mathcal{F}^v$ with respect to $\mathbf{u}$ and $\nabla\mathbf{u}$, respectively. Here $\mathbf{V}_A = \{\mathbf{w}\in V : \tilde{B}'[\mathbf{u}]\mathbf{w} = 0\}$ is the space of admissible variations which do not alter the boundary values, and $\tilde{B}'[\mathbf{u}]$ is the derivative with respect to $\mathbf{u}$ of boundary operator $\tilde{B}$ introduced in (231). In particular, on the wall boundary $\Gamma_W$ the variations in the velocity must vanish, i.e., $w_2 = w_3 = 0$. Furthermore, on an isothermal boundary $\Gamma_{\text{iso}}$ the variations of $T'w = \partial_{u_i}Tw_i$ of the temperature must vanish. According to (229) the temperature at the wall is given by $T = \frac{u_4}{c_v u_1}$, resulting in $0 = T'w = \frac{1}{c_v u_1^2}(u_1w_4 - u_4w_1)$. Thus, the variations $\mathbf{w}$ in (233) on $\Gamma_W$ must fulfill following conditions:

$$w_2 = w_3 = 0 \quad \text{on } \Gamma_W, \qquad\qquad u_1w_4 - u_4w_1 = 0 \quad \text{on } \Gamma_{\text{iso}}. \tag{234}$$

Note, that there are also restrictions on the variations $\mathbf{w}$ on the farfield boundary $\Gamma\setminus\Gamma_W$. However, the explicit form of these conditions are not further explored and thus obmitted for brevity. Employing another integration by parts on (233), we obtain the following variational formulation of the continuous adjoint problem: find $\mathbf{z}$ such that

$$\begin{aligned}&-\left(\mathbf{w},(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}})^\top\nabla\mathbf{z}\right)_\Omega - \left(\mathbf{w},\nabla\cdot\left((\mathcal{F}^v_{\nabla\mathbf{u}})^\top\nabla\mathbf{z}\right)\right)_\Omega + \left(\mathbf{w},\mathbf{n}\cdot\left((\mathcal{F}^v_{\nabla\mathbf{u}})^\top\nabla\mathbf{z}\right)\right)_\Gamma\\ &\qquad + \left(\mathbf{w},(\mathbf{n}\cdot(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}}))^\top\mathbf{z}\right)_\Gamma - \left(\nabla\mathbf{w},(\mathbf{n}\cdot\mathcal{F}^v_{\nabla\mathbf{u}})^\top\mathbf{z}\right)_\Gamma = J'[\mathbf{u}](\mathbf{w}) \quad \forall\mathbf{w}\in\mathbf{V}_A.\end{aligned} \tag{235}$$

Given that

$$\begin{aligned}J'[\mathbf{u}](\mathbf{w}) &= \int_{\Gamma_W} (p_{\mathbf{u}}[\mathbf{u}]\,\mathbf{n} - \underline{\tau}_{\mathbf{u}}[\mathbf{u}]\,\mathbf{n})\cdot\boldsymbol{\psi}\,\mathbf{w} - (\underline{\tau}_{\nabla\mathbf{u}}[\mathbf{u}]\,\mathbf{n})\cdot\boldsymbol{\psi}\,\nabla\mathbf{w}\,\mathrm{d}s\\ &= (\mathbf{w},(p_{\mathbf{u}}\,\mathbf{n} - \underline{\tau}_{\mathbf{u}}\,\mathbf{n})\cdot\boldsymbol{\psi})_{\Gamma_W} - (\nabla\mathbf{w},(\underline{\tau}_{\nabla\mathbf{u}}\,\mathbf{n})\cdot\boldsymbol{\psi})_{\Gamma_W},\end{aligned} \tag{236}$$

we see, that the adjoint solution $\mathbf{z}$ satisfies following equation,

$$-(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}})^\top\nabla\mathbf{z} - \nabla\cdot\left((\mathcal{F}^v_{\nabla\mathbf{u}})^\top\nabla\mathbf{z}\right) = 0 \qquad \text{on } \Omega \tag{237}$$

subject to the boundary conditions

$$(\mathbf{n}\cdot\mathcal{F}^v_{\nabla\mathbf{u}})^\top\mathbf{z} = (\underline{\tau}_{\nabla\mathbf{u}}\,\mathbf{n})\cdot\boldsymbol{\psi} \qquad \text{on } \Gamma_W = \Gamma_{\text{iso}}\cup\Gamma_{\text{adia}}, \tag{238}$$

obtained from (235) by comparing the terms on $\Gamma_W$ in (235) and (236) that are multiplied with $\nabla\mathbf{w}$. At wall boundaries $\Gamma_W$, where $\mathbf{v} = (v_1, v_2)^\top = 0$, the normal diffusive flux reduces to

$$\mathbf{n}\cdot\mathcal{F}^v(\mathbf{u},\nabla\mathbf{u}) = (0, (\tau\mathbf{n})_1, (\tau\mathbf{n})_2, \mathcal{K}\mathbf{n}\cdot\nabla T)^\top. \tag{239}$$

Hence, (238) is fulfilled provided $\mathbf{z}$ satisfies

$$\begin{aligned}
(\tau_{\nabla\mathbf{u}}\mathbf{n})_1\, z_2 &= (\tau_{\nabla\mathbf{u}}\mathbf{n})_1\, \psi_1 \quad \text{on } \Gamma_W,\\
(\tau_{\nabla\mathbf{u}}\mathbf{n})_2\, z_3 &= (\tau_{\nabla\mathbf{u}}\mathbf{n})_2\, \psi_2 \quad \text{on } \Gamma_W,\\
\mathcal{K}\mathbf{n}\cdot\nabla T_{\nabla\mathbf{u}}\, z_4 &= 0 \qquad\qquad\qquad \text{on } \Gamma_{\text{iso}},
\end{aligned}$$

which reduce to the conditions

$$z_2 = \psi_1, \quad z_3 = \psi_2 \quad \text{on } \Gamma_W, \qquad\qquad z_4 = 0 \quad \text{on } \Gamma_{\text{iso}}. \tag{240}$$

An additional condition can be obtained from (235) by comparing all boundary terms on $\Gamma_W$ which are multiplied with $\mathbf{w}$,

$$(\mathbf{w}, \mathbf{n}\cdot((\mathcal{F}^v_{\nabla\mathbf{u}})^\top\nabla\mathbf{z}))_{\Gamma_W} + \left(\mathbf{w}, (\mathbf{n}\cdot(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}}))^\top\mathbf{z}\right)_{\Gamma_W} = (\mathbf{w}, (p_{\mathbf{u}}\,\mathbf{n} - \underline{\tau}_{\mathbf{u}}\,\mathbf{n})\cdot\boldsymbol{\psi})_{\Gamma_W}. \tag{241}$$

Furthermore, on the wall boundary $\Gamma_W$ the normal convective and diffusive fluxes reduce to $\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}) = p(\mathbf{u})(0, n_1, n_2, 0)$ and (239). Together with (240) this gives $\mathbf{n}\cdot(\mathcal{F}^c - \mathcal{F}^v)^\top\mathbf{z} = (p\,\mathbf{n} - \underline{\tau}\,\mathbf{n})\cdot\boldsymbol{\psi}$ and $\mathbf{n}\cdot(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}})^\top\mathbf{z} = (p_{\mathbf{u}}\,\mathbf{n} - \underline{\tau}_{\mathbf{u}}\,\mathbf{n})\cdot\boldsymbol{\psi}$ on $\Gamma_W$. Thus (241) reduces to

$$\left(\mathbf{w}, \mathbf{n}\cdot\left((\mathcal{F}^v_{\nabla\mathbf{u}})^\top\nabla\mathbf{z}\right)\right)_{\Gamma_W} = 0.$$

Recalling, that $\mathcal{F}^v_{\nabla\mathbf{u}} = G(\mathbf{u})$, this condition can be rewritten as

$$\int_{\Gamma_W} \mathbf{w}\otimes\mathbf{n} : G^\top(\mathbf{u})\nabla\mathbf{z}\,\mathrm{d}s = 0, \tag{242}$$

where we use the standard notation $\mathbf{v}\otimes\mathbf{w}\in\mathbb{R}^{m\times n}$ with $(\mathbf{v}\otimes\mathbf{w})_{kl} = v_k\,w_l$ for vectors $\mathbf{v}\in\mathbb{R}^m, \mathbf{w}\in\mathbb{R}^n$. On the wall boundary $\Gamma_W$ many terms in the homogeneity tensors $G(\mathbf{u})$ defined in (226) vanish due to $\mathbf{v} = (v_1, v_2)^\top = 0$. Due to this the condition (242) can be reduced as follows.

**Lemma 6.2** *Given a vector* $\mathbf{w} = (w_1, 0, 0, w_4)$ *and* $\mathbf{u}$ *(cf.* (169)*) with* $\mathbf{v} = (v_1, v_2)^\top = 0$. *Then*

$$\mathbf{w}\otimes\mathbf{n} : G^\top(\mathbf{u})\nabla\mathbf{z} = \tfrac{\mu}{\rho^2}\tfrac{\gamma}{Pr}\mathbf{n}\cdot\nabla z_4\,(u_1 w_4 - u_4 w_1). \tag{243}$$

**Proof:** With notations $\mathbf{w}^{\mathbf{n}} = \mathbf{w}\otimes\mathbf{n}$ with $w^n_{jl} = w_j n_l$, and $z_{i,k} = (\nabla\mathbf{z})_{ik}$, we obtain

$$\begin{aligned}
\mathbf{w}\otimes\mathbf{n} : G^\top(\mathbf{u})\nabla\mathbf{z} &= \mathbf{w}^{\mathbf{n}} : (G^\top(\mathbf{u})\nabla\mathbf{z}) = G(\mathbf{u})\mathbf{w}^{\mathbf{n}} : \nabla\mathbf{z} = (G_{kl})_{4j} w_j n_l z_{4,k}\\
&= (G_{11})_{41} w_1 n_1 z_{4,1} + (G_{11})_{44} w_4 n_1 z_{4,1} + (G_{22})_{41} w_1 n_2 z_{4,2} + (G_{22})_{44} w_4 n_2 z_{4,2}\\
&= \tfrac{\mu}{\rho}\tfrac{\gamma}{Pr}\left(-E w_1 n_1 z_{4,1} + w_4 n_1 z_{4,1} - E w_1 n_2 z_{4,2} + w_4 n_2 z_{4,2}\right)\\
&= \tfrac{\mu}{\rho}\tfrac{\gamma}{Pr}\mathbf{n}\cdot\nabla z_4\,(-E w_1 + w_4) = \tfrac{\mu}{\rho^2}\tfrac{\gamma}{Pr}\mathbf{n}\cdot\nabla z_4\,(u_1 w_4 - u_4 w_1).
\end{aligned}$$

□

Due to (234) and (243) the left hand side of (242) vanishes on $\Gamma_{\text{iso}}$ giving no additional information. On $\Gamma_{\text{adia}}$, however, (242) reduces to the condition $\mathbf{n}\cdot\nabla z_4 = 0$. Thus, we arrive at following set of adjoint boundary conditions:

$$z_2 = \psi_1, \quad z_3 = \psi_2 \quad \text{on } \Gamma_W, \qquad z_4 = 0 \quad \text{on } \Gamma_{\text{iso}}, \qquad \mathbf{n}\cdot\nabla z_4 = 0 \quad \text{on } \Gamma_{\text{adia}}. \tag{244}$$

Finally, we deduce from (235) following adjoint conditions on the farfield boundary $\Gamma\backslash\Gamma_W$,

$$\begin{aligned} \left(\mathbf{w}, \mathbf{n}\cdot((\mathcal{F}^v_{\nabla\mathbf{u}})^\top \nabla\mathbf{z}) + (\mathbf{n}\cdot(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}}))^\top \mathbf{z}\right)_{\Gamma\backslash\Gamma_W} &= 0 \quad \forall \mathbf{w}\in\mathbf{V}_A, \\ -\left(\nabla\mathbf{w}, (\mathbf{n}\cdot\mathcal{F}^v_{\nabla\mathbf{u}})^\top \mathbf{z}\right)_{\Gamma\backslash\Gamma_W} &= 0 \quad \forall \mathbf{w}\in\mathbf{V}_A. \end{aligned} \tag{245}$$

## 6.3 Derivation of the DG discretization

In Section 5 we have concentrated on the discretization of the compressible Euler equations, i.e., on the discretization of the convective flux $\nabla\cdot\mathcal{F}^c(\mathbf{u}) = \frac{\partial}{\partial x_k}\mathbf{f}^c_k(\mathbf{u})$ which represents the convective part of the compressible Navier-Stokes equations (224). Therefore, for now, we can ignore the convective part and concentrate on the remaining diffusive part, i.e., we consider the discretization of

$$-\nabla\cdot\mathcal{F}^v(\mathbf{u},\nabla\mathbf{u}) = -\frac{\partial}{\partial x_k}\left(G_{kl}(\mathbf{u})\frac{\partial\mathbf{u}}{\partial x_l}\right) = -\partial_{x_k}\left((G(\mathbf{u})_{kl})_{ij}\,\partial_{x_l}u_j\right) = 0 \tag{246}$$

in $\Omega$, subject to the wall boundary conditions given in (227). The resulting discretization will then be combined with the discretization of the inviscid part at the end of this section.

The derivation of discontinuous Galerkin discretizations of (246) is similar to the derivation for Poisson's equation (cf. Section 4). Like for Poisson's equation we rewrite the problem (246) as a first-order system:

$$\underline{\sigma} = G(\mathbf{u})\nabla\mathbf{u}, \qquad -\nabla\cdot\underline{\sigma} = 0 \quad \text{in } \Omega,$$

i.e., $\sigma_{ik} = (G(\mathbf{u})_{kl})_{ij}\,\partial_{x_l}u_j$. Multiplying the first and second equations by test functions $\underline{\tau}$ and $v$, respectively, integrating on an element $\kappa\in\mathcal{T}_h$, and integrating by parts, we obtain

$$\begin{aligned} \int_\kappa \underline{\sigma} : \underline{\tau}\,\mathrm{d}\mathbf{x} &= -\int_\kappa \mathbf{u}\nabla\cdot\left(G^\top(\mathbf{u})\underline{\tau}\right)\,\mathrm{d}\mathbf{x} + \int_{\partial\kappa} \mathbf{u}\otimes\mathbf{n} : \left(G^\top(\mathbf{u})\underline{\tau}\right)\,\mathrm{d}s, \\ \int_\kappa \underline{\sigma} : \nabla\mathbf{v}\,\mathrm{d}\mathbf{x} &= \int_{\partial\kappa} \underline{\sigma} : \mathbf{v}\otimes\mathbf{n}\,\mathrm{d}s, \end{aligned} \tag{247}$$

where $\mathbf{n}$ is the unit outward normal vector to $\partial\kappa$. Here, we used

$$\int_\kappa \sigma_{ik}\tau_{ik}\,\mathrm{d}\mathbf{x} = \int_\kappa (G(\mathbf{u})_{kl})_{ij}\,\partial_{x_l}u_j\tau_{ik}\,\mathrm{d}\mathbf{x} = \int_\kappa \partial_{x_l}u_j\,(G(\mathbf{u})_{kl})_{ij}\,\tau_{ik}\,\mathrm{d}\mathbf{x} = \int_\kappa \nabla\mathbf{u} : \left(G^\top(\mathbf{u})\underline{\tau}\right)\,\mathrm{d}\mathbf{x}.$$

In addition to the vector-valued discrete function space $\mathbf{V}^p_h$ defined in (182) we now introduce the matrix-valued discrete function space $\underline{\Sigma}^p_h$ consisting matrix-valued polynomial functions of degree $p\geq 0$, defined by

$$\begin{aligned} \underline{\Sigma}^p_h = \{\underline{\tau}\in\left[L^2(\Omega)\right]^{4\times 2} :&\underline{\tau}|_\kappa\circ F_\kappa \in [Q_p(\hat\kappa)]^{4\times 2} \text{ if } \hat\kappa \text{ is the unit hypercube, and} \\ &\underline{\tau}|_\kappa\circ F_\kappa \in [P_p(\hat\kappa)]^{4\times 2} \text{ if } \hat\kappa \text{ is the unit simplex}, \kappa\in\mathcal{T}_h\}. \end{aligned}$$

We sum (247) over all elements $\kappa \in \mathcal{T}_h$, and replace the functions $\mathbf{u}$, $\mathbf{v}$, $\underline{\sigma}$ and $\underline{\tau}$ by discrete functions $\mathbf{u}_h, \mathbf{v}_h \in \mathbf{V}_h^p$ and $\underline{\sigma}_h, \underline{\tau}_h \in \underline{\Sigma}_h^p$. In addition, since the discrete functions $\mathbf{u}_h$ and $\underline{\sigma}_h$ may be discontinuous at interfaces $\partial\kappa \cap \partial\kappa'$ between elements $\kappa, \kappa' \in \mathcal{T}_h$, $\kappa' \neq \kappa$, we replace them by numerical flux functions $\hat{\mathbf{u}}_h$ and $\hat{\underline{\sigma}}_h$ which are approximations to $\mathbf{u}$ and $\underline{\sigma} = G(\mathbf{u})\nabla\mathbf{u}$, respectively. In particular,

$$\hat{\mathbf{u}}_h = \hat{\mathbf{u}}(\mathbf{u}_h) = \hat{\mathbf{u}}(\mathbf{u}_h^+, \mathbf{u}_h^-) \tag{248}$$

is a vector-valued numerical flux function depending on both the interior and the exterior traces, $\mathbf{u}_h^+$ and $\mathbf{u}_h^-$, of $\mathbf{u}_h$, and

$$\hat{\underline{\sigma}}_h = \hat{\underline{\sigma}}(\mathbf{u}_h, \nabla\mathbf{u}_h) = \hat{\underline{\sigma}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \nabla\mathbf{u}_h^+, \nabla\mathbf{u}_h^-) \tag{249}$$

numerical flux function depending on the interior and the exterior traces, $\mathbf{u}_h^+$ and $\mathbf{u}_h^-$, of $\mathbf{u}_h$, and the interior and the exterior traces, $\nabla\mathbf{u}_h^+$ and $\nabla\mathbf{u}_h^-$, of $\nabla\mathbf{u}_h$ on $\partial\kappa \setminus \Gamma$, $\kappa \in \mathcal{T}_h$. On the boundary $\Gamma$ we replace $\mathbf{u}$ by a numerical boundary flux function

$$\hat{\mathbf{u}}_h|_\Gamma = \hat{\mathbf{u}}_{\Gamma,h} = \hat{\mathbf{u}}_\Gamma(\mathbf{u}_h^+). \tag{250}$$

Furthermore, we replace $\underline{\sigma}$ on the boundary by a the numerical boundary flux function

$$\hat{\underline{\sigma}}_h|_\Gamma = \hat{\underline{\sigma}}_{\Gamma,h} = \hat{\underline{\sigma}}_\Gamma(\mathbf{u}_h^+, \nabla\mathbf{u}_h^+). \tag{251}$$

Thus we obtain the following discretization in *system flux formulation*: find $\mathbf{u}_h \in \mathbf{V}_h^p$ and $\underline{\sigma}_h \in \underline{\Sigma}_h^p$ such that

$$\int_\Omega \underline{\sigma}_h : \underline{\tau}_h \,\mathrm{d}\mathbf{x} = -\int_\Omega \mathbf{u}_h \nabla_h \cdot \left(G^\top(\mathbf{u}_h)\underline{\tau}_h\right) \mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\mathbf{u}}_h \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h)\underline{\tau}_h\right) \mathrm{d}s, \tag{252}$$

$$\int_\Omega \underline{\sigma}_h : \nabla_h\mathbf{v}_h \,\mathrm{d}\mathbf{x} = \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\underline{\sigma}}_h : \mathbf{v}_h \otimes \mathbf{n}\,\mathrm{d}s. \tag{253}$$

Like for Poisson's equation (cf. Section 4) there are many choices for the numerical flux functions $\hat{\mathbf{u}}_h$ and $\hat{\underline{\sigma}}_h$ resulting in several different DG discretizations. The numerical fluxes $\hat{\mathbf{u}}_h$ and $\hat{\underline{\sigma}}_h$ will be specified later.

The flux formulation (252) represents the discretization of a first order system with unknowns $\mathbf{u}_h \in \mathbf{V}_h^p$ and $\underline{\sigma}_h \in \underline{\Sigma}_h^p$. However, this is three times the size of a problem involving $\mathbf{u}_h \in \mathbf{V}_h^p$, only. In order to reduce the problem size, the auxiliary variable $\underline{\sigma}_h$ in (252) and (253) is usually eliminated to gain a *primal formulation* involving only the primal variable $\mathbf{u}_h$. To this end, we perform a second integration by parts on each element $\kappa$ in (252) and set $\underline{\tau}_h = \nabla_h\mathbf{v}_h$ which gives

$$\int_\Omega \underline{\sigma}_h : \nabla_h\mathbf{v}_h \,\mathrm{d}\mathbf{x} = \int_\Omega G(\mathbf{u}_h)\nabla_h\mathbf{u}_h : \nabla_h\mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{\mathbf{u}}_h - \mathbf{u}_h) \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\right) \mathrm{d}s.$$

Substituting this into (253) yields the *primal formulation*: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\begin{aligned}\int_\Omega G(\mathbf{u}_h)\nabla_h\mathbf{u}_h : \nabla_h\mathbf{v}_h \,\mathrm{d}\mathbf{x} &- \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\underline{\sigma}}_h : \mathbf{v}_h \otimes \mathbf{n}\,\mathrm{d}s \\ &+ \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{\mathbf{u}}_h - \mathbf{u}_h) \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\right) \mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \end{aligned} \tag{254}$$

We note that (87) is in *element-based* form, i.e., it is given in terms of $\sum_\kappa \int_{\partial\kappa}$. This means that each interior face $e = \Gamma_\mathcal{I}$ occurs twice in the sum over all elements $\kappa$ (once in $\int_{\partial\kappa}$ and once in $\int_{\partial\kappa'}$ for $\kappa' \neq \kappa$ and $e = \partial\kappa \cap \partial\kappa' \neq 0$). In extension of the mean-value and jump operators for scalar-valued and vector-valued functions $q_h \in V_h^p$ and $\boldsymbol{\phi}_h \in \mathbf{\Sigma}_h^p$ as defined in (88) and (89), in the following we define the respective operators, on interior faces $\Gamma_\mathcal{I}$ and on the boundary $\Gamma$, for vector-valued functions $\mathbf{v}_h \in \mathbf{V}_h^p$,

$$
\begin{aligned}
\{\!\{\mathbf{v}\}\!\} &= \tfrac{1}{2}(\mathbf{v}^+ + \mathbf{v}^-) && \text{on } \Gamma_\mathcal{I}, & \{\!\{\mathbf{v}\}\!\} &= \mathbf{v}^+ && \text{on } \Gamma,\\
\underline{[\![\mathbf{v}]\!]} &= \mathbf{v}^+ \otimes \mathbf{n}^+ + \mathbf{v}^- \otimes \mathbf{n}^- && \text{on } \Gamma_\mathcal{I}, & \underline{[\![\mathbf{v}]\!]} &= \mathbf{v}^+ \otimes \mathbf{n}^+ && \text{on } \Gamma,
\end{aligned}
\tag{255}
$$

and for matrix-valued functions $\underline{\tau}_h \in \underline{\Sigma}_h^p$,

$$
\begin{aligned}
\{\!\{\underline{\tau}\}\!\} &= \tfrac{1}{2}(\underline{\tau}^+ + \underline{\tau}^-) && \text{on } \Gamma_\mathcal{I}, & \{\!\{\underline{\tau}\}\!\} &= \underline{\tau}^+ && \text{on } \Gamma,\\
[\![\underline{\tau}]\!] &= \underline{\tau}^+\mathbf{n}^+ + \underline{\tau}^-\mathbf{n}^- && \text{on } \Gamma_\mathcal{I}, & [\![\underline{\tau}]\!] &= \underline{\tau}^+\mathbf{n}^+ && \text{on } \Gamma.
\end{aligned}
\tag{256}
$$

Then, the element-based primal formulation (254) can equivalently be written in following *face-based primal formulation* (cf. Appendix A.5 for the derivation): find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$
\begin{aligned}
\int_\Omega G(\mathbf{u}_h)\nabla_h \mathbf{u}_h : \nabla_h \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma} \underline{[\![\hat{\mathbf{u}}_h - \mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\}\!\} - \{\!\{\underline{\hat{\sigma}}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]} \,\mathrm{d}s&\\
+ \int_{\Gamma_\mathcal{I}} \{\!\{\hat{\mathbf{u}}_h - \mathbf{u}_h\}\!\} \cdot [\![G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h]\!] - [\![\underline{\hat{\sigma}}_h]\!] \cdot \{\!\{\mathbf{v}_h\}\!\} \,\mathrm{d}s = 0 \quad \mathbf{v}_h \in \mathbf{V}_h^p.&
\end{aligned}
\tag{257}
$$

We end this subsection by combining the discretization (187) of the compressible Euler equations (168) and the discretization (254) of the diffusive part (246) of the compressible Navier-Stokes equations. This results in following discretization of the compressible Navier-Stokes equations (224) in element-based form: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$
\begin{aligned}
\int_\Omega \left(-\mathcal{F}^c(\mathbf{u}_h) + \mathcal{F}^v(\mathbf{u}_h, \nabla_h \mathbf{u}_h)\right) : \nabla_h \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \left(\hat{\mathbf{h}}_h - \underline{\hat{\sigma}}_h \mathbf{n}\right) \cdot \mathbf{v}_h \,\mathrm{d}s&\\
+ \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{\mathbf{u}}_h - \mathbf{u}_h) \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\right) \,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p,&
\end{aligned}
\tag{258}
$$

where we used that $\underline{\hat{\sigma}} : \mathbf{v} \otimes \mathbf{n} = \underline{\hat{\sigma}}_{ik} v_i n_k = \underline{\hat{\sigma}}_{ik} n_k v_i = (\underline{\hat{\sigma}}\mathbf{n}) \cdot \mathbf{v}$. Rewriting the diffusive terms in face-based form according to (257) gives: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$
\begin{aligned}
\int_\Omega \left(-\mathcal{F}^c(\mathbf{u}_h) + \mathcal{F}^v(\mathbf{u}_h, \nabla_h \mathbf{u}_h)\right) : \nabla_h \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\mathbf{h}}_h \cdot \mathbf{v}_h \,\mathrm{d}s&\\
+ \int_{\Gamma_\mathcal{I}\cup\Gamma} \underline{[\![\hat{\mathbf{u}}_h - \mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\}\!\} - \{\!\{\underline{\hat{\sigma}}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]} \,\mathrm{d}s&\\
+ \int_{\Gamma_\mathcal{I}} \{\!\{\hat{\mathbf{u}}_h - \mathbf{u}_h\}\!\} \cdot [\![G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h]\!] - [\![\underline{\hat{\sigma}}_h]\!] \cdot \{\!\{\mathbf{v}_h\}\!\} \,\mathrm{d}s = 0 \quad \mathbf{v}_h \in \mathbf{V}_h^p.&
\end{aligned}
\tag{259}
$$

## 6.4 Consistency and conservation property

In Section 5.4 we discussed the consistency of the discretization of the compressible Euler equations. Thereby, here we can concentrate on analyzing consistency of the discretization (254) of the remaining diffusive part (246) of the compressible Navier-Stokes equations.

Integrating (254) back by parts we obtain following equivalent form of the discretization: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$-\int_\Omega \left(\nabla_h \cdot \mathcal{F}^v\left(\mathbf{u}_h, \nabla_h \mathbf{u}_h\right)\right) \cdot \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \left(\mathbf{n} \cdot \mathcal{F}^v\left(\mathbf{u}_h, \nabla_h \mathbf{u}_h\right) - \underline{\hat{\sigma}}_h \mathbf{n}\right) \cdot \mathbf{v}_h \,\mathrm{d}s$$
$$+ \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \left(\hat{\mathbf{u}}_h - \mathbf{u}_h\right) \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h) \nabla \mathbf{v}_h\right) \,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \quad (260)$$

This can be rewritten in following primal residual form: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega \mathbf{R}(\mathbf{u}_h) \cdot \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \mathbf{r}(\mathbf{u}_h) \cdot \mathbf{v}_h^+ + \underline{\rho}(\mathbf{u}_h) : \nabla_h \mathbf{v}_h^+ \,\mathrm{d}s$$
$$+ \int_\Gamma \mathbf{r}_\Gamma(\mathbf{u}_h) \cdot \mathbf{v}_h^+ + \underline{\rho}_\Gamma(\mathbf{u}_h) : \nabla_h \mathbf{v}_h^+ \,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p,$$

where the primal residuals are given by $\mathbf{R}(\mathbf{u}_h) = \nabla \cdot \mathcal{F}^v(\mathbf{u}_h, \nabla_h \mathbf{u}_h)$ on $\kappa \in \mathcal{T}_h$, and

$$\begin{aligned}
\mathbf{r}(\mathbf{u}_h) &= \underline{\hat{\sigma}}_h \mathbf{n} - \mathbf{n} \cdot \mathcal{F}^v(\mathbf{u}_h^+, \nabla_h \mathbf{u}_h^+), & \underline{\rho}(\mathbf{u}_h) &= G(\mathbf{u}_h)\left(\left(\hat{\mathbf{u}}_h - \mathbf{u}_h\right) \otimes \mathbf{n}\right) & &\text{on } \partial\kappa\setminus\Gamma, \kappa \in \mathcal{T}_h,\\
\mathbf{r}_\Gamma(\mathbf{u}_h) &= \underline{\hat{\sigma}}_{\Gamma,h} \mathbf{n} - \mathbf{n} \cdot \mathcal{F}^v(\mathbf{u}_h^+, \nabla_h \mathbf{u}_h^+), & \underline{\rho}_\Gamma(\mathbf{u}_h) &= G(\mathbf{u}_h)\left(\left(\hat{\mathbf{u}}_{\Gamma,h} - \mathbf{u}_h\right) \otimes \mathbf{n}\right) & &\text{on } \Gamma.
\end{aligned}$$

Note, that this is an extension of the primal residual form in (22) to include face and boundary residuals $\underline{\rho}(\mathbf{u}_h)$ and $\underline{\rho}_\Gamma(\mathbf{u}_h)$ multiplied by $\nabla_h \mathbf{v}_h$. Motivated by this we give the following definition and statement.

**Definition 6.3** *The numerical fluxes* $\hat{\mathbf{u}}$ *and* $\underline{\hat{\sigma}}$ *are said to* consistent *if*

$$\begin{aligned}
\hat{\mathbf{u}}(\mathbf{v}) &= \mathbf{v}, & \underline{\hat{\sigma}}(\mathbf{v}, \nabla\mathbf{v}) &= \mathcal{F}^v(\mathbf{v}, \nabla\mathbf{v}), & &\text{on } \partial\kappa \setminus \Gamma, \kappa \in \mathcal{T}_h,\\
\hat{\mathbf{u}}_\Gamma(\mathbf{v}) &= \mathbf{v}, & \underline{\hat{\sigma}}_\Gamma(\mathbf{v}, \nabla\mathbf{v}) &= \mathcal{F}^v(\mathbf{v}, \nabla\mathbf{v}), & &\text{on } \Gamma,
\end{aligned}$$

*whenever* $\mathbf{v}$ *is a smooth function satisfying the boundary conditions of the Navier-Stokes equations* (224).

**Lemma 6.4 (Consistency)** *Let the numerical fluxes* $\hat{\mathbf{u}}$ *and* $\hat{\mathbf{u}}_\Gamma$*, and* $\underline{\hat{\sigma}}$ *and* $\underline{\hat{\sigma}}_\Gamma$ *in* (254) *be* consistent*. Then,* (254) *is a* consistent *discretization of the diffusive part of the compressible Navier-Stokes equation* (246).

Combining Lemma 5.3 and Lemma 6.4 we arrive at the following statement.

**Theorem 6.5 (Consistency)** *Let the numerical fluxes* $\hat{\mathbf{h}}$ *and* $\hat{\mathbf{h}}_\Gamma$*,* $\hat{\mathbf{u}}$ *and* $\hat{\mathbf{u}}_\Gamma$*, and* $\underline{\hat{\sigma}}$ *and* $\underline{\hat{\sigma}}_\Gamma$ *in* (258) *be* consistent*. Then, the discretization* (258) *of the compressible Navier-Stokes equation* (224) *is* consistent*.*

In Section 4.5 we discussed the conservation property of the discontinuous Galerkin discretizations of Poisson's equation. Similar to Definitions 4.6 we give the following definition.

**Definition 6.6** *A numerical flux function* $\hat{\underline{\sigma}} = \hat{\underline{\sigma}}(\mathbf{u}^+, \mathbf{u}^-, \nabla\mathbf{u}^+, \nabla\mathbf{u}^-)$ *is said to be* conservative *(or* single-valued*") if*

$$\hat{\underline{\sigma}}(\mathbf{u}^+, \mathbf{u}^-, \nabla\mathbf{u}^+, \nabla\mathbf{u}^-) = \hat{\underline{\sigma}}(\mathbf{u}^-, \mathbf{u}^+, \nabla\mathbf{u}^-, \nabla\mathbf{u}^+).$$

Similar to Lemma 4.7 and combined with Lemma 5.5 we arrive at following statement.

**Lemma 6.7 (Conservation property)** *The discretization* (258) *is* conservative*, i.e., for any union $S$ of elements,*

$$\int_{\partial S} \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n}\, ds = 0,$$

*if and only if the numerical flux functions* $\hat{\mathbf{h}}$ *and* $\hat{\underline{\sigma}}_h$ *are* conservative. *In particular, we have the* global *and* local conservation property *for any* $\kappa \in \mathcal{T}_h$*,*

$$\int_{\Gamma} \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n}\, ds = 0, \qquad \int_{\partial\kappa} \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n}\, ds = 0.$$

We end this section by defining conservativity for the vector-valued numerical fluxes $\hat{\mathbf{u}}$ analogous to the conservativity of scalar numerical fluxes $\hat{u}$ in Definition 4.8.

**Definition 6.8** *A numerical flux function* $\hat{\mathbf{u}} = \hat{\mathbf{u}}(\mathbf{u}^+, \mathbf{u}^-)$ *is said to be* conservative *(or* single-valued*) if*

$$\hat{\mathbf{u}}(\mathbf{u}^+, \mathbf{u}^-) = \hat{\mathbf{u}}(\mathbf{u}^-, \mathbf{u}^+). \tag{261}$$

## 6.5 Adjoint consistency

In this section, we analyze the adjoint consistency of the discretization (258) of the compressible Navier-Stokes equations (224) in combination with specific discretizations of the force coefficients (232). This will be an extension of the analysis performed in Section 5.7.

While in previous work (cf. (Hartmann, 2007a)) a discretization of the force coefficients (232) was proposed which is adjoint consistent in combination with an SIPG discretization and the special case of the normal wall boundary condition (202) and a related discretization of the diffusive fluxes, in the following we generalize this to any *consistent* discretization $\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n})$ and $\hat{\underline{\sigma}}_\Gamma(\mathbf{u}_h^+, \nabla\mathbf{u}_h^+)$ of convective and diffusive boundary fluxes. This also includes the extension of the adjoint consistency analysis from the special case of the SIPG discretization (cf. (Hartmann, 2007a)) to *any* DG discretization of diffusive terms.

Given a consistent discretization $\hat{\mathbf{h}}_{\Gamma,h} = \hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n})$ and $\hat{\underline{\sigma}}_{\Gamma,h} = \hat{\underline{\sigma}}_\Gamma(\mathbf{u}_h^+, \nabla\mathbf{u}_h^+)$ of convective and diffusive boundary fluxes, we define the following discretization of the force coefficients (232),

$$J_h(\mathbf{u}_h) = \int_{\Gamma_W} \left(\hat{\mathbf{h}}_{\Gamma,h} - \hat{\underline{\sigma}}_{\Gamma,h}\mathbf{n}\right) \cdot \tilde{\boldsymbol{\psi}}\, \mathrm{d}s, \tag{262}$$

with $\tilde{\boldsymbol{\psi}} = (0, \psi_1, \psi_2, 0)^\top$ on $\Gamma_W$ for $\boldsymbol{\psi} = (\psi_1, \psi_2)^\top$ in (232), and $\tilde{\boldsymbol{\psi}} \equiv 0$ on $\Gamma \setminus \Gamma_W$. Note, that this is a *consistent* discretization of the target quantity (232), i.e., the values of the target quantity (232) and its discretization (262) evaluated for the exact solution $\mathbf{u}$ to (224) are identical, $J_h(\mathbf{u}) = J(\mathbf{u})$. To see this, recall that due to consistency of $\underline{\hat{\sigma}}_\Gamma$, the exact solution $\mathbf{u}$ to (224) satisfies $\underline{\hat{\sigma}}_\Gamma(\mathbf{u}, \nabla\mathbf{u}) = \mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u})$. Using $\mathbf{n}\cdot\mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u}) = (0, (\tau\mathbf{n})_1, (\tau\mathbf{n})_2, \mathcal{K}\mathbf{n}\cdot\nabla T)^\top$ on $\Gamma_W$ we obtain

$$(\underline{\hat{\sigma}}_\Gamma(\mathbf{u}, \nabla\mathbf{u})\mathbf{n}) \cdot \tilde{\boldsymbol{\psi}} = (\mathbf{n}\cdot\mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u})) \cdot \tilde{\boldsymbol{\psi}} = \tau\mathbf{n}\cdot\boldsymbol{\psi},$$

and together with (208) we arrive at $J_h(\mathbf{u}) = J(\mathbf{u})$.

**Theorem 6.9** *Let the numerical fluxes* $\hat{\mathbf{h}}$*,* $\underline{\hat{\sigma}}$ *and* $\hat{\mathbf{u}}$ *in* (258) *be conservative and the boundary fluxes* $\mathbf{h}_\Gamma$ *and* $\underline{\hat{\sigma}}_\Gamma$ *in* (258) *be consistent. Furthermore, let* $\hat{\mathbf{u}}_\Gamma = \mathbf{u}_\Gamma$ *where* $\mathbf{u}_\Gamma(\cdot)$ *is given by* (228) *on the wall boundary* $\Gamma_W$ *and as in Section 5.1 on the farfield boundary* $\Gamma\setminus\Gamma_W$*. Then, the discretization* (258) *of the compressible Navier-Stokes equations* (224) *in combination with the discretization* (262) *of the force coefficients* (232) *is* adjoint consistent*.*

**Proof:** We begin by considering the discrete adjoint problem: find $\mathbf{z}_h \in \mathbf{V}_h^p$ such that

$$\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h) = J_h'[\mathbf{u}_h](\mathbf{w}_h) \qquad \forall \mathbf{w}_h \in \mathbf{V}_h^p. \tag{263}$$

Here, $J_h'[\mathbf{u}_h](\mathbf{w}_h)$ is the Fréchet derivative of $J_h(\mathbf{u}_h)$ in (262) at $\mathbf{u}_h$ in the direction of $\mathbf{w}_h$, given by

$$J_h'[\mathbf{u}_h](\mathbf{w}_h) = \int_{\Gamma_W} \left( \left(\hat{\mathbf{h}}_{\Gamma,\mathbf{u}} - \underline{\hat{\sigma}}_{\Gamma,\mathbf{u}}\mathbf{n}\right)\, \mathbf{w}_h - \left(\underline{\hat{\sigma}}_{\Gamma,\nabla\mathbf{u}}\mathbf{n}\right)\, \nabla\mathbf{w}_h \right) \cdot \tilde{\boldsymbol{\psi}}\, \mathrm{d}s, \tag{264}$$

where we use the short notations

$$\hat{\mathbf{h}}_{\Gamma,\mathbf{u}} = \partial_\mathbf{u}\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h, \mathbf{n}), \qquad \underline{\hat{\sigma}}_{\Gamma,\mathbf{u}} = \partial_\mathbf{u}\underline{\hat{\sigma}}_\Gamma(\mathbf{u}_h, \nabla\mathbf{u}_h), \qquad \underline{\hat{\sigma}}_{\Gamma,\nabla\mathbf{u}} = \partial_{\nabla\mathbf{u}}\underline{\hat{\sigma}}_\Gamma(\mathbf{u}_h, \nabla\mathbf{u}_h).$$

Furthermore, $\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h)$ is the Fréchet derivative of $\hat{N}_h(\mathbf{u}_h, \mathbf{v})$ in (186) with respect to $\mathbf{u}_h$ in the direction of $\mathbf{w}_h$ given by

$$\begin{aligned}
\hat{N}_h'[\mathbf{u}_h](\mathbf{w}_h, \mathbf{z}_h) = &\int_\Omega \left(\left(-\mathcal{F}^c_\mathbf{u}[\mathbf{u}_h] + \mathcal{F}^v_\mathbf{u}[\mathbf{u}_h]\right)\mathbf{w}_h + \mathcal{F}^v_{\nabla\mathbf{u}}[\mathbf{u}_h]\nabla\mathbf{w}_h\right) : \nabla_h\mathbf{z}_h\,\mathrm{d}\mathbf{x} \\
&+ \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} \left(\hat{\mathbf{h}}_{\mathbf{u}^+}\mathbf{w}_h^+ - \underline{\hat{\sigma}}_{\mathbf{u}^+}\mathbf{n}\,\mathbf{w}_h^+ - \underline{\hat{\sigma}}_{\nabla\mathbf{u}^+}\mathbf{n}\,\nabla\mathbf{w}_h^+\right)\cdot\left(\mathbf{z}_h^+ - \mathbf{z}_h^-\right)\,\mathrm{d}s \\
&+ \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} (\hat{\mathbf{u}}_{\mathbf{u}^+}\mathbf{w}_h^+)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h^+)\nabla\mathbf{z}_h^+ - G^\top(\mathbf{u}_h^-)\nabla\mathbf{z}_h^-\right)\,\mathrm{d}s \\
&- \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa} \mathbf{w}_h^+\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{z}_h\right)\,\mathrm{d}s \\
&+ \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa} (\hat{\mathbf{u}}_h - \mathbf{u}_h)\otimes\mathbf{n} : \left(\left(G^\top\right)'[\mathbf{u}_h]\mathbf{w}_h\nabla\mathbf{z}_h\right)\,\mathrm{d}s \\
&+ \int_\Gamma (\hat{\mathbf{u}}_{\Gamma,\mathbf{u}}\mathbf{w}_h)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{z}_h\right)\,\mathrm{d}s \\
&+ \int_\Gamma \left(\left(\hat{\mathbf{h}}_{\Gamma,\mathbf{u}} - \underline{\hat{\sigma}}_{\Gamma,\mathbf{u}}\mathbf{n}\right)\,\mathbf{w}_h - \left(\underline{\hat{\sigma}}_{\Gamma,\nabla\mathbf{u}}\mathbf{n}\right)\,\nabla\mathbf{w}_h\right)\cdot\mathbf{z}_h\,\mathrm{d}s.
\end{aligned} \tag{265}$$

Here, we used $\int_e \hat{\mathbf{h}}_{\mathbf{u}^-}\mathbf{w}_h^-\mathbf{z}_h^+\,\mathrm{d}s = -\int_e \hat{\mathbf{h}}_{\mathbf{u}^+}\mathbf{w}_h^+\mathbf{z}_h^-\,\mathrm{d}s$ for a conservative flux $\hat{\mathbf{h}}$ (cf. (212)) and similar relations hold for conservative/single-valued fluxes $\underline{\hat{\sigma}}_{\mathbf{u}^+}\mathbf{n}$, $\underline{\hat{\sigma}}_{\nabla\mathbf{u}^+}\mathbf{n}$ and $\hat{\mathbf{u}}_{\mathbf{u}^+}\otimes\mathbf{n}$ (cf. Definitions 6.6 and 6.8):

$$\int_e \underline{\hat{\sigma}}_{\mathbf{u}^-}\mathbf{n}\mathbf{w}_h^-\mathbf{z}_h^+\,\mathrm{d}s = -\int_e \underline{\hat{\sigma}}_{\mathbf{u}^+}\mathbf{n}\mathbf{w}_h^+\mathbf{z}_h^-\,\mathrm{d}s,$$
$$\int_e \underline{\hat{\sigma}}_{\nabla\mathbf{u}^-}\mathbf{n}\nabla\mathbf{w}_h^-\mathbf{z}_h^+\,\mathrm{d}s = -\int_e \underline{\hat{\sigma}}_{\nabla\mathbf{u}^+}\mathbf{n}\nabla\mathbf{w}_h^+\mathbf{z}_h^-\,\mathrm{d}s,$$
$$\int_e (\hat{\mathbf{u}}_{\mathbf{u}^-}\mathbf{w}_h^-)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h^+)\nabla\mathbf{z}_h^+\right) = -\int_e (\hat{\mathbf{u}}_{\mathbf{u}^+}\mathbf{w}_h^+)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h^-)\nabla\mathbf{z}_h^-\right)\,\mathrm{d}s.$$

Integration by parts on the last term in the first line of (265) gives

$$\begin{aligned}
&\int_\Omega \left(\mathcal{F}^v_{\nabla\mathbf{u}}[\mathbf{u}_h]\nabla\mathbf{w}_h\right) : \nabla_h\mathbf{z}_h\,\mathrm{d}\mathbf{x} \\
&\quad = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa G(\mathbf{u}_h)\nabla\mathbf{w}_h : \nabla_h\mathbf{z}_h\,\mathrm{d}\mathbf{x} = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \nabla\mathbf{w}_h : G^\top(\mathbf{u}_h)\nabla_h\mathbf{z}_h\,\mathrm{d}\mathbf{x} \\
&\quad = -\sum_{\kappa\in\mathcal{T}_h}\int_\kappa \mathbf{w}_h\nabla\cdot\left(\left(\mathcal{F}^v_{\nabla\mathbf{u}}[\mathbf{u}_h]\right)^\top\nabla_h\mathbf{z}_h\right)\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa}\mathbf{w}_h^+\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{z}_h\right)\,\mathrm{d}s.
\end{aligned} \tag{266}$$

Note, that here the resulting face terms cancel with those in the fourth line in (265).

