# 1 Convection-Diffusion-Reaction

We will begin with solving the (linear) convection-diffusion-reaction equation with Dirichlet boundary conditions. The specific problem will be,

$$\partial_x \left[ F(u, \partial_x u) \right] + bu - f = 0, \qquad \forall x \in \Omega \tag{1}$$

$$F(u, \partial_x u) = cu - \nu \partial_x u \tag{2}$$

where $c(x)$, $\nu(x)$, $b(x)$, and $f(x)$ are all functions of $x$. We will chose the domain to be

$$\Omega = \{x : 0 < x < 1\} \tag{3}$$

And we will apply the following Dirichlet boundary conditions,

$$u(0) = g(0) \qquad u(1) = g(1) \tag{4}$$

where $g$ is a function which is technically only defined at the boundary points. We could define constants, say $g_0$ and $g_1$, to represent the boundary values, however, the notation later will be cleaner (and also consistent with multi-dimensional problems) if we choose to state the boundary values as a function.

# 2 Finite element mesh

We will discretize the domain $\Omega$ into a set of $N_e$ non-uniform elements where element $\kappa_i$ is defined as:

$$\kappa_i = \{x : x_i < x < x_{i+1}\} \qquad \forall\ 1 < i < N_e \tag{5}$$

And, $x_1 = 0$ and $x_{N_e+1} = 1$. We also define the element size $h_i$ as,

$$h_i = x_{i+1} - x_i \tag{6}$$

Also, we define the average element size as,

$$h = \frac{1}{N_e} \tag{7}$$

Finally. the discretization of the domain into elements, commonly referred to as the triangulation, is denoted as,

$$\mathcal{T}_h = \{\kappa_i \, \forall i\} \tag{8}$$

# 3 Some notation

We will introduce some notation which is commonly used to describe Finite Element methods.

We define the $L^2(\Omega)$ inner product over $\Omega$ of two function $u$ and $v$ as

$$(u, v) \equiv \int_\Omega uv \tag{9}$$

If the integration is over a subset of the domain, for example over an element $\kappa$, we will include a subscript,

$$(u, v)_\kappa \equiv \int_\kappa uv \tag{10}$$

Also, evaluation of boundary terms are often written as,

$$\langle u, v \rangle \equiv u(1)v(1) - u(0)v(0) \tag{11}$$

And similar over an element $\kappa$,

$$\langle u, v \rangle_\kappa \equiv u(x_1^\kappa)v(x_1^\kappa) - u(x_0^\kappa)v(x_0^\kappa) \tag{12}$$

where $x_0^\kappa = \min_\kappa x$ and $x_1^\kappa = \max_\kappa x$.

# 4 Continuous Galerkin formulation

## 4.1 Solution space and basis

In a continuous FEM formulation, the solution in each element is approximated as a set of polynomials which are required to be continuous from element to element. We will only consider solutions which are $C^0$ continuous between elements (i.e. the solution is continuous but not its derivatives). Also, we will only consider approximations with the same polynomial order $p$ in each element. Specifically, we will introduce the following notation for the continuous finite element solution space,

$$\mathcal{V}_h = \left\{ v \in C^0(\Omega) : v|_\kappa \in \mathcal{P}^p(\kappa) \; \forall \kappa \in \mathcal{T}_h \right\} \tag{13}$$

where $\mathcal{P}^p(\kappa)$ is the space of polynomials of order $p$ (e.g. $p = 1$ is a linear function).

We then need a basis for this space. In each element, we require $p + 1$ degrees of freedom to represent the polynomial functions. However, because of the continuity required between elements, the degrees of freedom to represent the entire approximation are reduced by one for each pair of elements. The result is that there are $N = pN_e + 1$ degrees of freedom needed to represent the solution. We write the approximation as,

$$u_h(x) = \sum_{i=1}^{N} u_i \phi_i(x) \tag{14}$$

where $\phi_i(x)$ are the basis functions. These notes do not include details on choice of basis, numerical integration, and other implementation issues.

## 4.2 Weak residual definition

Define the strong form of the residual for a function $v(x)$ as,

$$r(v) \equiv \partial_x \left[ F(v, \partial_x v) \right] + bv - f. \tag{15}$$

(This requires that $v(x)$ is an appropriately smooth function so that the required derivatives exist). The weak formulation then weights the residual $r(v)$, by a function $w$ and integrating by parts to

produce,

$$(w, r(v)) = \langle w, F(v, \partial_x v)\rangle - (\partial_x w, F(v, \partial_x v)) + (w, bv - f) \tag{16}$$

In the remainder of the document, we will not include the dependence of $r$ and $F$ (and similar) on $v$ and $\partial_x v$ to shorten the results, unless needed to clarify a possible ambiguity.

