# FORTIN OPERATORS FOR DPG ADVECTION DISCRETIZATIONS

PABLO CORTÉS CASTILLO AND JAY GOPALAKRISHNAN

Abstract. We construct Fortin operators for discontinuous Petrov–Galerkin (DPG) discretizations of the advection equation with a piecewise constant divergence-free advection vector $\beta$, on simplicial meshes of any spatial dimension $N$ and for any polynomial degree $k \geq 1$ of the trial space. A minimal test space is built on each element from facet and interior bubbles and later augmented. The Fortin operator is shown to be bounded, uniformly over shape-regular mesh families, in the natural $\beta$-weighted broken test graph norm built on $L_q$ for every $1 < q < \infty$, where $q$ is the exponent conjugate to the trial exponent $p$. A non-characteristic facet condition is assumed when $k \geq 2$, while the lowest-order case requires no such condition and admits characteristic facets. As applications we prove that the practical fully discrete residual minimization method is quasioptimal in the DPG energy norm for every $1 < p < \infty$, with a quasioptimality constant governed solely by the Fortin operator, that its computable residual is a globally reliable and efficient a posteriori error estimator, and that augmenting the test space and changing the test norm improves the convergence rate in the lowest-order case.

## 1. Introduction

The discontinuous Petrov–Galerkin (DPG) method [7, 8, 9] is a residual minimization method. Given a variational formulation over a trial space $X$ and a test space $Y$, one selects the discrete solution by minimizing the residual over a finite dimensional trial subspace $X_h \subset X$, measuring it in the dual norm of $Y$. So posed, the method is immediately stable and quasioptimal. The difficulty is that the dual norm computation requires a solve that is both global and infinite-dimensional. Variational DPG reformulation techniques (by breaking the test space [4, 9]) have shown how to replace the global solves by local element-by-element solves. The infinite-dimensionality is then eliminated by the *practical DPG method* which measures the residual in the local dual norm of a finite dimensional test space $Y_h \subset Y$ instead. The application of this method to advection is studied in this paper. The standard device [11, 14] for studying the price of replacing $Y$ by $Y_h$ is a *Fortin operator*: a linear map $\Pi_h : Y \to Y_h$, bounded independently of the mesh, whose deviation $v - \Pi_h v$ is not seen by the discrete trial space on the mesh. For advection, however, no such operator into a test space of polynomials on the same mesh elements has been available. We construct one in this paper.

We consider the model problem

$$\beta \cdot \nabla u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \Gamma_{\text{in}}, \tag{1}$$

where $\Omega \subset \mathbb{R}^N$ is a Lipschitz polyhedral domain, $\Gamma_{\text{in}} := \{x \in \partial\Omega : \beta(x) \cdot n(x) < 0\}$ is the inflow boundary, and $\beta$ is a piecewise constant, divergence-free advection field. For such problems, it is natural to adopt a functional setting beyond Hilbert spaces. Residual minimization in $L_p$-type graph norms is attractive for transport because it can reduce the over-smearing associated with $L_2$ settings and, as $p \to 1^+$, eliminate Gibbs phenomena near discontinuities [1, 15, 17, 18]. Such a residual minimization for advection–reaction in $L_p(\Omega)$

was carried out in [17], working with globally conforming test spaces. We adopt this avenue, but with broken test spaces. Our trial space pairs discontinuous piecewise polynomials of degree $k-1$ with interface variables of degree $k$, for arbitrary $k \geq 1$. Our minimal local test space on an $N$-simplex $K$ is $V(K) = b_s P_k^{\perp}(K) + b\, P_{k-2}(K)$. Here $b_s$ is a signed sum of facet bubbles tracking the inflow/outflow pattern of $\beta$, $b$ is the interior bubble, and $P_k^{\perp}(K)$ excludes the degree-$k$ interior bubbles (all precisely defined later). This space has the minimal dimension matching the active trial degrees of freedom, and the resulting Fortin operator is bounded uniformly in the mesh.