Then, the discrete adjoint problem (263) can be written in following adjoint residual form: find $\mathbf{z}_h\in\mathbf{V}_h^p$ such that

$$\begin{aligned}
&\int_\Omega \mathbf{w}_h\cdot\mathbf{R}^*[\mathbf{u}_h](\mathbf{z}_h)\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\mathbf{w}_h\cdot\mathbf{r}^*[\mathbf{u}_h](\mathbf{z}_h) + \nabla\mathbf{w}_h : \underline{\rho}^*[\mathbf{u}_h](\mathbf{z}_h)\,\mathrm{d}s \\
&\qquad + \int_\Gamma \mathbf{w}_h\cdot\mathbf{r}^*_\Gamma[\mathbf{u}_h](\mathbf{z}_h) + \nabla\mathbf{w}_h : \underline{\rho}^*_\Gamma[\mathbf{u}_h](\mathbf{z}_h)\,\mathrm{d}s = 0 \quad \forall\mathbf{w}_h\in\mathbf{V}_h^p.
\end{aligned} \tag{267}$$

Note, that this is an extension of the adjoint residual form in (42) to include face and boundary residuals $\underline{\rho}^*[\mathbf{u}_h](\mathbf{z}_h)$ and $\underline{\rho}^*_\Gamma[\mathbf{u}_h](\mathbf{z}_h)$ multiplied with $\nabla\mathbf{w}_h$. For proving adjoint consistency, we want to show that the exact (and assumingly smooth) solutions $\mathbf{u}$ and $\mathbf{z}$ to, respectively, the flow equations (224) and the adjoint equations (237) satisfy

$$\begin{aligned}
&\int_\Omega \mathbf{w}\cdot\mathbf{R}^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}}\int_{\partial\kappa\setminus\Gamma}\mathbf{w}\cdot\mathbf{r}^*[\mathbf{u}](\mathbf{z}) + \nabla\mathbf{w} : \underline{\rho}^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}s \\
&\qquad + \int_\Gamma \mathbf{w}\cdot\mathbf{r}^*_\Gamma[\mathbf{u}](\mathbf{z}) + \nabla\mathbf{w} : \underline{\rho}^*_\Gamma[\mathbf{u}](\mathbf{z})\,\mathrm{d}s = 0 \quad \forall\mathbf{w}\in\mathbf{V}_A.
\end{aligned} \tag{268}$$

To this end, note that the adjoint element residual is given by

$$\mathbf{R}^*[\mathbf{u}_h](\mathbf{z}_h) = \left(\mathcal{F}^c_{\mathbf{u}}[\mathbf{u}_h] - \mathcal{F}^v_{\mathbf{u}}[\mathbf{u}_h]\right)^\top\nabla_h\mathbf{z}_h + \nabla_h\cdot\left(\left(\mathcal{F}^v_{\nabla\mathbf{u}}[\mathbf{u}_h]\right)^\top\nabla_h\mathbf{z}_h\right), \tag{269}$$

and vanishes, $\mathbf{R}^*[\mathbf{u}](\mathbf{z}) = 0$, if evaluated for the exact solutions $\mathbf{u}$ and $\mathbf{z}$. Furthermore, the adjoint face residuals $\mathbf{r}^*[\mathbf{u}_h](\mathbf{z}_h)$ and $\underline{\rho}^*[\mathbf{u}_h](\mathbf{z}_h)$ vanish, $\mathbf{r}^*[\mathbf{u}](\mathbf{z}) = 0$ and $\underline{\rho}^*[\mathbf{u}](\mathbf{z}) = 0$, if evaluated for $\mathbf{u}$ and $\mathbf{z}$. To see this, note that all interior faces terms in (265), remaining after cancellation with the boundary term in (266), vanish due to the smoothness of $\mathbf{u}$ and $\mathbf{z}$ (thus $\mathbf{z}^+ = \mathbf{z}^-$ and $G^\top(\mathbf{u}^+)\nabla\mathbf{z}^+ = G^\top(\mathbf{u}^-)\nabla\mathbf{z}^-$), and due to consistency of $\hat{\mathbf{u}}$ (thus $\hat{\mathbf{u}}(\mathbf{u}) - \mathbf{u} = 0$).

Next, we consider the boundary terms on the farfield boundary $\Gamma \setminus \Gamma_W$. Due to consistency of $\mathbf{u}_\Gamma$, i.e., $\mathbf{u}_\Gamma(\mathbf{u}) - \mathbf{u} = 0$, the consistency of $\hat{\mathbf{h}}_\Gamma$, $\hat{\mathbf{h}}_\Gamma(\mathbf{u},\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u})$, and the consistency of $\underline{\hat{\sigma}}_\Gamma$, $\underline{\hat{\sigma}}_\Gamma(\mathbf{u},\nabla\mathbf{u}) = \mathcal{F}^v(\mathbf{u},\nabla\mathbf{u})$, the boundary terms on $\Gamma \setminus \Gamma_W$ evaluated for $\mathbf{u}$ and $\mathbf{z}$ reduce to

$$
\begin{aligned}
\int_{\Gamma\setminus\Gamma_W} \mathbf{w}\cdot\mathbf{r}_\Gamma^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}s &= -\left(\hat{\mathbf{u}}_{\Gamma,\mathbf{u}}\mathbf{w}, \mathbf{n}\cdot\left(G^\top(\mathbf{u})\nabla\mathbf{z}\right)\right)_{\Gamma\setminus\Gamma_W} - \left(\mathbf{w}, \left(\mathbf{n}\cdot\left(\mathcal{F}^c_{\mathbf{u}} - \mathcal{F}^v_{\mathbf{u}}\right)\right)^\top\mathbf{z}\right)_{\Gamma\setminus\Gamma_W}, \\
\int_{\Gamma\setminus\Gamma_W} \nabla\mathbf{w} : \underline{\rho}_\Gamma^*[\mathbf{u}](\mathbf{z})\,\mathrm{d}s &= \left(\nabla\mathbf{w}, \left(\mathbf{n}\cdot\mathcal{F}^v_{\nabla\mathbf{u}}\right)^\top\mathbf{z}\right)_{\Gamma\setminus\Gamma_W} \qquad \forall \mathbf{w}\in\mathbf{V}_A.
\end{aligned}
$$

Due to the assumption $\hat{\mathbf{u}}_\Gamma(\mathbf{u}) = \mathbf{u}_\Gamma(\mathbf{u})$, furthermore due to $\mathbf{u}_{\Gamma,\mathbf{u}}\mathbf{w} = \mathbf{w}$ for $\mathbf{w}\in\mathbf{V}_A$ on $\Gamma\setminus\Gamma_W$ (cf. Remark 6.1) and due to the adjoint condition (245) both boundary terms on $\Gamma\setminus\Gamma_W$ vanish.

Next, we consider the boundary term in the sixth line in (265) on the wall boundary $\Gamma_W$. If evaluated for the exact solutions $\mathbf{u}$ and $\mathbf{z}$ and with admissible variations $\mathbf{w}$, we obtain

$$
\int_{\Gamma_W} (\hat{\mathbf{u}}_{\Gamma,\mathbf{u}}\mathbf{w})\otimes\mathbf{n} : \left(G^\top(\mathbf{u})\nabla\mathbf{z}\right)\,\mathrm{d}s. \tag{270}
$$

Note, that due to Lemma 6.2 this term vanishes. To see this, recall that $w_2 = w_3 = 0$ on $\Gamma_W$ and $u_1w_4 - u_4w_1 = 0$ on $\Gamma_{\text{iso}}$ (cf. (234)). Also, note that by assumption $\hat{\mathbf{u}}_\Gamma(\mathbf{u}) = \mathbf{u}_\Gamma(\mathbf{u})$. Furthermore, we have $\mathbf{u}_{\Gamma,\mathbf{u}}\mathbf{w} = \mathbf{w}$ on $\Gamma_{\text{adia}}$ for all $\mathbf{w}\in\mathbf{V}_A$ (cf. Remark 6.1 or by explicitly computing $\mathbf{u}_{\Gamma,\mathbf{u}}\mathbf{w} = \mathrm{diag}(1,0,0,1)\mathbf{w} = \mathbf{w}$ for $\mathbf{w}$ with $w_2 = w_3 = 0$). Finally, due to $\tilde{\mathbf{w}} = \mathbf{u}_{\Gamma,\mathbf{u}}\mathbf{w} = (w_1, 0, 0, w_1c_vT)$ on $\Gamma_{\text{iso}}$ we have $\tilde{w}_2 = \tilde{w}_3 = 0$ and $u_1\tilde{w}_4 - u_4\tilde{w}_1 = u_1w_1c_vT_{\text{wall}} - u_4w_1 = (u_1c_vT_{\text{wall}} - u_4)w_1 = 0$ on $\Gamma_{\text{iso}}$. Thereby, $\tilde{\mathbf{w}} = \mathbf{u}_{\Gamma,\mathbf{u}}\mathbf{w} = (\tilde{w}_1, 0, 0, \tilde{w}_4)^\top$ on $\Gamma_W$, and we can apply Lemma 6.2 for $\mathbf{u}$ with vanishing velocity $\mathbf{v} = 0$, resulting in

$$
\int_{\Gamma_W} (\hat{\mathbf{u}}_{\Gamma,\mathbf{u}}\mathbf{w})\otimes\mathbf{n} : \left(G^\top(\mathbf{u})\nabla\mathbf{z}\right)\,\mathrm{d}s = \int_{\Gamma_W} \tfrac{\mu}{\rho^2}\tfrac{\gamma}{Pr}\mathbf{n}\cdot\nabla z_4\,(u_1\tilde{w}_4 - u_4\tilde{w}_1)\,\mathrm{d}s = 0, \tag{271}
$$

which vanishes due to $u_1\tilde{w}_4 - u_4\tilde{w}_1 = 0$ on $\Gamma_{\text{iso}}$ and due to $\mathbf{n}\cdot\nabla z_4 = 0$ on $\Gamma_{\text{adia}}$ (cf. (244)).

The remaining boundary terms on $\Gamma_W$ in (263) yield the adjoint boundary residuals,

$$
\begin{aligned}
\mathbf{r}_\Gamma^*[\mathbf{u}_h](\mathbf{z}_h) &= \left(\hat{\mathbf{h}}_{\Gamma,\mathbf{u}} - \underline{\hat{\sigma}}_{\Gamma,\mathbf{u}}\mathbf{n}\right)^\top\left(\tilde{\boldsymbol{\psi}} - \mathbf{z}_h\right) - (\hat{\mathbf{u}}_{\Gamma,h} - \mathbf{u}_h)\otimes\mathbf{n} : \left(\left(G^\top\right)'[\mathbf{u}_h]\nabla\mathbf{z}_h\right), \\
\underline{\rho}_\Gamma^*[\mathbf{u}_h](\mathbf{z}_h) &= -\left(\underline{\hat{\sigma}}_{\Gamma,\nabla\mathbf{u}}\mathbf{n}\right)^\top\left(\tilde{\boldsymbol{\psi}} - \mathbf{z}_h\right).
\end{aligned} \tag{272}
$$

Note, that these residuals vanish, $\mathbf{r}_\Gamma^*[\mathbf{u}](\mathbf{z}) = 0$ and $\underline{\rho}_\Gamma^*[\mathbf{u}](\mathbf{z}) = 0$, if evaluated for the exact solution $\mathbf{u}$ and $\mathbf{z}$. To see this, note that $\hat{\mathbf{u}}_\Gamma(\mathbf{u}) - \mathbf{u} = 0$ due to consistency of $\hat{\mathbf{u}}_\Gamma$. Furthermore, using consistency of $\hat{\mathbf{h}}_\Gamma$, $\hat{\mathbf{h}}_\Gamma(\mathbf{u},\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u})$, and $\mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}) = p(0,n_1,n_2,0)^\top$ on $\Gamma_W$ we obtain

$$
\left(\hat{\mathbf{h}}_{\Gamma,\mathbf{u}}\right)^\top\left(\tilde{\boldsymbol{\psi}} - \mathbf{z}\right) = p'[\mathbf{u}](0,n_1,n_2,0)^\top\cdot\left(\tilde{\boldsymbol{\psi}} - \mathbf{z}\right) = p'[\mathbf{u}]\mathbf{n}\cdot\left((\psi_1,\psi_2)^\top - (z_2,z_3)^\top\right) = 0,
$$

using $\tilde{\boldsymbol{\psi}} = (0, \psi_1, \psi_2, 0)^\top$ and due to $z_2 = \psi_1$ and $z_3 = \psi_2$ on $\Gamma_W$ (cf. (244)). Thus we have $\hat{\mathbf{h}}_{\Gamma,\mathbf{u}}(\mathbf{u}, \mathbf{n}) \cdot (\tilde{\boldsymbol{\psi}} - \mathbf{z}) = 0$ in (272). Furthermore, using consistency of $\hat{\underline{\sigma}}_\Gamma$, $\hat{\underline{\sigma}}_\Gamma(\mathbf{u}, \nabla\mathbf{u}) = \mathbf{n} \cdot \mathcal{F}^v(\mathbf{u})(\mathbf{u}, \nabla\mathbf{u})$, and $\mathbf{n} \cdot \mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u}) = (0, (\tau\mathbf{n})_1, (\tau\mathbf{n})_2, \mathcal{K}\mathbf{n} \cdot \nabla T)^\top$ on $\Gamma_W$ we obtain

$$-\left(\hat{\underline{\sigma}}_\Gamma \mathbf{n}\right)^\top \left(\tilde{\boldsymbol{\psi}} - \mathbf{z}\right) = -\underline{\tau}\mathbf{n} \cdot \left((\psi_1, \psi_2)^\top - (z_2, z_3)^\top\right) = 0, \tag{273}$$

again using $z_2 = \psi_1$ and $z_3 = \psi_2$ on $\Gamma_W$, and $z_4 = 0$ on $\Gamma_{\text{iso}}$ (cf. (244)) and $\mathbf{n} \cdot \nabla T = 0$ on $\Gamma_{\text{adia}}$. Thus we have $-\hat{\underline{\sigma}}_{\Gamma,\mathbf{u}}\mathbf{n} \cdot (\tilde{\boldsymbol{\psi}} - \mathbf{z}) = 0$ and $-\hat{\underline{\sigma}}_{\Gamma,\nabla\mathbf{u}}\mathbf{n} \cdot (\tilde{\boldsymbol{\psi}} - \mathbf{z}) = 0$ in (272).

Thereby, in summary, all terms in (268) vanish if evaluated for the exact and smooth solutions $\mathbf{u}$ and $\mathbf{z}$ to (224) and (237), respectively, proving adjoint consistency. $\square$

## 6.6 Numerical flux functions

Like in Section 4.7 for the discretization of Poisson's equation we now proceed in defining the numerical flux functions $\hat{\mathbf{u}}_h$ and $\underline{\sigma}_h$ for deriving various different DG discretization methods. However, in the following we concentrate on the adjoint consistent discretizations introduced in Section 4.7 while omitting the adjoint *in*consistent methods like the non-symmetric interior penalty (NIPG) method and the method of Baumann-Oden (cf. Section 4.7).

For the SIPG and BR2 discretization the fluxes $\hat{\mathbf{u}}_h$ and $\hat{\underline{\sigma}}_h$ on interior faces $\partial\kappa \setminus \Gamma$, $\kappa \in \mathcal{T}_h$, are given by

$$\hat{\mathbf{u}}_h = \{\!\{\mathbf{u}_h\}\!\}, \qquad \hat{\underline{\sigma}}_h = \{\!\{G(\mathbf{u}_h)\nabla_h \mathbf{u}_h\}\!\} - \underline{\delta}(\mathbf{u}_h) \qquad \text{on } \Gamma_{\mathcal{I}},$$

where the penalization term $\underline{\delta}(\mathbf{u}_h)$ for various IP and BR2 schemes is given by

$$\begin{aligned}
\underline{\delta}(\mathbf{u}_h) &= C_{\text{IP}}\tfrac{p^2}{h_e}\mu[\![\underline{\mathbf{u}_h}]\!] && \text{for IP (Hartmann and Houston, 2006a),}\\
\underline{\delta}(\mathbf{u}_h) &= C_{\text{IP}}\tfrac{p^2}{h_e}\{\!\{G(\mathbf{u}_h)\}\!\}[\![\underline{\mathbf{u}_h}]\!] && \text{for IP (Hartmann and Houston, 2008),}\\
\underline{\delta}(\mathbf{u}_h) &= C_{\text{BR2}}\{\!\{G(\mathbf{u}_h)\underline{L}_0^e(\mathbf{u}_h)\}\!\} && \text{for BR2 (Bassi et al., 1997, 2005),}\\
\underline{\delta}(\mathbf{u}_h) &= C_{\text{BR2}}\{\!\{\tilde{\underline{L}}_0^e(\mathbf{u}_h)\}\!\} && \text{for BR2 (Bassi and Rebay, 2000a, 2002),}
\end{aligned}$$

where the gradient-based local lifting operator $\underline{L}_0^e(\mathbf{u}_h) \in \underline{\Sigma}_h^p$ is given by

$$\int_{\Omega_e} \underline{L}_0^e(\mathbf{u}_h) : \underline{\tau}\,\mathrm{d}\mathbf{x} = \int_e [\![\underline{\mathbf{u}_h}]\!] : \{\!\{\underline{\tau}\}\!\}\,\mathrm{d}s \qquad \forall \underline{\tau} \in \underline{\Sigma}_h^p, \tag{274}$$

for $\Omega_e = \kappa_e^+ \cup \kappa_e^-$ with $e = \partial\kappa_e^+ \cap \partial\kappa_e^-$, and the flux-based local lifting operator $\tilde{\underline{L}}_0^e(\mathbf{u}_h) \in \underline{\Sigma}_h^p$ is given by

$$\int_{\Omega_e} \tilde{\underline{L}}_0^e(\mathbf{u}_h) : \underline{\tau}\,\mathrm{d}\mathbf{x} = \int_e [\![\underline{\mathbf{u}_h}]\!] : \{\!\{G^\top(\mathbf{u}_h)\underline{\tau}\}\!\}\,\mathrm{d}s \qquad \forall \underline{\tau} \in \underline{\Sigma}_h^p.$$

Then,

$$\begin{aligned}
[\![\underline{\hat{\mathbf{u}}_h}]\!] &= [\![\underline{\{\!\{\mathbf{u}_h\}\!\}}]\!] = 0,\\
\{\!\{\hat{\mathbf{u}}_h\}\!\} &= \{\!\{\{\!\{\mathbf{u}_h\}\!\}\}\!\} = \{\!\{\mathbf{u}_h\}\!\},\\
\{\!\{\hat{\underline{\sigma}}_h\}\!\} &= \{\!\{\{\!\{G(\mathbf{u}_h)\nabla_h\mathbf{u}_h\}\!\}\}\!\} - \{\!\{\underline{\delta}(\mathbf{u}_h)\}\!\} = \{\!\{G(\mathbf{u}_h)\nabla_h\mathbf{u}_h\}\!\} - \underline{\delta}(\mathbf{u}_h),\\
[\![\hat{\underline{\sigma}}_h]\!] &= [\![\{\!\{G(\mathbf{u}_h)\nabla_h\mathbf{u}_h\}\!\}]\!] - [\![\underline{\delta}(\mathbf{u}_h)]\!] = 0,
\end{aligned}$$

the last term in (259) vanishes and thus (259) reduces to: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\begin{aligned} N_h(\mathbf{u}_h, \mathbf{v}_h) \equiv \int_\Omega (-\mathcal{F}^c(\mathbf{u}_h) + \mathcal{F}^v(\mathbf{u}_h, \nabla_h \mathbf{u}_h)) : \nabla_h \mathbf{v}_h \, \mathrm{d}\mathbf{x} + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} \mathbf{h}_h \cdot \mathbf{v}_h \, \mathrm{d}s \\ - \int_{\Gamma_\mathcal{I}} \underline{[\![\mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla \mathbf{v}_h\}\!\} \, \mathrm{d}s - \int_{\Gamma_\mathcal{I}} \{\!\{G(\mathbf{u}_h)\nabla_h \mathbf{u}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]} \\ + \int_{\Gamma_\mathcal{I}} \underline{\delta}(\mathbf{u}_h) : \underline{[\![\mathbf{v}_h]\!]} \, \mathrm{d}s + N_{\Gamma,h}(\mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p, \end{aligned} \tag{275}$$

where the discretization $N_{\Gamma,h}(\mathbf{u}_h, \mathbf{v}_h)$ at the boundary will be specified in the following.

## 6.7 Numerical flux functions at the boundary

There are arbitrarily many different possible discretizations at the boundary which yield a consistent and adjoint consistent discretization according to Lemma 6.4 and Theorem 6.9. In the following, however, we restrict ourselves to discuss choices which are connected to the examples of numerical boundary flux functions given in Section 5.6. We split the discretization on the boundary in the part on the wall boundary $\Gamma_W$ and the part on the farfield boundary $\Gamma \setminus \Gamma_W$,

$$N_{\Gamma,h}(\mathbf{u}_h, \mathbf{v}_h) = N_{\Gamma_W,h}(\mathbf{u}_h, \mathbf{v}_h) + N_{\Gamma\setminus\Gamma_W,h}(\mathbf{u}_h, \mathbf{v}_h), \tag{276}$$

and consider them separately in the following subsections.

### 6.7.1 Discretization at the wall boundary based on normal boundary fluxes

In the following we give the generalization of the discretization as given in (202) for inviscid flows to the current case of viscous flows. In particular, we consider the numerical boundary fluxes

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+, \mathbf{n}) = \mathbf{n} \cdot \mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)), \ \hat{\mathbf{u}}_{\Gamma,h} = \mathbf{u}_\Gamma(\mathbf{u}_h^+), \ \underline{\hat{\sigma}}_{\Gamma,h} = \tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+), \nabla \mathbf{u}_h^+) - \underline{\delta}_\Gamma(\mathbf{u}_h^+), \tag{277}$$

where the boundary value function $\mathbf{u}_\Gamma(\cdot)$ is given according to (228), i.e., at adiabatic and isothermal no-slip wall boundaries, $\Gamma_{\text{adia}}$ and $\Gamma_{\text{iso}}$, we have

$$\begin{aligned} \mathbf{u}_\Gamma(\mathbf{u}_h) &= (u_{h,1}, 0, 0, u_{h,4})^\top && \text{on } \Gamma_{\text{adia}}, \\ \mathbf{u}_\Gamma(\mathbf{u}_h) &= (u_{h,1}, 0, 0, u_{h,1} c_v T)^\top && \text{on } \Gamma_{\text{iso}}. \end{aligned} \tag{278}$$

Furthermore, the diffusive flux $\tilde{\mathcal{F}}^v$ and the corresponding homogeneity tensor $\widetilde{G}$ is modified on $\Gamma_{\text{adia}}$ such that $\mathbf{n} \cdot \nabla T = 0$, i.e.,

$$\mathbf{n} \cdot \tilde{\mathcal{F}}^v(\mathbf{u}_h, \nabla \mathbf{u}_h) = \mathbf{n} \cdot \left(\widetilde{G}(\mathbf{u}_h)\nabla \mathbf{u}_h\right) = \begin{cases} (0, (\tau_h \mathbf{n})_1, (\tau_h \mathbf{n})_2, \mathbf{n} \cdot (\tau_h \mathbf{v}_h))^\top & \text{on } \Gamma_{\text{adia}}, \\ \mathbf{n} \cdot \mathcal{F}^v(\mathbf{u}_h, \nabla \mathbf{u}_h) & \text{on } \Gamma_{\text{iso}}. \end{cases} \tag{279}$$

Finally, the penalization term $\underline{\delta}_\Gamma(\mathbf{u}_h)$ on $\Gamma_W = \Gamma_{\text{adia}} \cup \Gamma_{\text{iso}}$ is given by

$$\begin{aligned} \underline{\delta}_\Gamma(\mathbf{u}_h) &= C_{\text{IP}} \tfrac{p^2}{h_e} \mu \, (\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)) \otimes \mathbf{n} && \text{for IP (Hartmann and Houston, 2006a)}, \\ \underline{\delta}_\Gamma(\mathbf{u}_h) &= C_{\text{IP}} \tfrac{p^2}{h_e} \widetilde{G}_\Gamma(\mathbf{u}_h) \, (\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)) \otimes \mathbf{n} && \text{for IP (Hartmann and Houston, 2008)}, \\ \underline{\delta}_\Gamma(\mathbf{u}_h) &= C_{\text{BR2}} \, \widetilde{G}_\Gamma(\mathbf{u}_h) \, \underline{L}_\Gamma^e(\mathbf{u}_h) && \text{for BR2 (Bassi et al., 1997, 2005)}, \\ \underline{\delta}_\Gamma(\mathbf{u}_h) &= C_{\text{BR2}} \, \underline{\tilde{L}}_\Gamma^e(\mathbf{u}_h) && \text{for BR2 (Bassi and Rebay, 2000a, 2002)}, \end{aligned}$$

with $\widetilde{G}_\Gamma(\mathbf{u}_h) = \widetilde{G}(\mathbf{u}_\Gamma(\mathbf{u}_h))$. Here, the gradient-based local lifting operator $\underline{L}_\Gamma^e(\mathbf{u}_h) \in \underline{\Sigma}_h^p$ on $e \subset \Gamma$ is defined by

$$\int_{\kappa_e} \underline{L}_\Gamma^e(\mathbf{u}_h) : \underline{\tau} \, \mathrm{d}\mathbf{x} = \int_e (\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)) \otimes \mathbf{n} : \underline{\tau} \, \mathrm{d}s \qquad \forall \underline{\tau} \in \underline{\Sigma}_h^p, \tag{280}$$

and the flux-based local lifting operator $\tilde{\underline{L}}_\Gamma^e(\mathbf{u}_h) \in \underline{\Sigma}_h^p$ on $e \subset \Gamma$ is defined by

$$\int_{\kappa_e} \tilde{\underline{L}}_\Gamma^e(\mathbf{u}_h) : \underline{\tau} \, \mathrm{d}\mathbf{x} = \int_e (\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)) \otimes \mathbf{n} : \left(\widetilde{G}_\Gamma^\top(\mathbf{u}_h)\underline{\tau}\right) \mathrm{d}s \qquad \forall \underline{\tau} \in \underline{\Sigma}_h^p \tag{281}$$

for $\kappa_e$ such that $\partial\kappa \cap \Gamma = e$. Thereby, the wall boundary term $N_{\Gamma_W,h}^v(\mathbf{u}_h, \mathbf{v}_h)$ is given by

$$\begin{aligned} N_{\Gamma_W,h}(\mathbf{u}_h, \mathbf{v}_h) = &\int_{\Gamma_W} \mathbf{n} \cdot \left(\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) - \tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+), \nabla\mathbf{u}_h^+) + \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \mathbf{v}_h^+ \, \mathrm{d}s \\ &- \int_{\Gamma_W} \left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right) \otimes \mathbf{n} : \left(G^\top(\mathbf{u}_h^+)\nabla\mathbf{v}_h^+\right) \mathrm{d}s \end{aligned} \tag{282}$$

Note, that the numerical fluxes in (277) are consistent. Thereby, according to Theorem 6.9, the discretization (275) with the discretization (282) at the wall boundary is *adjoint consistent* in combination with following discretization of the force coefficients (232),

$$\begin{aligned} J_h(\mathbf{u}_h) &= \int_{\Gamma_W} \left(\hat{\mathbf{h}}_{\Gamma,h} - \underline{\hat{\sigma}}_{\Gamma,h}\mathbf{n}\right) \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s \\ &= \int_{\Gamma_W} \mathbf{n} \cdot \left(\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) - \tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+), \nabla\mathbf{u}_h^+) + \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s \\ &= \int_{\Gamma_W} \left(p\left(\mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\mathbf{n} - \tau\left(\mathbf{u}_\Gamma(\mathbf{u}_h^+), \nabla\mathbf{u}_h^+\right)\mathbf{n}\right) \cdot \boldsymbol{\psi} \, \mathrm{d}s + \int_{\Gamma_W} \left(\mathbf{n} \cdot \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s \\ &= J(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) + \int_{\Gamma_W} \left(\mathbf{n} \cdot \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s. \end{aligned} \tag{283}$$

In contrast to that a simpler discretization of the force coefficients based on $J(\mathbf{u}_\Gamma(\mathbf{u}_h^+))$ is adjoint *in*consistent. Note, that the term

$$J_h(\mathbf{u}_h) - J(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) = \int_{\Gamma_W} \left(\mathbf{n} \cdot \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \, \mathrm{d}s \tag{284}$$

was derived in (Hartmann, 2007a) as a required additional term for the SIPG discretization to be added to the discretization of the target quantity based on $J(\mathbf{u}_\Gamma(\mathbf{u}_h^+))$. This term was called the *IP modification* in (Hartmann, 2007a). While the need of such a modification might be a surprising outcome of the analysis in (Hartmann, 2007a) it seems natural in view of (283), which is based on the numerical flux functions and does not require any additional modification. Furthermore, the formulation in (283) is also valid for discretizations other than SIPG.

**Remark 6.10** *From the proofs of Lemma 6.4 and Theorem 6.9 we see that the discretization is still* consistent *and* adjoint consistent*,*

- *if $G^\top(\mathbf{u}_h)$ in (282) is replaced by $G_\Gamma^\top(\mathbf{u}_h)$, and*
- *if $G^\top(\mathbf{u}_h)$ in (282) is replaced by $\widetilde{G}_\Gamma^\top(\mathbf{u}_h)$. Note, that the discretization resulting from this choice has been considered in (Hartmann, 2007a; Hartmann and Houston, 2008).*

### 6.7.2 Discretization at the wall boundary based on interior numerical fluxes

In the following we give the generalization of the discretization as given in (203) for inviscid flows to viscous flows. In particular, we employ the same numerical flux functions for defining the numerical boundary fluxes as employed on interior faces, i.e., we consider the numerical boundary fluxes

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = \hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^-(\mathbf{u}_h^+),\mathbf{n}),\;\; \hat{\mathbf{u}}_{\Gamma,h} = \mathbf{u}_\Gamma(\mathbf{u}_h^+),\;\; \hat{\underline{\sigma}}_{\Gamma,h} = \{\!\{\tilde{\mathcal{F}}^v(\mathbf{u}_h,\nabla\mathbf{u}_h)\}\!\}_\Gamma - \tilde{\underline{\delta}}_\Gamma(\mathbf{u}_h^+), \quad (285)$$

with the *boundary mean value*[2] $\{\!\{\cdot\}\!\}_\Gamma$ of the diffusive flux given by

$$\{\!\{\tilde{\mathcal{F}}^v(\mathbf{u}_h,\nabla\mathbf{u}_h)\}\!\}_\Gamma = \frac{1}{2}\left(\tilde{\mathcal{F}}^v(\mathbf{u}_h^+,\nabla\mathbf{u}_h^+) + \tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma^-,(\nabla\mathbf{u})_\Gamma^-)\right), \quad (286)$$

where $\tilde{\mathcal{F}}^v$ denotes the modified normal flux at $\Gamma_{\mathrm{adia}}$ according to (279). Furthermore, the *wall exterior state* $\mathbf{u}_\Gamma^- = \mathbf{u}_\Gamma^-(\mathbf{u}_h^+)$ is obtained by mirroring the interior state $\mathbf{u}_h^+$ at the wall boundary state $\mathbf{u}_\Gamma(\mathbf{u}_h^+)$ (cf. (278)) according to

$$\frac{1}{2}\left(\mathbf{u}_h^+ + \mathbf{u}_\Gamma^-(\mathbf{u}_h^+)\right) = \mathbf{u}_\Gamma(\mathbf{u}_h^+). \quad (287)$$

Finally, $(\nabla\mathbf{u})_\Gamma^-$ is the *wall exterior gradient* which might depend on the interior state and gradient, i.e., $(\nabla\mathbf{u})_\Gamma^- = (\nabla\mathbf{u})_\Gamma^-(\mathbf{u}_h^+,\nabla\mathbf{u}_h^+)$, and will be specified later. This boundary treatment of fluxes corresponds to introducing a ghost layer of elements at the wall boundary and evaluating the discretization on the wall boundary like on interior faces.

It remains to define the penalization term $\tilde{\underline{\delta}}_\Gamma$ in (285). It is given by

$$\tilde{\underline{\delta}}_\Gamma(\mathbf{u}_h) = C_{\mathrm{IP}}\tfrac{p^2}{h_e}\mu\left(\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)\right)\otimes\mathbf{n} \qquad \text{for IP (Hartmann and Houston, 2006a),}$$
$$\tilde{\underline{\delta}}_\Gamma(\mathbf{u}_h) = C_{\mathrm{IP}}\tfrac{p^2}{h_e}\{\!\{\widetilde{G}(\mathbf{u})\}\!\}_\Gamma\left(\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h)\right)\otimes\mathbf{n} \quad \text{for IP (Hartmann and Houston, 2008).}$$

For the gradient-based BR2 scheme (Bassi et al., 1997, 2005) it is given by

$$\begin{aligned}\tilde{\underline{\delta}}_\Gamma(\mathbf{u}_h^+) &= C_{\mathrm{BR2}}\{\!\{\widetilde{G}(\mathbf{u}_h)\underline{L}_\Gamma^e(\mathbf{u}_h)\}\!\}_\Gamma = \tfrac{C_{\mathrm{BR2}}}{2}\left(\widetilde{G}(\mathbf{u}_h^+)\underline{L}_\Gamma^e(\mathbf{u}_h^+) + \widetilde{G}(\mathbf{u}_\Gamma^-)\underline{L}_\Gamma^e(\mathbf{u}_\Gamma^-)\right)\\ &= C_{\mathrm{BR2}}\{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma\,\underline{L}_\Gamma^e(\mathbf{u}_h^+),\end{aligned}$$

where we used $\underline{L}_\Gamma^e(\mathbf{u}_h^+) = \underline{L}_\Gamma^e(\mathbf{u}_\Gamma^-)$ due to

$$\begin{aligned}\int_{\kappa_e}\underline{L}_\Gamma^e(\mathbf{u}_h^+):\underline{\tau}\,\mathrm{d}\mathbf{x} &= \int_e\left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes\mathbf{n}:\underline{\tau}\,\mathrm{d}s\\ &= \int_e\left(\mathbf{u}_\Gamma^- - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes(-\mathbf{n}):\underline{\tau}\,\mathrm{d}s = \int_{\kappa_e}\underline{L}_\Gamma^e(\mathbf{u}_\Gamma^-):\underline{\tau}\,\mathrm{d}\mathbf{x} \qquad \forall\underline{\tau}\in\underline{\Sigma}_h^p.\end{aligned}$$

And for the flux-based BR2 scheme (Bassi and Rebay, 2000a, 2002) it is given by

$$\tilde{\underline{\delta}}_\Gamma(\mathbf{u}_h^+) = C_{\mathrm{BR2}}\{\!\{\tilde{\underline{L}}_\Gamma^e(\mathbf{u}_h)\}\!\}_\Gamma = \tfrac{C_{\mathrm{BR2}}}{2}\left(\tilde{\underline{L}}_\Gamma^e(\mathbf{u}_h^+) + \tilde{\underline{L}}_\Gamma^e(\mathbf{u}_\Gamma^-)\right) = C_{\mathrm{BR2}}\tilde{\underline{L}}^e_{\Gamma,\{\!\{\cdot\}\!\}_\Gamma}(\mathbf{u}_h), \quad (288)$$

[2] We introduce the new notation of a *boundary mean value* $\{\!\{\underline{\tau}_h\}\!\}_\Gamma := \frac{1}{2}(\underline{\tau}_h^+ + \underline{\tau}_\Gamma^-)$ because the standard mean value $\{\!\{\underline{\tau}_h\}\!\}$ reduces to $\underline{\tau}_h^+$ at the boundary (cf. (256)).

with

$$
\begin{aligned}
\int_{\kappa_e} \underline{\tilde{L}}^e_{\Gamma,\{\!\{\cdot\}\!\}_\Gamma}(\mathbf{u}_h) : \underline{\tau}\,\mathrm{d}\mathbf{x} &= \int_{\kappa_e} \underline{\tilde{L}}^e_\Gamma(\mathbf{u}_h^+) + \underline{\tilde{L}}^e_\Gamma(\mathbf{u}_\Gamma^-) : \underline{\tau}\,\mathrm{d}\mathbf{x} \\
&= \int_e \Big( \big(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\big) \otimes \mathbf{n} : \widetilde{G}^\top(\mathbf{u}_h^+)\,\underline{\tau} + \big(\mathbf{u}_\Gamma^- - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\big) \otimes (-\mathbf{n}) : \widetilde{G}^\top(\mathbf{u}_\Gamma^-)\,\underline{\tau} \Big)\,\mathrm{d}s \\
&= \int_e \big(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\big) \otimes \mathbf{n} : \{\!\{\widetilde{G}^\top(\mathbf{u}_h)\}\!\}_\Gamma\,\underline{\tau}\,\mathrm{d}s \qquad \forall \underline{\tau} \in \underline{\Sigma}_h^p.
\end{aligned}
$$

**Remark 6.11** *Note, that here the penalization terms $\underline{\tilde{\delta}}_\Gamma$ include $\{\!\{G^\top(\mathbf{u}_h)\}\!\}_\Gamma$ instead of $G_\Gamma^\top(\mathbf{u}_h)$ as is included in the penalization terms $\underline{\delta}_\Gamma$ based on normal flux in Section 6.7.1.*

Motivated by this, we define the *wall exterior gradient* $(\nabla\mathbf{u})_\Gamma^-$ in (286) to be identical to the interior gradient, i.e., $(\nabla\mathbf{u})_\Gamma^- = \nabla\mathbf{u}_h^+$. This simple choice results in

$$
\{\!\{\tilde{\mathcal{F}}^v(\mathbf{u}_h,\nabla\mathbf{u}_h)\}\!\}_\Gamma = \frac{1}{2}\Big(\tilde{\mathcal{F}}^v(\mathbf{u}_h^+,\nabla\mathbf{u}_h^+) + \tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma^-,(\nabla\mathbf{u})_\Gamma^-)\Big) = \{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma \nabla u_h^+, \tag{289}
$$

which, similarly, includes $\{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma$ instead of $\widetilde{G}_\Gamma(\mathbf{u}_h)$ in $\tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+,\nabla\mathbf{u}_h^+) = \widetilde{G}_\Gamma(\mathbf{u}_h)\nabla u_h^+$ in (277).

**Remark 6.12** *Note, that there is a small difference between $\{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma$ and $\widetilde{G}_\Gamma(\mathbf{u}_h)$, only. As $G(\mathbf{u})$ is nonlinear in $\mathbf{u}$ (cf.* (226)*) we have*

$$
\{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma = \frac{1}{2}\Big(\widetilde{G}(\mathbf{u}_h^+) + \widetilde{G}(\mathbf{u}_\Gamma^-)\Big) \neq \widetilde{G}\left(\frac{1}{2}\left(\mathbf{u}_h^+ + \mathbf{u}_\Gamma^-\right)\right) = \widetilde{G}\left(\mathbf{u}_\Gamma\right). \tag{290}
$$

*In particular, there is a difference in the energy equation in* (226) *due to the velocity components squared.*

Thereby, the wall boundary term $N^v_{\Gamma_W,h}(\mathbf{u}_h,\mathbf{v}_h)$ is given by

$$
\begin{aligned}
N_{\Gamma_W,h}(\mathbf{u}_h,\mathbf{v}_h) = &\int_{\Gamma_W} \Big(\hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^-(\mathbf{u}_h^+),\mathbf{n}) - \{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma \nabla u_h^+ + \underline{\tilde{\delta}}_\Gamma(\mathbf{u}_h^+)\Big) \cdot \mathbf{v}_h^+\,\mathrm{d}s \\
&- \int_{\Gamma_W} \big(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\big) \otimes \mathbf{n} : \big(G^\top(\mathbf{u}_h^+)\nabla\mathbf{v}_h^+\big)\,\mathrm{d}s
\end{aligned} \tag{291}
$$

Note, that the numerical fluxes in (285) are consistent. Thereby, according to Theorem 6.9, the discretization (275) with the discretization (291) at the wall boundary is *adjoint consistent* in combination with following discretization of the force coefficients

$$
\begin{aligned}
J_h(\mathbf{u}_h) &= \int_{\Gamma_W} \Big(\hat{\mathbf{h}}_{\Gamma,h} - \underline{\hat{\sigma}}_{\Gamma,h}\mathbf{n}\Big) \cdot \tilde{\boldsymbol{\psi}}\,\mathrm{d}s \\
&= \int_{\Gamma_W} \Big(\hat{\mathbf{h}}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^-(\mathbf{u}_h^+),\mathbf{n}) - \{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma \nabla u_h^+ + \underline{\tilde{\delta}}_\Gamma(\mathbf{u}_h^+)\Big) \cdot \tilde{\boldsymbol{\psi}}\,\mathrm{d}s.
\end{aligned} \tag{292}
$$

**Remark 6.13** *First numerical tests indicate that the discretization based on* (277) *seems to be more accurate on coarse grids and low polynomial degree $p$ of $\mathbf{u}_h \in \mathbf{V}_{h,p}$ than that based on* (285) *while the discretization based on* (285) *seems to be more stable than that of* (277).