To enforce the boundary conditions, we will add a constraint weighted by a boundary weight $\lambda$,

$$(w, r) + \langle \lambda, v - g\rangle = \langle w, F\rangle + \langle \lambda, v - g\rangle - (\partial_x w, F) + (w, bv - f) \tag{17}$$

In the discrete setting, if we then attempt to determine the function $v$ for which this weighted residual is zero for all $w$ and $\lambda$, the problem will in general not be solvable as we will have more equations than degrees of freedom. In other words, the problem is over-constrained. To resolve this, we can add a perturbation to the weighted residual on the boundary resulting in the final form of the weighted residual statement,

$$R(v, \bar{F}, w, \lambda) \equiv \langle w, \bar{F}\rangle + \langle \lambda, v - g\rangle - (\partial_x w, F) + (w, bv - f) \tag{18}$$

where $\bar{F}$ is an unknown boundary state, actually the boundary flux, which will be determined as part of the entire solution.

### 4.3 Continuous Galerkin FEM

The continuous Galerkin (CG) FEM method is then stated as: Find $u_h \in \mathcal{V}_h$ and $\bar{F}_h = (\bar{F}_h(0), \bar{F}_h(1)) \in \mathbb{R}^2$ such that

$$R(u_h, \bar{F}_h, w_h, \lambda_h) = 0 \qquad \forall\, w_h \in \mathcal{V}_h,\ \forall\, \lambda_h \in \mathbb{R}^2 \tag{19}$$

A note about implementation of this method. There are $N + 2$ degrees of freedom. Specifically,

- $N$ for the weights $u_i$ of the solution basis functions
- 2 for the boundary fluxes $\bar{F}_h$

Therefore, there will be $N + 2$ residual equations. Specifically,

- $N$ residuals equations enforcing the PDE: $R(u_h, \bar{F}_h, \phi_i, 0) = 0$
- 2 residuals equations enforcing the boundary conditions: $(u_h, \bar{F}_h, 0, \lambda_h) = 0$.

## 5 Adjoint operator

An operator which will occur frequently in various aspects of our work is the adjoint operator. For now, since we are focusing on a linear PDE, we will define the adjoint operator assuming linearity. However, the concept can be generalized to nonlinear PDEs. For our governing PDE in Equation (1), which is often refered to as the primal equation, we can define the primal operator $\mathcal{L}$ acting on some function $v$ as,

$$\mathcal{L}v \equiv \partial_x(cv) - \partial_x(\nu \partial_x v) + bv \tag{20}$$

Thus, Equation (1) could be written simply as

$$\mathcal{L}u = f. \tag{21}$$

The adjoint operator is given the symbol $\mathcal{L}^*$ and is defined as,

$$(\mathcal{L}v, w) \equiv (v, \mathcal{L}^* w) \tag{22}$$

where integration by parts is used to derive $\mathcal{L}^*$ for a given $\mathcal{L}$ by moving the derivative on $v$ over to $w$, and all boundary terms that arise are ignored for now (later, they will become boundary conditions on the adjoint problem).

For our specific model problem, the adjoint operator can be derived as follows,

$$\begin{aligned}
(\mathcal{L}v, w) &= (\partial_x[cv] - \partial_x[\nu\partial_x v] + bv, w) && (23)\\
&= (\partial_x[cv], w) - (\partial[\nu\partial_x v], w) + (bv, w) && (24)\\
&= \langle cv, w\rangle - (cv, \partial_x w) - \langle \nu\partial_x v, w\rangle + (\nu\partial_x v, \partial_x w) + (bv, w) && (25)\\
&= \langle cv, w\rangle - (v, c\partial_x w) - \langle \nu\partial_x v, w\rangle + (\partial_x v, \nu\partial_x w) + (v, bw) && (26)\\
&= \langle cv, w\rangle - (v, c\partial_x w) - \langle \nu\partial_x v, w\rangle + \langle v, \nu\partial_x w\rangle - (v, [\nu\partial_x w]_x) + (v, bw) && (27)\\
&= (v, \mathcal{L}^* w) + \langle cv, w\rangle - \langle \nu\partial_x v, w\rangle + \langle v, \nu\partial_x w\rangle && (28)
\end{aligned}$$

where in the last step, we have gathered all of the volume terms to find that the adjoint operator acting on a function $w$ is given by,

$$\mathcal{L}^* w = -c\partial_x w - \partial_x[\nu\partial_x w] + bw \tag{29}$$

We note here that the even-order (diffusion and source) terms produce the same operator in the adjoint as was in the primal. When this occurs, the operator is referred to as self-adjoint. The first-order (convection) term however produces a sign change as well as not being in a "conservative" or "total derivative" form. That is, while the primal term was $\partial_x[cv]$, the corresponding adjoint term $-c\partial_x w$ (so this is not a self-adjoint term).

# 6 Stabilized Continuous Galerkin Methods

We will now consider how the continuous Galerkin methods defined in Section 4 can be stabilized to remove the global oscillations that are caused when the viscosity is small.