The main results have three parts. First, we construct this minimal Fortin operator for general $k$ and $N$ and prove its mesh-uniform boundedness in a $\beta$-weighted broken graph norm for every $1 < p < \infty$. Our lowest-order ($k = 1$) results are for any shape-regular mesh, while for $k \geq 2$ we assume a uniform non-characteristic facet condition. Second, the practical method is quasioptimal in the mesh-dependent DPG energy norm, converges at the expected rate in that norm, and admits a globally reliable and efficient residual estimator. Finally, we introduce a scale $\sigma$ into the test norm and show that a mean-preserving augmentation of the test space, costing at most two functions per element, makes the Fortin operator bounded uniformly for every fixed $\sigma$ above the intrinsic element-wise scale. At lowest order this yields optimal first-order convergence in the mesh-independent $L_2(\Omega)$ norm and direct estimator control up to a data oscillation, which vanishes for piecewise constant data (Theorem 9.1). To our knowledge this is the first analysis exhibiting the influence of such a scale on the error of any practical DPG method.

The closest antecedent of our test space is in the transport method of [7], one of the origins of the modern DPG methodology. Optimal $L_2$ estimates were obtained for constant $\beta$ using exactly tailored, flow-aligned test functions that are generally nonpolynomial and may be discontinuous within an element. Its two-dimensional error analysis relies on a flow-dependent mesh-layer condition. The present work instead constructs the explicit minimal polynomial test space $V(K)$ in each $N$-simplex with a Fortin bound for every $1 < p < \infty$, at the cost of weighted test norms and, for $k \geq 2$, a non-characteristic facet condition. The space $V(K)$ above may be viewed as a polynomial surrogate, licensed by the Fortin condition: its two summands play the roles of the trace and interior parts of the earlier space.

Our results complement those of [1]. In the Hilbert setting and for variable convection fields, they achieve discrete inf-sup stability uniformly in the relative orientation of the flow and mesh, in a mesh-independent graph norm, by taking test polynomials of one degree higher on a subgrid refinement of each element, of fixed but unspecified depth. Our construction instead provides an explicit minimal test space of degree $k+N$ on the original mesh, together with a Fortin operator and a posteriori estimates for every $1 < p < \infty$. The mesh-uniform Fortin bound we provide uses, in contrast, natural element-wise weights, and we recover mesh-independent optimality only at lowest order. Furthermore, while their analysis also accommodates variable convection fields, extending our scaling argument beyond piecewise constant fields remains open.

This distinction also clarifies the conclusions of [10]. For an element-wise constant convection field, their attempted local Fortin construction for the convection–reaction problem could not be bounded uniformly in the unweighted graph norm, particularly as element facets align with the convection direction. For pure advection, we show that a uniformly bounded minimal local operator does exist when the test norm carries certain weights (called $a_K$ later). The angle sensitivity is nevertheless genuine for $k \geq 2$. Somewhat surprisingly, at lowest order, our estimates are completely insensitive to the flow angle and to the presence of characteristic facets.

The rest of the paper is organized as follows. Section 2 fixes notation and states the practical DPG method. Section 3 constructs the local test space and Section 4 the Fortin operator, whose unisolvency is proved there. Section 5 develops the reference-element equivalences and scaling identities, which Section 6 uses to prove boundedness of $\Pi_h$ and of its augmented variants. Section 7 places the method in a functional setting, recalls the well-posedness of the undiscretized problem, and proves quasioptimality and convergence rates. Section 8 treats a posteriori error control and Section 9 the lowest-order case with a scaled test norm. Section 10 collects concluding remarks and open problems.