**Remark 6.14** *The discretizations based on the two different choices of boundary fluxes given in* (277) *and* (285) *can be viewed as extremes of a whole class of discretizations. Consider*

$$\begin{aligned}\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) &= \hat{\mathbf{h}}(\mathbf{u}_\Gamma^+(\mathbf{u}_h^+),\mathbf{u}_\Gamma^-(\mathbf{u}_h^+),\mathbf{n}),\\ \hat{\mathbf{u}}_h &= \tfrac{1}{2}\left((\mathbf{u}_\Gamma^+(\mathbf{u}_h^+)+\mathbf{u}_\Gamma^-(\mathbf{u}_h^+)\right) = \mathbf{u}_\Gamma(\mathbf{u}_h^+),\\ \hat{\underline{\sigma}}_{\Gamma,h} &= \tfrac{1}{2}\left(\tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma^+(\mathbf{u}_h^+),\nabla\mathbf{u}_h^+)+\tilde{\mathcal{F}}^v(\mathbf{u}_\Gamma^-(\mathbf{u}_h^+),\nabla\mathbf{u}_h^+)\right) - \underline{\delta}_\Gamma^{(3)}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-)\end{aligned} \tag{293}$$

*with*

$$\mathbf{u}_\Gamma^+(\mathbf{u}_h^+) = \xi\mathbf{u}_h^+ + (1-\xi)\mathbf{u}_\Gamma(\mathbf{u}_h^+), \qquad \mathbf{u}_\Gamma^-(\mathbf{u}_h^+) = 2\mathbf{u}_\Gamma(\mathbf{u}_h^+) - \mathbf{u}_\Gamma^+(\mathbf{u}_h^+), \tag{294}$$

*and*

$$\underline{\delta}_\Gamma^{(3)}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-) = C_{IP}\frac{p^2}{h_e}\mu\left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes\mathbf{n} \qquad \textit{for IP (Hartmann and Houston, 2006a)},$$
$$\underline{\delta}_\Gamma^{(3)}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-) = C_{IP}\frac{p^2}{h_e}\{\!\{\widetilde{G}(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma\left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes\mathbf{n} \quad \textit{for IP (Hartmann and Houston, 2008)},$$

*where* $\{\!\{\widetilde{G}(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma = \frac{1}{2}(\widetilde{G}(\mathbf{u}_\Gamma^+)+\widetilde{G}(\mathbf{u}_\Gamma^-))$. *Furthermore, for the gradient-based BR2 scheme (Bassi et al., 1997, 2005)*

$$\underline{\delta}_\Gamma^{(3)}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-) = C_{BR2}\{\!\{\widetilde{G}(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma\,\underline{L}_\Gamma^e(\mathbf{u}_h^+),$$

*and for the flux-based BR2 scheme (Bassi and Rebay, 2000a, 2002)*

$$\underline{\delta}_\Gamma^{(3)}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-) = C_{BR2}\underline{\tilde{L}}^{e,(3)}_{\Gamma,\{\!\{\cdot\}\!\}_\Gamma}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-), \tag{295}$$

*with*

$$\int_{\kappa_e}\underline{\tilde{L}}^{e,(3)}_{\Gamma,\{\!\{\cdot\}\!\}_\Gamma}(\mathbf{u}_h^+,\mathbf{u}_\Gamma^+,\mathbf{u}_\Gamma^-):\underline{\tau}\,d\mathbf{x} = \int_e\left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes\mathbf{n}:\{\!\{\widetilde{G}^\top(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma\,\underline{\tau}\,ds \qquad \forall\underline{\tau}\in\underline{\Sigma}_h^p.$$

*For* $\xi=0$ *we have* $\mathbf{u}_\Gamma^+ = \mathbf{u}_\Gamma^- = \mathbf{u}_\Gamma(\mathbf{u}_h^+)$, $\{\!\{\widetilde{G}(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma = \widetilde{G}_\Gamma(\mathbf{u}_h)$, *and thus* (293) *reduces to* (277). *For* $\xi=1$ *we have* $\mathbf{u}_\Gamma^+ = \mathbf{u}_h^+$, $\mathbf{u}_\Gamma^- = 2\mathbf{u}_\Gamma - \mathbf{u}_h^+$, $\{\!\{\widetilde{G}(\mathbf{u}_\Gamma^\pm)\}\!\}_\Gamma = \{\!\{\widetilde{G}(\mathbf{u}_h)\}\!\}_\Gamma$ *and thus* (293) *reduces to* (285). *The parameter* $\xi\in[0,1]$ *would allow to choose a discretization of boundary fluxes as a compromise between accuracy and stability.*

### 6.7.3 Discretization at the farfield boundary

In the following we give the generalization of the discretization at the farfield $\Gamma\setminus\Gamma_W$ as given in (5.6.1) for inviscid flows to the current case of viscous flows. In particular, we consider the numerical boundary fluxes

$$\hat{\mathbf{h}}_\Gamma(\mathbf{u}_h^+,\mathbf{n}) = \mathbf{n}\cdot\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)), \quad \hat{\mathbf{u}}_h = \mathbf{u}_\Gamma(\mathbf{u}_h^+), \quad \hat{\underline{\sigma}}_{\Gamma,h} = \mathcal{F}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+),\nabla\mathbf{u}_h^+) - \underline{\delta}_\Gamma(\mathbf{u}_h^+), \tag{296}$$

where the boundary value functions $\mathbf{u}_\Gamma(\cdot)$ for the various farfield boundary conditions are given according to Section 5.6.1. Furthermore, the penalization term $\underline{\delta}_\Gamma(\mathbf{u}_h)$ on $\Gamma\setminus\Gamma_W$ is given by

$$\begin{aligned}
&\underline{\delta}_\Gamma(\mathbf{u}_h) = C_{\text{IP}}\tfrac{p^2}{h_e}\mu\,(\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h))\otimes\mathbf{n} && \text{for IP (Hartmann and Houston, 2006a),}\\
&\underline{\delta}_\Gamma(\mathbf{u}_h) = C_{\text{IP}}\tfrac{p^2}{h_e}G_\Gamma(\mathbf{u}_h)\,(\mathbf{u}_h - \mathbf{u}_\Gamma(\mathbf{u}_h))\otimes\mathbf{n} && \text{for IP (Hartmann and Houston, 2008),}\\
&\underline{\delta}_\Gamma(\mathbf{u}_h) = C_{\text{BR2}}\,G_\Gamma(\mathbf{u}_h)\,\underline{L}^e_\Gamma(\mathbf{u}_h) && \text{for BR2 (Bassi et al., 1997, 2005),}\\
&\underline{\delta}_\Gamma(\mathbf{u}_h) = C_{\text{BR2}}\,\underline{\tilde{L}}^e_\Gamma(\mathbf{u}_h) && \text{for BR2 (Bassi and Rebay, 2000a, 2002).}
\end{aligned}$$

Thereby, the boundary term at the farfield boundary is given by

$$\begin{aligned}
N_{\Gamma\setminus\Gamma_W,h}(\mathbf{u}_h,\mathbf{v}_h) = &\int_{\Gamma\setminus\Gamma_W} \mathbf{n}\cdot\left(\mathcal{F}^c(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) - \mathcal{F}^v(\mathbf{u}_\Gamma(\mathbf{u}_h^+),\nabla\mathbf{u}_h^+) + \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right)\cdot\mathbf{v}_h^+\,\mathrm{d}s\\
&-\int_{\Gamma\setminus\Gamma_W}\left(\mathbf{u}_h^+ - \mathbf{u}_\Gamma(\mathbf{u}_h^+)\right)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h^+)\nabla\mathbf{v}_h^+\right)\,\mathrm{d}s.
\end{aligned}$$

## 6.8 Numerical results

### 6.8.1 Flow over a flat plate

We begin by investigating the accuracy of higher order DG discretizations in resolving laminar boundary layers. To this end, we consider a Mach 0.01 flow with Reynolds number 10 000 horizontally passing over a flat plate of length $l = 2$. The boundary layer solution to this problem can be approximated using Blasius' solution, see (Schlichting and Gersten, 2003), for example. In Figure 13, taken from (Hartmann and Houston, 2006a), we compare the numerical solution computed with the DG($p$) method for $1 \le p \le 3$, at $x = \frac{l}{2} = 1$ and a local Reynolds number $Re_x = 5\,000$, with the Blasius solution ($\eta = y\sqrt{u_\infty/(\nu x)} = \frac{y}{x}\sqrt{Re_x}$ versus $u/u_\infty$ (cf. (Schlichting and Gersten, 2003)) on a sequence of rather coarse computational meshes. On the coarsest mesh, which has about one or two elements within the boundary layer, we see that the DG solution computed with $p = 1, 2$ are not very close to the Blasius solution; increasing the polynomial order to $p = 3$ clearly yields a dramatic improvement in the underlying computed numerical solution. On the next finer mesh obtained by (equidistant) global mesh refinement, where three elements are placed within the boundary layer, the bilinear approximation is still not very accurate, though now both the computed solution with $p = 2, 3$ are in excellent agreement with the Blasius solution. On the subsequent two globally refined meshes we clearly observe that the DG approximation with bilinear elements ($p = 1$) finally starts to coincide with the Blasius solution, at least on a macroscopic level. A more detailed view of the numerical solution on these latter two finer meshes is shown in the zoom depicted in Figure 14. Here, we see that there is still a significant difference between the

| | DG(1) | DG(2) | DG(3) |
|---|---|---|---|
| elements | 36 | 5 | 3 |
| DoFs | 72 | 15 | 12 |

Table 1: Number of elements and degrees of freedom in the boundary layer required by DG($p$), $1 \le p \le 3$, discretizations for approximating the viscous force up to 5%, (Hartmann and Houston, 2006a).

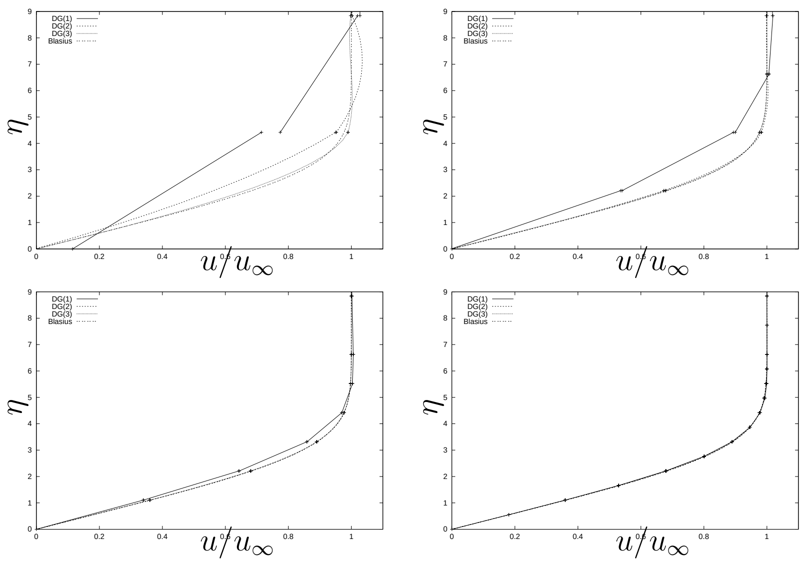



Figure 13: DG($p$), $1 \leq p \leq 3$, solutions in comparison with the Blasius solution ($\eta = y\sqrt{u_\infty/(\nu x)} = \frac{y}{x}\sqrt{Re_x}$ versus $u/u_\infty$) on a sequence of meshes with an increasing number of elements, (Hartmann and Houston, 2006a).

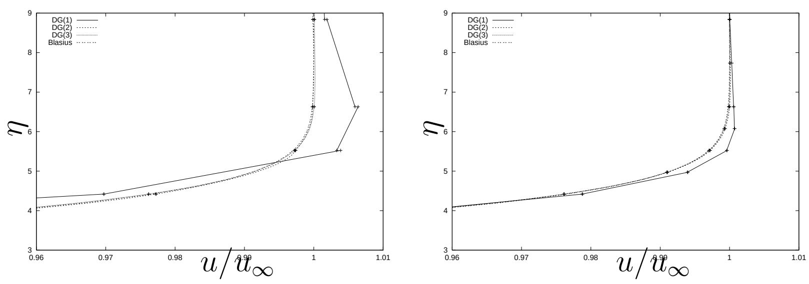



Figure 14: Zoom of the DG($p$), $1 \leq p \leq 3$, solutions on the two finest grids, (Hartmann and Houston, 2006a).

Blasius solution and the computed discontinuous Galerkin solution with $p = 1$. Indeed, these figures clearly highlight the substantial gains in accuracy attained when higher–order polynomial degrees are employed with the DG method. This is further highlighted in Table 1, where we summarize the number of elements and the number of degrees of freedom, orthogonal to the wall, which are required by the DG method for each polynomial degree in order to resolve the boundary layer to a sufficient accuracy that the error in computed viscous stress forces (based on a direct discretization of force coefficients) exerted on the wall are within 5% of that computed with the Blasius solution.

### 6.8.2 Viscous flow around a NACA0012 airfoil

In this example, we consider a subsonic viscous flow around a NACA0012 airfoil. At the farfield (inflow) boundary we specify a Mach 0.5 flow at a zero angle of attack, i.e., $\alpha = 0°$, with Reynolds number $\mathrm{Re} = 5\,000$; on the walls of the airfoil geometry, we impose a zero heat flux (adiabatic) no-slip boundary condition. This is a standard laminar test case which has been investigated by many other authors, cf. (Bassi and Rebay, 1997a;

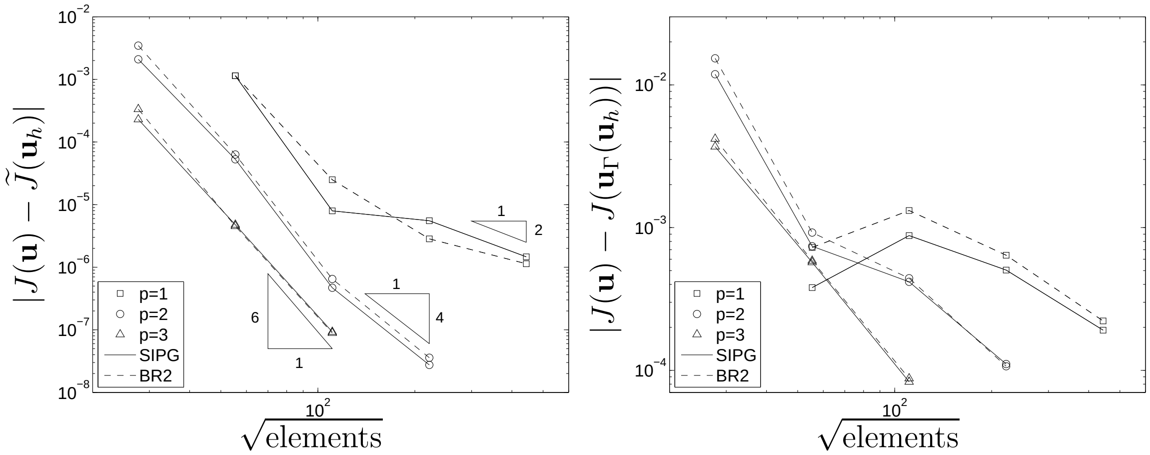



Figure 15: Viscous flow around NACA0012 airfoil. Comparison of the SIPG and BR2 methods employing: (left) Adjoint consistent discretization of the drag coefficient; (right) Adjoint *in*consistent discretization of the drag coefficient based on $J(\mathbf{u}_\Gamma(\mathbf{u}_h^+))$. Note the different scales on the error axis in the left and right plot, (Hartmann and Houston, 2008).

Hartmann and Houston, 2006a), for example. The solution to this problem consists of a strictly subsonic flow which is symmetric about the $x$-axis.

Here, we consider the evaluation of the drag coefficient $C_\mathrm{d}$; i.e., the target quantity is

$$J(\mathbf{u}) = \int_{\Gamma_W} (p\,\mathbf{n} - \underline{\tau}\,\mathbf{n}) \cdot \boldsymbol{\psi}_d \,\mathrm{d}s,$$

where $\boldsymbol{\psi}_d = \frac{1}{C_\infty}(\cos(\alpha), \sin(\alpha))^\top$, cf. (232) and (176). We remark that the adjoint consistency of the DG discretization is based on the adjoint consistent discretization $J_h(\cdot)$ given in (283), i.e.,

$$J_h(\mathbf{u}_h) = \int_{\Gamma_W} \left(\hat{\mathbf{h}}_{\Gamma,h} - \underline{\hat{\sigma}}_{\Gamma,h}\mathbf{n}\right) \cdot \tilde{\boldsymbol{\psi}} \,\mathrm{d}s = J(\mathbf{u}_\Gamma(\mathbf{u}_h^+)) + \int_{\Gamma_W} \left(\mathbf{n} \cdot \underline{\delta}_\Gamma(\mathbf{u}_h^+)\right) \cdot \tilde{\boldsymbol{\psi}} \,\mathrm{d}s. \qquad (297)$$

With this in mind, in Figure 15(a) we present a comparison of the error in the computed target quantity with the (square root of the) number of elements for $p = 1, 2, 3$, employing both the SIPG method with $C_\mathrm{IP} = 10$ and the (flux-based) Bassi–Rebay method (BR2) with $C_\mathrm{BR2} = 4$. In both cases, we observe that, asymptotically, at least, $|J(\mathbf{u}) - J_h(\mathbf{u}_h)|$ converges to zero at the expected optimal rate $\mathcal{O}(h^{2p})$ as the mesh is refined for each fixed $p$. Moreover, as before, we note that in terms of accuracy, for a given number of elements, or equivalently, for a fixed number of degrees of freedom, both the SIPG scheme and the BR2 method perform in a comparable manner. To highlight the necessity of the adjoint consistent discretization (297) of the target quantity including the term involving the penalty function $\underline{\delta}_\Gamma(\cdot)$, in Figure 15(b) we present a comparison of $|J(\mathbf{u}) - J(\mathbf{u}_\Gamma(\mathbf{u}_h))|$ with the (square root of the) number of elements for $p = 1, 2, 3$ employing both the SIPG and BR2 schemes. In this case, we now observe that there is a significant deterioration of the error for a given mesh size and polynomial order when compared to the corresponding results with the adjoint consistent discretization of the target functional. Indeed, comparing Figures 15(a) and 15(b), we see that the difference of $J(\mathbf{u}_\Gamma(\mathbf{u}_h))$ and $J_h(\mathbf{u}_h)$ as in (297) leads to around 2–3 orders of magnitude improvement in the computed error in the drag.

### 6.8.3 Laminar flow around a delta wing

As a final example we consider a laminar flow around a delta wing. At the farfield (inflow) boundary we specify a Mach 0.3 flow at an angle $\alpha = 12.5°$ of attack with Reynolds number $\mathrm{Re} = 4000$. On the walls of the delta wing we impose an isothermal wall boundary condition. This is the BTC3 test case of the EU-project ADIGMA (Kroll et al., 2010) and the C2.4 test case in the International Workshop on High-order CFD Methods (cf. Wang et al. (2013)). Figure 16 shows the Mach number isolines on several slices of the flow field. The flow is computed on a coarse mesh of 3264 elements. The corresponding surface mesh is depicted on the wing geometry. On the left part of the delta wing the flow field solution $\mathbf{u}_h \in \mathbf{V}_h^p$ for $p = 1$ is shown and on the right part the $p = 4$ solution is shown, i.e., Figure 16 compares the solutions of the (design) orders 2 and 5 on the same mesh. We see that the $5^{\text{th}}$ order flow solution provides a good resolution of the primary and secondary vortices. Furthermore, the vortices are tracked over some distance behind the wing. In contrast to that, the primary and secondary vortices are almost indistinguishable in the $2^{\text{nd}}$ order flow solution. Here, the vortices merge and are damped out far too early. Already after a short distance behind the wing the original vortex system is lost due to numerical viscosity.

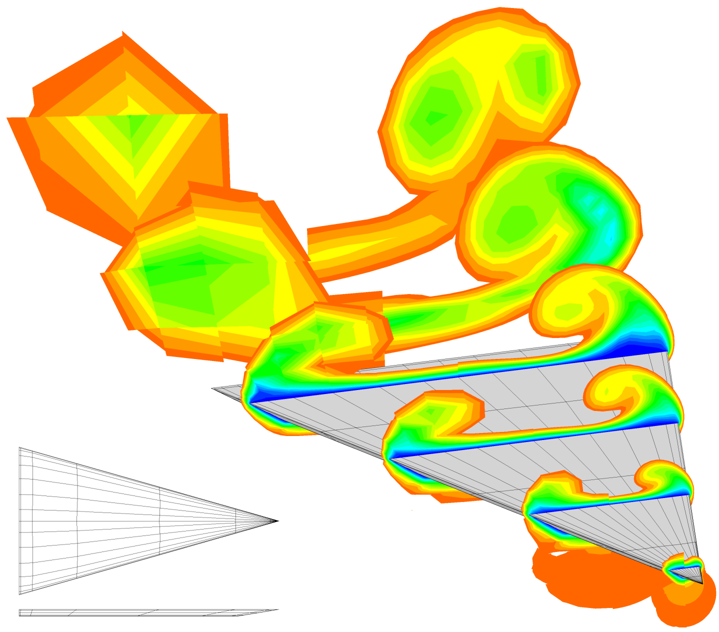


Figure 16: Laminar flow around a delta wing. Geometry and the Mach number isolines on several slices of the flow field solution $\mathbf{u}_h \in \mathbf{V}_h^p$ with $p = 1$ ($2^{\text{nd}}$ order by design) on the left part of the wing and with $p = 4$ ($5^{\text{th}}$ order by design) on the right part of the wing.

Note, that due to the non-smoothness of the flow solution at the sharp leading edge and the blunt trailing edge the order of convergence on a hierarchy of meshes is bounded by two irrespective of the design order of method (cf. Wang et al. (2013)). Note, however, that this bound can be circumvented by using high order polynomials with constant polynomial degree $p$ in combination with adaptive mesh refinement, or even better, by using $hp$-refinment, where non-smooth flow regions are resolved by mesh refinement ($h$-refinement) and smooth regions are resolved by increasing the polynomial degree $p$ ($p$-refinement).

# 7 DG discretization of the RANS-$k\omega$ equations

In this section, we follow (Hartmann, 2013b), and consider a discontinuous Galerkin discretization of turbulent flows as governed by the Reynolds averaged Navier-Stokes (RANS) equations and the $k$-$\omega$ turbulence model equations.

## 7.1 The RANS and $k$-$w$ turbulence equations

In particular, we consider the RANS and the Wilcox $k$-$\omega$ equations (Wilcox, 1988, 1993),

$$\nabla \cdot (\mathcal{F}^c(\mathbf{u}) - \mathcal{F}^v(\mathbf{u}, \nabla\mathbf{u})) - \mathcal{S}(\mathbf{u}, \nabla\mathbf{u}) = 0 \quad \text{in } \Omega, \tag{298}$$

on the domain $\Omega \subset \mathbb{R}^3$. Similar to Bassi *et al.*(Bassi et al., 2005, 2009) the equations are considered in terms of the auxiliary variable $\tilde{\omega} = \ln\omega$ instead of $\omega$ for a more moderate near-wall behavior of the variable. Additionally, this variable transformation guarantees positivity of $\omega$. Then, the vector of conservative variables $\mathbf{u} \in \mathbb{R}^7$ and the convective fluxes $\mathcal{F}^c = (\mathbf{f}_1^c, \mathbf{f}_2^c, \mathbf{f}_3^c)$ are given by

$$\mathbf{u} = \begin{bmatrix} \rho \\ \rho v_i \\ \rho E \\ \rho k \\ \rho\tilde{w} \end{bmatrix}, \qquad \mathbf{f}_j^c(\mathbf{u}) = \begin{bmatrix} \rho v_j \\ \rho v_i v_j + (p + \frac{2}{3}\rho k)\delta_{ij} \\ \left(\rho H + \frac{2}{3}\rho k\right) v_j \\ \rho k v_j \\ \rho\tilde{w} v_j \end{bmatrix}, \qquad j = 1, 2, 3,$$

and the viscous fluxes $\mathcal{F}^v = (\mathbf{f}_1^v, \mathbf{f}_2^v, \mathbf{f}_3^v)$ are given by

$$\mathbf{f}_j^v(\mathbf{u}, \nabla\mathbf{u}) = \begin{bmatrix} 0 \\ \bar{\tau}_{ij} \\ \bar{\tau}_{ji} v_i + \mathcal{K} T_{x_j} + (\mu + \sigma_k \mu_t) k_{x_j} \\ (\mu + \sigma_k \mu_t) k_{x_j} \\ (\mu + \sigma_\omega \mu_t)\tilde{\omega}_{x_j} \end{bmatrix}, \qquad j = 1, 2, 3.$$

Here the pressure is determined by the equation of state of an ideal gas,

$$p = (\gamma - 1)\rho(E_0 - \tfrac{1}{2}\mathbf{v}^2),$$

with $E_0 = E - k$. Furthermore, the thermal conductivity coefficient $\mathcal{K}$ is given by $\mathcal{K} = c_p(\frac{\mu}{\mathrm{Pr}} + \frac{\mu_t}{\mathrm{Pr}_t})$ and the temperature $T$ by $c_v T = e = E_0 - \frac{1}{2}\mathbf{v}^2$. Here $\mathrm{Pr} = 0.72$ and $\mathrm{Pr}_t = 0.9$ are the molecular and turbulent Prandtl numbers, and $\mu$ and $\mu_t = C_\mu \frac{\rho k}{\omega}$ with $C_\mu = 1$ are the molecular and turbulent viscosities.

The stress tensor $\bar{\underline{\tau}}$ is given by following relation

$$\underline{\tau} + \underline{\tau}^R = \bar{\underline{\tau}} - \tfrac{2}{3}\rho k \underline{I} = \underline{\tau} + \underline{\tau}_t - \tfrac{2}{3}\rho k \underline{I} = (\mu + \mu_t)\underline{S} - \tfrac{2}{3}\rho k \underline{I},$$

where $\underline{\tau} = \mu\underline{S}$ is the viscous stress tensor, and $\underline{\tau}^R = \mu_t\underline{S} - \frac{2}{3}\rho k\underline{I}$ is the Reynolds stress tensor with $\underline{S} = \nabla\mathbf{v} + (\nabla\mathbf{v})^\top - \frac{2}{3}(\nabla\cdot\mathbf{v})\underline{I}$. We note, that here, like in the laminar case (cf. Section 6.1), the viscous flux can be rewritten as $\mathbf{f}_k^v(\mathbf{u}, \nabla\mathbf{u}) = G_{kl}(\mathbf{u})\partial\mathbf{u}/\partial x_l$, $k = 1, 2, 3$, where the matrices $G_{kl}(\mathbf{u}) = \partial\mathbf{f}_k^v(\mathbf{u}, \nabla\mathbf{u})/\partial u_{x_l}$, for $k, l = 1, 2, 3$, are the homogeneity tensors defined by $(\mathbf{f}_k^v)_i(\mathbf{u}, \nabla\mathbf{u}) = (G_{kl}(\mathbf{u}))_{ij}\, \partial u_j/\partial x_l,\ k = 1, 2, 3,\ i = 1, \ldots, 7$. Finally, the source term $\mathcal{S}(\mathbf{u}, \nabla\mathbf{u})$ is given by

$$\mathcal{S}(\mathbf{u}, \nabla\mathbf{u}) = \left[0,\ \mathbf{0},\ 0,\ \tau_{ij}^R v_{i,x_j} - \beta_k \rho k e^{\tilde{\omega}_r},\ \frac{\alpha_\omega}{k}\tau_{ij}^R v_{i,x_j} - \beta_\omega \rho e^{\tilde{\omega}_r} + (\mu + \sigma_\omega \mu_t)\tilde{\omega}_{x_k}\tilde{\omega}_{x_k}\right]^\top,$$

with $\mu_t = C_\mu \rho k e^{-\tilde{\omega}_r}$ and $\tilde{\omega}_r = \max\{\tilde{\omega}, \tilde{\omega}_{r0}\}$, where $\tilde{\omega}_{r0}$ fulfills some realizability conditions for the turbulent stresses (Bassi et al., 2005, 2009; Hartmann et al., 2011). We note that here, like in (Hartmann, 2013b), the energy equation is formulated in terms of $\rho E$ whereas in (Bassi et al., 2005, 2009; Hartmann et al., 2011) the $\rho k$ equation has been subtracted resulting in an energy equation formulated in terms of $\rho E_0 = \rho(E - k)$. As a consequence, in (Bassi et al., 2005, 2009; Hartmann et al., 2011) the energy equation has no viscous term $\nabla\cdot((\mu + \sigma_k\mu_t)\nabla k)$, whereas here it has no extra source terms. While this change to previous works was motivated by a possible future extension to unsteady flows we note that for steady state flows we have seen marginal differences due to this change, only.

Furthermore, note that the use of the logarithm of turbulence variables has been introduced in (Ilinca and Pelletier, 1998) and results in an equivalent reformulation of the Wilcox $k$-$\omega$ equations (Wilcox, 1988, 1993). In the context of DG discretizations it has first been used in (Bassi and Rebay, 2000b). Furthermore, $k$ in the source and destruction terms as well as in the expression for $\mu_t$ is kept non-negative. We note, that the limitations on $k$ and $\tilde{\omega}$ avoid unphysical values and have been found in (Bassi et al., 2005, 2009) to have a stabilizing effect on the numerical scheme. Finally, the values of the $k$-$\omega$ closure parameters $\alpha_\omega = \frac{5}{9}$, $\beta_k = \frac{9}{100}$, $\beta_\omega = \frac{3}{40}$, $\sigma_k = \sigma_\omega = \frac{1}{2}$ are those of the high-Reynolds Wilcox $k$-$\omega$ model (Wilcox, 1988, 1993).

## 7.2 DG discretization

In contrast to the discontinuous Galerkin discretization of inviscid and laminar flows which was based on a discrete function space (cf. the definition of $\mathbf{V}_h^p$ in (182)) including tensor-product polynomials $Q_p$ on the unit square (or cube), the following DG discretization of the RANS-$k\omega$ equations will rely on the (complete) polynomial space $P_p$ for stability reasons. In particular, in the following the discrete function space is given by

$$\mathbf{V}_h^p = \{\mathbf{v}_h \in \left[L^2(\Omega)\right]^5 :\ \mathbf{v}_h|_\kappa \circ F_\kappa \in [P_p(\hat{\kappa})]^5, \kappa \in \mathcal{T}_h\}. \tag{299}$$

We consider an extension of the DG discretization (275) of the compressible Navier-Stokes equations to include the source terms $\mathcal{S}(\mathbf{u}, \nabla\mathbf{u})$ of the RANS-$k\omega$ equations as follows: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such

$$\begin{aligned} N_h(\mathbf{u}_h, \mathbf{v}_h) \equiv & \int_\Omega (-\mathcal{F}^c(\mathbf{u}_h) + \mathcal{F}^v(\mathbf{u}_h, \nabla_h\mathbf{u}_h)) : \nabla_h\mathbf{v}_h + \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h)\cdot\mathbf{v}_h \,\mathrm{d}\mathbf{x} \\ & - \int_{\Gamma_\mathcal{I}} \underline{[\![\mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\}\!\}\,\mathrm{d}s - \int_{\Gamma_\mathcal{I}} \{\!\{G(\mathbf{u}_h)\nabla_h\mathbf{u}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]} \\ & + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} \mathbf{h}_h\cdot\mathbf{v}_h\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} \underline{\delta}(\mathbf{u}_h) : \underline{[\![\mathbf{v}_h]\!]}\,\mathrm{d}s + N_{\Gamma,h}(\mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall\mathbf{v}_h \in \mathbf{V}_h^p, \end{aligned} \tag{300}$$

where the discretization $N_{\Gamma,h}(\mathbf{u}_h, \mathbf{v}_h)$ at the boundary is as specified in Section 6.7. Note, that this discretization is a *consistent* discretization of the RANS-$k\omega$ equations (298). In fact, like in Sections 5.4 and 6.4 the discretization (300) can be written in primal residual form: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega \mathbf{R}(\mathbf{u}_h) \cdot \mathbf{v}_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \mathbf{r}(\mathbf{u}_h) \cdot \mathbf{v}_h^+ + \underline{\rho}(\mathbf{u}_h) : \nabla_h \mathbf{v}_h^+ \,\mathrm{d}s$$
$$+ \int_\Gamma \mathbf{r}_\Gamma(\mathbf{u}_h) \cdot \mathbf{v}_h^+ + \underline{\rho}_\Gamma(\mathbf{u}_h) : \nabla_h \mathbf{v}_h^+ \,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p,$$

where the primal element residual is given by

$$\mathbf{R}(\mathbf{u}_h) = -\nabla\cdot(\mathcal{F}^c(\mathbf{u}_h) - \mathcal{F}^v(\mathbf{u}_h, \nabla_h\mathbf{u}_h)) + \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h) \quad \text{on } \kappa\in\mathcal{T}_h, \tag{301}$$

which vanishes, if evaluated for the exact solution $\mathbf{u}$ to (298), i.e., $\mathbf{R}(\mathbf{u}) = 0$. The face and boundary residuals are identical to those given in Sections 5.4 and 6.4 and also vanish for the exact solution $\mathbf{u}$.

Furthermore, note that based on the analysis in Sections 5.4 and 6.4 the parts of the discretization in (300) relating to the *convective and diffusive parts* of the RANS-$k\omega$ equations are *adjoint consistent* provided the force coefficients are discretized as in (262). It remains to discuss the adjoint consistency of the discretization of the source term $\mathcal{S}$.

**Lemma 7.1 (Adjoint consistency of source terms, (Oliver and Darmofal, 2009))** *The direct discretization of source terms as in* (300), *i.e.,*

$$\int_\Omega \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h) \cdot \mathbf{v}_h \, d\mathbf{x} \tag{302}$$

*is* adjoint inconsistent. *In contrast to that the following discretization (cf. Bassi et al. (1997, 2005)) of source terms*

$$\int_\Omega \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h - \underline{L}(\mathbf{u}_h)) \cdot \mathbf{v}_h \, d\mathbf{x}, \tag{303}$$

*where $\underline{L}(\mathbf{u}_h)$ denotes the global lifting operator*[3], *is* asymptotically adjoint consistent, *i.e., the adjoint residuals vanish as the resolution of the discrete function space $\mathbf{V}_h^p$ increases, i.e., the mesh size $h \to 0$ and/or the polynomial degree $p \to \infty$.*

Finally, for any union $S$ of elements, choosing $\mathbf{v}_h \in \mathbf{V}_h^0 \subset \mathbf{V}_h^p$ with $\mathbf{v}_h = 1$ on $S$ and $\mathbf{v}_h = 0$ on $\Omega\setminus S$ in (300), we arrive at following conservation properties.

**Lemma 7.2 (Conservation property)** *Let equations* (298) *be discretized based on* (300). *Then, the discretization* (300) *is* conservative, *i.e., for any union $S$ of elements,*

$$\int_{\partial S} \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n} \, ds = \int_S \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h) \, d\mathbf{x},$$

*if and only if the numerical flux functions $\hat{\mathbf{h}}$ and $\hat{\underline{\sigma}}_h$ are* conservative. *In particular, we have the* global *and* local conservation property *for any $\kappa\in\mathcal{T}_h$,*

$$\int_\Gamma \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n} \, ds = \int_\Omega \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h) \, d\mathbf{x}, \qquad \int_{\partial\kappa} \hat{\mathbf{h}}_h - \hat{\underline{\sigma}}_h \mathbf{n} \, ds = \int_\kappa \mathcal{S}(\mathbf{u}_h, \nabla_h\mathbf{u}_h) \, d\mathbf{x}.$$

[3]The global lifting operator $\underline{L}(\mathbf{u}_h)$ is given by $\underline{L}(\mathbf{u}_h) = \sum_{e\in\Gamma_\mathcal{I}} \underline{L}_0^e(\mathbf{u}_h) + \sum_{e\subset\Gamma} \underline{L}_\Gamma^e(\mathbf{u}_h)$ where $\underline{L}_0^e(\mathbf{u}_h)$ and $\underline{L}_\Gamma^e(\mathbf{u}_h)$ denote the local lifting operators defined in (274) and (280), respectively.

## 7.3 Shock capturing

In this section we follow (Hartmann, 2013b) and give details on the shock-capturing scheme used to stabilize the discretization (300) in non-smooth parts of the flow solution like near shocks. In particular, we augment the discretization with a stabilization term based on artificial viscosity that in general is given by

$$\mathcal{N}_{sc}(\mathbf{u}_h, \mathbf{v}_h) \equiv \sum_{\kappa \in \mathcal{T}_h} \int_\kappa \varepsilon(\mathbf{u}_h) \nabla \mathbf{u}_h : \nabla \mathbf{v}_h \,\mathrm{d}\mathbf{x} \equiv \sum_{\kappa \in \mathcal{T}_h} \int_\kappa \varepsilon_{klm}(\mathbf{u}_h) \partial_{x_l} u_h^m \, \partial_{x_k} v_h^m \,\mathrm{d}\mathbf{x}. \tag{304}$$

Many stabilization schemes based on artificial viscosity have been proposed, first by (Hughes and Mallet, 1986) in the context of SUPG and by (Johnson et al., 1990) for streamline diffusion (SD) finite element methods, by (Jaffre et al., 1995) for DG discretization of scalar hyperbolic conservation laws, later by (Hartmann, 2006) for DG discretizations of laminar compressible flows and by (Bassi et al., 2010, 2011) for DG discretizations of turbulent flows. The artificial viscosity schemes in (Jaffre et al., 1995; Hartmann, 2006; Bassi et al., 2010, 2011) are of the form (304); they only differ in the specific definition of the viscosity term $\varepsilon(\mathbf{u}_h)$.

We recall that the artificial viscosity term in (Hartmann, 2006) depends on the equation residuals. Therefore the shock-capturing term vanishes for the exact solution which is necessary for the scheme to be consistent. Furthermore, including directional element sizes it takes into account possible anisotropies of the elements in the mesh. However, numerical experiments showed that the natural extension of the scheme from the laminar compressible Navier-Stokes equation to the RANS-$k\omega$ equations is not usable. Furthermore, while working for second order DG discretizations it turned out that the scheme is not stable for high-order discretizations.

In contrast to that, the artificial viscosity term in (Bassi et al., 2010, 2011) works for the RANS-$k\omega$ equations as well as for higher order DG discretizations. Furthermore, due to a factor $\nabla \cdot \mathcal{F}^c(\mathbf{u}_h)$ it is consistent for the compressible Euler equations. However, the shock-capturing term is not consistent for laminar or turbulent flows; it is asymptotically consistent, though, because it vanishes as the element sizes tend to zero. Furthermore, it does not include directional mesh sizes which would be required for the high anisotropy of elements in the meshes typically employed for the computation of turbulent high Reynolds number flows.

In the following, we present the shock-capturing scheme of (Hartmann, 2013b) which consists of a suitable combination of the schemes in (Hartmann, 2006) and (Bassi et al., 2010, 2011). While consisting of ingredients of both schemes the combination overcomes the shortcomings described above. In particular, we consider an artificial viscosity given by

$$\varepsilon_{klm}(\mathbf{u}_h)|_\kappa = C_\varepsilon \, \delta_{kl} \tilde{h}_k^2 \, f_p(\mathbf{u}_h) \, |\mathcal{R}_p(\mathbf{u}_h)|, \quad k, l = 1, 2, 3\,, \; m = 1, \dots, 5, \tag{305}$$

and $\varepsilon_{klm}(\mathbf{u}_h) = 0$ for $m = 6, 7$, where $\delta_{kl}$ is the Kronecker symbol. Here, the scalar residual term $\mathcal{R}_p(\mathbf{u}_h)$ is

$$\mathcal{R}_p(\mathbf{u}_h) = \sum_{m=1}^{5} \frac{\partial p}{\partial u_m} \frac{R_m(\mathbf{u}_h)}{p},$$

where $\mathbf{R}(\mathbf{u}_h) = \{R_m(\mathbf{u}_h)\}_m$ denotes the element residual in strong form as defined in (301), and $\frac{\partial p}{\partial u_m}$ is the partial derivative of the pressure with respect to conservative variables

$u_m$. Furthermore, the directional element size is given by $\tilde{h}_k = h_k/(\mathrm{p}+1)$, $k = 1, 2, 3$, where $h_k$ is the dimension of the element $\kappa$ in the $x_k$-coordinate direction given by $h_k = s_k(V/\Pi_i s_i)^{1/3}$ with $s_k = 2V/\int_{\partial\kappa} |n_k|\, ds$ and the volume $V = \int_\kappa 1\, ds$ of the element $\kappa$, and p is the polynomial degree of $\mathbf{u}_h \in \mathbf{V}_{h,\mathrm{p}}$. Note, that $h_k$ accounts for the anisotropy of the element $\kappa$. Furthermore, as $\tilde{h}_k$ takes into account the resolution of the discrete solution $\mathbf{u}_h \in \mathbf{V}_{h,\mathrm{p}}$ with respect to the mesh size and the polynomial degree, it would be suitable for *hp*-adapted discretizations. In addition to that, the smoothness switch (or pressure sensor according to Bassi et al. (2010, 2011)),

$$f_p(\mathbf{u}_h) = \frac{\tilde{h}_\kappa |\nabla p|}{p + \varepsilon'}, \quad \varepsilon' = 10^{-12},$$

is small in smooth parts and large in non-smooth parts of the flow solution, and decreases the effect of the artificial viscosity in smooth parts with steep gradients. Again, through the factor $\tilde{h}_\kappa$ it takes into account anisotropies of the elements as well as the resolution of an *hp*-adapted solution. Finally, $C_\varepsilon$ is a tunable parameter of the artificial viscosity which is chosen as $C_\varepsilon = 0.2$ in Section 9.7 (i.e., the same value as used in (Bassi et al., 2010)) as a compromise between stability and accuracy.

We recall that the element residual $\mathbf{R}(\mathbf{u}_h)$ vanishes for the exact solution of the governing equations, i.e., $\mathbf{R}(\mathbf{u}) = 0$, and thus $\varepsilon(\mathbf{u}) = 0$ holds for the exact solution $\mathbf{u}$ to (298). As a consequence, the discretization (300) together with the shock-capturing term (304) and the artificial viscosity constant (305) is *consistent*.

However, being based on the element terms of the artificial viscosity only, while ignoring the face terms usually present in a DG discretization of diffusive terms (cf. Section 6.3), this shock capturing scheme is *not* adjoint consistent. It is, however, *asymptotically adjoint consistent*, i.e., the adjoint residuals vanish as the mesh size $h \to 0$ and/or the polynomial degree $p \to \infty$ due to the dependency of the artificial viscosity on $h/p$.

Finally, note that the shock-capturing term (304) vanishes for $\mathbf{v}_h \in \mathbf{V}_h^0$. Thereby, Lemma 7.2 on the conservation properties of (300) still holds when (300) is augmented with the shock-capturing term (304).

# 8 Adjoint-based error estimation and adaptive mesh refinement

Important quantities in aerodynamic flow simulations are the aerodynamic force coefficients like the pressure induced as well as the viscous stress induced drag, lift and moment coefficients. In addition to the exact approximation of these quantities, it is of increasing importance, in particular in the field of uncertainty quantification, to estimate the error in the computed values.

While local mesh refinement is required for obtaining reasonably accurate results in applications, the goal of the adaptive refinement is either to compute the force coefficients as accurately as possible within given computing resources or to compute these quantities up to a given tolerance with the minimum computing resources required. In both cases a goal-oriented refinement is needed, i.e., an adaptive refinement strategy specifically targeted to the efficient computation of the quantities of interest. Furthermore, in the latter case, an estimate is required on how accurate the force coefficients are approximated, i.e., an *a posteriori* error estimate that quantifies the error on the numerical solution measured in terms of the quantity of interest.

In the following, we proceed along the lines of (Hartmann and Houston, 2010). In particular, in Section 8.1 we outline the approach of *a posteriori* error estimation and adjoint-based mesh refinement for single target quantities. Then, in Section 8.2 we generalize this approach to multiple target quantities. In Section 8.4 we derive residual-based indicators which are targeted at resolving all flow features.

## 8.1 Error estimation and mesh refinement for single target quantities

We begin by recalling the general approach of duality based *a posteriori* error estimation for *single* target functionals; see e.g. (Becker and Rannacher, 2001; Hartmann, 2002; Hartmann and Houston, 2002a) among many others. Furthermore, we give the standard algorithm, as described in e.g. (Becker and Rannacher, 1996; Hartmann and Houston, 2002a), of goal-oriented (adjoint-based) adaptive mesh refinement tailored to the accurate and efficient computation of a single target quantity.

Let us consider the nonlinear problem

$$N\mathbf{u} = 0 \quad \text{in } \Omega, \qquad\qquad B\mathbf{u} = 0 \quad \text{on } \Gamma, \tag{306}$$

where $\Omega \in \mathbb{R}^d$, $d > 1$, is an open bounded domain with boundary $\Gamma = \partial\Omega$. $N$ is a nonlinear differential operator and $B$ is a possibly nonlinear boundary operator on $\Gamma$. Let $N_h : \mathbf{V} \times \mathbf{V} \to \mathbb{R}$ be a semi-linear form, nonlinear in its first argument and linear in its second argument, such that the nonlinear problem (306) is discretized as follows: find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$N_h(\mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \tag{307}$$

Furthermore, let us assume that the discretization (307) is *consistent*, i.e., the analytical solution $\mathbf{u} \in \mathbf{V}$ satisfies the following equation:

$$N_h(\mathbf{u}, \mathbf{v}) = 0 \quad \forall \mathbf{v} \in \mathbf{V}. \tag{308}$$

Here, $\mathbf{V}$ is some suitably chosen function space including the analytical solution $\mathbf{u} \in \mathbf{V}$ to the primal problem (306) and satisfying $\mathbf{V}_h^p \subset \mathbf{V}$, where $\mathbf{V}_h^p$ is a discrete function space on the mesh $\mathcal{T}_h = \{\kappa\}$ consisting of elements $\kappa$ covering the computational domain $\Omega$; cf. (Arnold et al., 2002; Hartmann, 2007a) for the choice of $\mathbf{V}$ in the case of DG methods. Subtracting (308) from (307) we then obtain the Galerkin orthogonality

$$N_h(\mathbf{u}, \mathbf{v}_h) - N_h(\mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p. \tag{309}$$

Let $J(\cdot)$ be a nonlinear and differentiable target functional. Furthermore, let $J_h(\cdot)$ be a consistent discretization of this target functional, i.e., the analytical solution $\mathbf{u} \in \mathbf{V}$ satisfies the following equation:

$$J_h(\mathbf{u}) = J(\mathbf{u}). \tag{310}$$

We define the mean–value linearization of $J_h(\cdot)$ as follows

$$\bar{J}_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h) = J_h(\mathbf{u}) - J_h(\mathbf{u}_h) = \int_0^1 J_h'[\theta \mathbf{u} + (1-\theta)\mathbf{u}_h](\mathbf{u} - \mathbf{u}_h)\, \mathrm{d}\theta, \tag{311}$$

where $J_h'[\mathbf{w}](\cdot)$ denotes the Fréchet derivative of $J_h(\cdot)$ evaluated at some $\mathbf{w}$ in $\mathbf{V}$.