## 6.1 Residual-stabilized methods

In this section, we present methods that add stabilization which depend on the (strong form) residual. Specifically, these methods all have the same basic form given by,

$$R(v, \bar{F}, w, \lambda) = \langle w, \bar{F}\rangle + \langle \lambda, v - g\rangle - (\partial_x w, F) + (w, bv - f) + \sum_\kappa (\bar{\mathcal{L}}w, \tau(\mathcal{L}v - f))_\kappa \tag{30}$$

where $\bar{\mathcal{L}}$ is an operator which depends on the specific method, in particular,

- $\bar{\mathcal{L}}w = (cw)_x$ is known as the SUPG (Streamwise-Upwind/Petrov-Galerkin) method
- $\bar{\mathcal{L}}w = \mathcal{L}w$ is known as the GLS (Galerkin Least Squares) method
- $\bar{\mathcal{L}}w = -\mathcal{L}^* w$ is known as the adjoint stabilized method, or also the Variational Multiscale (VMS) method.

And, $\tau$ is a parameter of these methods which has the following properties,

- It has units of time and is sometimes refered to as the instrinsic timescale.
- $\tau$ should be $o(h)$ to ensure that this stabilization does not degrade the accuracy of the FEM approach for arbitrary $p$ (see the discussion at the end of this section).
- In the convection-dominated limit where $Pe_h \to 0$, $\tau$ should scale with $h_i/|c|$, that is $\tau|c|/h = O(1)$. Note $Pe_h \equiv (|c|h_i)/(2\nu)$ is the local grid Peclet number. This behavior of $\tau$ is required to provide sufficient upwinding of the convection term.
- For viscous-dominated problems ($Pe_h^{-1} \to 0$), the desired behavior of $\tau$ depends on choice of $\bar{\mathcal{L}}$.
- A large variety of choices for $\tau$ are described in the literature. One fairly common choice is,

$$\tau = \frac{h_i}{2|c|}(\coth Pe_h - 1/Pe_h) \tag{31}$$

  which is known to provide nodally exact solution for constant convection and viscosity using linear ($p = 1$) elements.
- For linear elements, often $\tau$ is taken as a single value in an element. However, for $p > 1$, $\tau$ should probably vary inside an element since the solution can vary more significantly in the element.

We note that these stabilization all provide a (primal) consistent discretization which, in the FEM context, means that the residual of the method (i.e. Equation 30) is identically zero when evaluated with the exact solution $u$ for all $w$ and $\lambda$. This is because the exact solution satisfies $r(u) = \mathcal{L}u - f = 0$ and thus the stabilization is zero. Since the stabilization is zero, then the weighted residual of the residual-stabilized methods will be equal to the standard Galerkin weighted residual (i.e. as given in Equation 18). And, since this weighted residual is nothing more than an integrated-by-parts version of the weighted primal residual $\int_\Omega wr(u)$, then the exact solution will satisfy all of these residual-stabilized methods.

Residual stabilization has benefits in terms of achieving higher-order accuracy as the polynomial order $p$ is increased. Here's a rough idea for how this works. We can think of the stabilized residual as a sum of the standard Galerkin residual and a stabilization term,

$$R = R_{\text{gal}} + R_{\text{stab}} \tag{32}$$

For an order $p$ approximation space, an interpolant of the exact solution would have an error $u - u_h$ (in the $L^2$ norm) that is $O(h^{p+1})$. This is known as the *optimal* convergence rate. For the finite element approximation, therefore, the best we could hope for would be to achieve this optimal convergence rate. If the solution converges optimally at order $p+1$, then the $m^{\text{th}}$ derivative of $u_h$

would likely converge as $O(h^{p-m+1})$ (assuming a sufficiently smooth $u$). Thus, the residual $r(u_h)$ we might expect to also be $O(h^{p-m+1})$, where $m$ is the highest derivative of the residual operator. Since the Galerkin residual is a weighted integral of $r(u)$, then we expect that $R_{\text{gal}} = O(h^{p-m+1})$. Further, since the stabilized residual is again a direct function of $r(u_h)$ and is multiplied by $\tau$ which by design is at least $O(h)$, then this implies that the residual-based stabilization will be adding a higher-order effect to the residual, i.e. it will be an $R_{\text{stab}} = O(h^{p-m+2})$ contribution. Since this stabilization contribution should be dominated by the Galerkin residual, the intent is that the order of accuracy of $u_h$ will not be affected by this higher-order residual perturbation.

### 6.2 Artificial viscosity stabilization

In contrast to the residual-stabilized methods, consider a non-residual-based approach, e.g. an artificial viscosity approach would be to add a term of the form,

$$R = R_{\text{gal}} + (\partial_x w, \nu_{\text{art}} \partial_x v) \tag{33}$$

where $\nu_{\text{art}}$ is an artificial viscosity level introduced because the physical level of viscosity is not sufficient to stabilize the algorithm. Commonly, the $\nu_{\text{art}}$ is chosen to be proportional to $|c|h$. However, with this choice, the stabilization being added to the residual is always $O(h)$. Thus, this will cause the solution (and its derivatives) to have an $O(h)$ error. In other words, the best solution accuracy that can be achieved with an artificial viscosity approach in which $\nu_{\text{art}} = O(h)$ is $O(h)$ (i.e. first order) independent of $p$.