## 2. The DPG method for advection

Let $\Omega_h$ be a conforming simplicial mesh of $\Omega \subset \mathbb{R}^N$, $N \geq 1$. Let $\hat{K}$ denote the unit reference $N$-simplex with vertices $\hat{a}_0, \ldots, \hat{a}_N$ and barycentric coordinates $\hat{\lambda}_0, \ldots, \hat{\lambda}_N$. Each element $K \in \Omega_h$, with vertices $a_0, \ldots, a_N$, is the image of $\hat{K}$ under an affine homeomorphism

$$x = \Phi_K(\hat{x}) = M_K \hat{x} + g_K, \qquad M_K \in \mathbb{R}^{N\times N},\ g_K \in \mathbb{R}^N, \tag{2}$$

mapping $\hat{a}_i$ to $a_i$. Let $h_K = \operatorname{diam} K$. For a measurable set $D$, $|D|$ denotes its measure, and $|\cdot|$ also denotes the Euclidean norm on $\mathbb{R}^N$ and the associated matrix norm, when applied to a vector or matrix; the meaning will be clear from context. Recall the standard relations

$$|M_K| \leq C\, h_K, \qquad h_K \leq C\, |M_K|, \tag{3}$$

with constants depending only on $\hat{K}$. Throughout this paper, we consider families of meshes that are *shape regular*:

$$\kappa_K := |M_K|\, |M_K^{-1}| \leq \kappa_0 \qquad \text{for all } K \in \Omega_h, \tag{4}$$

for a fixed $\kappa_0$.

The graph space for the advection operator $\beta\cdot\nabla$, and a "broken" version of it, are defined, for $1 \leq r \leq \infty$ and any (sub)domain $D \subseteq \Omega$, by

$$W_r(\partial_\beta, D) := \{v \in L_r(D) \,:\, \beta\cdot\nabla v \in L_r(D)\}, \qquad W_r(\partial_\beta, \Omega_h) := \bigtimes_{K\in\Omega_h} W_r(\partial_\beta, K).$$

For an integer $s \geq 0$ and $1 \leq r \leq \infty$, we denote by $W_r^s(D)$ the standard Sobolev space of functions on $D$ whose distributional derivatives up to order $s$ lie in $L_r(D)$, with norm $\|\cdot\|_{W_r^s(D)}$ and seminorm $|v|_{W_r^s(D)} := \big(\sum_{|\alpha|=s} \|\partial^\alpha v\|_{L_r(D)}^r\big)^{1/r}$ (with the usual modification when $r = \infty$), and we write $H^s(D) := W_2^s(D)$. Using a prime ($'$) to denote the dual space, and letting

$$q = p/(p-1) \text{ when } p > 1, \quad \text{and} \quad q = \infty \text{ when } p = 1, \tag{5}$$

be the conjugate exponent of $p$, define for each $K \in \Omega_h$, the element boundary operator $D_K : W_p(\partial_\beta, K) \to \big(W_q(\partial_\beta, K)\big)'$,

$$\langle D_K w, v\rangle := (\beta\cdot\nabla w, v)_K + (w, \beta\cdot\nabla v)_K, \tag{6}$$

where $(\cdot,\cdot)_K$ denotes the $L_2(K)$ inner product, $w \in W_p(\partial_\beta, K)$ and $v \in W_q(\partial_\beta, K)$. Throughout the action of a functional $f \in X'$ on an $x \in X$ is denoted by $f(x)$ or $\langle f, x\rangle_X$ where the subscript is omitted when no confusion can arise.

Let $P_k(D)$ denote the space of polynomials of degree at most $k$ restricted to a subdomain $D$ of $\Omega$ (with the convention that $P_k(D) = \{0\}$ when $k < 0$) and let

$$P_k(\Omega_h) = \bigtimes_{K\in\Omega_h} P_k(K)$$

denote the space of piecewise polynomials of degree at most $k$ on $\Omega$ (which are possibly discontinuous across mesh facets). When $w$ and $v$ are smooth functions (or can be approximated arbitrarily closely by smooth functions) up to (including) the boundary of $K$, such as $v, w \in P_k(\Omega_h)$, integration by parts and the fact that $\beta|_K$ is constant shows that

$$\langle D_K w, v\rangle = \int_{\partial K} (\beta\cdot n)\, w\, v\, ds. \tag{7}$$

Hölder's inequality shows directly that (6) is continuous on the indicated graph spaces. Formula (7) is its boundary representation whenever $w$ and $v$ are smooth.