Analogously, for $\mathbf{v}$ in $\mathbf{V}$, we define the mean–value linearization of $N_h(\cdot, \mathbf{v})$

$$\begin{aligned} M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{v}) &= N_h(\mathbf{u}, \mathbf{v}) - N_h(\mathbf{u}_h, \mathbf{v}) \\ &= \int_0^1 N_h'[\theta \mathbf{u} + (1-\theta)\mathbf{u}_h](\mathbf{u} - \mathbf{u}_h, \mathbf{v})\, \mathrm{d}\theta. \end{aligned} \tag{312}$$

Here, $N_h'[\mathbf{w}](\cdot, \mathbf{v})$ denotes the Fréchet derivative of $\mathbf{u} \mapsto N_h(\mathbf{u}, \mathbf{v})$, for $\mathbf{v} \in \mathbf{V}$ fixed, at some $\mathbf{w}$ in $\mathbf{V}$. Let us now introduce the following adjoint problem: find $\mathbf{z} \in \mathbf{V}$ such that

$$M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{w}, \mathbf{z}) = \bar{J}_h(\mathbf{u}, \mathbf{u}_h; \mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V}. \tag{313}$$

Choosing $\mathbf{w} = \mathbf{u} - \mathbf{u}_h$ in (313) and recalling the linearization performed in (311) we get

$$\begin{aligned} J(\mathbf{u}) - J_h(\mathbf{u}_h) &= J_h(\mathbf{u}) - J_h(\mathbf{u}_h) = \bar{J}_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h) \\ &= M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{z}) = -N_h(\mathbf{u}_h, \mathbf{z}). \end{aligned}$$

Thereby, we have the following error representation formula

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) = R_h(\mathbf{u}_h, \mathbf{z}), \tag{314}$$

where $R_h(\mathbf{u}_h, \mathbf{z}) := -N_h(\mathbf{u}_h, \mathbf{z})$ includes the primal residuals multiplied by the adjoint solution $\mathbf{z}$. We note that the error representation formula (314) depends on the unknown analytical solution $\mathbf{z}$ to the adjoint problem (313) which in turn depends on the unknown analytical solution $\mathbf{u}$ to the primal problem (306). Thus, in order to render these quantities computable, both $\mathbf{u}$ and $\mathbf{z}$ must be replaced by suitable approximations. Here, the linearizations leading to $M_h(\mathbf{u}, \mathbf{u}_h; \cdot, \cdot)$ and $\bar{J}_h(\mathbf{u}, \mathbf{u}_h; \cdot)$ are performed about $\mathbf{u}_h$ and the adjoint solution $\mathbf{z}$ is replaced by the solution $\bar{\mathbf{z}}$ to the following linearized adjoint problem: find $\bar{\mathbf{z}} \in \mathbf{V}$ such that

$$N_h'[\mathbf{u}_h](\mathbf{w}, \bar{\mathbf{z}}) = J_h'[\mathbf{u}_h](\mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V}. \tag{315}$$

This is then approximated by the discrete adjoint problem: find $\bar{\mathbf{z}}_h \in \bar{\mathbf{V}}_h^p$ such that

$$N_h'[\mathbf{u}_h](\mathbf{w}_h, \bar{\mathbf{z}}_h) = J_h'[\mathbf{u}_h](\mathbf{w}_h) \quad \forall \mathbf{w}_h \in \bar{\mathbf{V}}_h^p. \tag{316}$$

Here, $\bar{\mathbf{V}}_h^p$ is an *adjoint* finite element space from which the approximate adjoint solution $\bar{\mathbf{z}}_h$ is sought. Rewriting the error representation (314) as follows

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) = R_h(\mathbf{u}_h, \mathbf{z}) = R_h(\mathbf{u}_h, \mathbf{z} - \bar{\mathbf{z}}) + R_h(\mathbf{u}_h, \bar{\mathbf{z}} - \bar{\mathbf{z}}_h) + R_h(\mathbf{u}_h, \bar{\mathbf{z}}_h), \tag{317}$$

we see that replacing the adjoint solution $\mathbf{z}$ in (314) by the solution $\bar{\mathbf{z}}_h$ to the discrete adjoint problem (316), we obtain the following approximate error representation

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) \approx R_h(\mathbf{u}_h, \bar{\mathbf{z}}_h). \tag{318}$$

This corresponds to ignoring in (317) the error $R_h(\mathbf{u}_h, \mathbf{z}-\bar{\mathbf{z}})$ due to the linearization of the adjoint problem and the error $R_h(\mathbf{u}_h, \bar{\mathbf{z}} - \bar{\mathbf{z}}_h)$ due to the approximation of the linearized adjoint problem. In fact, it can be shown (see e.g. Becker and Rannacher (2001)) that the linearization and the approximation errors of the adjoint problem are of higher order (quadratic) in the discretization error, $\mathbf{e} = \mathbf{u} - \mathbf{u}_h$, and may thus be neglected. In fact, in the series of publications (Hartmann and Houston, 2002a,b, 2006b), for example, among many others, it has been demonstrated that the approximate error representation in (318) is close to the true error in the target functional. Furthermore, the approximate error representation (320) can be used to enhance the computed target quantity $J(\mathbf{u}_h)$ by

$$\tilde{J}_h(\mathbf{u}_h) = J_h(\mathbf{u}_h) + \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h). \tag{319}$$

Finally, we note that (318) can be localized

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) \approx R_h(\mathbf{u}_h, \bar{\mathbf{z}}_h) \equiv \sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa, \tag{320}$$

where $|\bar{\eta}_\kappa|$ are local error indicators including the primal local residuals weighted with the discrete adjoint solution, denoted as adjoint-based indicators or as dual-weighted-residual (DWR) indicators (Becker and Rannacher, 2001).

**Remark 8.1** *We note that in Galerkin finite element methods the Galerkin orthogonality* (309) *can be used to subtract any discrete function* $\mathbf{z}_h \in \mathbf{V}_h^p$ *from the adjoint solution in the error estimate* (314) *resulting in* $J(\mathbf{u}) - J(\mathbf{u}_h) = -N_h(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h) = R_h(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h)$ *(see e.g. Becker and Rannacher (2001)). In order to avoid a vanishing error estimate* (318)*, the discrete adjoint solution* $\bar{\mathbf{z}}_h \in \bar{\mathbf{V}}_h^p$ *must be computed in a richer space (e.g.,* $\bar{\mathbf{V}}_h^p = \mathbf{V}_h^{\bar{p}} = \mathbf{V}_h^{p+1}$*) than the flow solution* $\mathbf{u}_h \in \mathbf{V}_h^p$*. In continuous finite element methods it is required to choose* $\mathbf{z}_h$ *to be an approximation of* $\mathbf{z}$ *in order to ensure that the local indicators in* (320) *are of the right order of convergence (Becker and Rannacher, 2001). For continuous finite element methods Galerkin orthogonality is a global property (i.e., it holds true on the whole domain). In addition to that discontinuous Galerkin methods satisfy a Galerkin orthogonality on a local, element-wise level. Therefore, for DG methods* $\mathbf{z}_h$ *can be omitted without changing the local indicators.*

The local indicators $\bar{\eta}_\kappa$ in (320) can be used to drive an adaptive refinement (and coarsening) algorithm specifically tailored to the accurate and efficient approximation of the target quantity $J(\mathbf{u})$. For example, suppose that the aim of the computation is to compute $J(\cdot)$ such that the error $|J(\mathbf{u}) - J_h(\mathbf{u}_h)|$ is less than some user–defined tolerance $\mathtt{TOL}$, i.e., $|J(\mathbf{u}) - J_h(\mathbf{u}_h)| \leq \mathtt{TOL}$, then in practice we may enforce the stopping criterion $|\sum_{\kappa\in\mathcal{T}_h} \bar{\eta}_\kappa| \leq \mathtt{TOL}$. If this condition is not satisfied on the current finite element mesh $\mathcal{T}_h$, then the local indicators $\eta_\kappa$ are employed as local error indicators to guide mesh refinement and coarsening. The cycle of the goal-oriented adaptive mesh refinement (Hartmann and Houston, 2002a) may be outlined as follows.

**Algorithm 8.1 (Single-target adaptive algorithm)** *Adaptive algorithm for the accurate and efficient approximation of a single target quantity* $J(\mathbf{u})$*:*

1. *Construct an initial mesh* $\mathcal{T}_h$.
2. *Compute* $\mathbf{u}_h \in \mathbf{V}_h^p$*, see* (307)*, on the current mesh* $\mathcal{T}_h$.
3. *Compute* $\bar{\mathbf{z}}_h \in \bar{\mathbf{V}}_h^p = \mathbf{V}_h^{p+1}$*, see* (316)*, on the same mesh employed for* $\mathbf{u}_h$.
4. *Evaluate the approximate error representation* $R_h(\mathbf{u}_h, \bar{\mathbf{z}}_h) = \sum_{\kappa\in\mathcal{T}_h} \bar{\eta}_\kappa$.
5. *If* $|\sum_{\kappa\in\mathcal{T}_h} \bar{\eta}_\kappa| \leq \mathtt{TOL}$*, where* $\mathtt{TOL}$ *is a given tolerance, then* `STOP`.
6. *Otherwise, refine and coarsen a fixed fraction of the total number of elements according to the size of* $|\bar{\eta}_\kappa|$ *and generate a new mesh* $\mathcal{T}_h$*;* `GOTO` *2.*

Again, in several publications, e.g. (Becker and Rannacher, 2001; Hartmann, 2002, 2006; Hartmann and Houston, 2002b; Leicht and Hartmann, 2010; Hartmann et al., 2011), the versatility of this adaptive algorithm has been demonstrated.

## 8.2 Error estimation for multiple target quantities

In the following we present an extension of this approach to the efficient and accurate computation of *multiple* target quantities. Given, $N$ target quantities we replace the computation of $N$ adjoint solutions as required in direct approaches by the solution of *two* auxiliary problems, namely one discrete adjoint problem and one discrete error equation where the latter can also be considered as the adjoint to the adjoint problem. In particular, the solution to the discrete error equation provides the *a posteriori* error estimation of arbitrarily many target quantities. Furthermore, the solution to the adjoint problem related to an appropriately defined combination of the original target functionals provides the adjoint-based refinement indicators required for goal-oriented refinement.

This approach has been developed and applied to the scalar inviscid Burgers equation considering point values in (Hartmann and Houston, 2003). It has later been extended to the treatment of laminar compressible flows considering multiple aerodynamic force coefficients in (Hartmann, 2008a).

### 8.2.1 The direct approach

Let us now consider the extension of the above analysis to the error estimation and goal-oriented mesh refinement for multiple target quantities. Given $N$ target functionals $J_i(\mathbf{u}), i = 1, \ldots, N$, with consistent discretizations $J_{i,h}(\mathbf{u}), i = 1, \ldots, N$, the direct approach for deriving an error representation formula analogous to (314) for each $J_i(\cdot)$ is to introduce the following $N$ adjoint problems: find $\mathbf{z}_i \in \mathbf{V}$ such that

$$M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{w}, \mathbf{z}_i) = \bar{J}_{i,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V}, \tag{321}$$

for $i = 1, \ldots, N$. Analogous to (314) we obtain the following error representation formulae

$$J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h) = M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{z}_i) = R_h(\mathbf{u}_h, \mathbf{z}_i), \tag{322}$$

for each $J_i(\cdot)$, $i = 1, \ldots, N$. In practice, the adjoint solutions $\mathbf{z}_i$, $i = 1, \ldots, N$, are unknown analytically and must be approximated numerically. After linearization and approximation we have: find $\bar{\mathbf{z}}_{i,h} \in \bar{\mathbf{V}}_h^p$ such that

$$N_h'[\mathbf{u}_h](\mathbf{w}_h, \bar{\mathbf{z}}_{i,h}) = J_{i,h}'[\mathbf{u}_h](\mathbf{w}_h) \quad \forall \mathbf{w}_h \in \bar{\mathbf{V}}_h^p; \tag{323}$$

this amounts to solving $N$ systems of linear equations with the same matrix but $N$ different right–hand side vectors. Based on the discrete adjoint solutions $\bar{\mathbf{z}}_{i,h}$, $i = 1, \ldots, N$, the following approximate error representation formulae and local error indicators can be evaluated

$$J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h) \approx R_h(\mathbf{u}_h, \bar{\mathbf{z}}_{i,h}) = \sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa^{(i)}, \tag{324}$$

for $i = 1, \ldots, N$.

### 8.2.2 A new approach

In view of the error representation formula (322) an alternative approach consists of considering the following error equation: find $\mathbf{e} \in \mathbf{V}$ such that

$$M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{e}, \mathbf{w}) = R_h(\mathbf{u}_h, \mathbf{w}) \qquad \forall \mathbf{w} \in \mathbf{V}, \tag{325}$$

whose solution is simply the discretization error $\mathbf{e} = \mathbf{u} - \mathbf{u}_h$. We remark that in the context of duality, (325) may be thought of as the adjoint of the adjoint problem and (322) the adjoint/adjoint-adjoint equivalence relating (321) to (325). Again after linearization, we obtain the following discrete error equation: find $\bar{\mathbf{e}}_h \in \bar{\mathbf{V}}_h^p$ such that

$$N_h'[\mathbf{u}_h](\bar{\mathbf{e}}_h, \mathbf{w}_h) = R_h(\mathbf{u}_h, \mathbf{w}_h) \quad \forall \mathbf{w}_h \in \bar{\mathbf{V}}_h^p. \tag{326}$$

Thereby, in practice, instead of solving $N$ discrete adjoint problems, cf. (323), for $\bar{\mathbf{z}}_{i,h} \in \bar{\mathbf{V}}_h^p$ with data $J_{i,h}'[\mathbf{u}_h](\cdot)$ and then evaluating $R_h(\mathbf{u}_h, \bar{\mathbf{z}}_{i,h})$ to determine the size of the error in the target functional $J_i(\cdot)$, $i = 1, \ldots, N$, one can simply solve the discrete error equation (325) for the approximate error $\bar{\mathbf{e}}_h \in \bar{\mathbf{V}}_h^p$ and evaluate

$$J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h) = \bar{J}_{i,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{e}) \approx J_{i,h}'[\mathbf{u}_h](\mathbf{e}) \approx J_{i,h}'[\mathbf{u}_h](\bar{\mathbf{e}}_h), \tag{327}$$

as an approximation to $J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)$, for $i = 1, \ldots, N$. Again, similar to (319), these error estimates can be be used to enhance the computed target quantities $J_{i,h}(\mathbf{u}_h)$ as follows

$$\tilde{J}_{i,h}(\mathbf{u}_h) = J_{i,h}(\mathbf{u}_h) + J'_{i,h}[\mathbf{u}_h](\bar{\mathbf{e}}_h). \tag{328}$$

When $N > 2$ this approach is clearly much more computationally efficient than the direct method. However, a disadvantage of this second approach is that while solving the discrete error equation (326) for $\bar{\mathbf{e}}_h$ gives information concerning the size of the error in the computed target functionals $J_i(\cdot)$, $i = 1, \ldots, N$, it does not provide the necessary local information on each element in the computational mesh to guide adaptive mesh refinement when the desired level of accuracy has not been achieved on the current mesh. On the other hand, computing the solution $\mathbf{z}_{i,h}$, $i = 1, \ldots, N$, to the $N$ discrete adjoint problems (323), the approximate error representation formulae in (324) provide not only information concerning the size of the error in the computed target functionals, but also local error indicators $|\bar{\eta}_\kappa^{(i)}|$ which can be employed for adaptive mesh design.

## 8.3 Adaptive refinement for multiple target quantities

In this section we propose a strategy based on solving only two auxiliary problems (the discrete error equation (326) and an adjoint problem subject to appropriate data which stems from a specific combined target functional (cf. (333) below) which provide all the necessary information needed to both estimate the size of the error in the computed target functionals, as well as provide local error indicators that can be used to drive an adaptive mesh refinement algorithm.

Given $N$ different target functionals $J_i(\cdot)$ with discretizations $J_{i,h}(\cdot)$, $i = 1, \ldots, N$, $N > 1$, we would like to compute each $J_{i,h}(\mathbf{u}_h)$ to within a given user–defined tolerance $\mathtt{TOL_i}$, $i = 1, \ldots, N$, respectively. More precisely, we consider the following problem: find $J_{i,h}(\mathbf{u}_h) \in \mathbb{R}$, $i = 1, \ldots, N$, such that

$$|J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)| \leq \mathtt{TOL_i}\,, \quad \text{for } i = 1, \ldots, N. \tag{329}$$

However, as we want to define a combined target quantity $J_c(\cdot)$ including all original target quantities $J_i(\cdot)$, $i = 1, \ldots, N$, we weaken the requirement (329), and simply insist that the sum of the *relative* errors in each of the target functionals $J_i(\cdot)$, $i = 1, \ldots, N$, is less than $\mathtt{TOL}$. In practice, since $J_i(\mathbf{u})$, $i = 1, \ldots, N$, is unknown, we approximate the sum of the relative errors by

$$\sum_{i=1}^{N} |J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)|/|J_{i,h}(\mathbf{u}_h)|, \tag{330}$$

see (Hartmann and Houston, 2003), assuming that $J_{i,h}(\mathbf{u}_h) \neq 0$, for $i = 1, \ldots, N$. As an alternative choice we might insist that the (weighted) sum of *absolute* errors in each of the target functionals $J_i(\cdot)$, $i = 1, \ldots, N$, is less than $\mathtt{TOL}$, i.e., considering

$$\sum_{i=1}^{N} \alpha_i |J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)|. \tag{331}$$

where $\alpha_i > 0$, $i = 1, \ldots, N$. Here, choosing $\alpha_i = 1$, $i = 1, \ldots, N$, represents the special case of considering the (unweighted) sum of absolute errors.

Let us begin by assuming that the sign of the error in each target functional $J_i(\cdot)$, $i = 1, \dots, N$, is known. For example, in some applications it may be known from either theoretical considerations or numerical experimentation that under mesh refinement the computed quantity of interest $J_{i,h}(\mathbf{u}_h)$ is always either smaller or greater than the exact value $J_i(\mathbf{u})$, for $i = 1, \dots, N$. This includes the special case of monotonically convergent target quantities. For the case that under mesh refinement the quantity $J_h(\mathbf{u}_h)$ converges to $J(\mathbf{u})$ from above, for example, then the error $J(\mathbf{u}) - J_h(\mathbf{u}_h)$ is always negative; analogously, when it converges from below the error is always positive.

Employing this *a priori* knowledge concerning the convergence of the target functionals, we introduce a combined target functional $J_c(\cdot)$ and its discretization $J_{c,h}(\cdot)$,

$$J_c(\mathbf{v}) = \sum_{i=1}^{N} \omega_i J_i(\mathbf{v}), \qquad J_{c,h}(\mathbf{v}) = \sum_{i=1}^{N} \omega_i J_{i,h}(\mathbf{v}), \tag{332}$$

where $\omega_i = s_i/|J_{i,h}(\mathbf{u}_h)|$ or $\omega_i = \alpha_i s_i$, depending on whether the relative and weighted absolute errors (330) and (331), respectively, are considered. Here, $s_i$ denotes the expected signs of the errors $J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)$, $i = 1, \dots, N$, respectively. Thereby, we may now proceed as in Section 8.1 to derive an error representation formula for the error in the combined target functional $J_c(\cdot)$. To this end, we introduce the following adjoint problem: find $\mathbf{z}_c \in \mathbf{V}$ such that

$$M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{w}, \mathbf{z}_c) = \bar{J}_{c,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{w}) \qquad \forall \mathbf{w} \in \mathbf{V}, \tag{333}$$

where $\bar{J}_{c,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{w}) = \sum_{i=1}^{N} \omega_i \bar{J}_{i,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{w})$ is the mean value linearization to $J_{c,h}$ analogous to (311). Thus, we now deduce the following error representation formula

$$\begin{aligned} J_c(\mathbf{u}) - J_{c,h}(\mathbf{u}_h) &= \sum_{i=1}^{N} \omega_i (J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h)) = \sum_{i=1}^{N} \omega_i \bar{J}_{i,h}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h) \\ &= M_h(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{z}_c) = R_h(\mathbf{u}_h, \mathbf{z}_c). \end{aligned} \tag{334}$$

In general, the signs $s_i$, $i = 1, \dots, N$, will not be known *a priori*. Thereby, we must first solve the discrete error equation (326) for $\bar{\mathbf{e}}_h$ and evaluate $\bar{s}_i = \operatorname{sgn}(J'_{i,h}[\mathbf{u}_h](\bar{\mathbf{e}}_h))$, $i = 1, \dots, N$. Then, the adjoint problem (333) may be solved computationally using the predicted values of $s_i$, $i = 1, \dots, N$, in $J_c(\cdot)$: find $\bar{\mathbf{z}}_{c,h} \in \bar{\mathbf{V}}_h^p$ such that

$$N'_h[\mathbf{u}_h](\mathbf{w}_h, \bar{\mathbf{z}}_{c,h}) = J'_{c,h}[\mathbf{u}_h](\mathbf{w}_h) \quad \forall \mathbf{w}_h \in \bar{\mathbf{V}}_h^p. \tag{335}$$

Then the approximate error representation formula can be evaluated as follows

$$J_c(\mathbf{u}) - J_c(\mathbf{u}_h) = R_h(\mathbf{u}_h, \mathbf{z}_c) \approx R_h(\mathbf{u}_h, \bar{\mathbf{z}}_{c,h}) \equiv \sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa. \tag{336}$$

This now provides both global information concerning the size of the error in the combined target functional $J_c(\cdot)$, as well as local information necessary for adaptive mesh refinement. Thus, the cycle of the adaptive algorithm can be outlined as follows.

**Algorithm 8.2 (Multi-target adaptive algorithm)** *Adaptive algorithm for the accurate and efficient approximation of multiple target quantities $J_i(\mathbf{u})$, $i = 1, \dots, N$:*

1. *Construct an initial mesh* $\mathcal{T}_h$.
2. *Compute* $\mathbf{u}_h \in \mathbf{V}_h^p$, *see* (307), *on the current mesh* $\mathcal{T}_h$.
3. *Compute* $\bar{\mathbf{e}}_h \in \bar{\mathbf{V}}_h^p = \mathbf{V}_h^{p+1}$, *see* (326), *on the same mesh employed for* $\mathbf{u}_h$.
4. *Evaluate* $J_i(\mathbf{u}) - J_{i,h}(\mathbf{u}_h) \approx J'_{i,h}[\mathbf{u}_h](\bar{\mathbf{e}}_h) =: \psi_i$, $i = 1, \ldots, N$.
5. *If* $|\psi_i| \leq \mathtt{TOL_i}$ *for all* $i = 1, \ldots, N$, *then* `STOP`.
6. *Build the target quantity* $J_{c,h}$ *based on* $\bar{s}_i = sgn(\psi_i)$, $i = 1, \ldots, N$.
7. *Compute* $\bar{\mathbf{z}}_{c,h} \in \bar{\mathbf{V}}_h^p = \mathbf{V}_h^{p+1}$, *see* (335), *on the same mesh employed for* $\mathbf{u}_h$.
8. *Evaluate the approximate error representation* $\sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa$, *see* (336).
9. *If* $|\sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa| \leq \mathtt{TOL}$, *where* `TOL` *is a given tolerance, then* `STOP`.
10. *Otherwise, refine and coarsen a fixed fraction of the total number of elements according to the size of* $|\bar{\eta}_\kappa|$ *and generate a new mesh* $\mathcal{T}_h$; `GOTO` *2.*

Here, the stopping criterion in line (5) of Algorithm 8.2 corresponds to enforcing (329); on the other hand, the stopping criterion in line (9) corresponds to enforcing either equation (330) or (331) to be less than `TOL`, depending on the choice of weights in the combined target functional. This approach leads to the solution of only two auxiliary problems, in comparison to the $N$ required for the direct approach.

We note that this approach has first been developed for and applied to the DG discretization of the inviscid 1d Burgers equation in (Hartmann and Houston, 2003) considering the sum of relative errors of point values of the solution. Then in (Hartmann and Houston, 2008) it has been applied to the SIPG discretization of the 2d compressible Navier-Stokes equations considering sums of relative and absolute errors of aerodynamic force coefficients including pressure induced and viscous drag, lift and moment coefficients. The application to 3d laminar and turbulent flow problems is shown in Sections 9.4 and 9.5, taken from (Hartmann et al., 2010b) and (Hartmann et al., 2011), respectively.

## 8.4 Derivation of residual-based indicators

Provided the adjoint solution related to an arbitrary target functional is sufficiently smooth the corresponding error representation can be bounded from above by an error estimate which includes the primal residuals but is independent of the adjoint solution. By localizing this error estimate so-called residual-based indicators can be derived. Mesh refinement based on these indicators leads to meshes which resolve *all* flow features irrespective of any specific target quantity. We recall the derivation of these indicators from (Hartmann and Houston, 2002a, 2006b; Hartmann, 2008a). Furthermore, we note that the residual-based indicators have been extended to include symmetry boundary conditions in (Leicht and Hartmann, 2010).

Let $\mathbf{u}$ and $\mathbf{u}_h$ denote the solutions to (306) and (307), respectively. We now recall the error representation formula in (314), and apply Galerkin orthogonality (309),

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) = -N_h(\mathbf{u}_h, \mathbf{z}) = -N_h(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h) = R_h(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h), \tag{337}$$

for any $\mathbf{z}_h \in \mathbf{V}_h^p$. In particular, we can choose $\mathbf{z}_h := \Pi_h \mathbf{z} \in \mathbf{V}_h^p$ in (337), i.e.,

$$J(\mathbf{u}) - J_h(\mathbf{u}_h) = R_h(\mathbf{u}_h, \mathbf{z} - \Pi_h \mathbf{z}), \tag{338}$$

where $\Pi_h \mathbf{z}$ denotes an appropriate interpolation/projection of $\mathbf{z}$ into the discrete function space $\mathbf{V}_h^p$. Indeed, here we select $\Pi_h$ so that the following approximation property holds: given $\kappa \in \mathcal{T}_h$, suppose that $\mathbf{z}|_\kappa$ in $[H^{s_\kappa+1}(\kappa)]^5$, $0 \le s_\kappa \le p$. Then, there exists a constant $C$ dependent on $s_\kappa$, $p$, and the shape regularity of $\mathcal{T}_h$, but is independent of the local mesh size $h_\kappa$, such that for $0 \le m \le s_\kappa + 1$,

$$\|\mathbf{z} - \Pi_h \mathbf{z}\|_{H^m(\kappa)} \le C h_\kappa^{s_\kappa+1-m} \|\mathbf{z}\|_{H^{s_\kappa+1}(\kappa)}. \tag{339}$$

Then, by employing the trace theorem, we have

$$\begin{aligned} \|\mathbf{z} - \Pi_h \mathbf{z}\|_{L^2(\partial\kappa)} &\le C h_\kappa^{s_\kappa+1/2} \|\mathbf{z}\|_{H^{s_\kappa+1}(\kappa)}, & 0 \le s_\kappa \le p, \\ \|\mathbf{z} - \Pi_h \mathbf{z}\|_{H^1(\partial\kappa)} &\le C h_\kappa^{s_\kappa-1/2} \|\mathbf{z}\|_{H^{s_\kappa+1}(\kappa)}, & 1 \le s_\kappa \le p; \end{aligned} \tag{340}$$

cf. (Babuška and Suri, 1987), for example. Following the derivation in Sections 5.4 and 6.4 we rewrite (338) as follows

$$\begin{aligned} J(\mathbf{u}) - J(\mathbf{u}_h) \;=\; & \int_\Omega \mathbf{R}(\mathbf{u}_h) \cdot (\mathbf{z} - \Pi_h \mathbf{z}) \, \mathrm{d}\mathbf{x} \\ & + \sum_{\kappa \in \mathcal{T}_h} \int_{\partial\kappa \setminus \Gamma} \mathbf{r}(\mathbf{u}_h) \cdot (\mathbf{z} - \Pi_h \mathbf{z})^+ + \underline{\rho}(\mathbf{u}_h) : \nabla (\mathbf{z} - \Pi_h \mathbf{z})^+ \, \mathrm{d}s \\ & + \int_\Gamma \mathbf{r}_\Gamma(\mathbf{u}_h) \cdot (\mathbf{z} - \Pi_h \mathbf{z})^+ + \underline{\rho}_\Gamma(\mathbf{u}_h) : \nabla (\mathbf{z} - \Pi_h \mathbf{z})^+ \, \mathrm{d}s, \end{aligned} \tag{341}$$

where the primal element residuals $\mathbf{R}(\mathbf{u}_h)$, the interior face residuals $\mathbf{r}(\mathbf{u}_h)$ and $\underline{\rho}(\mathbf{u}_h)$, and the boundary residuals $\mathbf{r}_\Gamma(\mathbf{u}_h)$ and $\underline{\rho}_\Gamma(\mathbf{u}_h)$ are given for the compressible Euler and the diffusive part of the Navier-Stokes equations in Sections 5.4 and 6.4, respectively.

Assuming $\mathbf{z}|_\kappa \in [H^{s_\kappa+1}(\kappa)]^5$, $1 \le s_\kappa \le p$, for each $\kappa \in \mathcal{T}_h$, and applying Cauchy-Schwarz inequality and the approximation estimates (339) and (340) in (341) we obtain

$$|J(\mathbf{u}) - J(\mathbf{u}_h)| \le C \left( \sum_{\kappa \in \mathcal{T}_h} \left(\eta_\kappa^{(\mathrm{res})}\right)^2 \right)^{1/2}, \tag{342}$$

where $\eta_\kappa^{(\mathrm{res})}$ is given by

$$\eta_\kappa^{(\mathrm{res})} = h_\kappa^{s_\kappa+1} \|\mathbf{R}(\mathbf{u}_h)\|_{L^2(\kappa)} + h_\kappa^{s_\kappa+1/2} \|\mathbf{r}_{\partial\kappa}(\mathbf{u}_h)\|_{L^2(\partial\kappa)} + h_\kappa^{s_\kappa-1/2} \|\underline{\rho}_{\partial\kappa}(\mathbf{u}_h)\|_{L^2(\partial\kappa)}. \tag{343}$$

Here, we use the short notation $\mathbf{r}_{\partial\kappa} = \mathbf{r}$ on $\partial\kappa \setminus \Gamma$ and $\mathbf{r}_{\partial\kappa} = \mathbf{r}_\Gamma$ on $\Gamma$, i.e.,

$$\|\mathbf{r}_{\partial\kappa}(\mathbf{u}_h)\|^2_{L^2(\partial\kappa)} = \|\mathbf{r}(\mathbf{u}_h)\|^2_{L^2(\partial\kappa\setminus\Gamma)} + \|\mathbf{r}_\Gamma(\mathbf{u}_h)\|^2_{L^2(\Gamma)},$$

and analogously for $\underline{\rho}_{\partial\kappa}$, i.e.,

$$\|\underline{\rho}_{\partial\kappa}(\mathbf{u}_h)\|^2_{L^2(\partial\kappa)} = \|\underline{\rho}(\mathbf{u}_h)\|^2_{L^2(\partial\kappa\setminus\Gamma)} + \|\underline{\rho}_\Gamma(\mathbf{u}_h)\|^2_{L^2(\Gamma)}.$$

We point out that the *a posteriori* error bound (342) places severe regularity constraints on the adjoint solution $\mathbf{z}$, which are typically not fulfilled in practice. On the basis of numerical experimentation, and stimulated by the estimate (342), we employ following so-called *residual-based indicators*

$$\eta_\kappa^{\text{res}} = h_\kappa \|\mathbf{R}(\mathbf{u}_h)\|_{L^2(\kappa)} + h_\kappa^{1/2} \|\mathbf{r}_{\partial\kappa}(\mathbf{u}_h)\|_{L^2(\partial\kappa)} + h_\kappa^{-1/2} \|\underline{\rho}_{\partial\kappa}(\mathbf{u}_h)\|_{L^2(\partial\kappa)}, \qquad (344)$$

in subsequent numerical examples.

# 9 Numerical results

In this section we demonstrate the performance of the adjoint-based error estimation, the adjoint-based (goal-oriented) mesh refinement and the residual-based mesh refinement for a range of aerodynamic test cases. The computations have been performed with the DLR-PADGE code (Hartmann et al., 2010a) which is based on a modified version of the `deal.II` library (Bangerth et al., 2007). After starting with a model problem based on the linear advection equation we consider a number of aerodynamic test cases of increasing complexity. In particular, we consider following test cases:

- Discontinuities advected along curved vector field (Hartmann, 2001);
- Supersonic inviscid flow past the BAC3-11 airfoil (Hartmann, 2002);
- Supersonic viscous flow past the NACA0012 airfoil (Hartmann, 2006);
- Laminar flow around delta wing (Hartmann et al., 2010b);
- Turbulent flow around streamlined body (Hartmann et al., 2011);
- Turbulent flow around the DLR-F6 wing-body (Hartmann et al., 2011);
- Turbulent flow around the VFE-2 delta wing (Hartmann, 2012, 2013b).

## 9.1 Linear advection equation

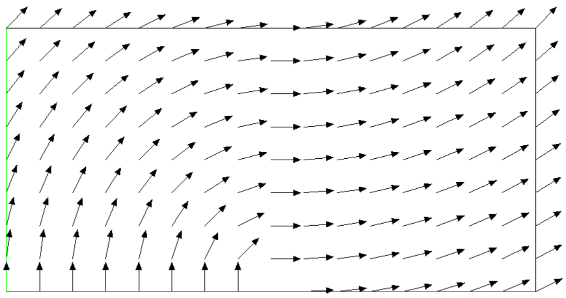
As first numerical example, taken from (Hartmann, 2001, 2002), we consider the linear advection equation (45) on $\Omega = [0,2] \times [0,1] \in \mathbb{R}^2$ with a vector field $\mathbf{b}$ as shown in Figure 17.1, $c = 0$ and $f = 0$. For this problem and the prescribed boundary values on the inflow boundary ($u(x,0) = 1$ for $\frac{1}{8} < x < \frac{3}{4}$ and zero boundary values elsewhere) the solution is shown in Figure 17.3. Here, the two jumps of the discontinuous boundary function are transported along the characteristic directions given by the vector field.

Assume that we are interested in the values of the solution on the part $\frac{1}{4} < y < 1$ of the right outflow boundary $\Gamma_+ = \Gamma \setminus \Gamma_-$. Let us consider, for example, the target quantity $J(u) = \int_{\Gamma_+} j_\Gamma \, u \, ds$, where $j_\Gamma$ is chosen to be following smooth function

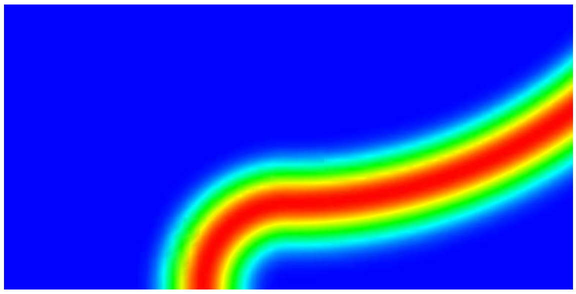
$$j_\Gamma(2,y) = \exp\left((\tfrac{3}{8})^{-2} - ((y-\tfrac{5}{8})^2 - \tfrac{3}{8})^{-2}\right) \quad \text{for } \tfrac{1}{4} < y < 1,$$

and 0 elsewhere. Then the corresponding adjoint solution (cf. Figure 17.2) is also smooth. Note, that the boundary value distribution $j_\Gamma$ of the adjoint solution is transported along the characteristics of the flow field $\mathbf{b}$ but in *opposite* direction.

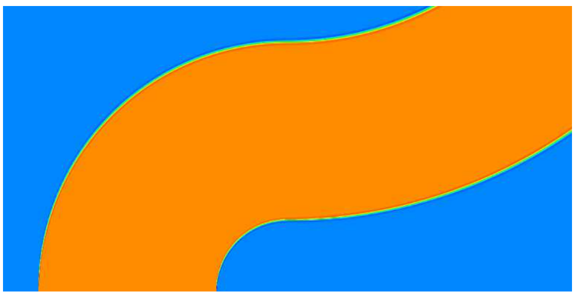
Figure 17.4 shows the numerical solution on the adaptively refined mesh, see Figure 17.6, which has been refined using the adjoint-based indicators. Note, that the refinement takes place at the position of only one of the discontinuities present in $u$. Indeed, the second discontinuity is not resolved at all, as it is outside of the support of the adjoint solution, and hence does not belong to the domain of influence of the target quantity. Thereby, the residuals in the neighborhood of this discontinuity do *not* contribute to the error in the target quantity. Comparing the two meshes, Figure 17.6 and Figure 17.5, it is obvious that the mesh in Figure 17.6 is more cost-efficient for evaluating the value of the target quantity than the mesh refined with residual-based indicators which do not include the adjoint solution.



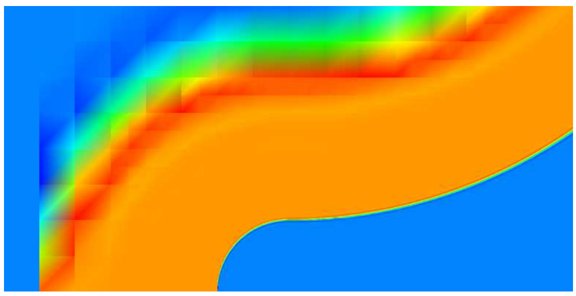
17.1: vector field $\mathbf{b}$



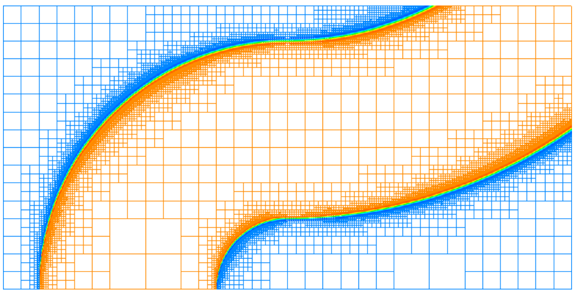
17.2: adjoint solution conntect to the target functional $\int_{\Gamma_+} j_\Gamma \, u \, ds$ with smooth $j_\Gamma$



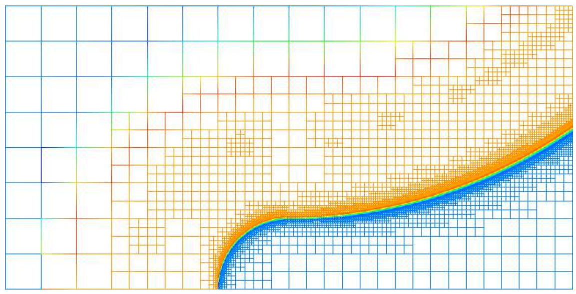
17.3: primal solution on mesh in Fig. 17.5



17.4: primal solution on mesh in Fig. 17.6



17.5: residual-based refined mesh



17.6: adjoint-based refined mesh

Figure 17: Linear advection equation: Comparison of residual-based and adjoint-based mesh refinement (Hartmann, 2001, 2002).

## 9.2 Supersonic inviscid flow past the BAC3-11 airfoil

In this example, taken from (Hartmann, 2002; Hartmann and Houston, 2002b), we study a supersonic flow around a BAC3-11 airfoil; this unsymmetric airfoil, see Figure 18, was originally specified in the AGARD Report AR-303 (1994). Here, we consider an inviscid flow at Mach number $M = 1.2$ and an angle of attack $\alpha = 5°$.

The solution to this problem includes two shocks: one located in front of the leading edge of the airfoil and one originating from the trailing edge; see Figure 20(b) which shows a mesh that is refined at the position of the two shocks. Here, Figure 19(a) shows the

Mach 1 isolines of the solution; the Mach $M = 1$ isoline to the left of the airfoil indicates the position of the first shock. The $M = 1$ isolines that originate from the upper and lower surfaces of the airfoil represent the transonic lines of the flow. The flow left of the first shock is supersonic; it is simply the $M = 1.2$ flow prescribed on the inflow boundary of the computational domain. The flow in between the shock and the transonic lines is subsonic; we note that the leading edge of the airfoil is located within this subsonic part of the flow. Finally, the flow behind the transonic lines is supersonic again.

In this example we take the target quantity to be the value of the pressure at the leading edge, i.e.,

$$J(u) = p(0,0),$$

cf. Figure 18. A computation on a fine mesh gives a reference value of $J(u) = 2.393$.

The structure of the solution $\bar{\mathbf{z}}_h$ to the discrete adjoint problem (316) corresponding to this point evaluation is displayed in Figure 19(b). This figure illustrates some principles of information transport in supersonic as well as in subsonic flow regions. To the right–hand side of the transonic lines the adjoint solution is zero as no information, neither by material transport nor even by information transport due to sound waves, can enter the subsonic region from the supersonic one. Within the whole subsonic region the adjoint solution is non-zero corresponding to the fact that sound waves can reach the point of evaluation from any point in the subsonic area and that all numerical errors which occur within this subsonic region can (even though possibly to a small portion) affect the value of the solution at the point of evaluation. However, the adjoint solution in the subsonic region is concentrated in a thin spike that is transported upstream from the point of evaluation in direction of the flow. This spike corresponds to the path of material transport and represents the main path of information transport. To the left of the airfoil, this spike crosses the shock and splits into three spikes while entering the supersonic region left of the shock. These spikes are transported upstream along the characteristics corresponding to the three eigenvalues $v$ and $v \pm c$. We recall that the characteristic corresponding to $v$ represents the path of material transport, that in this example is given by the line inclined at 5 degrees, whereas the characteristics corresponding to $v \pm c$ represent the paths of information transport due to sound waves.

In Figure 20 we show the meshes produced using the adjoint-based and the residual-based error indicators. Here, we see that the mesh constructed using residual-based indicators is concentrated in the neighborhood of the two shocks. In contrast, the mesh produced using the adjoint-based indicators only refines the mesh in the vicinity of the point of evaluation and the part of the shock where the spike of the adjoint solution, i.e., where the main part of information, crosses the shock. The other parts of the shock are not resolved, as the numerical error in these regions only has a small affect on the

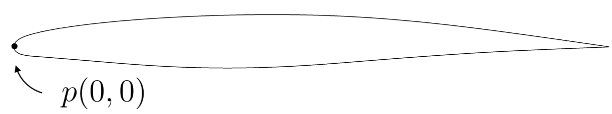



Figure 18: Profile of the BAC3-11 airfoil. Target quantity: pressure $p$ at leading edge (Hartmann, 2002; Hartmann and Houston, 2002b).

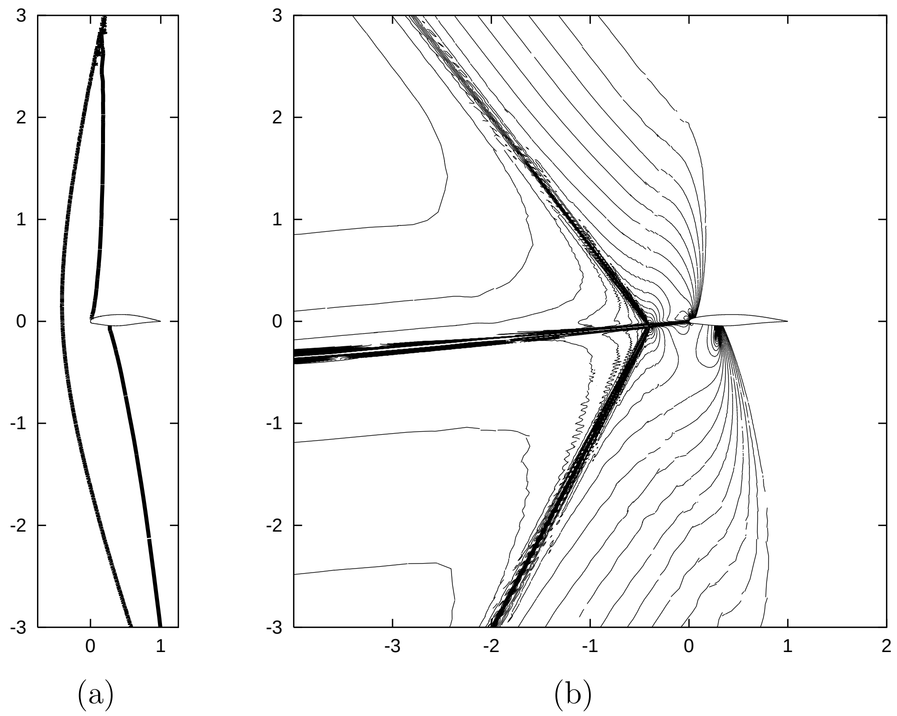


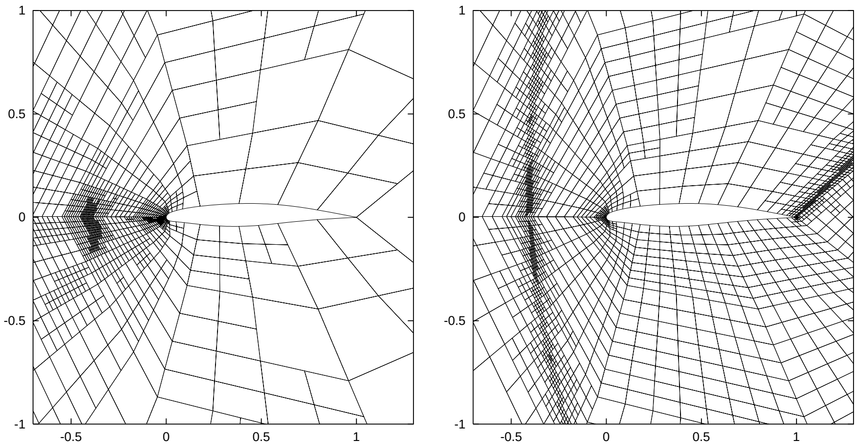

Figure 19: Supersonic BAC3-11 flow. (a) Mach 1 isolines of the flow solution; (b) $z_1$ isolines of adjoint solution (Hartmann, 2002; Hartmann and Houston, 2002b).

accuracy of the solution at the point of evaluation. Also there is no refinement in the vicinity of the shock emanating from the trailing edge of the airfoil; thereby, this shock is not well resolved at all. Nevertheless, the solution at the leading edge of the airfoil is not affected by this as no information is transported upstream from the trailing edge, located in a supersonic part of the flow, to the leading edge, located in the subsonic region. As in the previous example, we see that the adaptively refined meshes generated by employing the adjoint-based indicators are much more economical than those produced using the residual-based indicators. Indeed, in Figure 21 we clearly observe the superiority of the former error indicator; on the final mesh the true error in the computed functional is over two orders of magnitude smaller when the weighted error indicator is employed.