## 7 Discontinuous Galerkin formulations

### 7.1 Solution space and basis

In a discontinuous FEM formulation, the solution in each element is approximated as a set of polynomials which are not required to be continuous from element to element. As before in the CG, we will only consider approximations with the same polynomial order $p$ in each element. Specifically, we will introduce the following notation for the discontinuous finite element solution space,

$$\mathcal{V}_h = \left\{ v \in L^2(\Omega) : v|_\kappa \in \mathcal{P}^p(\kappa) \; \forall \kappa \in \mathcal{T}_h \right\} \tag{34}$$

where $\mathcal{P}^p(\kappa_i)$ is the space of polynomials of order $p$ (e.g. $p = 1$ is a linear function). The only difference between this definition and the definition for the CG space (see Equation 13) is the change from $C^0(\Omega)$ to $L^2(\Omega)$. Since $L^2(\Omega)$ is the space of functions that are square-integrable, then discontinuous solutions are allowed. Further, since the space uses polynomials inside an element, then the discontinuities are only between elements.

We then need a basis for this space. In each element, we require $p+1$ degrees of freedom to represent the polynomial functions. The result is that there are $N = (p+1)N_e$ degrees of freedom needed to represent the solution. We write the approximation as,

$$u_h(x) = \sum_{i=1}^{N} u_i \phi_i(x) \tag{35}$$

where $\phi_i(x)$ are the basis functions.

In particular, the common approach for DG is to pick a basis such that each function $\phi_i(x)$ is only non-zero inside a specific element $\kappa$ and zero for any other element. Thus, we frequently think of a basis function as being associated with a specific element, and for an order $p$ approximation, each element has $p+1$ basis functions.

### 7.2 Weak residual definition: a first attempt

The discontinuous Galerkin approach then constructs the weighted residual on a single element using a weight function $w \in \mathcal{V}_h$ and integrates by parts to produce,

$$R_\kappa(v,w) \equiv \int_\kappa wr(v) = \langle w, F\rangle_\kappa - (\partial_x w, F)_\kappa + (w, bv - f)_\kappa \tag{36}$$

Since $w$, $v$, and $F$ are multi-valued at the boundaries of an element, we need to be clear about the intrepretation of the above weak residual. Specifically, we interpret the relevant boundary values as being taken as the limiting value from within the element $\kappa$. These element boundary values are often referred to as the trace values of the function. Taking this view that the boundary values are the traces from within the element, then we note that this weighted residual on element $\kappa$ does not depend in any way on values of $w$, $v$, $F$, etc. from outside of $\kappa$.

We can also sum the elemental weighted residuals over all elements to produce a global weighted residual,

$$R(v,w) \equiv \sum_\kappa R_\kappa(v,w) \tag{37}$$

Following the same approach as before to enforce boundary conditions, then we produce the following weighted residual statement

$$R(v,\bar{F},w,\lambda) \equiv \sum_\kappa R_\kappa(v,w) + \langle w, \bar{F} - F\rangle + \langle \lambda, v - g\rangle \tag{38}$$

In the continuous Galerkin finite element formulation, we then set this weight residual to zero for all $w \in \mathcal{V}_h$ and we have a method which works (assuming sufficient viscosity). In the DG formulation, this is not sufficient because of the complete lack of coupling between the elements. As a result, we will need to modify this residual to achieve a workable discretization.

Before making any modifications, however, we will look at some other common manipulations to the residual as it currently stands. Specifically, we can gather the element boundary terms (which are presently expressed as a sum over elements) and re-write this as a sum of the element boundaries, such that,

$$\sum_\kappa \langle w, f\rangle_\kappa = \sum_{\Gamma_i} [\![wF]\!] + \langle w, F\rangle \tag{39}$$

where $\Gamma_i$ is the set of element faces that are in the interior of the domain and $[\![\phi]\!]$ is the jump operator which for a one-dimensional problem is defined as,

$$[\![\phi]\!] \equiv \phi^- - \phi^+ \tag{40}$$

where

$$\phi^\pm(x) = \lim_{\epsilon \to 0^+} \phi(x \pm \epsilon) \tag{41}$$

With this notation, a common way to write the global DG weak form is,

$$R(v,\bar{f},w,\lambda) = \langle w,\bar{f}\rangle + \langle \lambda, v-g\rangle - (\partial_x w, f) + (w, bv-s) + \sum_{\Gamma_i} [\![ wf ]\!] \tag{42}$$