For a fixed degree $k \geq 1$, let $S_h^k := P_k(\Omega_h) \cap C(\overline{\Omega})$ denote the standard Lagrange finite element space and let $S_{h,0}^k := \{z_h \in S_h^k : z_h|_{\Gamma_{\text{in}}} = 0\}$. Then any $z_h \in S_h^k$ has single-valued traces on mesh interfaces. Abbreviate

$$\langle D_h z, v\rangle := \sum_{K\in\Omega_h} \langle D_K(z|_K), v|_K\rangle, \qquad v \in Y := W_q(\partial_\beta, \Omega_h). \tag{8}$$

(For now, this defines a linear operator $D_h : S_h^k \to Y'$, but we will extend its domain later.) Write $\gamma$ for the element-wise trace, i.e., $\gamma(z_h)|_{\partial K} = z_h|_{\partial K}$. Define the discrete interface spaces

$$\hat{X}_h^k := D_h(S_h^k), \qquad \hat{X}_{h,0}^k := D_h(S_{h,0}^k). \tag{9}$$

The DPG method uses the trial space

$$X_h := P_{k-1}(\Omega_h) \times \hat{X}_{h,0}^k, \tag{10}$$

whose second component incorporates the inflow boundary condition, together with a finite dimensional discrete test space $Y_h \subset P_r(\Omega_h) \subset W_q(\partial_\beta, \Omega_h)$ for some degree $r$ to be determined. Multiplying (1) by a test function $v$ element-wise and integrating by parts gives the bilinear form

$$\mathscr{b}\big((w_h, D_h z_h),\, v\big) := \sum_{K\in\Omega_h} \left[-\int_K w_h\, \beta\cdot\nabla v\, dx + \int_{\partial K} \beta\cdot n\, z_h\, v\, ds\right] \tag{11}$$

for any $w_h \in P_{k-1}(\Omega_h)$, $z_h \in S_h^k$ and $v \in Y_h$. Here we have used (7), which also shows that $\langle D_h z_h, v\rangle$ depends only on $\gamma(z_h)$. The same formula (7) shows that, on replacing the second term of the sum in (11) by $\langle D_h z_h, v\rangle$ of (8), the bilinear form $\mathscr{b}$ allows for test functions $v$ in the larger space $W_q(\partial_\beta, \Omega_h)$, this extension is recorded in (63) of Section 7.

Given $1 \leq p < \infty$ and data $f \in L_p(\Omega)$, set $\ell(v) := (f, v)_\Omega$, a linear functional on $W_q(\partial_\beta, \Omega_h)$. The *practical DPG method* then finds

$$\boldsymbol{u}_h := \operatorname*{arg\,min}_{\boldsymbol{w}_h\in X_h} \big\|\ell - \mathscr{b}(\boldsymbol{w}_h, \cdot)\big\|_{Y_h'} \tag{12}$$

which is the practical DPG solution of the advection problem. The only requirement we shall place on $Y_h$ is that it contain an advection-adapted space $V_h$ of minimal dimension, constructed in Section 3, on which the Fortin operator of this paper is built; see (22). Specification of the dual norm $\|\cdot\|_{Y_h'}$ and the space $V_h$ (done later in (60) and (21), respectively), completes the mathematical definition of the method. The method, as stated in (12), performs a residual minimization over the discrete trial space. For $p = 2$, (12) is equivalent, as is well known [9], to a linear system computable element by element (via the discrete Riesz map of the Hilbert subspace $Y_h$ or via "discrete optimal test functions"), and also to a saddle point problem. For $p \neq 2$ it is equivalent to a mixed system with monotone nonlinearity, in which the Riesz map is replaced by the duality map of the Banach subspace $Y_h$ as shown in [17, eq. (4.2)] and [18], where they also call it "DDMRes method." We will not need these reformulations in our analysis. We will work directly with (12).

## 3. The discrete local test space

The test space of the practical DPG method is constructed element by element. Hence throughout this section, we fix a $K \in \Omega_h$. Using the notation of (2), we write $\lambda_i = \hat{\lambda}_i \circ \Phi_K^{-1}$ for the barycentric coordinates of $K$, $F_i$ for the facet of $K$ opposite to $a_i$, and $n_i$ for the outward unit normal on $F_i$.