Motivated by the structure of the mesh generated by the adjoint-based error indicator, here we also consider the performance of an alternative *ad hoc* error indicator based on a modification of the residual indicator, whereby only elements in a neighborhood of a region upstream of the point of interest are marked for refinement. More precisely, we write $C$ to denote the cone depicted in Figure 22(a) with apex half angle $\beta$, located in the center of the airfoil with symmetry axes inclined at $\alpha = 5°$ according to the direction of the inflow. We now define the modified residual-based indicator $\eta_\kappa^{\mathrm{res,C}}$ as follows:

$$\eta_\kappa^{\mathrm{res,C}} = \begin{cases} \eta_\kappa^{\mathrm{res}}, & \text{if centroid}(\kappa) \in C, \\ 0, & \text{otherwise.} \end{cases}$$

This modification takes into account that we are not interested in the flow field in the whole domain, but only in the point value of the pressure at the leading edge. Thereby, adaptive mesh refinement is inhibited in the region downstream of the airfoil including the

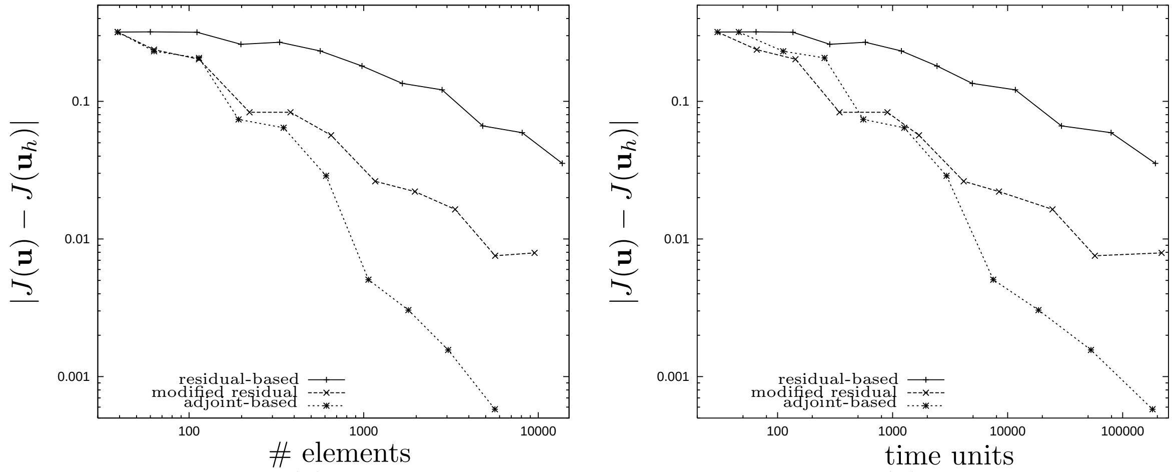



(a) (b)

Figure 20: Supersonic flow around BAC3-11 airfoil: (a) Adjoint-based refined mesh with 1803 elements and $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 3.042 \times 10^{-3}$; (b) Residual-based refined mesh with 13719 elements and $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 3.542 \times 10^{-2}$, taken from (Hartmann, 2002; Hartmann and Houston, 2002b).

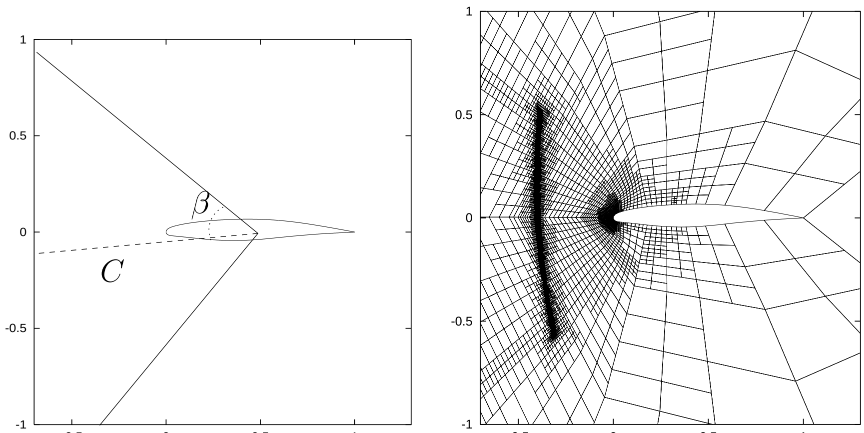
neighbourhood of the shock emanating from the trailing edge. Furthermore, refinement of the shock in front of the leading edge of the airfoil is prevented in regions that are placed too far above or below the airfoil since a low resolution of this shock in these areas is believed to not significantly degrade the accuracy of the pressure value at the leading edge, cf. Figure 20(a). In Figure 22(b) we show the mesh produced by employing $\eta_\kappa^{\mathrm{res,C}}$ with $\beta = 45°$.

Finally, in Figure 21 we see that the modified residual indicator produces meshes that are much more efficient for computing the value of the pressure at the leading edge of the airfoil in comparison to the (unmodified) residual-based indicator $\eta_\kappa^{\mathrm{res}}$. Nevertheless, the meshes produced using the adjoint-based indicators are even more efficient than those designed by $\eta_\kappa^{\mathrm{res,C}}$; on the final mesh the true error in the computed functional is an order of magnitude smaller when the adjoint-based error indicator is employed. We note that the chosen shape and size of the subdomain $C$ and the resulting modified indicator only represents an 'attempt' to find a reasonable modification of the residual indicator $\eta_\kappa^{\mathrm{res}}$ that is capable of efficiently computing the pressure at the leading edge of the airfoil and to provide a 'fair' comparison with the goal–oriented adjoint-based indicator $|\bar{\eta}_\kappa|$. Indeed, the value of the angle $\beta$ may be chosen differently, though *a priori* it is unclear which parts of the shock in front of the leading edge of the airfoil will influence the target functional. The angle $\beta$ should not be chosen too small as otherwise the lack of resolution of the shock in front of the leading edge of the airfoil will impact on the computed value of the pressure at the point of interest; on the other hand choosing $\beta$ too large may lead to over–refinement. In contrast, the adjoint-based indicator provides all the necessary information in order to decide which regions of the shock should be refined, and to what extent.




(a) (b)

Figure 21: Supersonic flow around BAC3-11. Target quantity $J(\mathbf{u})$: pressure at leading edge. Use of the residual-based, the modified residual-based (*ad hoc*) and the adjoint-based indicators. Convergence of $|J(\mathbf{u}) - J(\mathbf{u}_h)|$ vs. (a) number of elements and (b) time units (Hartmann, 2002).

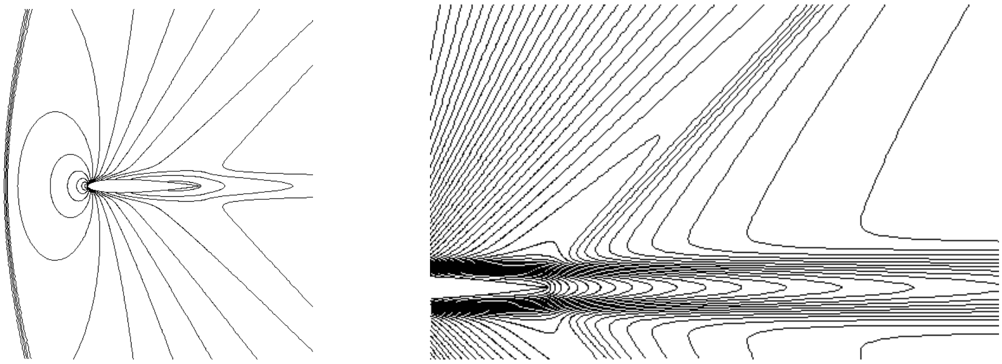



(a) (b)

Figure 22: Supersonic flow around BAC3-11. (a) Cone $C$: domain where the modified residual (*ad hoc*) indicator is active; (b) Mesh constructed using the modified residual (*ad hoc*) indicator with 9516 elements and $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 7.924 \times 10^{-3}$, (Hartmann, 2002; Hartmann and Houston, 2002b).

## 9.3 Supersonic viscous flow past the NACA0012 airfoil

In this example, taken from (Hartmann, 2006), we consider a symmetric viscous flow at $\alpha = 0°$, $M = 1.2$ and $\mathrm{Re} = 1000$, with an adiabatic no-slip boundary condition imposed on the airfoil, see Figure 23.1. Due to the only slightly supersonic Mach number, the bow shock is located at some distance in front of the airfoil. Furthermore, there are two weak shocks emanating from the trailing edge of the airfoil, see Figure 23.2.



23.1: density isolines

23.2: Zoom of density isolines at trailing edge

Figure 23: Supersonic viscous flow (Hartmann, 2006).

In the following, we consider the numerical approximation of the pressure induced, viscous and total drag coefficients, $J_{C_{\mathrm{dp}}}(\mathbf{u})$, $J_{C_{\mathrm{df}}}(\mathbf{u})$ and $J_{C_{\mathrm{d}}}(\mathbf{u})$, given by

$$J_{C_{\mathrm{dp}}}(\mathbf{u}) = \int_{\Gamma_W} p\,\mathbf{n}\cdot\boldsymbol{\psi}_d\,\mathrm{d}s, \quad J_{C_{\mathrm{df}}}(\mathbf{u}) = -\int_{\Gamma_W} \tau\mathbf{n}\cdot\boldsymbol{\psi}_d\,\mathrm{d}s, \quad J_{C_{\mathrm{d}}}(\mathbf{u}) = J_{C_{\mathrm{dp}}}(\mathbf{u}) + J_{C_{\mathrm{df}}}(\mathbf{u}),$$

where $\boldsymbol{\psi}_d$ is as defined in Section 5.2. Furthermore, we consider the corresponding lift coefficients, $J_{C_{\mathrm{lp}}}(\mathbf{u})$, $J_{C_{\mathrm{lp}}}(\mathbf{u})$ and $J_{C_{\mathrm{l}}}(\mathbf{u})$. Given that the flow is symmetric about the $x$-axis, both lift coefficients, $C_{\mathrm{lp}}$ and $C_{\mathrm{lp}}$, vanish. On the basis of fine grid computations the reference values of the pressure induced drag, $C_{\mathrm{dp}}$, and the viscous drag, $C_{\mathrm{df}}$, are given by $J_{C_{\mathrm{dp}}}(\mathbf{u}) \approx 0.10109$ and $J_{C_{\mathrm{df}}}(\mathbf{u}) \approx 0.10773$, respectively.

In the following, we consider the numerical approximation of the pressure induced force coefficient, $J(\mathbf{u}) = J_{C_{\mathrm{dp}}}(\mathbf{u})$, discretized as follows

$$J(\mathbf{u}_h) = \int_{\Gamma_W} p(\mathbf{u}_h)\,\mathbf{n}\cdot\boldsymbol{\psi}\,\mathrm{d}s. \tag{345}$$

In particular, we demonstrate that the approximate error representation $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h) = \sum_{\kappa\in\mathcal{T}_h} \bar{\eta}_\kappa$, cf. (320), which was derived from the (exact) error representation (314) by replacing the (exact) adjoint solution $\mathbf{z}$ by a computed adjoint solution $\bar{\mathbf{z}}_h$, gives a good approximation to the true error measured in terms of the target quantity $J(\mathbf{u})$ under consideration. Furthermore, as in previous examples we highlight the advantages of designing an adaptive finite element algorithm based on adjoint-based indicators (320) in comparison to residual-based indicators (344).

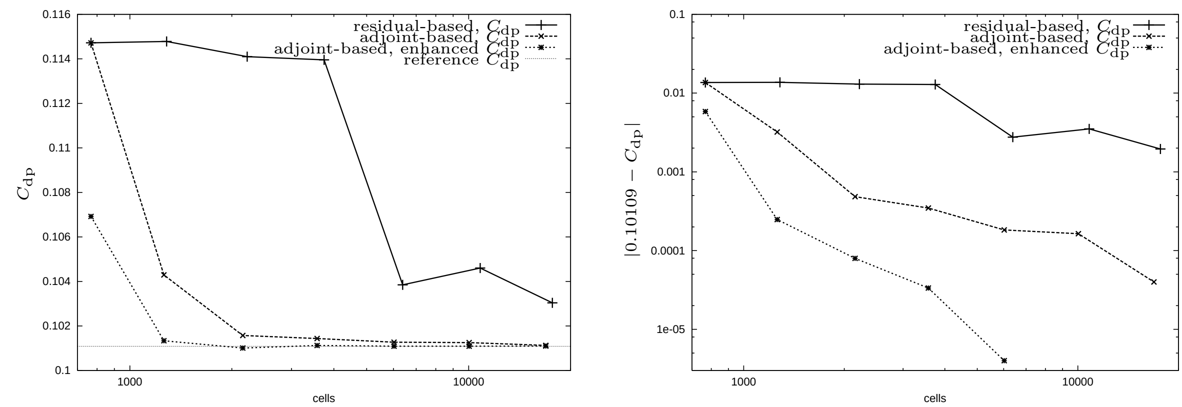
In Table 2, we collect the data of the adaptive algorithm based on employing the adjoint-based indicators. Here, we show the number of elements and degrees of freedom (DoF) for $p = 1$ (bilinear elements), the true error in the target quantity, $J(\mathbf{u}) - J(\mathbf{u}_h)$, the approximate error representation formula $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h - \mathbf{z}_h) := \sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa$ and the effectivity index $\theta = \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h - \mathbf{z}_h)/(J(\mathbf{u}) - J(\mathbf{u}_h))$ of the error estimation. First, we note that on all meshes the sign of the error is predicted correctly, which is always negative in this computation, i.e., the computed $C_{\rm dp}$ values converge to the reference value from above. Furthermore, from the second mesh onwards, the approximate error representation represents a very good approximation to the true errors, which is indicated by the effectivity indices $\theta$ being very close to one.

Note, that here the discretization of the considered target quantity $C_{\rm dp}$ is *adjoint inconsistent* due to following two reasons: First, the direct discretization of the target quantity as given in (345) is adjoint inconsistent. To see this, compare it with the adjoint consistent discretization in (283). Second, for viscous flows only the total force coefficients, i.e., the pressure induced plus the viscous force coefficients, may lead to an adjoint consistent discretization, whereas any discretization of only a part of the force coefficients is adjoint inconsistent. However, note, that dispite of the fact, that the discretization of the target quantity is adjoint inconsistent, and thus the adjoint solution is irregular near the wall boundary (cf. Section 5.8), the resulting error estimation (cf. Table 2) is still very accurate.

In Figure 24 we compare the true error in the target quantity based on refining the computational mesh employing either the adjoint-based or residual-based indicators. Here, we see that for the first three refinement steps, when employing the residual-based indicator, the accuracy in the target quantity is hardly improved. In contrast to that, when using adjoint-based indicators, the error decreases significantly faster, being a factor of more than three smaller already after the second refinement step than the error on the finest residual-based refined mesh. Furthermore, the computed values of the target quantity $J(\mathbf{u}_h)$ can be enhanced by employing the approximate error representation $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h) = \sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa$ to yield an enhanced value of the target quantity, $\tilde{J}(\mathbf{u}_h) = J(\mathbf{u}_h) + \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$. In Figure 24 we see, that the improved values, $\tilde{J}(\mathbf{u}_h)$, are significantly more accurate than the (baseline) $J(\mathbf{u}_h)$ values, and even show a higher rate of convergence. In fact, it can be shown, see (Hartmann and Houston, 2006b), that this value has a higher order of convergence than $J(\mathbf{u}_h)$, provided the primal and the adjoint solutions are smooth and the adjoint solution is approximated using higher-order polynomials. Furthermore, the approximate error representation is close to the true error

| # Elements | # DoF | $J(\mathbf{u}) - J(\mathbf{u}_h)$ | $\sum_{\kappa \in \mathcal{T}_h} \bar{\eta}_\kappa$ | $\theta$ |
|---|---|---|---|---|
| 768 | 12288 | -1.363e-02 | -6.312e-03 | 0.46 |
| 1260 | 20160 | -3.203e-03 | -2.995e-03 | 0.94 |
| 2154 | 34464 | -4.844e-04 | -5.368e-04 | 1.11 |
| 3570 | 57120 | -3.474e-04 | -3.333e-04 | 0.96 |
| 6021 | 96336 | -1.835e-04 | -1.856e-04 | 1.01 |
| 10038 | 160608 | -1.644e-04 | -1.653e-04 | 1.01 |

Table 2: Supersonic viscous flow: Adaptive algorithm for the accurate approximation of pressure induced drag coefficient $C_{\rm dp}$ (Hartmann, 2006).




(a) (b)

Figure 24: Supersonic viscous flow with target quantity $J(\mathbf{u}) = J_{C_{\rm dp}}(\mathbf{u})$: (a) $J(\mathbf{u}_h)$ values on residual-based refined meshes, $J(\mathbf{u}_h)$ and the enhanced values, $\tilde{J}(\mathbf{u}_h) = J(\mathbf{u}_h) + \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$, on adjoint-based refined meshes versus number of elements; (b) Error of these values versus number of elements (Hartmann, 2006).

even in cases of smooth adjoint solutions but possibly non-smooth primal solutions. The large difference in the performance, see Figure 24, of the adjoint-based indicator and the residual-based indicator in producing adaptively refined meshes for the accurate approximation of the target quantity $C_{\rm dp}$, is due to the very different parts of the computational meshes being marked for refinement by the two types of indicators. Figures 25 (a) & (b) show the finest mesh produced by employing the residual-based indicator. We see that this refinement criterion aims at resolving all flow features: the extensive bow shock, the wake of the flow behind the airfoil as well as the weak shocks emanating from the trailing edge of the airfoil. In contrast to that, the refinement of the mesh produced by employing the adjoint-based indicator, see Figures 25 (c) & (d), is very concentrated close to the airfoil. In particular, the bow shock is mainly resolved in a small region upstream of the airfoil only, and there is even no refinement at all at the position of the bow shock beyond six chord lengths above and below the airfoil. Furthermore, the weak shocks emanating from the trailing edge are not resolved and there is no refinement in the wake of the flow beyond three chord lengths behind the airfoil. Instead, the refinement of the mesh is concentrated near the leading edge of the airfoil and in the boundary layer of the flow. All other parts of the computational domain are recognized by the adjoint-based indicator to be of minor importance for the accuracy of the $C_{\rm dp}$ target quantity. In fact, the adjoint solution, see Figures 26 and 27, includes the crucial information concerning which local residuals contribute to the error in the target quantity and to what extent. Herewith, it offers all necessary information of error transport and accumulation. Finally, the adjoint-based indicators mark only those parts of the domain for refinement where residuals of the flow solution significantly contribute to the error of the target quantity, i.e., all parts which are important for the accurate approximation of the target quantity.

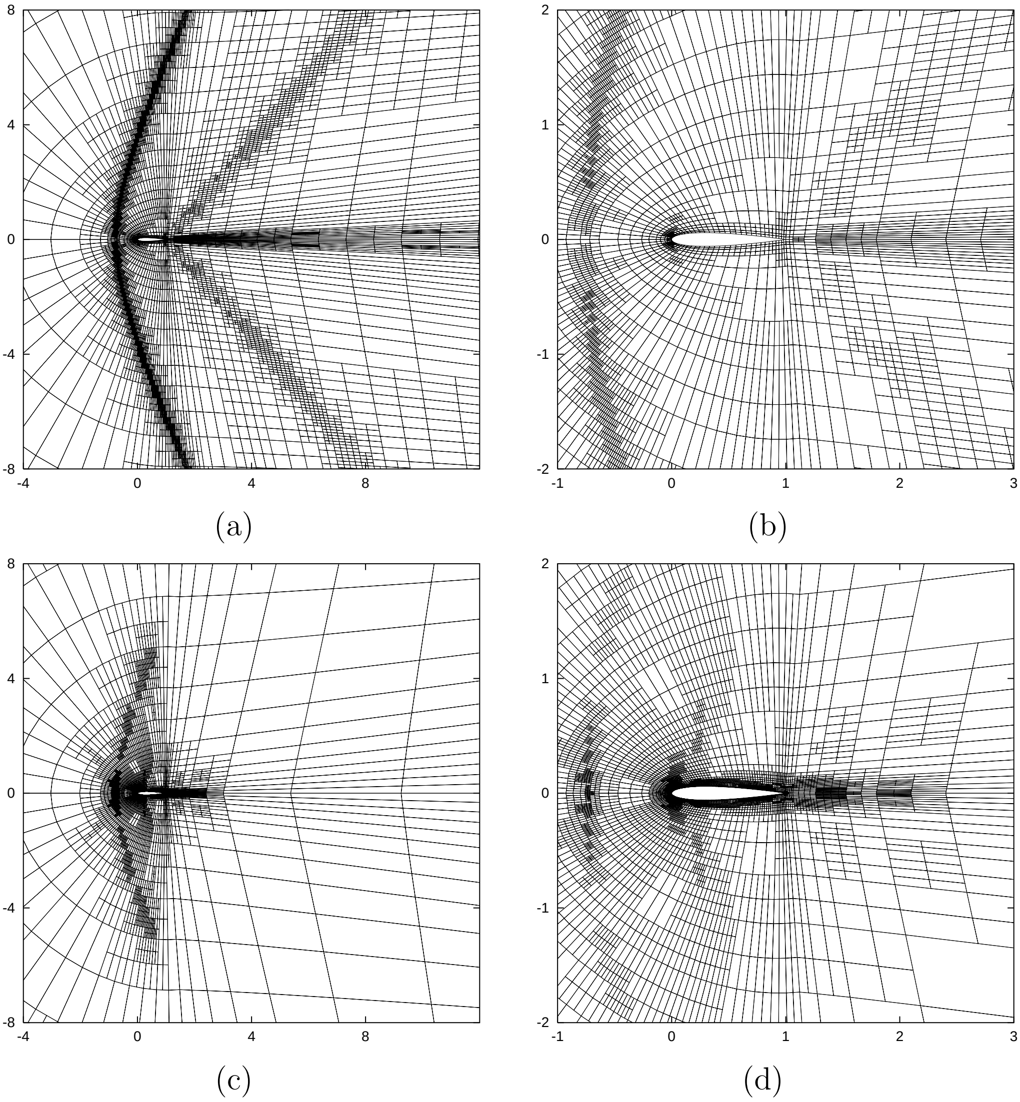
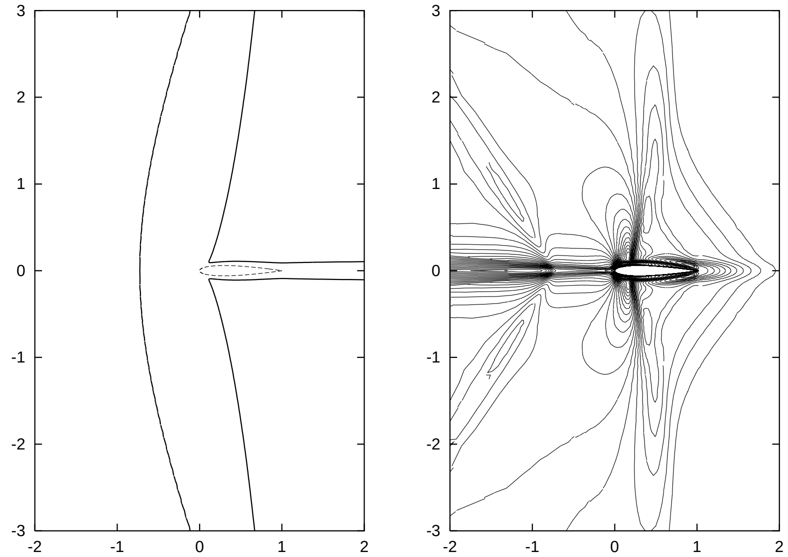
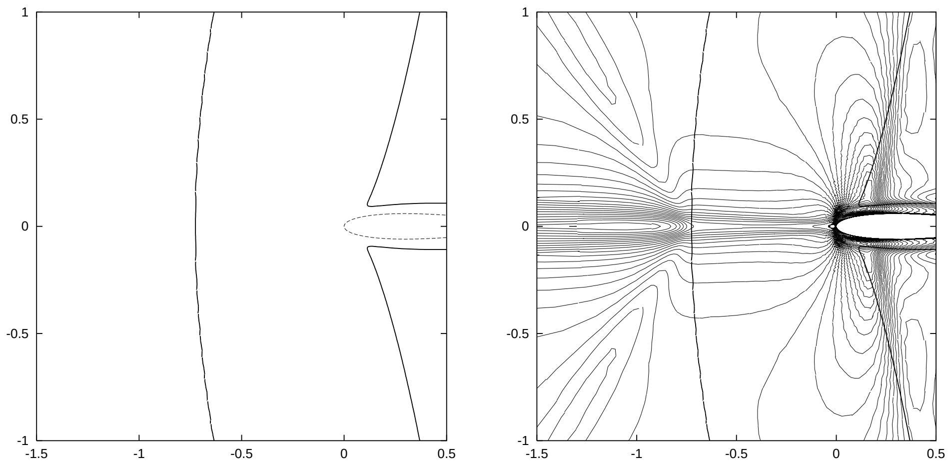



Figure 25: Supersonic viscous flow: (a) & (b) residual-based refined mesh of 17670 elements with 282720 degrees of freedom and $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 1.9 \times 10^{-3}$ ; (c) & (d) goal-oriented refined mesh for $C_{\mathrm{dp}}$: mesh of 10038 elements with 160608 degrees of freedom and $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 1.6 \times 10^{-4}$ (Hartmann, 2006).



(a) (b)

Figure 26: Supersonic viscous flow: (a) Sonic isolines of the flow solution; (b) isolines of the first component of the computed adjoint solution $\bar{\mathbf{z}}_h$ (Hartmann, 2006).

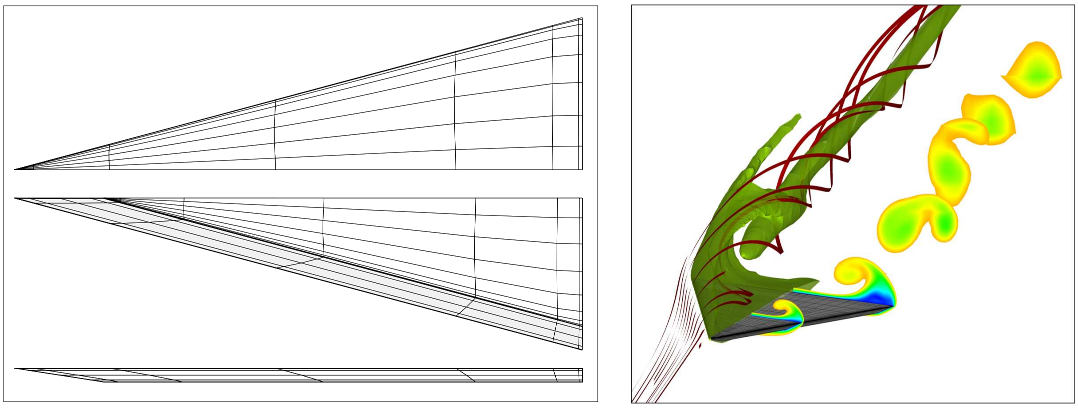


(a) (b)

Figure 27: Supersonic viscous flow: Zoom of (a) sonic ($M = 1$) isolines of the flow solution; (b) together with isolines of first component of the discrete adjoint solution $\bar{\mathbf{z}}_h$ (Hartmann, 2006).

## 9.4 Laminar flow around delta wing

In this example, taken from (Hartmann et al., 2010b), we consider a laminar flow around a delta wing. The delta wing has a sloped and sharp leading edge and a blunt trailing edge. A similar case has previously been considered in (Klaij et al., 2006). The geometry of the delta wing can be seen from the initial surface mesh in Figure 28(a). The delta wing



(a) (b)

Figure 28: Laminar delta wing: a) initial surface mesh: Top, bottom and side view of the half delta wing with straight leading edges, b) solution plot showing streamlines and a Mach number isosurface over the left half of the wing as well as Mach number slices over the right half (Leicht and Hartmann, 2010).

is considered at laminar conditions with inflow Mach number equal to 0.3, at an angle of attack $\alpha = 12.5°$, and Reynolds number $\mathrm{Re} = 4\,000$ with isothermal no-slip wall boundary condition imposed on the wing surface. This is the BTC3 test case of the EU-project ADIGMA (Kroll et al., 2010) and the C2.4 test case in the International Workshop on High-order CFD Methods (cf. Wang et al. (2013)) that we already considered in Section 6.8. As the flow passes the leading edge it rolls up, creates a vortex and a secondary vortex. The resulting vortex system remains over long distances behind the wing, see Figure 28(b). In the following the total drag, lift, and moment coefficients, $C_\mathrm{d}$, $C_\mathrm{l}$ and $C_\mathrm{m}$, will be computed up to a predefined error tolerance $\mathtt{TOL}$. The following industrial accuracy requirements have been defined in the EU-project ADIGMA (Kroll et al., 2010):

$$\begin{aligned} |J_{C_\mathrm{l}}(\mathbf{u}) - J_{C_\mathrm{l},h}(\mathbf{u}_h)| &\le \mathtt{TOL}_{C_\mathrm{l}} &= 10^{-2}, \\ |J_{C_\mathrm{d}}(\mathbf{u}) - J_{C_\mathrm{d},h}(\mathbf{u}_h)| &\le \mathtt{TOL}_{C_\mathrm{d}} &= 10^{-3}, \\ |J_{C_\mathrm{m}}(\mathbf{u}) - J_{C_\mathrm{m},h}(\mathbf{u}_h)| &\le \mathtt{TOL}_{C_\mathrm{m}} &= 10^{-3}. \end{aligned} \tag{346}$$

Reference values of the force coefficients have been obtained by performing high order computations on fine meshes (cf. Hartmann et al. (2010b)).

In the following we compare the performance of various refinement strategies in meeting these accuracy requirements. In particular, we consider the single-target error estimation and mesh refinement approach for each of the $C_\mathrm{l}$, $C_\mathrm{d}$, and $C_\mathrm{m}$ coefficients, separately. This results in three different sequences of locally refined meshes where on each mesh a flow problem (307) and a discrete adjoint problem (316) are solved and the error estimate

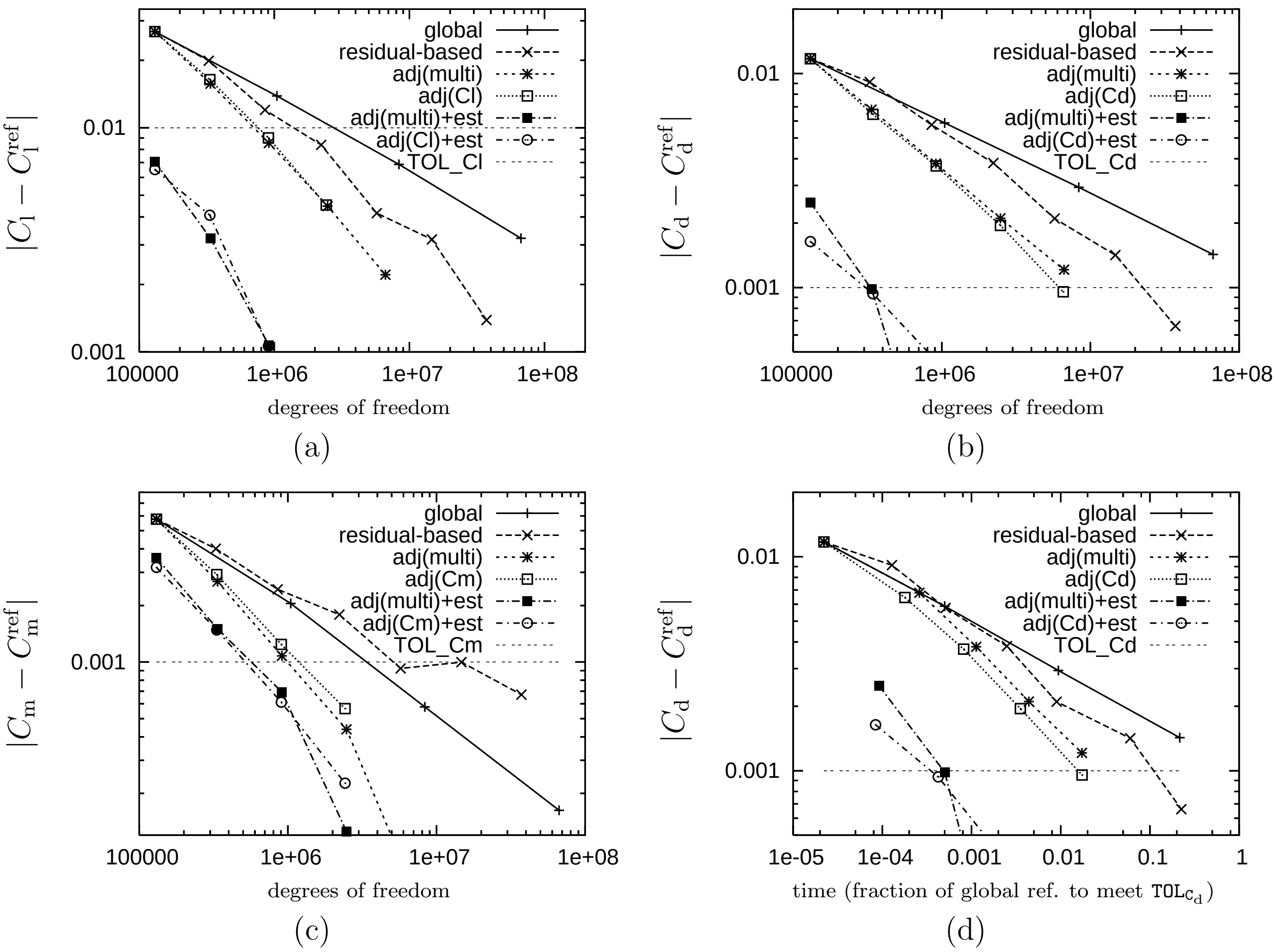
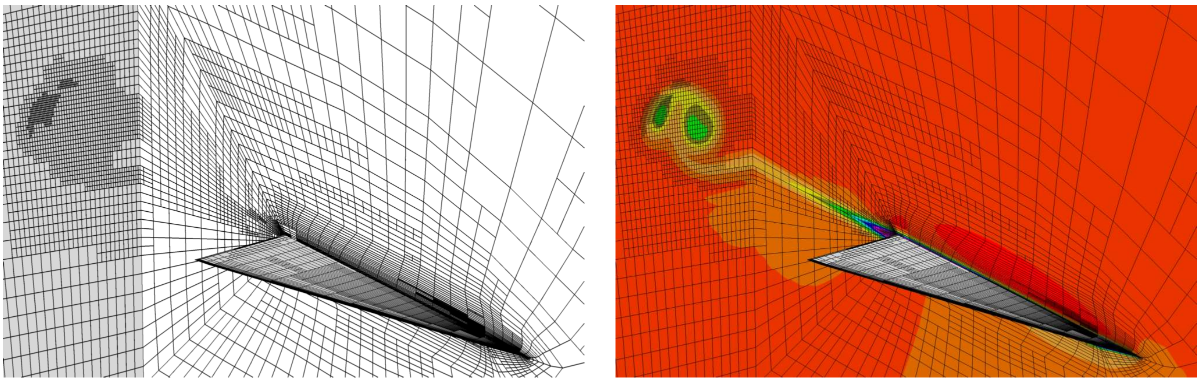



Figure 29: Error in the (a) lift, (b) drag, and (c) moment coefficient for global, residual-based, adjoint-based(single-target) and adjoint-based(multi-target) mesh refinement vs. number of degrees of freedom. On the adjoint-based refined meshes also the enhanced force coefficients $\tilde{C}_{\mathrm{l/d/m}}$ (cf. (328)) are given. (d) Error in the drag vs. computing time relative to the extrapolated time required for global mesh refinement to meet the tolerances (346) (Hartmann et al., 2010b).

(320) is evaluated. This is compared to residual-based and to global mesh refinement. Furthermore, we consider a multi-target error estimation and mesh refinement approach for reducing the sum of relative errors of the $C_\mathrm{l}$, $C_\mathrm{d}$ and $C_\mathrm{m}$ coefficients. This results in one sequence of locally refined meshes which is targeted at reducing the error in all three coefficients, simultaneously. Here, on each mesh a flow problem (307), a discrete error equation (326), and a discrete adjoint problem (335) are solved and the error estimates (327) and (336) are evaluated.

In Figure 29(a)-(c) we see that for $C_\mathrm{l}$ and $C_\mathrm{d}$ the residual-based refinement is more efficient than global mesh refinement which, however, is not the case for $C_\mathrm{m}$. Whereas the residual-based indicators target at resolving all flow features, see the resolution of the vortex system in Fig. 30, they do not necessarily result in meshes suitable for accurately approximating force coefficients. In contrast to that we see that the adjoint-based refinement is significantly more accurate than both, residual-based and global mesh refinement. Furthermore, we see that the accuracy of the single-target and the multi-target adjoint-based mesh refinement is comparable. Finally, we see that the enhanced force coefficients, $\tilde{C}_{\mathrm{d/l/m}} = C_{\mathrm{d/l/m}} + \sum_{\kappa\in\mathcal{T}_h} \bar{\eta}_\kappa$, in case of the single-target algorithm and $\tilde{C}_{\mathrm{d/l/m}} = C_{\mathrm{d/l/m}} + J_i'[\mathbf{u}_h](\bar{\mathbf{e}}_h)$ in case of the multi-target algorithm, are significantly more accurate than the original $C_\star$ values on the adjoint-based refined meshes. This demonstrates that the error estimation for single as well as for multiple target quantities is accurate and reliable.



(a) (b)

Figure 30: (a) Grid after 5 residual-based mesh refinement steps; (b) Mach number isolines of the flow solution on this grid with $14.7 \times 10^6$ DoFs.

Figure 29(d) shows the error in the drag coefficient vs. the computing time relative to the extrapolated time required for global mesh refinement to meet the tolerances (346). For meeting the tolerances (346) the residual-based mesh refinement requires about 10% of the time required for global mesh refinement. The adjoint-based mesh refinement requires about 2% and the adjoint-based mesh refinement including error estimation requires in the range of 0.1%. These time measurements include the time for solving the flow problem and possibly the adjoint problem and the discrete error accumulated for the solutions on coarser meshes. The time comparison clearly demonstrates the advantage of using error estimation and adjoint-based mesh refinement.

## 9.5 Turbulent flow around streamlined body

In this example, taken from (Hartmann et al., 2011), we consider a turbulent flow around a streamlined three-dimensional body based on a 10 percent thick airfoil with boundaries constructed by a surface of revolution. In particular, we consider the streamlined body at a Mach number $M = 0.5$, an angle of attack $\alpha = 5°$, and a Reynolds number $\mathrm{Re} = 10\times 10^6$ with adiabatic no-slip wall boundary conditions. This is the ADIGMA BTC0 test case which has been defined in the ADIGMA project (Kroll et al., 2010) in order to enable grid convergence studies. It is also considered in the International Workshop on High-order CFD Methods (cf. Wang et al. (2013)) as test case C2.3. Based on extrapolation of higher order computational results and of the results on very fine locally adapted meshes, following reference values of the total drag, lift, and pitching moment coefficients, $C_\mathrm{d}$, $C_\mathrm{l}$ and $Cmy$ have been obtained: $J_{C_\mathrm{l}}(\mathbf{u}) = 0.00663$, $J_{C_\mathrm{d}}(\mathbf{u}) = 0.00858$ and $J_{C_\mathrm{my}}(\mathbf{u}) = 0.00588$ (cf. Hartmann et al. (2011)).

In the following the total drag, lift, and pitching moment coefficients will be computed up to a predefined error tolerance $\mathtt{TOL}$. Let us consider the following accuracy requirements:

$$\begin{aligned} |J_{C_\mathrm{l}}(\mathbf{u}) - J_{C_\mathrm{l}}(\mathbf{u}_h)| &\leq \mathtt{TOL}_{C_\mathrm{l}} &&= 3\times 10^{-4}, \\ |J_{C_\mathrm{d}}(\mathbf{u}) - J_{C_\mathrm{d}}(\mathbf{u}_h)| &\leq \mathtt{TOL}_{C_\mathrm{d}} &&= 1\times 10^{-4}, \\ |J_{C_\mathrm{my}}(\mathbf{u}) - J_{C_\mathrm{my}}(\mathbf{u}_h)| &\leq \mathtt{TOL}_{C_\mathrm{my}} &&= 2\times 10^{-4}. \end{aligned} \tag{347}$$

In the following we compare the performance in meeting these accuracy requirements for higher order discretizations against second order discretizations on globally refined meshes. We also compare against the performance of the two goal-oriented refinement strategies presented in Section 8. In particular, we consider the single-target error estimation and mesh refinement approach for each of the $C_\mathrm{l}$, $C_\mathrm{d}$, and $C_\mathrm{my}$ coefficients, separately. This results in three different sequences of locally refined meshes where on each mesh a flow problem (300) and a discrete adjoint problem (316) are solved, and the error estimate (314) is evaluated. Furthermore, we consider a multi-target error estimation and mesh refinement approach for reducing a weighted sum of absolute errors of the $C_\mathrm{l}$, $C_\mathrm{d}$ and $C_\mathrm{my}$ values. This yields one sequence of locally refined meshes that targets at reducing the error in all three coefficients simultaneously. Here, on each mesh a flow problem (300), a discrete error equation (326) and a discrete adjoint problem (335) are solved, and the error estimates (327) and (336) are evaluated. The weighting factors in the weighted sum (331) might account for the different tolerances in (347). In fact, in the following, the multi-target approach will be based on the weighted sum

$$2|J_{C_\mathrm{l}}(\mathbf{u}) - J_{C_\mathrm{l}}(\mathbf{u}_h)| + 6|J_{C_\mathrm{d}}(\mathbf{u}) - J_{C_\mathrm{d}}(\mathbf{u}_h)| + 3|J_{C_\mathrm{my}}(\mathbf{u}) - J_{C_\mathrm{my}}(\mathbf{u}_h)|.$$

Figure 31 shows the starting mesh of this computation with 6 656 curved elements. The edges are given by polynomials of degree 4 based on additional points taken from finer grids with straight edges. On this mesh we first compute the flow solutions $\mathbf{u}_h \in \mathbf{V}_h^p$ for the polynomial degrees $p = 1,\dots,4$. Additionally, for the lower polynomial degrees, we compute the solutions on globally refined meshes. The resulting force coefficients $C_\mathrm{l}$, $C_\mathrm{d}$ and $C_\mathrm{my}$ plotted over the number of degrees of freedom are given in Figures 32(a), 32(c) and 32(e), respectively. We note that here and throughout this work numbers of degrees of freedom always refer to the total number of degrees of freedom (including all components) of flow solutions $\mathbf{u}_h \in \mathbf{V}_h^p$. In Figure 32(a) we see that the $p = 2, 3$ and

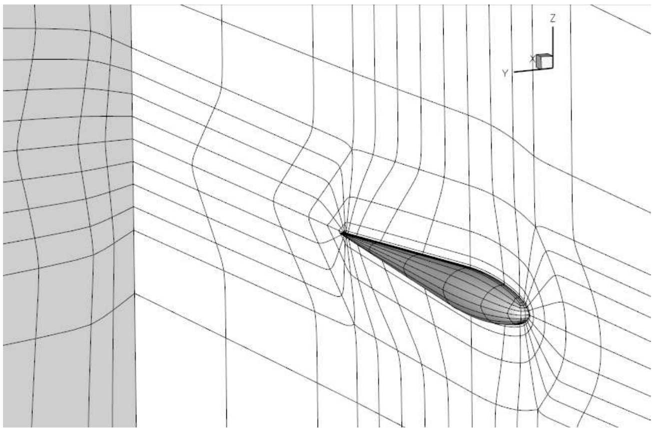


Figure 31: Turbulent flow around streamlined body: Body surface, symmetry plane and cut through the domain of the coarse mesh with 6 656 curved elements. The lines are given by polynomials of degree 4.

4 solutions are within the prescribed accuracy tolerance of the $C_\mathrm{l}$ value on the coarsest mesh, and they require significantly less degrees of freedom than the $p = 1$ solution on the twice globally refined mesh. A similar behaviour is observed in Figure 32(e) for the $C_\mathrm{m}$ value. Furthermore, from Figure 32(c) we see that the $p = 3$ and 4 solutions on the coarsest mesh are within the prescribed accuracy tolerance of the $C_\mathrm{d}$ value.