In addition to the jump operator, DG algorithms frequently use a face average operator which is written as,

$$\{\phi\} \equiv \frac{1}{2}\left(\phi^- + \phi^+\right) \tag{43}$$

### 7.3 DG fluxes

The method used to provide coupling between elements in the standard DG formulation is to replace the traces of the fluxes (which are double-valued) with a unique flux that is in some way a function of the elemental solutions, and in particular the trace values. That is, we replace $f^\pm$ on a face with a unique flux,

$$F^\pm(v^\pm, \partial_x v^\pm) \rightarrow \hat{F}(v^+, v^-, \partial_x v^+, \partial_x v^-) \tag{44}$$

With this modification, the local (elemental) DG residual is then,

$$R_\kappa(v,w) \equiv \int_\kappa w r(v) = \langle w, \hat{F}\rangle_\kappa - (\partial_x w, F)_\kappa + (w, bv-f)_\kappa \tag{45}$$

and the global DG residual is,

$$R(v,\bar{f},w,\lambda) = \langle w,\bar{F}\rangle + \langle \lambda, v-g\rangle - (\partial_x w, F) + (w, bv-f) + \sum_{\Gamma_i} \hat{F}\,[\![ w ]\!] \tag{46}$$

Then, separate the flux into an (inviscid) convection part $F_I = av$ and the (viscous) diffusion part $F_V = -\nu\partial_x v$, thus, $F = F_I + F_V$ and the unique DG flux will be split in a similar manner,

$$\hat{F}(v^+, v^-, \partial_x v^+, \partial_x v^-) = \hat{F}_I(v^+, v^-) + \hat{F}_V(v^+, v^-, \partial_x v^+, \partial_x v^-) \tag{47}$$

#### 7.3.1 Convective flux

For the convection term, $\hat{F}_I$ is then evaluated using an upwind flux function. Specifically for the linear convection problem we are studying,

$$\hat{F}_I = \frac{1}{2}c(v^+ + v^-) - \frac{1}{2}|c|(v^+ - v^-) \tag{48}$$

#### 7.3.2 Diffusive flux: a simple average

For the diffusion part, a good solution is much more complex, and so we will first motivate the difficulty with a simple (but, in the end, ineffective) choice. Specifically, consider the simple average of the diffusive flux,

$$\hat{F}_V = -\frac{\nu}{2}(\partial_x v^+ + \partial_x v^-) \tag{49}$$

To see why this will not work in general, consider the case of pure diffusion for which $b = c = 0$ and $\nu = 1$. Suppose there exists a solution $v^0 \in \mathcal{V}_h$ which satisfies the DG weak form (given in

Equation 46) for all $w \in \mathcal{V}_h$ and $\lambda_h \in \mathbb{R}^2$. Specifically,

$$\langle w, \bar{F}^0 \rangle + \langle \lambda, v^0 - g \rangle + (\partial_x w, v_x^0) + \sum_{\Gamma_i} \hat{F}^0 \, [\![w]\!] = 0 \tag{50}$$

where

$$\hat{F}^0 = -\frac{1}{2}(v_x^{0^+} + v_x^{0^-}) \tag{51}$$

and $\bar{F}^0$ is the boundary flux required to satisfy the (weakly-imposed) Dirichlet conditions.

Now, consider a perturbation to $v^0$ created by adding arbitrary constant values to the solution in each element, that is,

$$v(x) = v^0(x) + v_\kappa \tag{52}$$

where $v_\kappa$ is a constant function in the element $\kappa$ but can vary from element-to-element and the only restriction is that $v_\kappa = 0$ for the first and last element (so that the boundary values of $v = v^0 = g$). Note that this $v$ is still in $\mathcal{V}_h$. Further, $\partial_x v = v_x^0$ and this also implies for our simple choice of $\hat{f}_V$ that,

$$\hat{F}(\partial_x v^+, \partial_x v^-) = -\frac{1}{2}(\partial_x v^+ + \partial_x v^-) = -\frac{1}{2}(v_x^{0^+} + v_x^{0^-}) = \hat{F}^0 \tag{53}$$

As a result, the DG weak form (Equation 46) is satisfied by this perturbed solution. In other words, this choice of diffusive flux results in a singular discretization (i.e. possessing a zero eigenvalue) with infinitely many solutions. We note that the fundamental problem is that the simple average diffusive flux does not "see" jumps in the solution from element-to-element. Thus, this suggests that the DG diffusive flux must somehow depend on the jump in $v$, i.e. $[\![v]\!]$.