**Assumption 3.1.** *The advection field $\beta$ is constant on each element ($\beta|_K \equiv \beta_K \in \mathbb{R}^N$ for all $K \in \Omega_h$) with continuous normal component across element interfaces (so that* $\operatorname{div}\beta = 0$ *on $\Omega$), and $\beta_K \neq 0$ for every $K \in \Omega_h$.*

Fields satisfying Assumption 3.1 arise canonically as the lowest-order Raviart–Thomas interpolant of a divergence-free velocity field. For $\beta \in \mathbb{R}^N \setminus \{0\}$, define its *sign vector* (relative to the facets of $K$)

$$s(\beta) := (s_0, \dots, s_N), \qquad s_i := \operatorname{sgn}(\beta \cdot n_i) \in \{-1, 0, +1\}. \tag{13}$$

Under Assumption 3.1, write $s := s(\beta_K)$ for the sign vector of $\beta_K$ on each $K$ (suppressing its dependence on $K$ from the notation). A facet $F_i$ is called a *characteristic facet* if $s_i = 0$ (and non-characteristic otherwise).

**Lemma 3.2.** *For any $N$-simplex $K$ and any $\beta \in \mathbb{R}^N \setminus \{0\}$, the sign vector $s(\beta)$ of (13) has at least two nonzero entries. In particular, under Assumption 3.1, every $K \in \Omega_h$ has at least two non-characteristic facets.*

*Proof.* Any $N$ of the $N+1$ outward normals $n_i$ of a simplex are linearly independent. Hence the nonzero vector $\beta$ can be orthogonal to at most $N-1$ of them. ☐

**Definition 3.3** (Bubbles)**.** Define the interior and facet bubbles of $K$ by

$$b := \prod_{j=0}^{N} \lambda_j \in P_{N+1}(K), \qquad b_i := \prod_{j \neq i} \lambda_j = \frac{b}{\lambda_i} \in P_N(K). \tag{14}$$

Note that $b > 0$ in the interior of $K$ and $b = 0$ on $\partial K$, while $b_i$ vanishes on $\partial K$ except in the relative interior of $F_i$, denoted by $\operatorname{int} F_i = F_i \setminus \partial F_i$, where it is positive. Using the signs $s_i$ of (13), set

$$b_s := \sum_{i=0}^{N} s_i\, b_i. \tag{15}$$

For $j > N$ let

$$\mathring{P}_j(K) := \{z \in P_j(K) : z|_{\partial K} = 0\} = b\, P_{j-N-1}(K),$$
$$P_j^{\perp}(K) := \{z \in P_j(K) : (z, \mathring{p})_K = 0 \text{ for all } \mathring{p} \in \mathring{P}_j(K)\} = \mathring{P}_j(K)^{\perp} \cap P_j(K),$$

and

$$\mathring{P}_j(K) = \{0\}, \qquad P_j^{\perp}(K) = P_j(K), \qquad \text{when } 0 \leq j \leq N. \tag{16}$$

When $s = s(\beta)$, two properties of $b_{s(\beta)}$ are used repeatedly. First,

$$(\beta \cdot n)\, b_{s(\beta)}\big|_{F_i} = (\beta \cdot n_i)\, s_i\, b_i\big|_{F_i} = |\beta \cdot n_i|\, b_i\big|_{F_i} \geq 0, \qquad i = 0, \dots, N, \tag{17}$$

since on $F_i$ all other facet bubbles vanish. Second, $b_{s(\beta)}$ is nonzero on $\operatorname{int} F_i$ for every non-characteristic facet $F_i$, of which there are at least two for the $\beta$ we consider (see Lemma 3.2), so

$$b_{s(\beta)} \not\equiv 0 \tag{18}$$

is a nonzero polynomial.