Let us now consider the adjoint-based error estimation and goal-oriented mesh refinement that targets the lift coefficient $C_\mathrm{l}$ (i.e., the single target quantity is $J(\mathbf{u}) = J_{C_\mathrm{l}}(\mathbf{u})$). To this end we compute the solution $\mathbf{u}_h \in \mathbf{V}_h^1$ to the flow problem (307) and the solution $\bar{\mathbf{z}}_h \in \mathbf{V}_h^2$ to the discrete adjoint problem (316). We then evaluate the approximate error representation (320) and obtain the adjoint-based indicators $\bar{\eta}_\kappa$ which we employ for adaptive mesh refinement. Starting on the coarse mesh of 6 656 curved elements shown in Figure 31 a sequence of locally refined meshes specifically tailored to the accurate approximation of the $C_\mathrm{l}$ value is obtained. In Table 3 we collect the number of elements, the (total) number of degrees of freedom (DoF) of $\mathbf{u}_h \in \mathbf{V}_h^1$, the true error $J_{C_\mathrm{l}}(\mathbf{u}) - J_{C_\mathrm{l}}(\mathbf{u}_h)$ in the lift coefficient, the estimated error $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$, and the quotient $\theta = \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h) / \left(J_{C_\mathrm{l}}(\mathbf{u}) - J_{C_\mathrm{l}}(\mathbf{u}_h)\right)$ of the estimated and the true error. Here, we see that the estimated errors are very close to the true errors. This is also indicated by the effectivity indices which are very close to one. In fact, from the third mesh onwards the rounded index equals one. We note that this represents an almost perfect error estimation. Given the complexity of the governing RANS-$k\omega$ flow equations this might seem surprising. However, the turbulent flow considered here is particularly smooth. So we expect the error estimation to work very well in this case.

Figure 32(b) shows the $C_\mathrm{l}$ values on the sequence of adaptively refined meshes. For comparison it also includes the $C_\mathrm{l}$ values of the $\mathbf{u}_h \in \mathbf{V}_h^1$ flow solutions on the globally refined meshes already shown in Figure 32(a). Furthermore, Figure 32(b) includes the enhanced target quantities (319) evaluated based on the computed $C_\mathrm{l}$ values and the error estimates $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$ given in Table 3. Here, we see that the enhanced quantity $\tilde{J}_{C_\mathrm{l}}(\mathbf{u}_h) = J_{C_\mathrm{l}}(\mathbf{u}_h) + \mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$ on the coarsest mesh is already very close to the reference $C_\mathrm{l}$ value. Due to the high accuracy of the error estimates in Table 3, we see that the

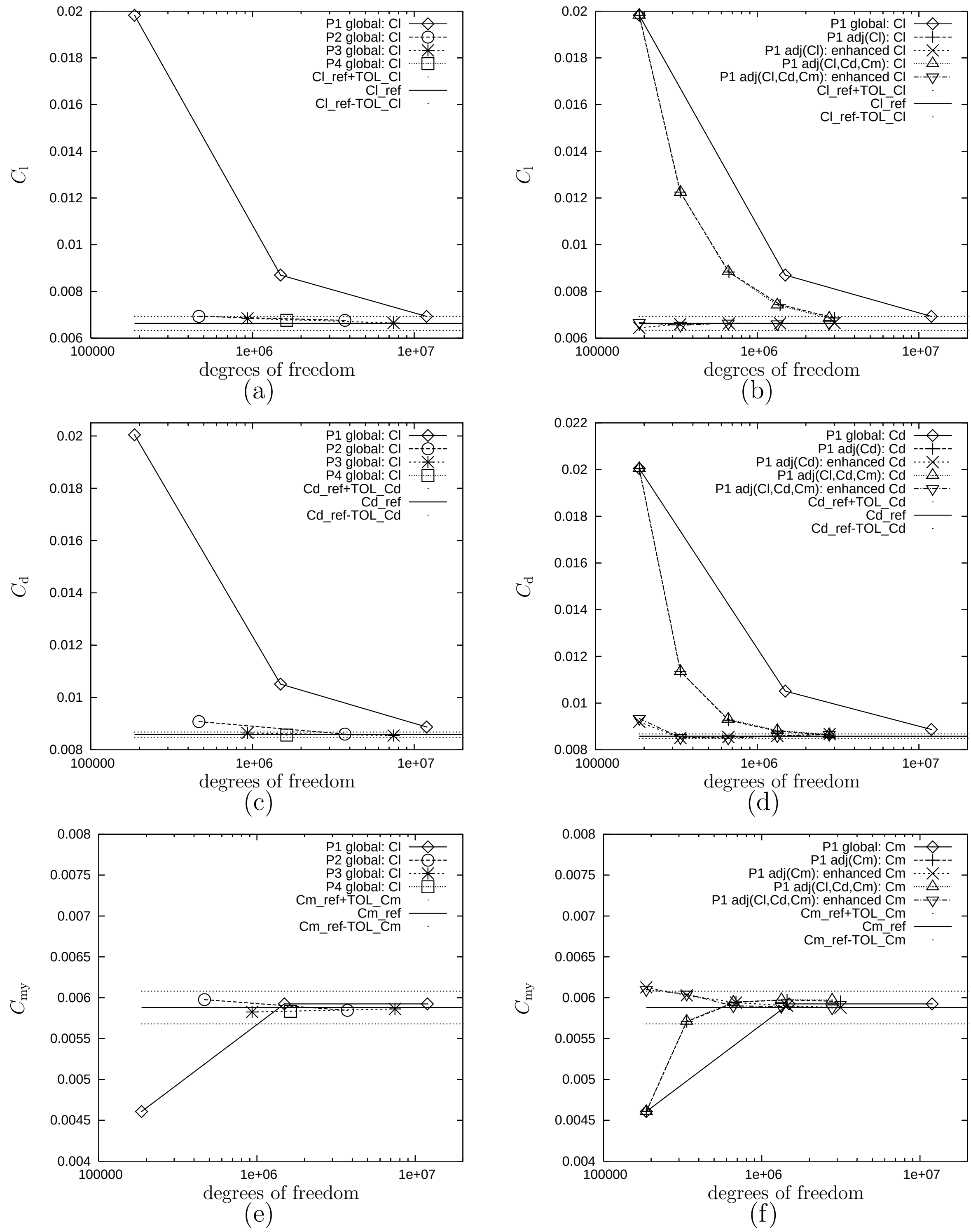



Figure 32: Turbulent flow around streamlined body: $C_{\mathrm{l}}$, $C_{\mathrm{d}}$ and $C_{\mathrm{my}}$ values in the top, middle and bottom row, respectively. (left) The $p = 1, 2, 3$ solutions on globally refined meshes; (right) The force coefficients $J(\mathbf{u}_h)$ and the enhanced force coefficients, (319) and (328), on the single-target and multi-target adjoint-based refined meshes (Hartmann et al., 2011).

| # Elements | # DoF | $J_{C_l}(\mathbf{u}) - J_{C_l}(\mathbf{u}_h)$ | $\mathcal{R}(\mathbf{u}_h, \bar{\mathbf{z}}_h)$ | $\theta$ |
|---|---|---|---|---|
| 6656 | 186368 | -1.320e-02 | -1.338e-02 | 1.01 |
| 11919 | 333732 | -5.622e-03 | -5.661e-03 | 1.01 |
| 23894 | 669032 | -2.200e-03 | -2.203e-03 | 1.00 |
| 49478 | 1385384 | -8.007e-04 | -8.011e-04 | 1.00 |

Table 3: ADIGMA BTC0 case at turbulent conditions: Error estimation for the $C_l$ value (Hartmann et al., 2011).

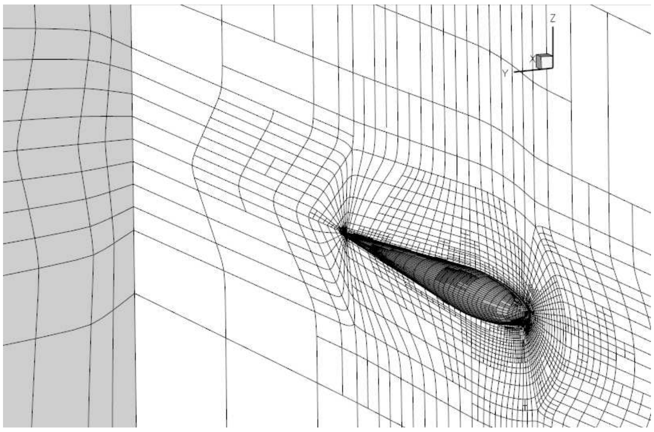


Figure 33: Turbulent flow around streamlined body: Mesh with 47 497 curved elements after 3 multi-target adjoint-based refinement steps (Hartmann et al., 2011).

error estimation significantly improves the computed force coefficients. In fact, while the adjoint-based refinement reduces the number of degrees of freedom, which is required for meeting the accuracy tolerance, by a factor of 4 compared to global mesh refinement, there is another factor of about 16 gained by using the error estimation. Figures 32(d) and 32(f) show the respective plots for the error estimation and goal-oriented mesh that targets the drag and pitching moment coefficients, $C_d$ and $C_{my}$, respectively. Here, the behaviour is similar to that described before for the lift coefficient. In fact, here the enhanced force coefficients meet the accurarcy requirements on the second but coarsest mesh, which again corresponds to a very good error estimation.

Additionally, Figures 32(b), 32(d) and 32(f) show the convergence of the force coefficients on the sequence of multi-target refined meshes. Targeted at reducing the errors in $C_l$, $C_d$ and $C_{my}$ simultaneously, the resulting values are not expected to be as accurate as for the single-target adapted meshes. However, we see that the differences for this test case are marginal. This demonstrates that by the multi-target mesh refinement about the same accuracy in the force coefficients is achieved for this test case as for the single-target mesh refinements (see also Figure 33). However, this is accomplished with significantly reduced computing. In fact, the multi-target approach requires only the flow solution, the adjoint solution and the solution to the discrete error equation on *one* sequence of adaptively refined meshes. In contrast to that, the single-target approach requires the flow solutions and the adjoint solutions on three different sequences of refined meshes. This results in a factor of 3 in the number of flow solutions and a factor of 1.5 in the number of auxiliary problems to be solved. Note, that these factors and thus the gain of using the multi-target approach increases with the number of target quantities under consideration. Finally, we note that the multi-target error estimation is comparable to the single-target error estimation since the enhanced $C_{\mathrm{d}}$, $C_{\mathrm{l}}$ and $C_{\mathrm{my}}$ values of (328) in Figures 32(b), 32(d) and 32(f) are close to the single-target enhanced values of (319).

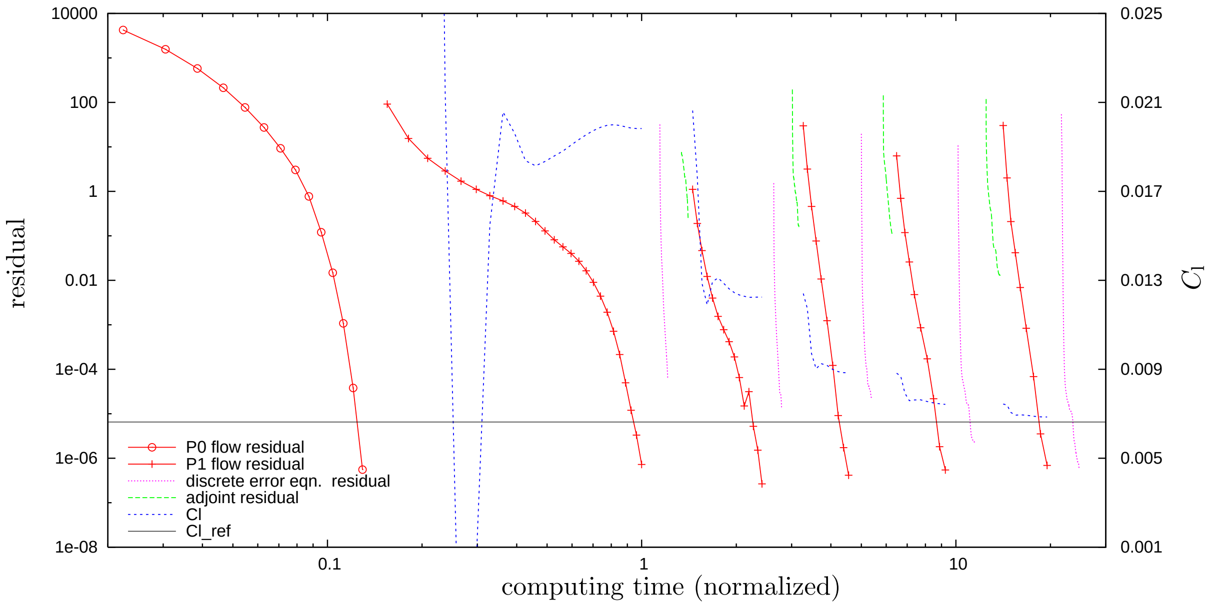



Figure 34: Turbulent flow around streamlined body: Convergence of the solver on the sequence of goal-oriented refined meshes that target the $C_{\mathrm{l}}$, $C_{\mathrm{d}}$ and $C_{\mathrm{m}}$ values: the non-linear residuals and $C_{\mathrm{l}}$ values of the flow solutions and the linear residuals of the discrete error equations and the adjoint problems are shown (Hartmann et al., 2011).

In the following, we investigate the performance of the flow solver and of the solver for the additional discrete error and adjoint equations. In particular, we are interested in the computing time required for solving the additional problems relative to the time spent for solving the flow problems in a multi-target adjoint-based mesh refinement algorithm. After initialization of freestream values on the coarsest mesh a flow solution $\mathbf{u}_h \in \mathbf{V}_h^p$ for $p = 0$ is computed which serves as starting solution for $p = 1$. On finer meshes the flow solution and the solutions to the discrete error and adjoint equations are initialized with interpolations of the solutions on the previous mesh. The nonlinear residual of the flow solution is decreased based on an implicit/backward Euler method where the CFL number and thus the local time step size is driven by the switched evolution relaxation (SER) method (Mulder and van Leer, 1985; Essers et al., 1995). In each nonlinear solution step a linear system is solved with the restarted and block-ILU-preconditioned GMRES method with 60 Krylov vectors. The linear solver is stopped once the linear residual is reduced by a factor of $10^{-6}$ or a maximum number of linear iterations is reached. On the coarsest mesh the linear solver performs at most 120 iterations; under mesh refinement this number increases with the cube root of the number of elements. The nonlinear solution process is stopped once the nonlinear residual is reduced to $10^{-6}$. Similarly, the linear problems arising from the discrete error equations and the adjoint problems are solved using 120 GMRES iteration steps with a restart after 60 iterations. Again, under mesh refinement the number of linear iteration steps increases with the cube root of the number of elements.

Figure 34 shows the convergence of the flow solver on the sequence of multi-target adjoint-based refined meshes that target the $C_{\mathrm{l}}$, $C_{\mathrm{d}}$ and $C_{\mathrm{m}}$ values, i.e., on the sequence of the 5 adaptively refined meshes considered in the right plots of Figure 32. In particular, Figure 34 shows the convergence of the nonlinear residuals and $C_{\mathrm{l}}$ values of the flow solutions $\mathbf{u}_h \in \mathbf{V}_h^1$. As the flow solution to this test case is particularly smooth, the flow solver takes very few iteration steps, only. In fact, the solver of the $p = 1$ solution on the coarsest mesh requires 25 steps for convergence. This number decreases to less than 10 on subsequently refined meshes. Figure 34 also includes the convergence of the linear residuals of the solutions $\bar{\mathbf{e}}_h$ and $\bar{\mathbf{z}}_h \in \bar{\mathbf{V}}_h^2$ to the discrete error and adjoint equations. Although the discrete error equation is solved with a higher polynomial degree than the flow problems, it takes less computing time than the flow solver due to the linearity of the problem. In fact, approximating the discrete error on the coarsest mesh takes about 18% of the computing time of the flow solution. The additional computing time for the adjoint problem is of similar magnitude. We note that on finer meshes the computing time for the additional problems increases relative to the time taken by the flow solver due to the decreasing effort of the flow solver. However, already the additional 18% computing time spent on the solution to the discrete error equation on top of the flow solver on the coarsest mesh yields an enhanced $C_{\mathrm{l}}$ value that is within the prescribed accuracy tolerance (see Figure 32(b)). Furthermore, it results in significantly improved $C_{\mathrm{d}}$ and $C_{\mathrm{m}}$ values (see Figure 32(d)&(f)). In contrast to that a higher-order flow solution $\mathbf{u}_h \in \mathbf{V}_h^2$ on the coarsest mesh requires more than 5 times the time required for the $\mathbf{u}_h \in \mathbf{V}_h^1$ flow solution.

## 9.6 Turbulent flow around the DLR-F6 wing-body

In this example, taken from (Hartmann et al., 2011), we consider a turbulent flow at a Mach number $M = 0.5$, a Reynolds number $\mathrm{Re} = 5 \times 10^6$ and an angle of attack $\alpha = -0.141°$ around the DLR-F6 wing-body configuration without fairing (see Figure 35). This is a modification of the DPW III test case, where a fixed angle of attack has been assumed instead of a given target lift. Also, the Mach number has been reduced from the originally $M = 0.75$ to $M = 0.5$ in order to obtain a subsonic flow.

The original DPW mesh of about 3.2 million hexahedral elements has been agglomerated twice resulting in a coarse mesh of 50 618 hexahedral elements. The additional points of the original mesh have been used to define 50 618 curved elements (see Figure 36), where the curved lines are represented by polynomials of degree 4.

On this mesh we first compute the flow solutions $\mathbf{u}_h \in \mathbf{V}_h^p$ for the polynomial degrees $p = 1, 2$ and 3. Additionally, for $p = 1$ and 2 we compute the solutions on a once globally refined mesh. The resulting drag coefficients are given in Figure 37. Due to the complexity of the problem, no rigorous convergence study and thus no reference value is available for this case. Nevertheless, we clearly see the advantage in terms of accuracy and degrees of freedom of using discretizations with higher polynomial degrees $p = 2$ and 3 over the discretization with the low polynomial degree $p = 1$.

Figure 38 shows the surface mesh near the wing-body junction, the $c_p$ distribution, and wall streamlines of the $p = 1, 2$ and 3 solutions on the coarse mesh. For the $p = 2$ and $p = 3$ solutions we clearly recognize the separation of the flow. The resolution of the

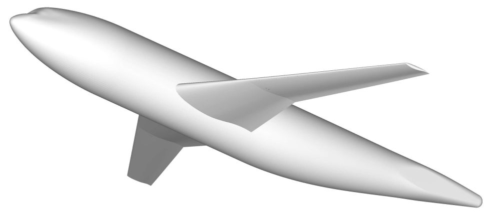


Figure 35: Geometry of the DLR-F6 wing-body configuration without fairing.

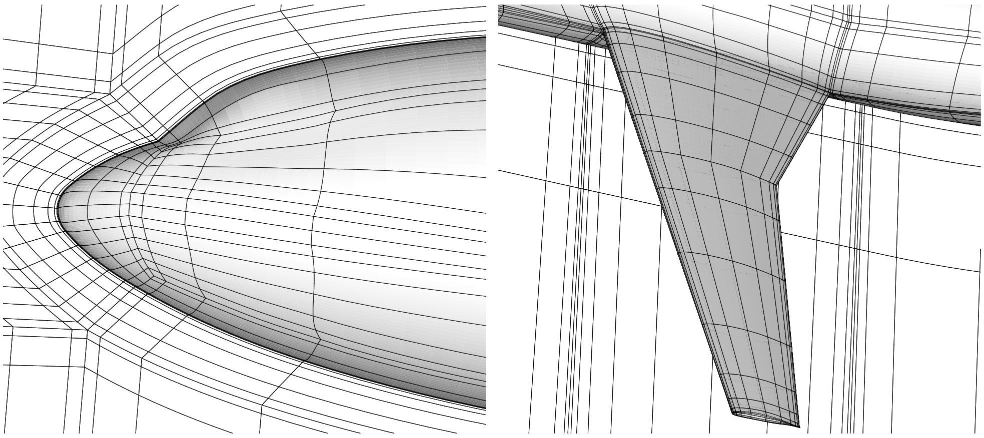


Figure 36: The DLR-F6 wing-body configuration: Mesh with 50 618 curved elements shown close to the nose and the wing (Hartmann et al., 2011).

$p = 1$ solution on the coarse mesh is too low to capture the separation.

Let us now consider the adjoint-based error estimation and goal-oriented mesh refinement that targets the drag coefficient $C_\mathrm{d}$. As an example, Figure 39 shows the density adjoint (i.e., the first component of the discrete adjoint solution $\bar{\mathbf{z}}_h$) on the locally adapted mesh after two adjoint-based mesh refinement steps. Finally, the $C_\mathrm{d}$ values and the enhanced $C_\mathrm{d}$ values on this sequence of adjoint-based refined meshes are given in Figure 37. As in the previous examples, we see a significant decrease in the number of degrees of freedom required for computing the force coefficient up to a specific accuracy. A further significant improvement can be seen in the enhanced force coefficient, which again corresponds to a good error estimation.

As already seen in Figure 38(b) the $p = 1$ solution on the coarse mesh does not capture the separation. Without showing details, we note that the $p = 1$ solution on the twice adjoint-based refined mesh has an onset of separation which is then fully captured on the three times adjoint-based refined mesh.

## 9.7 Turbulent flow around the VFE-2 delta wing

In the following, we consider a turbulent flow around the 65° swept delta wing configuration with medium rounded leading edge of the second Vortex Flow Experiment (Hummel and Redeker, 2001). The geometry of the VFE-2 configuration is shown in Figure 40. We note, that the sting has been extended in the geometry model to about 1.5 cord lengths behind the wing and has a rounded trailing edge, see Figure 41.

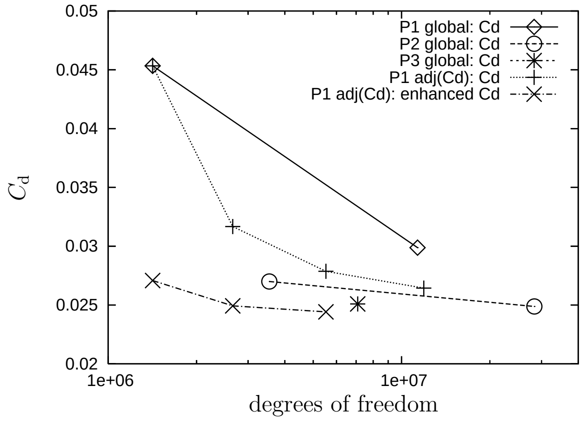



Figure 37: Turbulent flow around the DLR-F6 wing-body configuration: Convergence of the drag coefficient $C_\mathrm{d}$ for the $p = 1, 2$ and 3 solutions on the coarse and a once globally refined mesh. Convergence of the $C_\mathrm{d}$ and the enhanced $C_\mathrm{d}$ values under adjoint-based mesh refinement targeted at the $C_\mathrm{d}$ value (Hartmann et al., 2011).

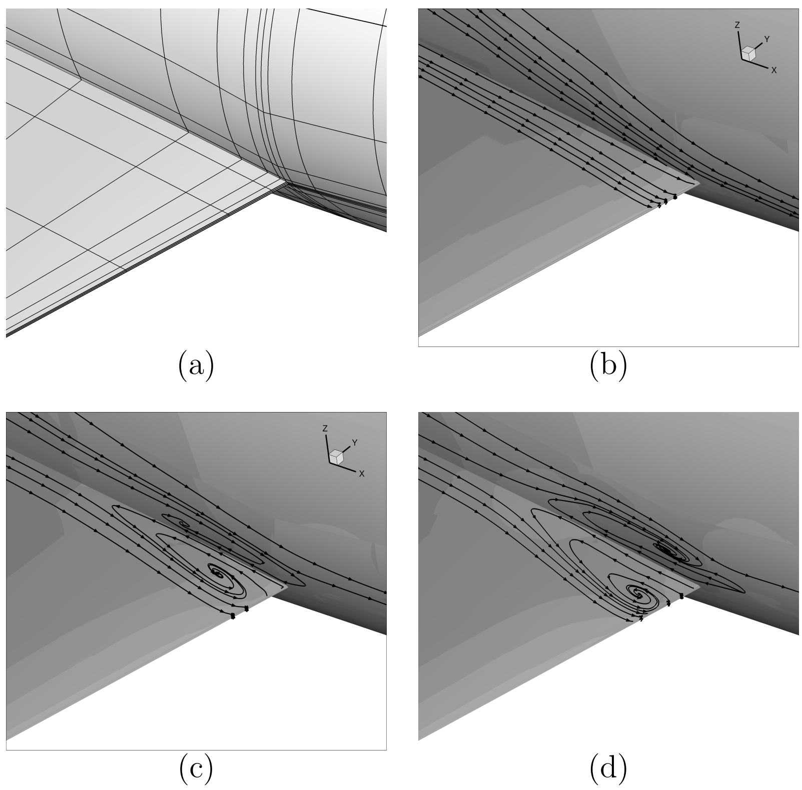


Figure 38: Turbulent flow around the DLR-F6 wing-body configuration: (b)-(d) $c_p$ distributions and wall streamlines of the $p = 1, 2$ and 3 solutions on (a) the coarse mesh of 50 618 curved elements (Hartmann et al., 2011).

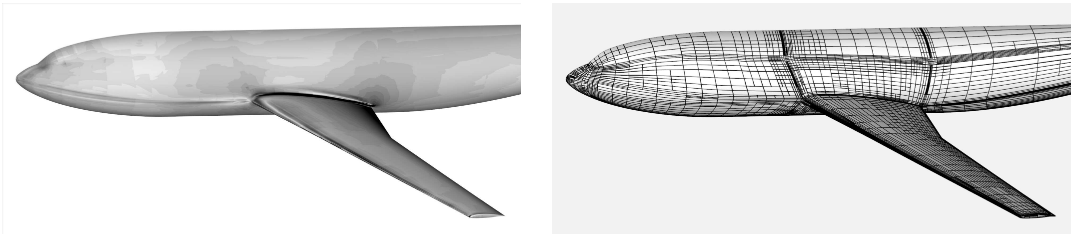


Figure 39: Turbulent flow around the DLR-F6 wing-body configuration: Density adjoint distribution, i.e., the distribution of the first component of the discrete adjoint solution $\bar{\mathbf{z}}_h$, on a twice adjoint-based refined mesh (Hartmann et al., 2011).

An original block-structured mesh with 884 224 hexahedral elements has been agglomerated twice resulting in a coarse mesh of 13 816 elements. The additional points of the original mesh have been used to define 13 816 curved elements, where the curved lines are represented by polynomials of degree 4 and interpolate the original points. The surface mesh of the coarse mesh with curved elements is depicted on the wing and the symmetry plane in Figure 40. Figure 42 shows the approximation of the rounded leading edge close to the symmetry plane for the original straight-sided mesh and the coarse curved mesh, respectively. The original mesh shows clear kinks in the approximation of the leading edge. In contrast to that, the coarse mesh with curved elements is smooth and approximates the curved boundary very well.

### 9.7.1 Subsonic turbulent flow around the VFE-2 delta wing

In this example, taken from (Hartmann, 2013a), we consider a subsonic fully turbulent flow at a Mach number $M = 0.4$, a Reynolds number $Re = 3 \times 10^6$ and an angle of attack of $\alpha = 13.3°$ around the VFE-2 configuration with medium rounded leading edge. This is the U1.b test case considered in the EU-project IDIHOM (Kroll, 2012). According to the description in (Schütte and Lüdeke, 2009) we expect the following flow topology:

> "Because of the leading edge geometry with rounded leading edges the wing tip can be assumed as a blunt body. The flow around the leading edge is

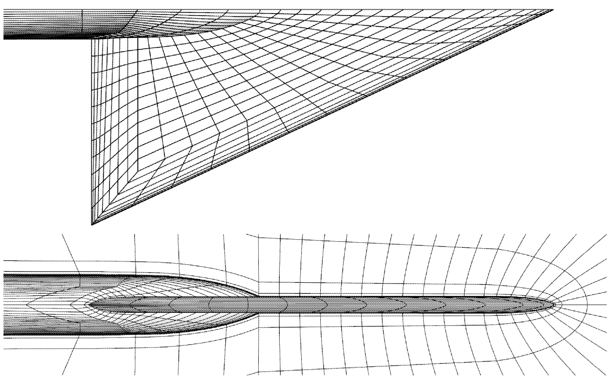


Figure 40: The VFE-2 configuration: Top and side view of the 65° swept delta wing with medium rounded leading edge (Hartmann, 2013a,b).

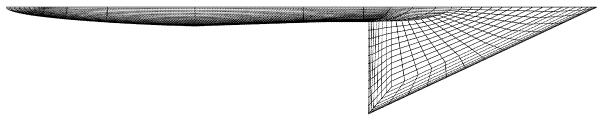


Figure 41: The VFE-2 configuration: Full view on sting and wing.

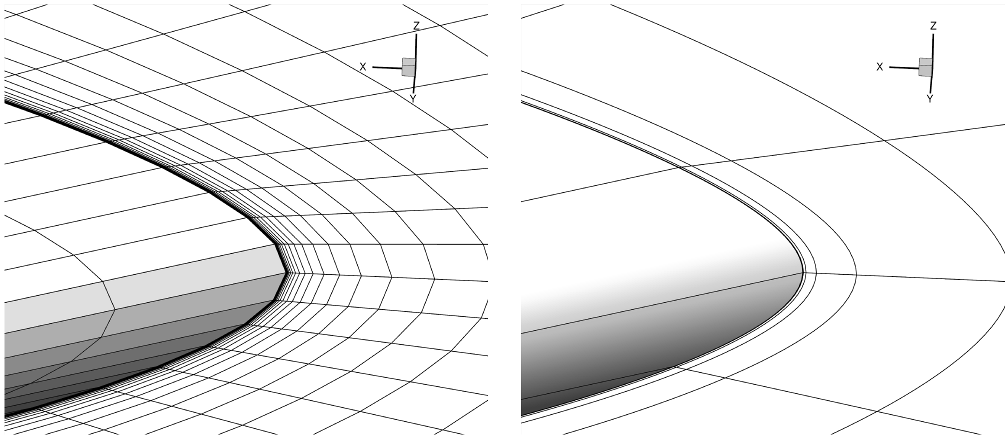


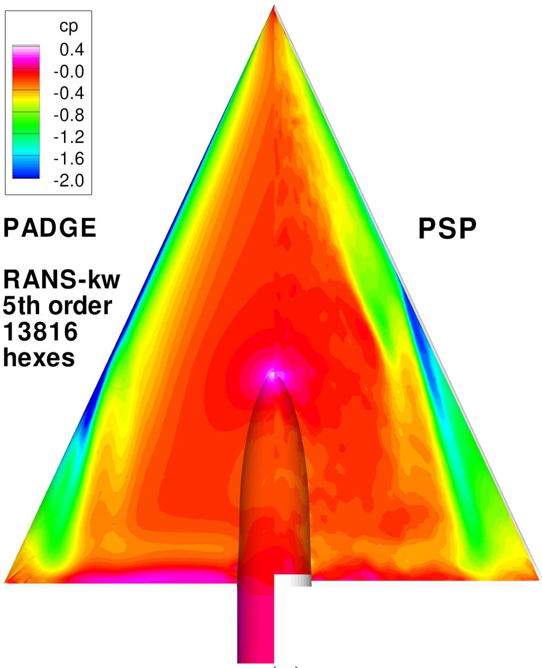

Figure 42: Zoom of the rounded leading edge of the delta wing close to the symmetry plane. Left: the straight-sided mesh; Right: the coarse curved mesh.

> initially attached. Further downstream the flow separates and the inner vortex is generated. Regarding the ratio between leading edge radius and wing span the leading edge is getting sharper relative to the downstream position. This leads eventually to a separation of the shear layer further downstream at the leading edge and the outer vortex is generated."

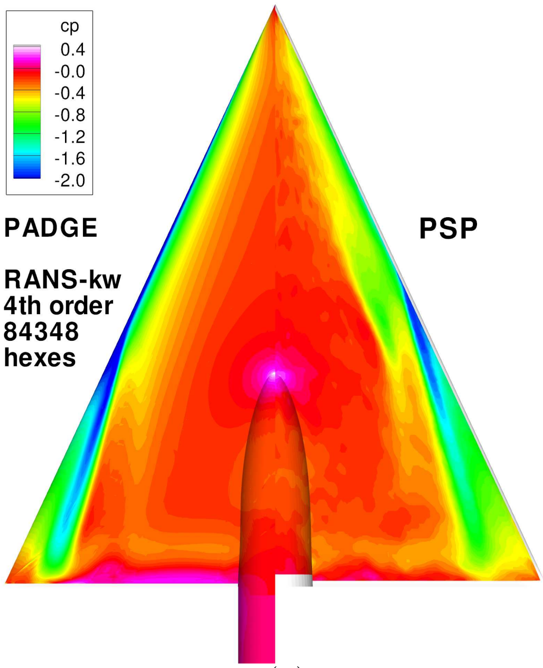
While typically only one primary vortex is encountered on flows around delta wings with sharp leading edges, the delta wing with a rounded leading edge considered here creates two primary vortices which makes the flow field particularly interesting and challenging for numerical simulation.

First we consider a discrete solution $\mathbf{u}_h \in \mathbf{V}_h^4$ of (307), i.e., a (by design) 5$^{\text{th}}$-order flow solution, on the coarse mesh of 13 816 curved elements depicted in Figures 40 and 41. The resulting $c_p$-distribution on the upper side of the wing is shown in the left part of Figure 43(a) compared to the "pressure sensitive paint" (PSP) measurements (Konrath et al., 2006) in the right part of Figure 43(a). In the measurements as well as in the 5$^{\text{th}}$-order flow solution on this extremely coarse mesh we recognize the suction trace of the stronger outer vortex as well as the suction trace of the weaker inner vortex.

Next, we consider residual-based mesh refinement. In particular, the adaptation will be based on the residuals of the mean-flow equations in (344), while ignoring the residuals of the $k$ and $\omega$ equations. Furthermore, the adaptation will be combined with an anisotropic element subdivision based on jump indicators (cf. Leicht and Hartmann (2010)). Figure 44 shows a 2$^{\text{nd}}$-order DG solution on a 6 times residual-based adapted mesh with 562 892 curved elements. We see that the stronger outer vortex is resolved and tracked over long distances. In fact, it is resolved until it crosses the farfield outflow boundary about 14 cord lengths behind the wing.




(a)




(b)

Figure 43: Subsonic turbulent flow around the VFE-2 delta wing: The $c_p$-distribution on the upper side of the wing computed with the PADGE code in comparison to PSP measurements: (a) $5^{\text{th}}$-order DG solution on the mesh with 13 816 curved elements and (b) the $4^{\text{th}}$-order DG solution on a residual-based adapted mesh with 84 348 curved elements (Hartmann, 2013a).

Finally, we combine the higher-order DG discretization with residual-based adaptation. In particular, we consider a $4^{\text{th}}$-order DG solution on a locally adapted mesh of 84 348 curved elements. Figure 43(b) shows its $c_p$-distribution in comparison to the PSP measurements. Furthermore, Figure 45 shows the $c_p$-distribution on the wing together with slices of the $\lambda_2$-criterion. The two primary vortices are clearly visible: the inner vortex that weakens and the outer vortex that strenghtens while being advected downstream.

In the following we compare the performance of the residual-based adaptation in predicting aerodynamic forces with adjoint-based adaptation and with computations performed by Cassidian with the DLR-TAU code in the EU-project IDIHOM (Kroll, 2012). In particular, we compare the performance in predicting the drag and lift coefficients $C_{\mathrm{d}}$ and $C_{\mathrm{l}}$ for

- the residual-based mesh refinement,
- the adjoint-based mesh refinement,
- with global mesh refinement, and
- with computations performed by Cassidian with the DLR-TAU Code.

In Table 4 we collect some characteristic data of this comparison. Figures 46(a) and (b) show the convergence of respectively, the drag and the lift coefficient under mesh refinement. Here, residual-based, adjoint-based and global mesh refinement of $3^{\text{rd}}$ order


Figure 44: Subsonic turbulent flow around the VFE-2 delta wing: $c_p$ of a 2$^{\text{nd}}$-order DG solution on a residual-based adapted mesh with 562 892 elements: The outer vortex is resolved until the outflow boundary where the $\lambda_2$-criterion is shown (Hartmann, 2013a).

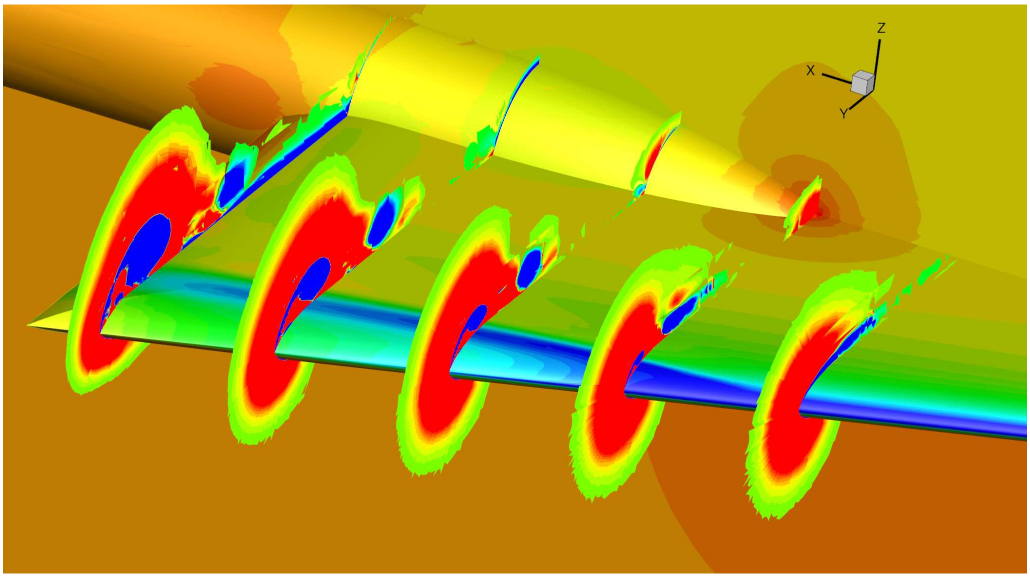



Figure 45: Subsonic turbulent flow around the VFE-2 delta wing: The $c_p$-distribution and slices of the $\lambda_2$-criterion of a 4$^{\text{th}}$-order DG solution on a residual-based adapted mesh of 84 348 curved elements (Hartmann, 2013a).

DG solutions (PADGE) is compared with the 2$^{\text{nd}}$ order FV solutions (TAU) on a sequence of meshes. We see that adjoint-based mesh refinement leads to “grid converged“ force coefficients on significantly coarser meshes than the other computations.

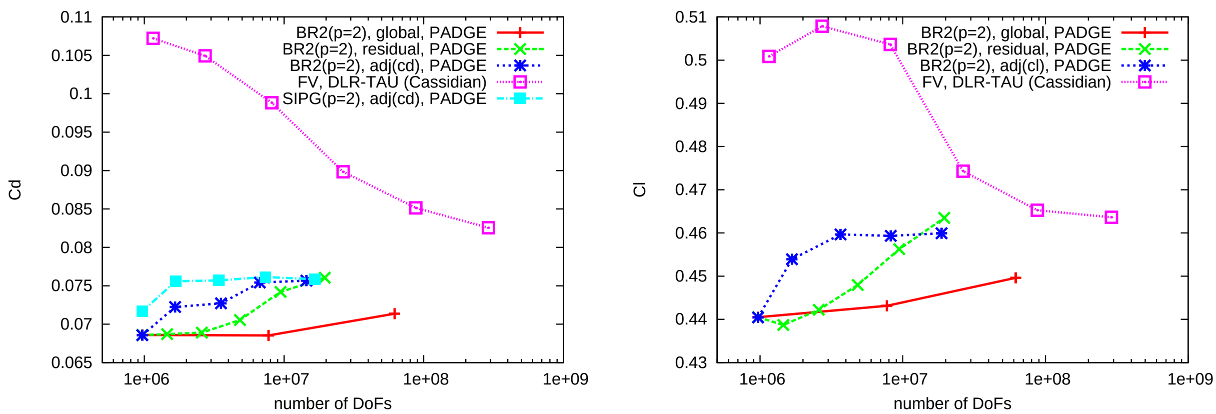
### 9.7.2 Transonic turbulent flow around the VFE-2 delta wing

In this example, taken from (Hartmann, 2013b), we consider a transonic fully turbulent flow at a Mach number $M = 0.8$, a Reynolds number $Re = 2 \times 10^6$ and an angle of attack of $\alpha = 20.5°$ around the VFE-2 configuration with medium rounded leading edge. This is the U1.c test case considered in the EU-project IDIHOM (Kroll, 2012). This vortex dominated flow features a system of vortices and a shock. The following numerical results are based on the discretization (300) augmented with the artificial viscosity based shock-capturing described in Section 7.3.

First we consider a discrete solution $\mathbf{u}_h \in V_h^3$ of the discretization, i.e., a 4$^{\text{th}}$-order flow solution, on a mesh with 110 528 curved elements which is obtained after one global mesh refinement step of the coarse mesh with 13 816 curved elements depicted in Figure 40.

| IDIHOM test case U1.b | DLR-TAU Code | DLR-PADGE Code |
|---|---|---|
| numerical scheme | finite volume | discontinuous Galerkin |
| design order | 2 | 3 |
| grids | hybrid unstructured<br>linear elements<br>grid sequence | hexahedral<br>q4 elements<br>refinement of starting grid |
| # elements | $0.6 - 146 \times 10^6$ | $14 - 884 \times 10^3$ (global ref.)<br>$14 - 280 \times 10^3$ (local ref.) |
| degrees of freedom | 7 per node | 70 per element |
| $\sum$ degrees of freedom | $1.2 - 290 \times 10^6$ | $1.6 - 62 \times 10^6$ (global ref.)<br>$1.6 - 20 \times 10^6$ (local ref.) |

Table 4: Subsonic turbulent flow around the VFE-2 delta wing: Characteristic data of the comparison of local mesh refinement (residual-based and adjoint-based) with global mesh refinement and the computations with the DLR-TAU Code.




(a) (b)

Figure 46: Subsonic turbulent flow around the VFE-2 delta wing(IDIHOM test case U1.b): Convergence of (a) the drag and (b) the lift coefficient under mesh refinement. Residual-based, adjoint-based and global mesh refinement of 3$^{rd}$ order DG solutions (PADGE), and 2$^{nd}$ order FV solutions (TAU) on a sequence of meshes.

The resulting $c_p$-distribution on the upper side of the wing is shown in the left part of Figure 47(a) compared to the pressure sensitive paint (PSP) measurements (Konrath et al., 2006) in the right part of Figure 47(a). In the measurements as well as in the $4^{\text{th}}$-order flow solution on this rather coarse mesh we recognize the suction trace of the strong primary vortex and an onset of a weaker secondary vortex. We note, that the suction trace of the computed primary vortex is much stronger than that in the measurements.

Again starting from the coarse mesh with 13 816 curved elements we now perform 3 residual-based mesh refinement steps. The left part of Figure 47(b) shows the $c_p$-distribution of the discrete solution $\mathbf{u}_h \in V_h^3$ on the resulting locally refined mesh of 93 433 curved elements in comparison to the PSP measurements in the right part of the Figure 47(b). Although the number of elements is smaller than in the case of the globally refined mesh we see that the solution quality is significantly increased. In particular, the strength of the suction trace of the primary vortex is reduced and that of the secondary vortex is increased. Overall, the strength of both suction traces under local mesh refinement in Figure 47(b) is closer to that of the experiment than in case of the globally refined mesh in Figure 47(a). Furthermore, more similarities of the $c_p$-distribution near the trailing edge is seen on the locally refined mesh than on the globally refined mesh.

Finally, Figure 47(c) shows the corresponding plot for a once more locally refined mesh. After in total 4 residual-based mesh refinement steps the mesh includes 201 259 curved elements. In comparison to the $c_p$-distribution in Figure 47(b) we see that solution features in Figure 47(c) are sharpened. Furthermore, the strength of the suction trace of the primary vortex is further reduced getting closer to that of the experiments.

Figure 48 shows the $c_p$-distribution on the wing together with slices of the $\lambda_2$-criterion. The primary and secondary vortices are clearly visible. The resolution of the numerical solution is high enough such that some vorticity even at the position of a possible tertiary vortex can be recognized. At the symmetry plane the Mach number distribution is shown. Furthermore, on the symmetry plane and the wing the isoline of the critical $c_p$-value is shown. In the symmetry plane one clearly recognizes the region of supersonic flow over the wing which decelerates through a shock. As the flow passes the front of the sting it accelerates again to supersonic flow and then decelerates smoothly, i.e., without shock, to subsonic flow.