### 7.3.3 Face lifting operator

A commonly used technique in DG methods is to re-write integrals on element faces in terms of volume integrals. This requires that an integrand on an element face be related to an integrand on the element volume. This process is referred to as a "lifting" operation. Note that for one-dimensional problems, the face integrals are not actually integrals but rather simply a value associated with the face. Specifically, we will define a one-dimensional face lifting operator $r_f(\varphi_f)$ which takes a value $\varphi_f \in \mathbb{R}$ on face $f$ and produces a function in $\mathcal{V}^p$, i.e.

$$r_f(\varphi_f) : \mathbb{R} \to \mathcal{V}^p \tag{54}$$

The particular face lifting operator we use is defined as,

$$(r_f, w) = -\varphi_f \, \{w\}|_f \qquad \forall w \in \mathcal{V}_h^p \tag{55}$$

Next, we will consider how to numerically calculate $r_f(\varphi_f)$ and then show what $r_f(\varphi_f)$ looks like for varying $p$. Since $r_f \in \mathcal{V}_h^p$ we may write it using our basis functions $\phi_i(x)$,

$$r_f(x) = \sum_{i=1}^{N} a_i \phi_i(x) \tag{56}$$

Then, picking $w = \phi_j(x)$, the left-hand side integral can be written as,

$$(r_f, \phi_j) = \sum_{i=1}^{N} M_{ij} a_i \tag{57}$$

where $M_{ij} = (\phi_i, \phi_j)$ are the entries of the usual mass matrix.

The lifting operator for $\varphi = 1$ is shown in Figure 1 for $p = 0 - 3$.

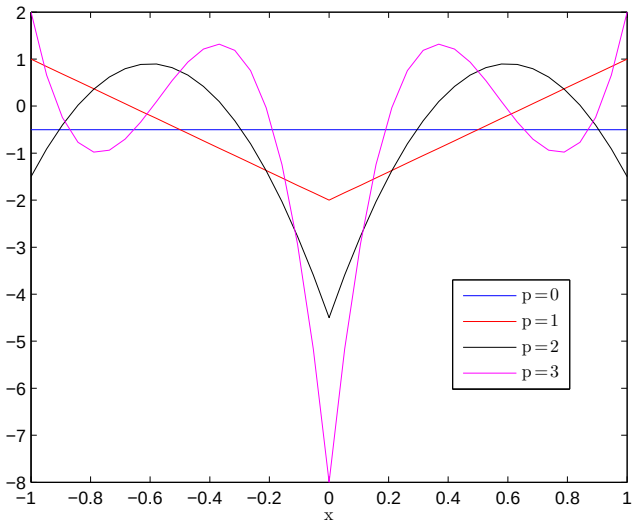


Figure 1: Lifting operator $r_f(1)$ for different $p$ (note $r_f = 0$ for elements not touching $f$)

### 7.3.4 BR2 method

The second method of Bassi and Rebay, commonly referred to as BR2, has a number of desirable properties and as a result is one of the more frequently used DG approximations for diffusive (second-order) operators. The weak form associated with BR2 method can be written as,

$$R(v, \bar{F}, w, \lambda) = \langle w, \bar{F} \rangle + \langle \lambda, v - g \rangle - (\partial_x w, F) + (w, bv - f) + \sum_{\Gamma_i} \hat{F}_I [\![w]\!] \tag{58}$$

$$- \sum_{\Gamma_i} \{\nu \partial_x v\} [\![w]\!] - \sum_{\Gamma_i} [\![v]\!] \{\nu \partial_x w\} \tag{59}$$

$$- \sum_{\Gamma_i} \eta_f \{\nu r_f([\![v]\!])\} [\![w]\!] \tag{60}$$

The first row of this equation are the standard boundary condition, element interior, and convective DG flux terms. The second row of this equation are terms needed for primal and adjoint consistency of the discretization, respectively. And, the last term (on the third row) is required for stability. In particular, stability analysis shows that $\eta_f$ should be greater than or equal to the number of faces in an element. So, for this one-dimensional problem, $\eta_f \geq 2$. We will consider more about these issues of consistency and stability later.

Finally, we note that the adjoint consistency term (the sum over the faces of $[\![v]\!]\{\nu\partial_x w\}$) cannot be written in a flux-like form on the faces (i.e. in the form $\hat{G}[\![w]\!]$). Thus, it is difficult (impossible?) to think of the BR2 scheme as being a specific diffusive flux. And, since DG methods which are adjoint consistent must have the same term, that is a conclusion about any adjoint-consistent DG method.

### 7.4 Mixed forms

Another common approach for approximating second-order PDE is to express the second-order PDE as a system of first-order PDE. Thus, Equations 1 and 2 are written as,

$$[F(u,q)]_x + bu - f = 0, \qquad \forall x \in \Omega \tag{61}$$

$$q - u_x = 0, \qquad \forall x \in \Omega \tag{62}$$

$$F(u,q) = cu - \nu q \tag{63}$$

In mixed approaches, both Equation 61 and 62 are discretized. In fact, the BR2 method was originally derived using this mixed approach. Let's look at that derivation now.

The elemental DG residual for Equation 61 is essentially unchanged from the functional form written in Equation 45 except that $\partial_x v$ is replaced with $q$ in the evaluation of $F$ and $\hat{F}$. To be clear, we will explicitly write out these differences in this derivation, so the elemental DG residual in a mixed form derivation is,

$$R_\kappa(v,q,w) = \langle w, \hat{F}\rangle_\kappa - (\partial_x w, cv - \nu q)_\kappa + (w, bv - f)_\kappa \tag{64}$$

$$\hat{F}(v^+,v^-,q^+,q^-) = \hat{F}_I(v^+,v^-) + \hat{F}_V(v^+,v^-,q^+,q^-) \tag{65}$$

For element faces that are on the domain boundary, we will also need to enforce the boundary conditions (using the Lagrange multiplier approach as before) but do not include those terms for now.