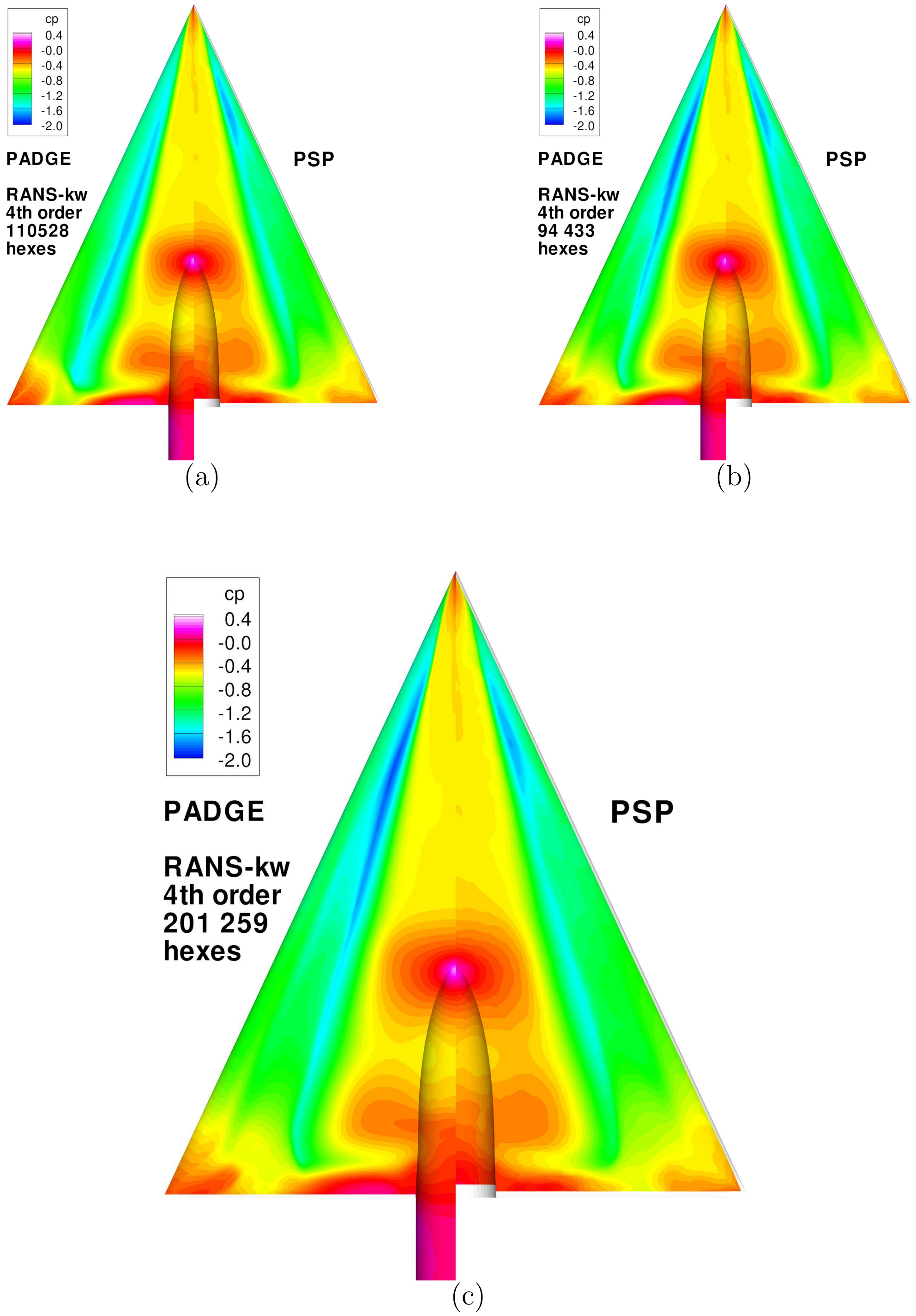



Figure 47: Transonic turbulent flow around the VFE-2 delta wing: The $c_p$-distribution on the upper side of the wing at $M = 0.8$, $\alpha = 20.5°$, $Re = 2 \times 10^6$ computed with the PADGE code in comparison to PSP measurements: (a) the 4$^{\text{th}}$-order DG solution on a once globally refined mesh with 110 528 curved elements; (b) the 4$^{\text{th}}$-order DG solution on a 3 times residual-based adapted mesh with 93 433 curved elements, and (c) the 4$^{\text{th}}$-order DG solution on a 4 times residual-based adapted mesh with 201 259 curved elements (Hartmann, 2013b).

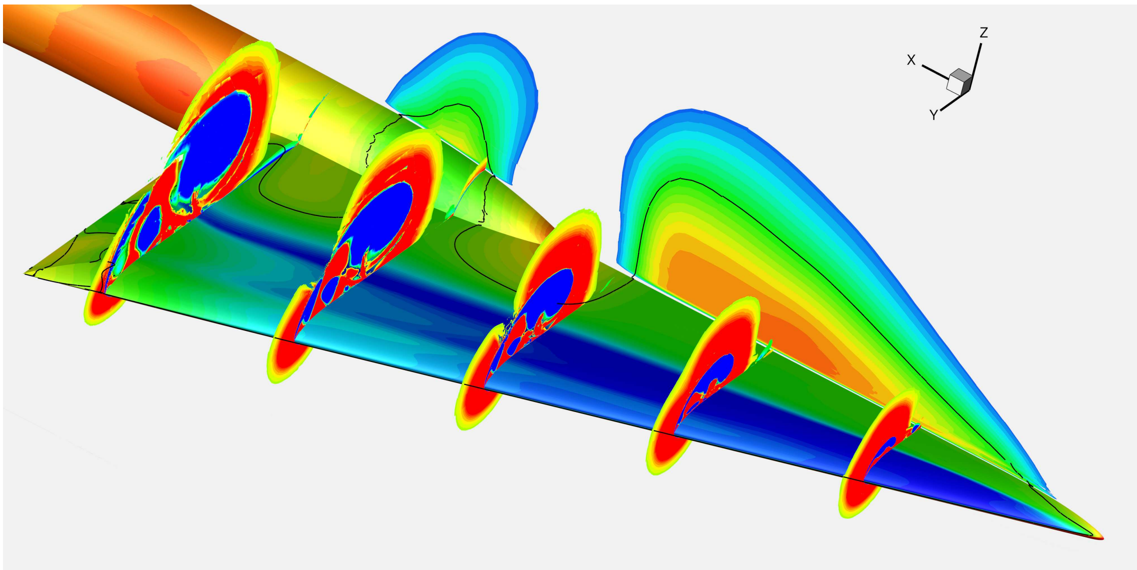



Figure 48: Transonic turbulent flow around the VFE-2 delta wing: The $c_p$-distribution and slices of the $\lambda_2$-criterion of a $4^{\text{th}}$-order DG solution on a 4 times residual-based adapted mesh with 201 259 curved elements (Hartmann, 2013b).

# References


AGARD Report AR-303 (1994). A selection of experimental test cases for the validation of CFD codes.

Arnold, D., Brezzi, F., Cockburn, B., and Marini, L. (2002). Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779.

Babuška, I. and Suri, M. (1987). The *hp*–version of the finite element method with quasiuniform meshes. *$M^2AN$ Mathematical Modeling and Numerical Analysis*, 21:199–238.

Bangerth, W., Hartmann, R., and Kanschat, G. (2007). deal.II – A general purpose object oriented finite element library. *ACM Transactions on Mathematical Software*, 33(4):24:1–24:27.

Bassi, F., Botti, L., Colombo, A., Crivellini, A., Franchina, N., Ghidoni, A., and Rebay, S. (2010). Very high-order accurate discontinuous Galerkin computation of transonic turbulent flows on aeronautical configurations. In Kroll, N., Bieler, H., Deconinck, H., Couallier, V., van der Ven, H., and Sorensen, K., editors, *ADIGMA - A European Initiative on the Development of Adaptive Higher-Order Variational Methods for Aerospace Applications*, volume 113 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, pages 25–38. Springer.

Bassi, F., Botti, L., Colombo, A., Ghidoni, A., and Rebay, S. (2011). Discontinuous Galerkin for turbulent flows. In Wang, Z. J., editor, *Adaptive High-Order Methods in Computational Fluid Dynamics*, volume 2 of *Advances in Computational Fluid Dynamics*, chapter 1, pages 1–32. World Science Books.

Bassi, F., Crivellini, A., Ghidoni, A., and Rebay, S. (2009). High-order discontinuous Galerkin discretization of transonic turbulent flows. 47th AIAA Aerospace Sciences Meeting, AIAA 2009-180.

Bassi, F., Crivellini, A., Rebay, S., and Savini, M. (2005). Discontinuous Galerkin solution of the Reynolds-averaged Navier-Stokes and $k - \omega$ turbulence model equations. *Computers & Fluids*, 34:507–540.

Bassi, F. and Rebay, S. (1997a). A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *J. Comp. Phys.*, 131:267–279.

Bassi, F. and Rebay, S. (1997b). High-order accurate discontinuous finite element solution of the 2d Euler equations. *J. Comp. Phys.*, 138:251–285.

Bassi, F. and Rebay, S. (2000a). GMRES discontinuous Galerkin solution of the compressible Navier-Stokes equations. In Cockburn, B., Karniadakis, G., and Shu, C.-W., editors, *Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Comput. Sci. Engrg.*, pages 197–208. Springer.

Bassi, F. and Rebay, S. (2000b). A high order discontinuous Galerkin method for compressible turbulent flows. In Cockburn, B., Karniadakis, G., and Shu, C.-W., editors, *Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Comput. Sci. Engrg.*, pages 77–79. Springer.

Bassi, F. and Rebay, S. (2002). Numerical evaluation of two discontinuous Galerkin methods for the compressible Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 40:197–207.

Bassi, F., Rebay, S., Mariotti, G., Pedinotti, S., and Savini, M. (1997). A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows. In Decuypere, R. and Dibelius, G., editors, *2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, Antwerpen, Belgium, March 5–7, 1997*, pages 99–108. Technologisch Instituut.

Becker, R. and Rannacher, R. (1996). A feed-back approach to error control in finite element methods: Basic analysis and examples. *East–West J. Numer. Math.*, 4:237–264.

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

Brezzi, F., Manzini, G., Marini, D., Pietra, P., and Russo, A. (2000). Discontinuous Galerkin approximations for elliptic problems. *Num. Meth. Part. Diff. Eq.*, 16(4):365–378.

Brezzi, F., Marini, L. D., and Süli, E. (2004). Discontinuous Galerkin methods for first-order hyperbolic problems. *Math. Models and Methods in Appl. Sci.*, 14(12):1893–1903.

Chiocchia, G. (1985). Exact solutions to transonic and supersonic flows. Technical Report AR-211, AGARD.

Essers, J. A., Delanaye, M., and Rogiest, P. (1995). Upwind-biased finite volume technique solving Navier-Stokes equations on irregular meshes. *AIAA Journal*, 33.

Fidkowski, K. J. and Darmofal, D. L. (2007). A triangular cut-cell adaptive method for high-order discretizations of the compressible Navier-Stokes equations. *J. Comput. Physics*, 225:1653–1672.

Fidkowski, K. J., Oliver, T. A., Lu, J., and Darmofal, D. L. (2005). $p$-multigrid solution of high-order discontinuous Galerkin discretizations of the compressible Navier-Stokes equations. *J. Comput. Phys.*, 207(1):92–113.

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

Harriman, K., Gavaghan, D., and Süli, E. (2004). The importance of adjoint consistency in the approximation of linear functionals using the discontinuous Galerkin finite element method. Technical report, Oxford University Computing Laboratory.

Harriman, K., Houston, P., Senior, B., and Süli, E. (2003). *hp*–Version discontinuous Galerkin methods with interior penalty for partial differential equations with nonnegative characteristic form. In *Recent Advances in Scientific Computing and Partial Differential Equations*, volume 330 of *Contemporary Mathematics*, pages 89–119. AMS.

Harten, A. and Hyman, J. (1983). Self adjusting grid methods for one-dimensional hyperbolic conservation laws. *J. Comput. Phys.*, 50(2):416–433.

Hartmann, R. (2001). Adaptive FE Methods for Conservation Equations. In Freistühler, H. and Warnecke, G., editors, *Hyperbolic Problems: theory, numerics, applications: eighth international conference in Magdeburg, February, March 2000*, volume 141 of *International series of numerical mathematics*, pages 495–503. Birkhäuser, Basel.

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

Hartmann, R. (2006). Adaptive discontinuous Galerkin methods with shock-capturing for the compressible Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 51(9–10):1131–1156.

Hartmann, R. (2007a). Adjoint consistency analysis of discontinuous Galerkin discretizations. *SIAM J. Numer. Anal.*, 45(6):2671–2696.

Hartmann, R. (2007b). Error estimation and adjoint based refinement for an adjoint consistent DG discretization of the compressible Euler equations. *Int. J. Computing Science and Mathematics*, 1(2–4):207–220.

Hartmann, R. (2008a). Multitarget error estimation and adaptivity in aerodynamic flow simulations. *SIAM J. Sci. Comput.*, 31(1):708–731.

Hartmann, R. (2008b). Numerical analysis of higher order discontinuous Galerkin finite element methods. In Deconinck, H., editor, *VKI LS 2008-08: 35$^{th}$ CFD/ADIGMA course on very high order discretization methods, Oct. 13-17, 2008*. Von Karman Institute for Fluid Dynamics, Rhode Saint Genèse, Belgium.

Hartmann, R. (2012). Higher-order and adaptive discontinuous Galerkin methods with shock-capturing applied to transonic turbulent delta wing flow. 50$^{\text{th}}$ AIAA Aerospace Sciences Meeting, AIAA 2012-459.

Hartmann, R. (2013a). Higher-order and adaptive discontinuous Galerkin methods applied to turbulent delta wing flow. In Dillmann, A., Heller, G., Kreplin, H.-P., Nitsche, W., and Peltzer, I., editors, *New Results in Numerical and Experimental Fluid Mechanics VIII: Contributions to the 17th STAB/DGLR Symposium Berlin, Germany, 2010*, volume 121 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, pages 497–505. Springer.

Hartmann, R. (2013b). Higher-order and adaptive discontinuous Galerkin methods with shock-capturing applied to transonic turbulent delta wing flow. *Int. J. Numer. Meth. Fluids*, 72(8):883–894.

Hartmann, R., Held, J., and Leicht, T. (2011). Adjoint-based error estimation and adaptive mesh refinement for the RANS and $k$-$\omega$ turbulence model equations. *J. Comput. Phys.*, 230(11):4268–4284.

Hartmann, R., Held, J., Leicht, T., and Prill, F. (2010a). Discontinuous Galerkin methods for computational aerodynamics – 3D adaptive flow simulation with the DLR PADGE code. *Aerosp. Sci. Technol.*, 14:512–519.

Hartmann, R., Held, J., Leicht, T., and Prill, F. (2010b). Error estimation and adaptive mesh refinement for aerodynamic flows. In Kroll, N., Bieler, H., Deconinck, H., Couallier, V., van der Ven, H., and Sorensen, K., editors, *ADIGMA - A European Initiative on the Development of Adaptive Higher-Order Variational Methods for Aerospace Applications*, volume 113 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, pages 339–353. Springer.

Hartmann, R. and Houston, P. (2002a). Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 24(3):979–1004.

Hartmann, R. and Houston, P. (2002b). Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations. *J. Comput. Phys.*, 183(2):508–532.

Hartmann, R. and Houston, P. (2003). Goal-oriented a posteriori error estimation for multiple target functionals. In Hou, T. Y. and Tadmor, E., editors, *Hyperbolic problems: theory, numerics, applications*, pages 579–588. Springer.

Hartmann, R. and Houston, P. (2006a). Symmetric interior penalty DG methods for the compressible Navier–Stokes equations I: Method formulation. *Int. J. Num. Anal. Model.*, 3(1):1–20.

Hartmann, R. and Houston, P. (2006b). Symmetric interior penalty DG methods for the compressible Navier–Stokes equations II: Goal–oriented a posteriori error estimation. *Int. J. Num. Anal. Model.*, 3(2):141–162.

Hartmann, R. and Houston, P. (2008). An optimal order interior penalty discontinuous Galerkin discretization of the compressible Navier–Stokes equations. *J. Comput. Phys.*, 227(22):9670–9685.

Hartmann, R. and Houston, P. (2010). Error estimation and adaptive mesh refinement for aerodynamic flows. In Deconinck, H., editor, *VKI LS 2010-01: 36$^{th}$ CFD/ADIGMA course on hp-adaptive and hp-multigrid methods, Oct. 26-30, 2009*. Von Karman Institute for Fluid Dynamics, Rhode Saint Genèse, Belgium.

Hillewaert, K. (2013). *Development of the Discontinuous Galerkin Method for high-resolution, large scale CFD and acoustics in industrial geometries*. PhD thesis, Ecole polytechnique de Louvain/iMMC.

Houston, P. and Süli, E. (2001). *hp*–adaptive discontinuous Galerkin finite element methods for first–order hyperbolic problems. *SIAM J. Sci. Comp.*, 23(4):1226–1252.

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

Hummel, D. and Redeker, G. (2001). A new vortex flow experiment for computer code validation. In *RTO-AVT Symposium on "Vortex fow and high angle of attack"*, Loen, Norway.

Ilinca, F. and Pelletier, D. (1998). Positivity preservation and adaptive solution for the $k-\epsilon$ model of turbulence. *AIAA J.*, 36(1):44–50.

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

Johnson, C., Szepessy, A., and Hansbo, P. (1990). On the convergence of shock–capturing streamline diffusion finite element methods for hyperbolic conservation laws. *Math. Comp.*, 54:107–129.

Klaij, C. M., van der Vegt, J. J. W., and van der Ven, H. (2006). Space–time discontinuous Galerkin method for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 217(2):589–611.

Konrath, R., Klein, C., Engler, R., and Otter, D. (2006). Analysis of PSP results obtained for the VFE-2 65° delta wing configuration at sub- and transonic speeds. 44th AIAA Aerospace Sciences Meeting and Exhibit. AIAA 2006-59-624.

Kroll, N. (2012). IDIHOM - european project on industrialization of high-order methods for aeronautical applications. In *Proceedings of the ECCOMAS 2012, September 10-14, Vienna, Austria.*

Kroll, N., Bieler, H., Deconinck, H., Couallier, V., van der Ven, H., and Sorensen, K., editors (2010). *ADIGMA - A European Initiative on the Development of Adaptive Higher-Order Variational Methods for Aerospace Applications*, volume 113 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*. Springer.

Leicht, T. and Hartmann, R. (2010). Error estimation and anisotropic mesh refinement for 3d laminar aerodynamic flow simulations. *J. Comput. Phys.*, 229(19):7344–7360.

Leicht, T. and Hartmann, R. (2011). Error estimation and *hp*–adaptive mesh refinement for discontinuous Galerkin methods. In Wang, Z. J., editor, *Adaptive High-Order Methods in Computational Fluid Dynamics*, volume 2 of *Advances in Computational Fluid Dynamics*, chapter 3, pages 67–94. World Science Books.

Lu, J. (2005). *An a posteriori Error Control Framework for Adaptive Precision Optimization using Discontinuous Galerkin Finite Element Method*. PhD thesis, M.I.T.

Mulder, W. A. and van Leer, B. (1985). Experiments with implicit upwind methods for the Euler equations. *J. Comput. Phys.*, 59:232–246.

Oliver, T. A. and Darmofal, D. L. (2009). Analysis of dual-consistency for discontinuous Galerkin discretizations of source terms. *SIAM Journal of Numerical Analysis*, 47(5):3507–3525.

Peterson, T. E. (1991). A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation. *SIAM J. Numer. Anal.*, 28(1):133–140.

Prudhomme, S., Pascal, F., Oden, J., and Romkes, A. (2000). Review of *a priori* error estimation for discontinuous Galerkin methods. TICAM Report 00-27, University of Texas.

Schlichting, H. and Gersten, K. (2003). *Boundary-Layer Theory*. Springer.

Schütte, A. and Lüdeke, H. (2009). Numerical investigations on the VFE-2 65-degree rounded leading edge delta wing using the unstructured DLR-TAU-Code. 46th AIAA Aerospace Sciences Meeting and Exhibit. AIAA 2009-398-883.

Schwab, C. (1998). *p- and hp-Finite Element methods. Theory and Applications to Solid and Fluid Mechanics*. Oxford University Press.

Shahbazi, K. (2005). An explicit expression for the penalty parameter of the interior penalty method. *J. Comput. Phys.*, 205:401–407.

van der Vegt, J. and van der Ven, H. (2002). Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows, I. General formulation. *J. Comp. Phys.*, 182:546–585.

van der Zee, K. G. (2004). An $H^1(P^h)$-coercive discontinuous Galerkin formulation for the Poisson problem: 1-d analysis. Master's thesis, TU Delft.

Wallraff, M., Leicht, T., and Lange-Hegermann, M. (2013). Numerical flux functions for Reynolds-averaged Navier-Stokes and $k\omega$ turbulence model computations with a line-preconditioned p-multigrid discontinuous Galerkin solver. *Int. J. Numer. Meth. Fluids*, 71(8):1055–1072.

Wang, Z. J., Fidkowski, K., Abgrall, R., Bassi, F., Caraeni, D., Cary, A., Deconinck, H., Hartmann, R., Hillewaert, K., Huynh, H. T., Kroll, N., May, G., Persson, P.-O., van Leer, B., and Visbal, M. (2013). High-order CFD methods: current status and perspective. *Int. J. Num. Meth. Fluids*, 72(8):811–845.

Wilcox, D. C. (1988). Reassessment of the scale-determining equation for advanced turbulence models. *AIAA J.*, 26(11):1299–1310. 1988.

Wilcox, D. C. (1993). *Turbulence Modeling for CFD*. DCW Industries, Inc., La Canada CA.

# A Appendix

## A.1 Mesh related function spaces

We assume that the domain $\Omega$ can be subdivided into a shape regular mesh $\mathcal{T}_h = \{\kappa\}$ consisting of elements $\kappa$. Here, $h$ denotes the piecewise constant mesh function defined by $h|_\kappa \equiv h_\kappa = \text{diam}(\kappa)$ for all $\kappa \in \mathcal{T}_h$.

In the following we define some broken (mesh related) function spaces on $\mathcal{T}_h$:

**Definition A.1 (Broken Sobolev space $H^m(\mathcal{T}_h)$)** *By $H^m(\mathcal{T}_h)$ we denote the space of $L^2$ functions on $\Omega$ whose restriction to each element $\kappa$ belongs to the Sobolev space $H^m(\kappa)$, i.e.,*

$$H^m(\mathcal{T}_h) = \{v \in L^2(\Omega) : v|_\kappa \in H^m(\kappa), \kappa \in \mathcal{T}_h\}. \tag{348}$$

**Definition A.2 (Interior faces: $\Gamma_\mathcal{I}$)** *Let $\kappa$ and $\kappa'$ be two adjacent elements of $\mathcal{T}_h$ with common edge (interior face) $e = \partial\kappa \cap \partial\kappa'$. We define $\Gamma_\mathcal{I}$ to be the union of all interior faces of $\mathcal{T}_h$.*

**Definition A.3 (Traces $u_\kappa^+$, $u_\kappa^-$ and the space $T(\mathcal{T}_h)$)** *Suppose that $u \in H^1(\mathcal{T}_h)$, i.e., $u|_\kappa \in H^1(\kappa)$ for each $\kappa \in \mathcal{T}_h$. By $u_\kappa^\pm$ (simplified notation: $u^\pm$) we denote the traces of $u$ taken from within the interior of $\kappa$ and of the neighboring element, see Figure 49. For simplicity of notation we can obmit the $\kappa$-subscript in case it is clear relative to which element the traces are defined. We note, that for $u|_\kappa \in H^1(\kappa)$ the trace $u_\kappa^+$ belongs to $L^2(\partial\kappa)$, and traces of $u \in H^1(\mathcal{T}_h)$ belong to $T(\mathcal{T}_h) := \prod_{\kappa\in\mathcal{T}_h} L^2(\partial\kappa)$.*

Finally, we define mesh related (or broken) gradient, divergence and Laplace operators.

**Definition A.4 ($\nabla_h$, $\nabla_h\cdot$ and $\Delta_h$)** *We define broken operators by restriction to each element $\kappa \in \mathcal{T}_h$ as follows:*

- *The broken gradient operator $\nabla_h : H^1(\mathcal{T}_h) \to [L^2(\mathcal{T}_h)]^d$ is defined by*

$$(\nabla_h v)|_\kappa := \nabla(v|_\kappa), \quad \kappa \in \mathcal{T}_h, \tag{349}$$

  *for $v \in H^1(\mathcal{T}_h)$, where $(\nabla v)_i = \partial_{x_i} v, i = 1, \dots, d$.*

- *The broken divergence operator $\nabla_h\cdot : [H^1(\mathcal{T}_h)]^d \to L^2(\mathcal{T}_h)$ is defined by*

$$(\nabla_h \cdot \tau)|_\kappa = \nabla \cdot (\tau|_\kappa), \quad \kappa \in \mathcal{T}_h, \tag{350}$$

  *for $\tau \in [H^1(\mathcal{T}_h)]^d$, where $\nabla \cdot \tau = \sum_{1\le i\le d} \partial_{x_i}\tau_i$.*

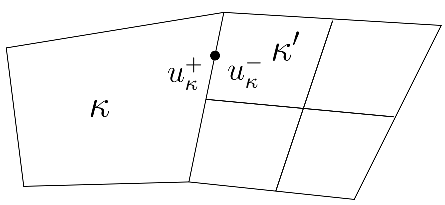



Figure 49: Definition of the interior and exterior traces $u_\kappa^\pm$ wrt. element $\kappa$.

- *Finally, the broken Laplace operator* $\Delta_h : H^2(\mathcal{T}_h) \to L^2(\mathcal{T}_h)$ *is defined by*

$$(\Delta_h u)|_\kappa := \Delta(u|_\kappa), \quad \kappa \in \mathcal{T}_h, \tag{351}$$

*for* $u \in H^2(\mathcal{T}_h)$, *where* $\Delta u = \nabla \cdot \nabla u = \sum_{1 \le i \le d} \partial^2_{x_i} u$.

## A.2 The local $L^2$-projection and approximation estimates

**Definition A.5 (Local $L^2$-projection)** *Let* $p \geq 0$ *and* $V_h^p$ *be the discontinuous finite element space defined in* (52). *Then, by* $P_h^p$ *we denote the* $L^2$*-projection onto* $V_h^p$, *i.e., given a* $u \in L^2(\Omega)$ *we define* $P_h^p u \in V_h^p$ *by*

$$\int_\Omega (u - P_h^p u)\, v_h \, d\mathbf{x} = 0 \quad \forall v_h \in V_h^p. \tag{352}$$

*We use the short notation* $P_h u$ *instead of* $P_h^p u$ *when it is clear which projection is meant.*

Given a $\kappa \in \mathcal{T}_h$, we set $v_h \equiv 0$ on $\kappa' \in \mathcal{T}_h$ with $\kappa' \neq \kappa$ in (352) and see that $P_h$ has following *local projection* property: For any $\kappa \in \mathcal{T}_h$ we have

$$\int_\kappa (u - P_h u)\, v_h \, \mathrm{d}\mathbf{x} = 0 \quad \forall v_h \in V_h^p. \tag{353}$$

$P_h$ restricted to $\kappa \in \mathcal{T}_h$ is an $L^2(\kappa)$-projection which is why $P_h$ is also called *local* $L^2$-projection. There are following approximation estimates for the $L^2$-projection:

**Corollary A.6 (Local approximation estimates for the $L^2$-projection)** *Let* $p \geq 0$ *and* $P_h u := P_h^p u$ *be the* $L^2$*-projection defined in Definition A.5. Suppose* $u|_\kappa$ *in* $H^{s_\kappa+1}(\kappa)$, $s_\kappa \geq 0$, *for* $\kappa \in \mathcal{T}_h$. *Then*

$$\|u - P_h u\|_{H^m(\kappa)} \leq C h_\kappa^{t_\kappa+1-m} |u|_{H^{t_\kappa+1}(\kappa)}, \tag{354}$$

*where* $t_\kappa = \min(s_\kappa, p)$, $\kappa \in \mathcal{T}_h$.

Again, for sufficiently smooth functions $u \in H^{p+1}(\kappa)$, i.e., $s_\kappa \geq p$, this estimate reduces to

$$\|u - P_h u\|_{H^m(\kappa)} \leq C h_\kappa^{p+1-m} |u|_{H^{p+1}(\kappa)}, \tag{355}$$

while for functions with a lower smoothness, i.e., $u \in H^{s_\kappa+1}(\kappa)$ with $s_\kappa < p$, we have

$$\|u - P_h u\|_{H^m(\kappa)} \leq C h_\kappa^{s_\kappa+1-m} |u|_{H^{s+1}(\kappa)}, \tag{356}$$

for $m \leq s_\kappa + 1$. Furthermore, an analogous estimate holds in the $L^\infty$-norm:

$$\|u - P_h u\|_{L^\infty(\kappa)} \leq C h_\kappa^{p+1} |u|_{H^{p+1,\infty}(\kappa)}, \tag{357}$$

Furthermore, we have following approximation estimate in the $L^2(\partial\Omega)$-norm,

$$\|u - P_h u\|_{L^2(\partial\kappa)} \leq C h_\kappa^{p+1/2} |u|_{H^{p+1}(\kappa)}. \tag{358}$$

## A.3 DG discretization of the linear advection equation: Details

In this section we give the proofs which were left out in Section 3 for brevity.

**Theorem A.7 (Coercivity, (Brezzi et al., 2004))** *Let $L_h(\cdot,\cdot)$ be given by*

$$\begin{aligned} L_h(u,v) = &-\int_\Omega (\mathbf{b}u)\cdot\nabla_h v \, d\mathbf{x} + \int_\Omega cuv \, d\mathbf{x} \\ &+ \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa\setminus\Gamma} \hat{h}_{b_0}(u^+,u^-,\mathbf{n}) v \, ds + \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, uv \, ds, \end{aligned} \tag{359}$$

*where $\hat{h}_{b_0}$ as defined in* (65) *represents the mean value flux or the upwind flux depending on $b_0 = 0$ or $b_0 = \frac{1}{2}|\mathbf{b}\cdot\mathbf{n}|$, respectively. Then for all $v \in H^{1,\mathbf{b}}(\mathcal{T}_h)$, with*

$$H^{1,\mathbf{b}}(\mathcal{T}_h) = \{v \in L^2(\Omega) : \nabla\cdot(\mathbf{b}v)\,|_\kappa \in L^2(\kappa), \kappa\in\mathcal{T}_h\},$$

*we have*

$$L_h(v,v) = \|c_0 v\|^2 + \sum_{e\in\Gamma_\mathcal{I}} \int_e b_0\,[v]^2 \, ds + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\, v^2 \, ds. \tag{360}$$

**Proof:** First we rewrite

$$\begin{aligned} -\int_\kappa (\mathbf{b}v)\cdot\nabla v \, \mathrm{d}\mathbf{x} &= -\frac{1}{2}\int_\kappa \mathbf{b}\cdot\nabla v^2 \, \mathrm{d}\mathbf{x} \\ &= -\frac{1}{2}\int_\kappa \nabla\cdot\left(\mathbf{b}v^2\right) \, \mathrm{d}\mathbf{x} + \frac{1}{2}\int_\kappa (\nabla\cdot\mathbf{b})\, v^2 \, \mathrm{d}\mathbf{x} \\ &= -\frac{1}{2}\int_{\partial\kappa} \mathbf{b}\cdot\mathbf{n}(v^+)^2 \, \mathrm{d}s + \frac{1}{2}\int_\kappa (\nabla\cdot\mathbf{b})\, v^2 \, \mathrm{d}\mathbf{x}. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} &-\frac{1}{2}\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa} \mathbf{b}\cdot\mathbf{n}(v^+)^2 \, \mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} \mathbf{b}\cdot\mathbf{n}\,\{\!\{v\}\!\}\, v \, \mathrm{d}s \\ &\quad = -\frac{1}{2}\sum_{e\in\Gamma_\mathcal{I}}\int_e \mathbf{b}\cdot\mathbf{n}^+\left((v^+)^2-(v^-)^2\right) \, \mathrm{d}s + \sum_{e\in\Gamma_\mathcal{I}} \mathbf{b}\cdot\mathbf{n}^+\frac{1}{2}(v^++v^-)(v^+-v^-) \, \mathrm{d}s = 0. \end{aligned}$$

Thereby,

$$\begin{aligned} L_h(v,v) &= -\int_\Omega (\mathbf{b}v)\cdot\nabla_h v \, \mathrm{d}\mathbf{x} + \int_\Omega cv^2 \, \mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} \left(\mathbf{b}\cdot\mathbf{n}\,\{\!\{v\}\!\} + b_0\,[v]\right) v \, \mathrm{d}s + \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, v^2 \, \mathrm{d}s \\ &= \int_\Omega \left(c + \frac{1}{2}\nabla\cdot\mathbf{b}\right) v^2 \, \mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma} b_0\,[v] v \, \mathrm{d}s + \int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, v^2 \, \mathrm{d}s - \frac{1}{2}\int_\Gamma \mathbf{b}\cdot\mathbf{n}\, v^2 \, \mathrm{d}s \\ &= \int_\Omega c_0^2 v^2 \, \mathrm{d}\mathbf{x} + \sum_{e\in\Gamma_\mathcal{I}} b_0\,[v][v] \, \mathrm{d}s + \frac{1}{2}\int_{\Gamma_+} \mathbf{b}\cdot\mathbf{n}\, v^2 \, \mathrm{d}s - \frac{1}{2}\int_{\Gamma_-} \mathbf{b}\cdot\mathbf{n} v^2 \, \mathrm{d}s \\ &= \|c_0 v\|^2 + \sum_{e\in\Gamma_\mathcal{I}}\int_e b_0\,[v]^2 \, \mathrm{d}s + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\, v^2 \, \mathrm{d}s, \end{aligned}$$

where we used hypothesis (47). □

**Definition A.8** *Motivated by the coercivity* (360) *we define the DG-norm* $|||\cdot|||_{b_0}$ *by*

$$|||v|||_{b_0}^2 = \|c_0 v\|^2 + \sum_{e\in\Gamma_{\mathcal{I}}} \int_e b_0\,[v]^2\,ds + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\,v^2\,ds. \tag{361}$$

From the coercivity of $L_h$, (360), we immediately obtain the stability of the discontinuous Galerkin discretization in the DG-norm $|||\cdot|||_{b_0}$ as follows:

$$\begin{aligned}|||v|||_{b_0}^2 = L_h(v,v) = F(v) &= \int_\Omega f v\,\mathrm{d}\mathbf{x} + \int_{\Gamma_-} |\mathbf{b}\cdot\mathbf{n}|\,gv\,\mathrm{d}s \\ &\le \|f\|_{L^2(\Omega)}\,\|v\|_{L^2(\Omega)} + \left(\int_{\Gamma_-} |\mathbf{b}\cdot\mathbf{n}|\,g^2\,\mathrm{d}s\right)^{1/2}\left(\int_{\Gamma_-} |\mathbf{b}\cdot\mathbf{n}|\,v^2\,\mathrm{d}s\right)^{1/2} \le C|||v|||_{b_0},\end{aligned}$$

for $f\in L^2(\Omega)$ and $g\in L^2(\Gamma_-)$. After division by $|||v|||_{b_0}$ we obtain $|||v|||_{b_0}\le C$, and hence

$$|||v|||_{b_0}^2 = \|c_0 v\|^2 + \sum_{e\in\Gamma_{\mathcal{I}}} \int_e b_0\,[v]^2\,\mathrm{d}s + \frac{1}{2}\int_\Gamma |\mathbf{b}\cdot\mathbf{n}|\,v^2\,\mathrm{d}s \le C^2. \tag{362}$$

**Theorem A.9 (A priori error estimate, (Brezzi et al., 2004))** *Let* $u\in H^{p+1}(\Omega)$ *be the exact solution to the linear advection equation* (45)*. Furthermore, let* $u_h\in \tilde{V}_h^p := \{v_h\in L^2(\Omega) : v_h|_\kappa \in P_p(\kappa), \kappa\in\mathcal{T}_h\}$ *be the solution to*

$$L_h(u_h,v_h) = \int_\Omega f v_h\,d\mathbf{x} - \int_{\Gamma_-} \mathbf{b}\cdot\mathbf{n}\,g v_h\,ds \qquad \forall v_h\in\tilde{V}_h^p,$$

*where* $L_h(\cdot,\cdot)$ *is the bilinear form given in* (67)*.*

*Then, for* $b_0 = c_e|\mathbf{b}\cdot\mathbf{n}|$ *with* $c_e>0$ *(which for* $c_e = 1/2$ *gives the* upwind flux*) we have*

$$|||u-u_h|||_{b_0} \le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}, \tag{363}$$

*and for* $b_0 = 0$*, i.e., when using the* mean value flux*, we have*

$$|||u-u_h|||_{b_0} \le Ch^{p}|u|_{H^{p+1}(\Omega)}, \tag{364}$$

*where* $|||\cdot|||_{b_0}^2$ *is the DG-norm defined in Definition A.8.*

**Proof:** Let $e = u-u_h = \eta-\xi$ with $\eta = u - P_h u$ and $\xi = u_h - P_h u$ where $P_h := P_h^p$ is the $L^2$-projection onto $V_h := \tilde{V}_h^p$ as defined in Definition (A.5). Then, by triangle inequality,

$$|||e|||_{b_0} \le |||\eta|||_{b_0} + |||\xi|||_{b_0}. \tag{365}$$

For the first term we use approximation results for $P_h u$:

$$\begin{aligned}\|c_0\eta\|_{L^2(\Omega)} &\le C\|u-P_h u\|_{L^2(\Omega)} \le Ch^{p+1}|u|_{H^{p+1}(\Omega)},\\ \left(\sum_{e\in\Gamma_{\mathcal{I}}}\int_e b_0\,[\eta]^2\,\mathrm{d}s\right)^{1/2} &\le C\sum_{e\in\Gamma_{\mathcal{I}}}\|u-P_h u\|_{L^2(e)} \le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)},\\ \left(\frac{1}{2}\int_{\Gamma_-}|\mathbf{b}\cdot\mathbf{n}|\,\eta^2\,\mathrm{d}s\right)^{1/2} &\le C\|u-P_h u\|_{L^2(\partial\Omega)} \le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)},\end{aligned} \tag{366}$$

see (355) for $m = 0$ and (358), and hence

$$|||\eta|||_{b_0} \leq h^{p+1/2}|u|_{H^{p+1}(\Omega)}. \tag{367}$$

The second term in (365) we rewrite as follows

$$|||\xi|||_{b_0}^2 = L_h(\xi, \xi) = L_h(\eta - e, \xi) = L_h(\eta, \xi), \tag{368}$$

where we used coercivity (360) of $L_h$ and the Galerkin orthogonality property $L_h(e, \xi) = 0$ for $\xi \in V_h$. Using the definition of $L_h(\cdot, \cdot)$ we obtain

$$|||\xi|||_{b_0}^2 = \int_\Omega \eta\left(-\mathbf{b}\cdot\nabla_h\xi + c\xi\right)\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\left(\mathbf{b}\cdot\mathbf{n}\,\{\!\{\eta\}\!\} + b_0\,[\eta]\right)\xi\,\mathrm{d}s + \int_{\Gamma_+}\mathbf{b}\cdot\mathbf{n}\,\eta\xi\,\mathrm{d}s.$$

Next observe that $\nabla_h\xi \in V_h$, so that, by the definition of the projector $P_{h,0}^d$,

$$\int_\kappa \left(P_{h,0}^d\mathbf{b}\cdot\nabla\xi\right)\eta\,\mathrm{d}\mathbf{x} = 0.$$

Using this, together with the approximation estimate (357) for $p = 0$, the Cauchy-Schwarz inequality, the inverse inequality $\|\xi\|_{H^1(\kappa)} \leq Ch_\kappa^{m-1}\|\xi\|_{H^m(\kappa)}$ for $\xi \in V_h$, and the approximation estimate (355) for $m = 0$, we deduce that

$$\begin{aligned}
\int_\Omega \eta\left(-\mathbf{b}\cdot\nabla_h\xi + c\xi\right)\,\mathrm{d}\mathbf{x} &= \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \eta\left(-\mathbf{b}\cdot\nabla\xi + c\xi\right)\,\mathrm{d}\mathbf{x} \\
&= \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \eta\left(\left(P_{h,0}^d\mathbf{b} - \mathbf{b}\right)\cdot\nabla\xi + c\xi\right)\,\mathrm{d}\mathbf{x} \\
&\leq C\|\eta\|_{L^2(\Omega)}\left(\sum_{\kappa\in\mathcal{T}_h}\left(\|P_h^0\mathbf{b} - \mathbf{b}\|_{L^\infty(\kappa)}|\xi|_{H^1(\kappa)} + \|\xi\|_{L^2(\kappa)}\right)^2\right)^{\frac{1}{2}} \\
&\leq C\|\eta\|_{L^2(\Omega)}\left(\sum_{\kappa\in\mathcal{T}_h}\left(h_\kappa|\mathbf{b}|_{H^{1,\infty}(\kappa)}h_\kappa^{-1}\|\xi\|_{L^2(\kappa)} + \|\xi\|_{L^2(\kappa)}\right)^2\right)^{\frac{1}{2}} \\
&\leq C\|\eta\|_{L^2(\Omega)}\|\xi\|_{L^2(\Omega)} \\
&\leq Ch^{p+1}|u|_{H^{p+1}(\Omega)}\,|||\xi|||_{b_0}.
\end{aligned} \tag{369}$$

Furthermore, using Cauchy-Schwarz inequality and approximation estimate (358) we find

$$\begin{aligned}
\int_{\Gamma_+}\mathbf{b}\cdot\mathbf{n}\,\eta\xi\,\mathrm{d}s &\leq \left(\int_{\Gamma_+}|\mathbf{b}\cdot\mathbf{n}|\,\eta^2\,\mathrm{d}s\right)^{\frac{1}{2}}\left(\int_{\Gamma_+}|\mathbf{b}\cdot\mathbf{n}|\,\xi^2\,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\leq Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}|||\xi|||_{b_0}.
\end{aligned} \tag{370}$$

Finally, we have

$$\sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa\setminus\Gamma}\left(\mathbf{b}\cdot\mathbf{n}\,\{\!\{\eta\}\!\} + b_0\,[\eta]\right)\xi\,\mathrm{d}s = \sum_{e\in\Gamma_\mathcal{I}}\int_e\left(\mathbf{b}\cdot\mathbf{n}\,\{\!\{\eta\}\!\} + b_0\,[\eta]\right)[\xi]\,\mathrm{d}s.$$

If $b_0 = c_e|\mathbf{b}\cdot\mathbf{n}|$ with $c_e > 0$ we have $\mathbf{b}\cdot\mathbf{n} \le |\mathbf{b}\cdot\mathbf{n}| = \frac{b_0}{c_e}$ and obtain

$$\begin{aligned}\int_e (\mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\} + b_0\,[\eta])\,[\xi]\,\mathrm{d}s &\le \int_e b_0^{1/2}\left(\tfrac{1}{c_e}\{\!\{\eta\}\!\} + [\eta]\right) b_0^{1/2}|[\xi]|\,\mathrm{d}s \\ &\le \left(\int_e b_0\left(\tfrac{1}{c_e}\{\!\{\eta\}\!\} + [\eta]\right)^2\mathrm{d}s\right)^{\frac{1}{2}}\left(\int_e b_0[\xi]^2\,\mathrm{d}s\right)^{\frac{1}{2}} \\ &\le Ch_\kappa^{p+1/2}|u|_{H^{p+1}(\kappa)}\left(\int_e b_0[\xi]^2\,\mathrm{d}s\right)^{\frac{1}{2}},\end{aligned}$$

and hence

$$\begin{aligned}\sum_{e\in\Gamma_{\mathcal{I}}}\int_e (\mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\} + b_0\,[\eta])\,[\xi]\,\mathrm{d}s &\le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}\left(\sum_{e\in\Gamma_{\mathcal{I}}}\int_e b_0[\xi]^2\,\mathrm{d}s\right)^{\frac{1}{2}} \\ &\le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}|||\xi|||_{b_0}.\end{aligned} \tag{371}$$

However, if $b_0 = 0$, the norm $|||\xi|||_{b_0}$ does not include $\int_e[\xi]^2\,\mathrm{d}s$ which is why we cannot bound $\int_e \mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\}[\xi]\,\mathrm{d}s$ in terms of $|||\xi|||_{b_0}$. Thereby, we are forced to use the inverse inequality

$$\|[\xi]\|_{L^2(e)} \le Ch_\kappa^{-\frac{1}{2}}\|\xi\|_{L^2(\kappa)},$$

to bound the $L^2(e)$-norm by the $L^2(\kappa)$-norm. Hence, instead of (371) we obtain

$$\int_e \mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\}[\xi]\,\mathrm{d}s \le C\|[\eta]\|_{L^2(e)}\|[\xi]\|_{L^2(e)} \le Ch^{p+1/2}|u|_{H^{p+1}(\Omega)}h_\kappa^{-\frac{1}{2}}\|\xi\|_{L^2(\kappa)},$$

and hence,

$$\begin{aligned}\sum_{e\in\Gamma_{\mathcal{I}}}\int_e \mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\}[\xi]\,\mathrm{d}s &\le Ch^p|u|_{H^{p+1}(\Omega)}\|\xi\|_{L^2(\kappa)} \\ &\le Ch^p|u|_{H^{p+1}(\Omega)}|||\xi|||_{b_0}.\end{aligned} \tag{372}$$

Combining (369), (370) and (371) we obtain (363), while (369), (370) and (372) gives (364). □

From the proof above we see that it is essential whether the term $\sum_{e\in\Gamma_{\mathcal{I}}}\int_e b_0\,[v]^2\,\mathrm{d}s$ is included in the DG-norm with $b_0 \ne 0$ or with $b_0 = 0$. In the former case we can bound the interior face terms

$$\sum_{e\in\Gamma_{\mathcal{I}}}\int_e (\mathbf{b}\cdot\mathbf{n}\{\!\{\eta\}\!\} + b_0\,[\eta])\,[\xi]\,\mathrm{d}s \tag{373}$$

in terms of $|||\xi|||_{b_0}$ whereas in the latter case we are forced to use the inverse inequality due to which we loose half an order of $h$.

## A.4 DG discretization of Poisson's equation: Details

In this section we give the details and proofs which were left out in Section 4 for brevity.

### A.4.1 Derivation of the face-based primal form

For rewriting the element-based primal formulation (cf. (87)): find $u_h \in V_h^p$ such that

$$\int_\Omega \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\boldsymbol{\sigma}}_h \cdot \mathbf{n}\, v_h \,\mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} (\hat{u}_h - u_h)\mathbf{n} \cdot \nabla_h v_h \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} \quad (374)$$

for all $v_h \in V_h^p$, into a face-based primal form, we recall the definition of mean value and jump operators in (88) and (89) for discrete functions $q_h \in V_h^p$ and $\boldsymbol{\phi}_h \in \boldsymbol{\Sigma}_h^p$, respectively. Note, that the according operators can be defined for $q \in T(\mathcal{T}_h)$ and $\boldsymbol{\phi} \in [T(\mathcal{T}_h)]^d$.

We begin by showing following result which will frequently be used to transfer between element-based and face-based forms.

**Lemma A.10** *Again, let $q \in T(\mathcal{T}_h)$ and $\boldsymbol{\phi} \in [T(\mathcal{T}_h)]^d$, then*

$$\sum_\kappa \int_{\partial\kappa\setminus\Gamma} \boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+ \, ds = \int_{\Gamma_\mathcal{I}} \{\!\{\boldsymbol{\phi}\}\!\} \cdot [\![q]\!] \, ds + \int_{\Gamma_\mathcal{I}} [\![\boldsymbol{\phi}]\!] \{\!\{q\}\!\} \, ds, \quad (375)$$

$$\sum_\kappa \int_{\partial\kappa} \boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+ \, ds = \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\boldsymbol{\phi}\}\!\} \cdot [\![q]\!] \, ds + \int_{\Gamma_\mathcal{I}} [\![\boldsymbol{\phi}]\!] \{\!\{q\}\!\} \, ds. \quad (376)$$

**Proof:** On $\Gamma_\mathcal{I}$ we have

$$\begin{aligned}
\{\!\{\boldsymbol{\phi}\}\!\} \cdot [\![q]\!] + [\![\boldsymbol{\phi}]\!] \{\!\{q\}\!\} =& \tfrac{1}{2}(\boldsymbol{\phi}^+ + \boldsymbol{\phi}^-) \cdot (q^+\mathbf{n}^+ + q^-\mathbf{n}^-) + \\
& \tfrac{1}{2}(\boldsymbol{\phi}^+ \cdot \mathbf{n}^+ + \boldsymbol{\phi}^- \cdot \mathbf{n}^-)(q^+ + q^-) \\
=& \tfrac{1}{2}(\boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+ + \boldsymbol{\phi}^- \cdot \mathbf{n}^+ q^+ + \boldsymbol{\phi}^+ \cdot \mathbf{n}^- q^- + \boldsymbol{\phi}^- \cdot \mathbf{n}^- q^-) + \\
& \tfrac{1}{2}(\boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+ + \boldsymbol{\phi}^- \cdot \mathbf{n}^- q^+ + \boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^- + \boldsymbol{\phi}^- \cdot \mathbf{n}^- q^-) \\
=& \boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+ + \boldsymbol{\phi}^- \cdot \mathbf{n}^- q^-
\end{aligned}$$

using $\mathbf{n}^- = -\mathbf{n}^+$ in the last identity. On $\Gamma$ we have $\{\!\{\boldsymbol{\phi}\}\!\} \cdot [\![q]\!] = \boldsymbol{\phi}^+ \cdot \mathbf{n}^+ q^+$. □

Using (376) and the Gauss integral formula we obtain following result.

**Corollary A.11** *Let $v \in H^1(\mathcal{T}_h)$ and $\tau \in [H^1(\mathcal{T}_h)]^d$, then*

$$\int_\Omega \tau \cdot \nabla_h v \, d\mathbf{x} = -\int_\Omega \nabla_h \cdot \tau \, v \, d\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\tau\}\!\} \cdot [\![v]\!] \, ds + \int_{\Gamma_\mathcal{I}} [\![\tau]\!] \{\!\{v\}\!\} \, ds. \quad (377)$$

**Proof:** Using the Gauss integral formula on each $\kappa \in \mathcal{T}_h$,

$$\int_\kappa \nabla \cdot \psi \,\mathrm{d}\mathbf{x} = \int_{\partial\kappa} \psi \cdot \mathbf{n} \,\mathrm{d}s, \quad (378)$$

for $\psi := \tau v \in [H^1(\mathcal{T}_h)]^d$, and summing over all $\kappa \in \mathcal{T}_h$ we obtain

$$\int_\Omega \nabla_h \cdot \tau v \,\mathrm{d}\mathbf{x} + \int_\Omega \tau \cdot \nabla_h v \,\mathrm{d}\mathbf{x} = \sum_\kappa \int_{\partial\kappa} \tau \cdot \mathbf{n}\, v \,\mathrm{d}s = \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\tau\}\!\} \cdot [\![v]\!] \,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} [\![\tau]\!] \{\!\{v\}\!\} \,\mathrm{d}s, \quad (379)$$

which shows (377). □

We now proceed in transferring the element-based form (374) into a face-based form. To this end, we use equation (376) twice (once for $\boldsymbol{\phi} = \hat{\boldsymbol{\sigma}}$ and $q = v$, and once for $\boldsymbol{\phi} = \nabla_h v$ and $q = \hat{u} - u$), and rewrite (374) as follows

$$\begin{aligned}\hat{L}_h(u,v) = &\int_\Omega \nabla_h u \cdot \nabla_h v \,\mathrm{d}\mathbf{x} - \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\hat{\boldsymbol{\sigma}}\}\!\} \cdot [\![v]\!] \,\mathrm{d}s - \int_{\Gamma_\mathcal{I}} [\![\hat{\boldsymbol{\sigma}}]\!] \{\!\{v\}\!\} \,\mathrm{d}s \\ &+ \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\nabla_h v\}\!\} \cdot [\![\hat{u} - u]\!] \,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} [\![\nabla_h v]\!] \{\!\{\hat{u} - u\}\!\} \,\mathrm{d}s,\end{aligned}$$

which results in following face-based *primal flux form*,

$$\begin{aligned}\hat{L}_h(u,v) = &\int_\Omega \nabla_h u \cdot \nabla_h v \,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma} \left([\![\hat{u} - u]\!] \cdot \{\!\{\nabla_h v\}\!\} - \{\!\{\hat{\boldsymbol{\sigma}}\}\!\} \cdot [\![v]\!]\right) \mathrm{d}s \\ &+ \int_{\Gamma_\mathcal{I}} \left(\{\!\{\hat{u} - u\}\!\} [\![\nabla_h v]\!] - [\![\hat{\boldsymbol{\sigma}}]\!] \{\!\{v\}\!\}\right) \mathrm{d}s,\end{aligned} \tag{380}$$

i.e., we have shown (90).

### A.4.2 Continuity of the bilinear forms

In the following we show that the bilinear form corresponding to the method by Baumann-Oden, and the bilinear forms of the SIPG, the NIPG and the BR2 methods are continuous.

**Lemma A.12 (Continuity of Baumann-Oden)** *Let $L_h$ be given by*

$$L_h(u,v) = \int_\Omega \nabla_h u \cdot \nabla_h v \, d\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \left(\theta [\![u]\!] \cdot \{\!\{\nabla_h v\}\!\} - \{\!\{\nabla_h u\}\!\} \cdot [\![v]\!]\right) \, ds, \tag{381}$$

*with $\theta = 1$. Then,*

$$|L_h(u,v)| \le |||u|||_\delta \, |||v|||_\delta, \qquad \forall u, v \in H^2(\mathcal{T}_h), \tag{382}$$

*for any $\delta > 0$, where*

$$|||v|||_\delta^2 = \|\nabla_h v\|^2_{L^2(\Omega)} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta^{-1} \left(\mathbf{n} \cdot \{\!\{\nabla v\}\!\}\right)^2 \, ds + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta [v]^2 \, ds. \tag{383}$$

*Furthermore,* (382) *holds also for $\theta = -1$.*

**Proof:** We have

$$\left| \int_\Omega \nabla_h u \cdot \nabla_h v \,\mathrm{d}\mathbf{x} \right| \le \|\nabla_h u\|_{L^2(\Omega)} \, \|\nabla_h v\|_{L^2(\Omega)}.$$

Furthermore, we use

$$\begin{aligned}\left| \int_{\Gamma_\mathcal{I}\cup\Gamma_D} [\![u]\!] \cdot \{\!\{\nabla_h v\}\!\} \,\mathrm{d}s \right| &= \left| \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta^{1/2} [u] \, \delta^{-1/2} \mathbf{n} \cdot \{\!\{\nabla_h v\}\!\} \,\mathrm{d}s \right| \\ &\le \left( \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta [u]^2 \,\mathrm{d}s \right)^{1/2} \left( \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta^{-1} \left(\mathbf{n} \cdot \{\!\{\nabla_h v\}\!\}\right)^2 \,\mathrm{d}s \right)^{1/2},\end{aligned} \tag{384}$$

and (384) with $u$ and $v$ exchanged, and obtain

$$
\begin{aligned}
|L_h(u,v)| &\leq \left|\int_\Omega \nabla_h u\cdot\nabla_h v\,\mathrm{d}\mathbf{x}\right| + \left|\int_{\Gamma_\mathcal{I}\cup\Gamma_D} [\![u]\!]\cdot\{\!\{\nabla_h v\}\!\}\,\mathrm{d}s\right| + \left|\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\{\!\{\nabla_h u\}\!\}\cdot[\![v]\!]\,\mathrm{d}s\right| \\
&\leq \|\nabla_h u\|_{L^2(\Omega)}\,\|\nabla_h v\|_{L^2(\Omega)} \\
&\quad + \left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u]^2\,\mathrm{d}s\right)^{1/2}\left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}\right)^2\,\mathrm{d}s\right)^{1/2} \\
&\quad + \left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h u\}\!\}\right)^2\,\mathrm{d}s\right)^{1/2}\left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s\right)^{1/2} \\
&\leq \left(\|\nabla_h u\|^2_{L^2(\Omega)} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h u\}\!\}\right)^2\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u]^2\,\mathrm{d}s\right)^{1/2} \\
&\quad\ \left(\|\nabla_h v\|^2_{L^2(\Omega)} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}\right)^2\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s\right)^{1/2} \\
&\leq |||u|||_\delta\,|||v|||_\delta,
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality $\sum_i a_i b_i \leq \left(\sum_i a_i^2\right)^{1/2}\left(\sum_i b_i^2\right)^{1/2}$. □

**Lemma A.13 (Continuity of NIPG and SIPG)** *Let $L_h$ be given as in* (120) *with $\delta > 0$ and $\theta = 1$ for NIPG and $\theta = -1$ for SIPG. Then there is a constant $1 < C \leq 2$ such that*

$$|L_h(u,v)| \leq C|||u|||_\delta\,|||v|||_\delta, \qquad \forall u,v\in H^2(\mathcal{T}_h), \tag{385}$$

*where the norm $|||\cdot|||_\delta$ is as defined* (383).

**Proof:** Using the Cauchy-Schwarz inequality $\int ab \leq \left(\int a^2\right)^{1/2}\left(\int b^2\right)^{1/2}$ for $a = \delta^{1/2}[u]$ and $b = \delta^{1/2}[v]$ we obtain

$$\left|\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u][v]\,\mathrm{d}s\right| \leq \left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u]^2\,\mathrm{d}s\right)^{1/2}\left(\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s\right)^{1/2}.$$

Thereby, using (382) for $\theta = 1$ and $\theta = -1$ we obtain

$$
\begin{aligned}
|L_h(u,v)| &\leq |||u|||_\delta\,|||v|||_\delta + \left|\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u][v]\,\mathrm{d}s\right| \\
&\leq \left(|||u|||^2_\delta + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[u]^2\,\mathrm{d}s\right)^{1/2}\left(|||v|||^2_\delta + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s\right)^{1/2} \\
&\leq 2|||u|||_\delta\,|||v|||_\delta.
\end{aligned}
$$

and hence (385). □

We note that $\delta > 0$ in (382) may be any positive constant. In contrast to that $\delta > 0$ in (385) is the constant of the interior penalty term (109).

**Lemma A.14 (Continuity of BR2, (Brezzi et al., 2000))** *Let $L_h$ be given as in* (142). *Then there is a constant $C > 1$ such that*

$$|L_h(u,v)| \leq C|||u|||_{L_0^e}|||v|||_{L_0^e} \qquad \forall u,v\in H^2(\mathcal{T}_h), \tag{386}$$

*where the* $|||\cdot|||_{L_0^e}$*-norm is given by*

$$|||v|||_{L_0^e}^2 = \|\nabla_h v\|_{L^2(\Omega)}^2 + \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma} \|\mathbf{L}_0^e(v)\|_{L^2(\Omega)}^2. \tag{387}$$

**Proof:** First, we note that $\mathbf{L}_0(v) = \sum_{e\subset\partial\kappa}\mathbf{L}_0^e(v)$ on $\kappa$. Furthermore, since the support of each $L_0^e$ is the union of (one or two) elements sharing the edge $e$, we have

$$\sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma} \|\mathbf{L}_0^e(v)\|_{L^2(\Omega)}^2 = \sum_{\kappa\in\mathcal{T}_h}\sum_{e\subset\partial\kappa} \|\mathbf{L}_0^e(v)\|_{L^2(\kappa)}^2. \tag{388}$$

Thereby, using Cauchy-Schwarz inequality, $\left(\sum_{i=1}^N a_i\right)^2 \le \sum_{i=1}^N 1 \sum_{i=1}^N a_i^2 = N\sum_{i=1}^N a_i^2$,

$$\|\mathbf{L}_0(v)\|_{L^2(\kappa)}^2 = \int_\kappa \left(\sum_{e\subset\partial\kappa}\mathbf{L}_0^e(v)\right)^2 \mathrm{d}\mathbf{x} \le N\int_\kappa \sum_{e\subset\partial\kappa}(\mathbf{L}_0^e(v))^2\mathrm{d}\mathbf{x} = N\sum_{e\subset\partial\kappa}\|\mathbf{L}_0^e(v)\|_{L^2(\kappa)}^2, \tag{389}$$

where $N$ is the number of faces $e\subset\partial\kappa$ of an element $\kappa$. And, due to (388)

$$\|\mathbf{L}_0(v)\|_{L^2(\Omega)}^2 \le N\sum_{\kappa\in\mathcal{T}_h}\sum_{e\subset\partial\kappa}\|\mathbf{L}_0^e(v)\|_{L^2(\kappa)}^2 = N\sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma}\|\mathbf{L}_0^e(v)\|_{L^2(\Omega)}^2. \tag{390}$$

Given $L_h$ as in (142), using Cauchy-Schwarz, (390) and again Cauchy-Schwarz we obtain

$$\begin{aligned}
L_h(u,v) =& \int_\Omega (\nabla_h u\cdot\nabla_h v - \mathbf{L}_0(u)\cdot\nabla_h v - \nabla_h u\cdot\mathbf{L}_0(v))\,\mathrm{d}\mathbf{x} + \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\int_\Omega \mathbf{L}_0^e(u)\cdot\mathbf{L}_0^e(v)\,\mathrm{d}\mathbf{x}\\
\le& \|\nabla_h u\|_{L^2(\Omega)}\|\nabla_h v\|_{L^2(\Omega)} + \|\mathbf{L}_0(u)\|_{L^2(\Omega)}\|\nabla_h v\|_{L^2(\Omega)} + \|\nabla_h u\|_{L^2(\Omega)}\|\mathbf{L}_0(v)\|_{L^2(\Omega)}\\
&+ \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\|\mathbf{L}_0^e(u)\|_{L^2(\Omega)}\|\mathbf{L}_0^e(v)\|_{L^2(\Omega)}\\
\le& \left(2\|\nabla_h u\|_{L^2(\Omega)}^2 + \|\mathbf{L}_0(u)\|_{L^2(\Omega)}^2 + \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\|\mathbf{L}_0^e(u)\|_{L^2(\Omega)}^2\right)\\
&\times\left(2\|\nabla_h v\|_{L^2(\Omega)}^2 + \|\mathbf{L}_0(v)\|_{L^2(\Omega)}^2 + \sum_{e\subset\Gamma_{\mathcal{I}}\cup\Gamma_D} C_{\mathrm{BR2}}\|\mathbf{L}_0^e(u)\|_{L^2(\Omega)}^2\right) \le C|||u|||_{L_0^e}|||v|||_{L_0^e}
\end{aligned}$$

with $C = (N + C_{\mathrm{BR2}})^2$. □

### A.4.3 Coercivity of the bilinear forms

Based on the relations shown in the proof of Lemma A.14 coercivity of BR2 is easily obtained.

**Theorem A.15 (Coercivity of BR2, (Brezzi et al., 2000))** *Let* $L_h$ *be given as in* (142). *Then there is a constant* $C_{BR2}^0 > 0$ *(*$C_{BR2}^0 = 3$ *on triangles,* $C_{BR2}^0 = 4$ *on quadrilaterals) such that for all* $C_{BR2} > C_{BR2}^0$ *we have following coercivity property: There is a constant* $\gamma > 0$ *such that*

$$L_h(v,v) \ge \gamma|||v|||_{L_0^e}^2 \qquad \forall v\in H^2(\mathcal{T}_h), \tag{391}$$

*where the* $|||\cdot|||_{L_0^e}$*-norm is as defined in* (387).

**Proof:** Using $2ab \le \epsilon a^2 + \frac{1}{\epsilon} b^2$ and (389) we obtain

$$\begin{aligned} 2\int_\Omega \nabla_h v \cdot \mathbf{L}_0(v)\,\mathrm{d}\mathbf{x} &\le \epsilon \|\nabla_h v\|^2_{L^2(\Omega)} + \frac{1}{\epsilon}\|\mathbf{L}_0(v)\|^2_{L^2(\Omega)} \\ &\le \epsilon \|\nabla_h v\|^2_{L^2(\Omega)} + \frac{N}{\epsilon} \sum_{e\subset\Gamma_\mathcal{I}\cup\Gamma_D} \|\mathbf{L}_0^e(v)\|^2_{L^2(\Omega)}, \end{aligned} \tag{392}$$

Then, using the definition (142) of $L_h$ we have

$$\begin{aligned} L_h(v,v) &= \|\nabla_h v\|^2_{L^2(\Omega)} + 2\int_\Omega \mathbf{L}_0(v)\cdot\nabla_h v\,\mathrm{d}\mathbf{x} + \sum_{e\subset\Gamma_\mathcal{I}\cup\Gamma_D} C_{\mathrm{BR2}} \|\mathbf{L}_0^e(v)\|^2_{L^2(e)} \\ &\ge (1-\epsilon)\|\nabla_h v\|^2_{L^2(\Omega)} + \sum_{e\subset\Gamma_\mathcal{I}\cup\Gamma_D} \left(C_{\mathrm{BR2}} - \frac{N}{\epsilon}\right) \|\mathbf{L}_0^e(v)\|^2_{L^2(e)}, \end{aligned} \tag{393}$$

and hence (391) with $\gamma = \min(1-\epsilon, C_{\mathrm{BR2}} - \frac{N}{\epsilon})$ which is positive whenever $\frac{N}{C_{\mathrm{BR2}}} < \epsilon < 1$, i.e., whenever $C_{\mathrm{BR2}} > N$. Thereby, (391) holds provided $C_{\mathrm{BR2}} > C^0_{\mathrm{BR2}} := N$, where $N$ is the number of faces $e \subset \partial\kappa$ of an element $\kappa$ □

Also the coercivity of the method of Baumann-Oden is easily shown.

**Lemma A.16 (Coercivity of Baumann-Oden)** *Let $L_h$ be given as in* (120) *with $\theta = 1$ and $\delta = 0$. Then,*

$$L_h(v,v) = \|\nabla_h v\|^2_{L^2(\Omega)} \qquad \forall v \in H^2(\mathcal{T}_h). \tag{394}$$

**Proof:**

$$L_h(v,v) = \int_\Omega \nabla_h v\cdot\nabla_h v\,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} (\theta-1)\{\!\{\nabla_h v\}\!\}\cdot[\![v]\!]\,\mathrm{d}s = \|\nabla_h v\|^2_{L^2(\Omega)} \tag{395}$$

for $\theta = 1$. □

We see, that the bilinear form $L_h$ for Baumann-Oden is coercive only with respect to the $H^1(\mathcal{T}_h)$-seminorm. In particular, for any $v_h \in V^d_{h,0}$ we have $L_h(v_h, v_h) = 0$, i.e., the method of Baumann-Oden is unstable. However, considering the discretization of $-\Delta u + cu = f$ with $c \ge c_0 > 0$ we obtain $L_h(v,v) \ge c_0\|v\|^2_{L^2(\Omega)} + \|\nabla_h v\|^2_{L^2(\Omega)}$, hence coercivity in the $H^1(\mathcal{T}_h)$-norm.

Finally, in order to show coercivity of the NIPG and SIPG discretization we first recall the following standard inverse estimate: There is a constant $C > 0$ such that

$$\|\nabla v_h\|_{L^2(\kappa)} \le C h_\kappa^{-1} \|v_h\|_{L^2(\kappa)} \quad \forall v_h \in V_h, \tag{396}$$

In the following we quote from (Schwab, 1998), p. 208, a generalization of this estimate to $v_h \in V_h^p$.

**Lemma A.17 (Inverse estimate on $V_h^p$)** *Let $\mathcal{T}_h$ be a shape regular mesh. Then, there is a constant $C \ge 0$ such that for any $\kappa \in \mathcal{T}_h$ we have*

$$\|\nabla v_h\|_{L^2(\kappa)} \le C\frac{p_\kappa^2}{h_\kappa}\|v_h\|_{L^2(\kappa)}, \qquad \forall v_h \in V_h^p. \tag{397}$$

Furthermore, let us recall following trace inequality:

**Lemma A.18 (Multiplicative trace inequality, (Prudhomme et al., 2000))** *Let $\kappa \in \mathcal{T}_h$, with diameter $h_\kappa$ and radius $r_\kappa$ of an inscribed circle, with $ch_\kappa < r_\kappa < h_\kappa$, $c > 0$, then*

$$\|v\|^2_{L^2(\partial\kappa)} \leq C\left(h_\kappa^{-1}\|v\|^2_{L^2(\kappa)} + \|v\|_{L^2(\kappa)}\|\nabla v\|_{L^2(\kappa)}\right) \qquad \forall v \in H^1(\kappa). \tag{398}$$

**Theorem A.19 (Coercivity of NIPG and SIPG, (Prudhomme et al., 2000))** *Let $L_h$ be given as in (120) with $\delta = C_{IP}\frac{p^2}{h}$. Then there is a constant $C^0_{IP} \geq 0$ ($C^0_{IP} = 0$ for NIPG, i.e., $\theta = 1$, and $C^0_{IP} > 0$ for SIPG, i.e., $\theta = -1$), such that for all $C_{IP} > C^0_{IP}$ we have following coercivity property: There is a constant $\gamma > 0$ such that*

$$L_h(v_h, v_h) \geq \gamma|||v_h|||^2_\delta \qquad \forall v_h \in V_h^p, \tag{399}$$

*where*

$$|||v|||^2_\delta = \|\nabla_h v\|^2_{L^2(\Omega)} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla v\}\!\}\right)^2\, ds + \int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta[v]^2\, ds. \tag{400}$$

**Proof:** This proofs follows the proof given in (Prudhomme et al., 2000). We begin by considering the term $\int_e (\mathbf{n}\cdot\nabla v)^2\,\mathrm{d}s$ on $e \in \Gamma_\mathcal{I}\cup\Gamma_D$. By employing the trace inequality (398) and the inverse estimate (397) we obtain for all $v_h \in V_h^p$,

$$\begin{aligned}
\int_e (\mathbf{n}\cdot\nabla v_h)^2\,\mathrm{d}s &\leq C\left(h_\kappa^{-1}\|\nabla v_h\|^2_{L^2(\kappa)} + \|\nabla v_h\|_{L^2(\kappa)}\|\nabla^2 v_h\|_{L^2(\kappa)}\right)\\
&\leq C\left(\frac{1}{h_\kappa} + \frac{p_\kappa^2}{h_\kappa}\right)\|\nabla v_h\|^2_{L^2(\kappa)}\\
&\leq C\frac{p_\kappa^2}{h_\kappa}\|\nabla v_h\|^2_{L^2(\kappa)} \leq \frac{C}{C_{\mathrm{IP}}}\delta\|\nabla v_h\|^2_{L^2(\kappa)},
\end{aligned} \tag{401}$$

and hence

$$-\int_{\Gamma_\mathcal{I}\cup\Gamma_D} \delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla v_h\}\!\}\right)^2\,\mathrm{d}s \geq -\frac{C}{C_{\mathrm{IP}}}\|\nabla_h v_h\|^2_{L^2(\Omega)}, \tag{402}$$

where we used $(a+b)^2 \leq 2(a^2+b^2)$ for $a = \mathbf{n}\cdot\nabla v_h^+$ and $b = \mathbf{n}\cdot\nabla v_h^-$. From (120) we have

$$L_h(v,v) = \int_\Omega \nabla_h v\cdot\nabla_h v\,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}(\theta-1)\{\!\{\nabla_h v\}\!\}\cdot[\![v]\!]\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s. \tag{403}$$

For $\theta = 1$ the second term vanishes and we obtain using (402)

$$L_h(v_h,v_h) - \gamma|||v_h|||^2_\delta \geq \left(1 - \gamma - \gamma\frac{C}{C_{\mathrm{IP}}}\right)\|\nabla v_h\|^2_{L^2(\Omega)} + (1-\gamma)\int_{\Gamma_\mathcal{I}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s.$$

Thereby, for any $C_{\mathrm{IP}} > C^0_{\mathrm{IP}} = 0$ we find a $0 < \gamma \leq 1/(1 + C/C_{\mathrm{IP}})$ such that (399) holds.

For $\theta = -1$ the second term in (403) is bounded using $ab \leq \frac{\epsilon}{4}a^2 + \frac{1}{\epsilon}b^2$,

$$\begin{aligned}
2\int_e\{\!\{\nabla_h v\}\!\}\cdot[\![v]\!]\,\mathrm{d}s &\leq 2\int_e \mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}[v]\,\mathrm{d}s \leq 2\left(\int_e\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}\right)^2\,\mathrm{d}s\right)^{1/2}\left(\int_e\delta[v]^2\,\mathrm{d}s\right)^{1/2}\\
&\leq \frac{\epsilon}{4}4\int_e\delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}\right)^2\,\mathrm{d}s + \frac{1}{\epsilon}\int_e\delta[v]^2\,\mathrm{d}s,
\end{aligned}$$

and hence

$$-2\int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} \{\!\{\nabla_h v\}\!\}\cdot [\![v]\!]\,\mathrm{d}s \geq -\epsilon\int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} \delta^{-1}\left(\mathbf{n}\cdot\{\!\{\nabla_h v\}\!\}\right)^2\,\mathrm{d}s - \frac{1}{\epsilon}\int_{\Gamma_{\mathcal{I}}\cup\Gamma_D} \delta[v]^2\,\mathrm{d}s. \qquad (404)$$

We then obtain

$$L_h(v_h,v_h)-\gamma|||v_h|||_\delta^2 \geq \left(1-\gamma-(\gamma+\epsilon)\frac{C}{C_{\mathrm{IP}}}\right)\|\nabla v_h\|^2_{L^2(\Omega)}+(1-\gamma-\frac{1}{\epsilon})\int_{\Gamma_{\mathcal{I}}\cup\Gamma_D}\delta[v]^2\,\mathrm{d}s.$$

Hence, we require

$$1-\gamma-(\gamma+\epsilon)\frac{C}{C_{\mathrm{IP}}}\geq 0 \quad\text{and}\quad 1-\gamma-\frac{1}{\epsilon}\geq 0.$$

The second inequality is fulfilled if $0<\gamma\leq 1-1/\epsilon$ and $\epsilon>1$. On the other hand the first inequality requires that $1-\epsilon C/C_{\mathrm{IP}}\geq\gamma(1+C/C_{\mathrm{IP}})$ and hence

$$0<\gamma\leq\frac{1-\epsilon C/C_{\mathrm{IP}}}{1+C/C_{\mathrm{IP}}}<\frac{1-C/C_{\mathrm{IP}}}{1+C/C_{\mathrm{IP}}}=\frac{C_{\mathrm{IP}}-C}{C_{\mathrm{IP}}+C},$$

for $C_{\mathrm{IP}}>C^0_{\mathrm{IP}}>0$, e.g. $C^0_{\mathrm{IP}}=C$ where $C$ is the constant in (402). □

We emphasize that for the NIPG method any choice of the interior penalty constant $C_{\mathrm{IP}}>0$ gives a stable scheme. In contrast to that the SIPG method requires $C_{\mathrm{IP}}>C^0_{\mathrm{IP}}>0$ for stability with a constant $C^0_{\mathrm{IP}}$ which is in general not known. However, numerical experiments showed that $C_{\mathrm{IP}}=10-20$ is a good choice a for large variety of problems.

We note that whereas continuity of $L_h$ could be shown on $H^2(\mathcal{T}_h)$ coercivity of $L_h$ on $H^2(\mathcal{T}_h)$ does not hold, see Prop. 4.4 in (van der Zee, 2004). However, coercivity of $L_h$ on the discrete function space $V_h^p$ as shown in Theorem A.19 is sufficient for proving existence and uniqueness of the discrete solution $u_h\in V_h^p$.

**Remark A.20** *The estimate* (401) *motivates the particular choice of* $\delta=C_{IP}\frac{p^2}{h}$.

### A.4.4 *A priori* error estimates

In this section we give an *a priori* error estimates for the NIPG and SIPG discretization.

**Lemma A.21 (*A priori* error estimate for NIPG and SIPG)** *Let* $u\in H^{p+1}(\Omega)$ *be the exact solution to Poisson's equation* (79). *Furthermore, let* $u_h\in V_h^p$ *be the solution to*

$$L_h(u_h,v_h)=F_h(v_h)\quad\forall v_h\in V_h^p,$$

*where* $L_h$ *is as given in* (120) *with* $\theta=1$ *(NIPG) or* $\theta=-1$ *(SIPG) and* $\delta=C_{IP}\frac{p^2}{h}$, $C_{IP}>C^0_{IP}$, *cf. Theorem A.19. Then*

$$|||u-u_h|||_\delta\leq Ch^p|u|_{H^{p+1}(\Omega)}, \qquad (405)$$

*where* $|||\cdot|||_\delta$ *is the norm as defined in* (383).

**Proof:** Let the error $e = u - u_h$ be split as follows

$$e = u - u_h = (u - P_h u) - (u_h - P_h u) = \eta - \xi,$$

with $\eta = u - P_h u$ and $\xi = u_h - P_h u$. Here, $P_h := P_h^p$ is the $L^2$-projector onto $V_h := V_h^p$ given in Definition A.5. Applying coercivity (399) of $L_h$ for $\xi \in V_h$ we obtain

$$\gamma |||\xi|||_\delta^2 \leq L_h(\xi, \xi) = L_h(\eta - e, \xi) = L_h(\eta, \xi),$$

where we used Galerkin orthogonality (4). Using continuity of $L_h$, (385), we obtain

$$\gamma |||\xi|||_\delta^2 \leq L_h(\eta, \xi) \leq C |||\eta|||_\delta \, |||\xi|||_\delta.$$

In summary, we obtain

$$|||u - u_h|||_\delta \leq |||\eta|||_\delta + |||\xi|||_\delta \leq C |||\eta|||_\delta.$$

Thereby, employing (383) and the approximation estimates in Section A.2 we obtain

$$\begin{aligned} |||u - u_h|||_\delta^2 \leq & C \left( \|\nabla_h \eta\|_{L^2(\Omega)}^2 + \int_{\Gamma_\mathcal{I} \cup \Gamma_D} \delta^{-1} \left( \mathbf{n} \cdot \{\!\{\nabla \eta\}\!\} \right)^2 \, \mathrm{d}s + \int_{\Gamma_\mathcal{I} \cup \Gamma_D} \delta [\eta]^2 \, \mathrm{d}s \right) \\ \leq & C \left( C h^{2p} + \tfrac{1}{C_{\mathrm{IP}}} \tfrac{h}{p^2} \, C h^{2(p-1/2)} + C h^{2(p+1/2)} \right) |u|_{H^{p+1}(\Omega)}^2, \end{aligned}$$

and thus (144). □

**Remark A.22** *We note that estimate* (144) *is of optimal order p which corresponds to the order of approximation of polynomials of degree p in the $H^1$-norm, cf. estimate* (355).

Having an estimate in the $H^1$-norm one can obtain an estimate in the $L^2$-norm by using a duality argument (Aubin-Nitsche) which is based on the definition of an appropriate adjoint (or dual) problem. We will use this technique also for the interior penalty discontinuous Galerkin discretization. However, as we will see in the following, application of an duality argument requires an adjoint consistent discretization.

**Lemma A.23 (*A priori* error estimates in the $L^2$-norm for NIPG and SIPG)** *Let $u \in H^{p+1}(\Omega)$ be the exact solution to Poisson's equation* (79). *Furthermore, let $u_h \in V_h^p$ be the solution to*

$$L_h(u_h, v_h) = F_h(v_h) \quad \forall v_h \in V_h^p,$$

*where $L_h$ is as given in* (120) *with $\theta = 1$ (NIPG) or $\theta = -1$ (SIPG) and $\delta = C_{IP} \frac{p^2}{h}$, $C_{IP} > C_{IP}^0$, cf. Theorem A.19. Then, for NIPG:*

$$\|u - u_h\|_{L^2(\Omega)} \leq C h^p |u|_{H^{p+1}(\Omega)}, \tag{406}$$

*and for SIPG:*

$$\|u - u_h\|_{L^2(\Omega)} \leq C h^{p+1} |u|_{H^{p+1}(\Omega)}. \tag{407}$$

**Proof:** Let us consider the adjoint problem (81) with homogeneous Dirichlet and Neumann boundary conditions $j_D = j_N = 0$, i.e.,

$$\begin{aligned} -\Delta z &= j_\Omega \quad \text{in } \Omega, \\ -z &= 0 \quad \text{on } \Gamma_D, \\ \mathbf{n}\cdot\nabla z &= 0 \quad \text{on } \Gamma_N. \end{aligned} \tag{408}$$

We recall from Section 4.6 that the bilinear form $\hat{L}_h$, (cf. (94) where the numerical fluxes $\hat{u}_h$ and $\hat{\boldsymbol{\sigma}}_h$ are still unspecified), can be rewritten as

$$\begin{aligned} \hat{L}_h(w,z) = &\int_\Omega j_\Omega w \,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma} [\![\hat{u}(w)]\!]\cdot\nabla z \,\mathrm{d}s - \int_{\Gamma_\mathcal{I}} [\![\hat{\boldsymbol{\sigma}}(w)]\!] z \,\mathrm{d}s \\ &- \int_{\Gamma_D} \hat{\boldsymbol{\sigma}}(w)\cdot\mathbf{n}\, z \,\mathrm{d}s + \int_{\Gamma_N} \hat{u}(w)\mathbf{n}\cdot\nabla_h z \,\mathrm{d}s + \int_{\Gamma_D} (\theta+1) w\,\mathbf{n}\cdot\nabla_h z \,\mathrm{d}s, \end{aligned} \tag{409}$$

see Equation (102). In particular, for the SIPG method, both, $\hat{\boldsymbol{\sigma}}$ and $\hat{u}$ are conservative, i.e., single-valued, and thus $[\![\hat{\boldsymbol{\sigma}}]\!] = 0$ and $[\![\hat{u}]\!] = 0$. Furthermore with $\theta = -1$, and with (408) we find

$$B_h^s(w,z) = \int_\Omega j_\Omega w \,\mathrm{d}\mathbf{x}, \tag{410}$$

where $B_h^s(\cdot,\cdot)$ denotes the bilinear form of the symmetric interior penalty DG discretizations given in (120) with $\theta = -1$. We define $z^s$ to be the solution to (408) for $j_\Omega := e$. We assume that $z^s \in H^2(\Omega)$ and $\|z^s\|_{H^2(\Omega)} \le C\|e\|_{L^2(\Omega)}$ which is satisfied if $\Omega$ is a convex polygon, for example. In view of (410) we have

$$B_h^s(w,z^s) = \int_\Omega e w \,\mathrm{d}\mathbf{x}. \tag{411}$$

Now choosing $w = e$ we obtain

$$\|e\|^2_{L^2(\Omega)} = \int_\Omega e^2 \,\mathrm{d}\mathbf{x} = B_h^s(e,z^s) = B_h^s(e, z^s - z_h) \le |\!|\!| e |\!|\!|_\delta \, |\!|\!| z^s - z_h |\!|\!|_\delta, \tag{412}$$

where we used Galerkin orthogonality (4) for $z_h = P_h^p z^s \in V_h^p$ and continuity (385) of $L_h$. Thus using (144) and approximation estimates for $z^s - z_h$ we obtain

$$\|e\|^2_{L^2(\Omega)} \le |\!|\!| e |\!|\!|_\delta \, |\!|\!| z^s - z_h |\!|\!|_\delta \le Ch^p |u|_{H^{p+1}(\Omega)} Ch |z|_{H^2(\Omega)} \le Ch^{p+1} |u|_{H^{p+1}(\Omega)} \|e\|_{L^2(\Omega)},$$

and hence (407).

For the NIPG discretization the above argument fails because it is not adjoint consistent. In fact, for the NIPG discretization, we have $[\![\hat{\boldsymbol{\sigma}}_h]\!] = 0$ and $[\![\hat{u}_h]\!] = 2[\![u_h]\!]$, and $\theta = 1$, thus

$$B_h^n(w,z) = \int_\Omega j_\Omega w \,\mathrm{d}\mathbf{x} + 2\int_{\Gamma_\mathcal{I}\cup\Gamma} [\![w]\!]\cdot\nabla z \,\mathrm{d}s + 2\int_{\Gamma_D} w\,\mathbf{n}\cdot\nabla_h z \,\mathrm{d}s, \tag{413}$$

where $B_h^n(\cdot,\cdot)$ denotes the bilinear form of the symmetric interior penalty DG discretizations given in (120) with $\theta = 1$. Hence, the analytical solution $z^n$ to

$$B_h^n(w,z^n) = \int_\Omega e w \,\mathrm{d}\mathbf{x}, \tag{414}$$

is mesh-dependent. Furthermore, $z^n$ is not regular which is why we do not obtain an additional order of $h$ from $z^n - z_h$ as we do in the case of the SIPG method. □

## A.5 DG discretization of the compressible Navier-Stokes equations: Details

In this section we give some details which were left out in Section 6 for brevity.

### A.5.1 Derivation of the face-based primal form

For rewriting the element-based primal formulation (254): find $\mathbf{u}_h \in \mathbf{V}_h^p$ such that

$$\int_\Omega G(\mathbf{u}_h)\nabla_h \mathbf{u}_h : \nabla_h \mathbf{v}_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa} \underline{\hat{\sigma}}_h : \mathbf{n}\otimes\mathbf{v}_h \,\mathrm{d}s + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa} (\hat{\mathbf{u}}_h - \mathbf{u}_h)\otimes\mathbf{n} : \left(G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\right)\,\mathrm{d}s = 0 \quad \forall \mathbf{v}_h \in \mathbf{V}_h^p, \quad (415)$$

into a face-based primal form, we recall the definition of mean value and jump operators in (255) and (256) for $\mathbf{v}_h \in \mathbf{V}_h^p$ and $\underline{\tau}_h \in \underline{\Sigma}_h^p$, respectively. Note, that the according operators can be defined for $\mathbf{v} \in [T(\mathcal{T}_h)]^m$ and $\underline{\tau} \in [T(\mathcal{T}_h)]^{m\times d}$. Then, we can generalize Lemma A.10 to systems of equations:

**Lemma A.24** *Let $\mathbf{v} \in [T(\mathcal{T}_h)]^m$ and $\underline{\tau} \in [T(\mathcal{T}_h)]^{m\times d}$, then*

$$\sum_\kappa \int_{\partial\kappa\setminus\Gamma} \left(\underline{\tau}^+\mathbf{n}^+\right)\cdot\mathbf{v}^+\,ds = \int_{\Gamma_\mathcal{I}} \{\!\{\underline{\tau}\}\!\} : \underline{[\![\mathbf{v}]\!]}\,ds + \int_{\Gamma_\mathcal{I}} [\![\underline{\tau}]\!]\cdot\{\!\{\mathbf{v}\}\!\}\,ds, \quad (416)$$

$$\sum_\kappa \int_{\partial\kappa} \left(\underline{\tau}^+\mathbf{n}^+\right)\cdot\mathbf{v}^+\,ds = \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\underline{\tau}\}\!\} : \underline{[\![\mathbf{v}]\!]}\,ds + \int_{\Gamma_\mathcal{I}} [\![\underline{\tau}]\!]\cdot\{\!\{\mathbf{v}\}\!\}\,ds. \quad (417)$$

**Proof:** Employing Lemma A.10 for any $i = 1,\dots,m$ we obtain

$$\begin{aligned}\sum_\kappa \int_{\partial\kappa\setminus\Gamma} \tau_{ik}^+ n_k^+\, v_i^+\,\mathrm{d}s &= \int_{\Gamma_\mathcal{I}} \{\!\{\tau_{ik}\}\!\}\left(v_i^+n_k^+ + v_i^-n_k^-\right)\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} \left(\tau_{ik}^+n_k^+ + \tau_{ik}^-n_k^-\right)\{\!\{v_i\}\!\}\,\mathrm{d}s \\ &= \int_{\Gamma_\mathcal{I}} \{\!\{\underline{\tau}\}\!\} : \underline{[\![\mathbf{v}]\!]}\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} [\![\underline{\tau}]\!]\cdot\{\!\{\mathbf{v}\}\!\}\,\mathrm{d}s,\end{aligned}$$

thus (416). Use the definition of mean value and jump operators on $\Gamma$ for (417). □

We now proceed in transferring the cell-based form (415) into a face-based form. To this end, we use Equation (417) twice (once for $\underline{\tau} = \underline{\hat{\sigma}}_h$ and $\mathbf{v} = \mathbf{v}_h$, and once for $\underline{\tau} = G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h$, i.e., $\tau_{jl} = (G(\mathbf{u})_{kl})_{ij}\,\partial_{x_k} v_i$, and $\mathbf{v} = \hat{\mathbf{u}}_h - \mathbf{u}_h$) and rewrite (415) as follows

$$\int_\Omega G(\mathbf{u}_h)\nabla_h\mathbf{u}_h : \nabla_h\mathbf{v}_h\,\mathrm{d}\mathbf{x} - \int_{\Gamma_\mathcal{I}\cup\Gamma} \{\!\{\underline{\hat{\sigma}}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]}\,\mathrm{d}s - \int_{\Gamma_\mathcal{I}} [\![\underline{\hat{\sigma}}_h]\!]\cdot\{\!\{\mathbf{v}_h\}\!\}\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}\cup\Gamma} \underline{[\![\hat{\mathbf{u}}_h - \mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\}\!\}\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} \{\!\{\hat{\mathbf{u}}_h - \mathbf{u}_h\}\!\}\cdot[\![G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h]\!]\,\mathrm{d}s = 0 \quad \forall\mathbf{v}_h\in\mathbf{V}_h^p,$$

which results in following *face-based* primal formulation:

$$\int_\Omega G(\mathbf{u}_h)\nabla_h\mathbf{u}_h : \nabla_h\mathbf{v}_h\,\mathrm{d}\mathbf{x} + \int_{\Gamma_\mathcal{I}\cup\Gamma} \underline{[\![\hat{\mathbf{u}}_h - \mathbf{u}_h]\!]} : \{\!\{G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h\}\!\} - \{\!\{\underline{\hat{\sigma}}_h\}\!\} : \underline{[\![\mathbf{v}_h]\!]}\,\mathrm{d}s + \int_{\Gamma_\mathcal{I}} \{\!\{\hat{\mathbf{u}}_h - \mathbf{u}_h\}\!\}\cdot[\![G^\top(\mathbf{u}_h)\nabla\mathbf{v}_h]\!] - [\![\underline{\hat{\sigma}}_h]\!]\cdot\{\!\{\mathbf{v}_h\}\!\}\,\mathrm{d}s = 0 \quad \forall\mathbf{v}_h\in\mathbf{V}_h^p.$$