We can similarly derive a weak form of Equation (62) by multiplying by a weighting function $\zeta \in \mathcal{V}_h^p$ and integrating by parts to produce,

$$R_\kappa^q(v,q,\zeta) = -\langle \zeta, \hat{u}\rangle_\kappa + (\zeta_x, v)_\kappa + (\zeta, q)_\kappa \tag{66}$$

The BR2 scheme is given by the following choices for $\hat{u}$ and $\hat{F}_V$:

$$\hat{u} = \{v\} \tag{67}$$

$$\hat{F}_V = -\{\nu\partial_x v\} - \eta_f\{\nu r_f([\![v]\!])\} \tag{68}$$

In fact, these choices lead to a method that is slightly different from the one given in Equation (60). However, in the case where $\nu$ is constant, they can be shown to be identical. This proof though is very tedious and has been done in Section 3.2 of Arnold et. al.[1].

Finally, this mixed form suggests a different implementation of the BR2 (and many other similar DG methods). Specifically, we note that Equation (64) requires $q$ to be available in order to evaluate the residual. One possibility would be to include extra unknowns in the global problem (i.e. for the $q$'s). However, that would be a significant cost. Alternatively, we can express $q$ in terms of $v$. Specifically, from the residuals in Equation (66) (upon setting them to zero), $q$ will require a mass

matrix inversion in each element and the right-hand side will depend on $v$ within the element as well as the face-neighboring values of $v$ (through $\hat{u}$).

# 8 Hybridized DG formulations

## 8.1 Solution space

Hybridized DG (HDG) formulations introduce new unknowns (compared to the DG formulations described above) representing the trace values of the solution ($v$) on the element faces. However, HDG is constructed so that the elemental solution variables $v$ and, for problems involving diffusion, the solution gradients $q$, only depend on the trace values for that element. Thus, $v$ and $q$ can be removed from the global problem so that HDG methods solve a global problem that only for the trace unknowns. After the trace unknowns are solved, then the elemental solutions can be determined (in a completely parallel fashion).

HDG formulations are generally derived from the mixed form as described in Section 7.4. The dependent states determined by the HDG formulation will be $(v, q, \hat{v})$ which approximate, respectively, the solution and solution gradient in the elements and the trace values of the solution on element faces. In our one-dimensional setting, the trace values are real numbers associated with each (unique) element face. So, for $N_e$ elements, then there will be $N_e+1$ trace values. Specifically, we define the space of the trace values as $\mathcal{M}_h$ where in this one-dimensional setting,

$$\mathcal{M}_h = \mathbb{R}^{N_e+1} \tag{69}$$

Thus, HDG solutions live in the following spaces $(v, q, \hat{v}) \in (\mathcal{V}_h^p, \mathcal{V}_h^p, \mathcal{M}_h)$.

## 8.2 Elemental and face residuals

For all $\kappa$, the HDG formulation requires,

$$R_\kappa^v(v, q, \hat{v}, w) \equiv \langle w, \hat{F}\rangle_\kappa - (\partial_x w, cv - \nu q)_\kappa + (w, bv - f)_\kappa = 0, \quad \forall w \in \mathcal{P}^p(\kappa) \tag{70}$$

$$R_\kappa^q(v, q, \hat{v}, \zeta) \equiv -\langle \zeta, \hat{v}\rangle_\kappa + (\zeta_x, v)_\kappa + (\zeta, q)_\kappa = 0, \quad \forall \zeta \in \mathcal{P}^p(\kappa) \tag{71}$$

In HDG, the flux $\hat{f}$ is allowed to depend on $v$, $q$, and $\hat{v}$ and is commonly assumed to have the following form,

$$\hat{F}^\pm = c\hat{v} - \nu q^\pm \pm \tau^\pm (v^\pm - \hat{v}) \tag{72}$$

where $\tau$ is a stabilization parameter that remains to be chosen. Note that $q$ and $v$ in the expression for $\hat{F}$ are evaluated from within the element, thus, on a given face, $\hat{F}$ could have different values when evaluated from the two elements containing that face. Since this would lead to a violation of conservation, HDG then requires that the flux evaluated from either element at a given interior face be the same. Specifically,

$$R_f^{\hat{F}} \equiv \left[\!\left[\hat{F}\right]\!\right]_f = 0, \quad \forall f \in \Gamma_i \tag{73}$$

Finally, the Dirichlet boundary conditions are set on the boundary trace values,

$$R_0^{\hat{F}} \equiv \hat{v}(0) - g(0) = 0, \quad R_{N_e}^{\hat{F}} \equiv \hat{v}(1) - g(1) = 0 \tag{74}$$

### 8.3 Choice of stabilization parameter $\tau$

A variety of choices exist for $\tau$. Most involve adding a contribution for the convection and diffusion terms such that desired behavior (e.g. optimal $O(h^{p+1})$ accuracy for both $v$ and $q$) in the limit of pure convection or pure diffusion. Discussion of the options and their properties can be found in the literature[2, 3, 4].

A simple option which achieves $p+1$ accuracy for both $v$ and $q$ is,

$$\begin{aligned} \tau^{\pm} &= \tau^{\pm}_{\text{inv}} + \tau^{\pm}_{\text{vis}} && (75)\\ \tau^{\pm}_{\text{inv}} &= |c| && (76)\\ \tau^{\pm}_{\text{vis}} &= \nu/l_{\text{ref}} && (77) \end{aligned}$$

where $l_{\text{ref}}$ is a reference length that should be $O(1)$, i.e. independent of the mesh size. For the 1D problem consider here where the domain is of unit length, then $l_{\text{ref}} = 1$ is a reasonable choice. Note that for this choice, $\tau^+ = \tau^-$ is single-valued. Some properties of this choice are discussed in the above referenced literature. In particular, in the convective limit, this choice returns a purely upwinded flux, i.e. $\hat{F} = cv^-$ for $c > 0$ and $\hat{F} = cv^+$ otherwise. Somewhat surprisingly, $\hat{v}$ can be shown to be equal to the average of the elemental traces, i.e. $(v^- + v^+)/2$.

Another option which again achieves $p+1$ accuracy for both $v$ and $q$ is,

$$\begin{aligned} \eta^{\pm} &= \frac{|c| \mp c}{2|c|} && (78)\\ \tau^{\pm}_{\text{inv}} &= |c|\eta^{\pm} && (79)\\ \tau^{\pm}_{\text{vis}} &= \nu/l_{\text{ref}}\eta^{\pm} && (80) \end{aligned}$$

Thus, for this choice, $\tau$ is double-valued. In the convective limit, the flux is again purely upwinded, and so is $\hat{v}$, i.e. $\hat{v} = v^-$ for $c > 0$ and $\hat{v} = v^+$ otherwise.

### 8.4 Solution of HDG

The special structure of the HDG formulation permits the elemental variables to be first removed from the linear system such that the global linear system only involves the traces $\hat{v}$. To see this, let the solution degrees of freedom be organized with $q$ from the first element, then $v$ from the first element, then $q$ and $v$ from the second element, and so on with the trace values placed at the end. Also, organizing the residuals in a similar manner the linear system has the following form,

$$\begin{bmatrix} A_1^{qq} & A_1^{qv} & & 0 & & \cdots & B_1^{q\hat{v}} \\ A_1^{vq} & A_1^{vv} & & & & & B_1^{v\hat{v}} \\ & 0 & & A_2^{qq} & A_2^{qv} & \cdots & B_2^{q\hat{v}} \\ & & & A_2^{vq} & A_2^{vv} & & B_2^{v\hat{v}} \\ \vdots & & & \vdots & & \vdots & \vdots \\ C_1^{\hat{v}q} & & & C_2^{\hat{v}q} & & \cdots & D^{\hat{v}\hat{v}} \end{bmatrix} \begin{bmatrix} q_1 \\ v_1 \\ q_2 \\ v_2 \\ \vdots \\ \hat{v} \end{bmatrix} = \begin{bmatrix} G_1^q \\ G_1^v \\ G_1^q \\ G_1^v \\ \vdots \\ \hat{G} \end{bmatrix} \tag{81}$$

The right-hand side is related to the boundary conditions and the source terms. Because the $A$ portion of the matrix is block diagonal, it can be efficiently inverted to produce the Schur

complement matrix for $\hat{v}$, i.e.

$$K\hat{v} = \hat{K} \qquad K = D - CA^{-1}B \qquad \hat{K} = \hat{G} - CA^{-1}G \tag{82}$$

Further, note that the $A_i^{qq}$ blocks are mass matrices.

## References

[1] Douglas N. Arnold, Franco Brezzi, Bernardo Cockburn, and L. Donatella Marini. Unified analysis of discontinuous Galerkin methods for elliptical problems. *SIAM Journal on Numerical Analysis*, 39(5):1749–1779, 2002.

[2] N.C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous Galerkin method for linear convection-diffusion equations. *Journal of Computational Physics*, 228(9):3232–3254, 2009.

[3] N.C. Nguyen, J. Peraire, and B. Cockburn. Hybridizable discontinuous galerkin methods. In Jan S. Hesthaven and Einar M. Rnquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*, pages 63–84. Springer Berlin Heidelberg, 2011.

[4] R.M. Kirby, S.J. Sherwin, and B. Cockburn. To cg or to hdg: a comparative study. *Journal of Scientific Computing*, 51(1):183–212, 2012.