**Lemma 3.4** (Trace extension)**.** *The trace map $\gamma(u) := u|_{\partial K}$ restricted to $P_k^\perp(K)$ is a bijection onto $\gamma(P_k(K))$.*

*Proof.* The map $\gamma : P_k(K) \to \gamma(P_k(K))$ has kernel $\mathring{P}_k(K)$. Since $P_k(K) = \mathring{P}_k(K) \oplus P_k^\perp(K)$, the restriction of $\gamma$ to $P_k^\perp(K)$ is injective. It also remains onto, since any $u \in P_k(K)$ splits as $u = \mathring{u} + u^\perp$ with $\mathring{u} \in \mathring{P}_k(K)$ and $u^\perp \in P_k^\perp(K)$, whence $\gamma(u) = \gamma(u^\perp)$. □

**Definition 3.5** (Local and global test spaces)**.** Let $\mathscr{S}$ denote the set of sign vectors $s = (s_0, \dots, s_N) \in \{-1, 0, +1\}^{N+1}$ with at least two nonzero entries. For $k \geq 1$ and any sign vector $s \in \mathscr{S}$, set

$$V_s(K) := b_s\, P_k^\perp(K) \,+\, b\, P_{k-2}(K) \;\subseteq\; P_{k+N}(K), \tag{19}$$

with $b_s$ built from $s$ as in (15). We emphasize that $V_s(K)$ depends on $s$ (through the boundary bubble $b_s$, and only through it). Define the advection-adapted local space by

$$V(K) := V_{s(\beta_K)}(K), \tag{20}$$

whose sign vector $s(\beta_K)$, defined in (13), has at least two nonzero entries by Lemma 3.2. The local spaces (20) are assembled into the Cartesian product

$$V_h := \bigtimes_{K \in \Omega_h} V(K), \tag{21}$$

identifiable as a subspace of $W_\rho(\partial_\beta, \Omega_h)$ for every $1 \leq \rho \leq \infty$. Choose $Y_h$ in (12) to be any computationally convenient space satisfying

$$V_h \;\subseteq\; Y_h \;\subset\; W_q(\partial_\beta, \Omega_h), \qquad \dim Y_h < \infty. \tag{22}$$

(Since $V(K) \subset P_{k+N}(K)$, a simple choice is $Y_h = P_r(\Omega_h)$ for an $r \geq k + N$.)

Our next assumption excludes characteristic facets, uniformly over the mesh family. *It is needed only for degrees $k \geq 2$.* All our results in the lowest-order case $k = 1$ hold without it.

**Assumption 3.6** (Uniformly non-characteristic facets when $k \geq 2$)**.** *There is a $\theta_0 \in (0, 1]$ such that for every $K \in \Omega_h$ and every facet $F_i$ of $K$,*

$$|\beta_K \cdot n_i| \;\geq\; \theta_0\, |\beta_K|. \tag{23}$$

Assumption 3.6 forces the signs $s_i$ in (13) to lie in $\{-1, +1\}$. It is automatically satisfied, e.g., when $\beta$ is a fixed constant vector on $\Omega$ and the mesh family avoids facets parallel to it. Note also that on any single fixed mesh with no characteristic facets, some $\theta_0 > 0$ exists trivially. But the content of the assumption is its uniformity over a family of meshes. All invocations of Assumption 3.6 in this section use only its implication that $s_i \in \{-1, +1\}$, but the uniformity will be needed later. Throughout this paper, whenever $k \geq 2$ we require Assumption 3.6, while for $k = 1$ only Assumption 3.1 is required and *characteristic facets are allowed.*

**Lemma 3.7.** *Let $\beta \in \mathbb{R}^N \setminus \{0\}$ and let $s = s(\beta)$ determine the characteristic facets of $K$. Suppose that either Assumption 3.6 holds or $k = 1$. Then any $u \in P_k^\perp(K)$ that vanishes on all non-characteristic facets must vanish everywhere.*

*Proof.* There are two cases:

(i) When Assumption 3.6 holds, no facet is characteristic, and this gives $u|_{\partial K} = 0$, i.e. $u \in \mathring{P}_k(K) \cap P_k^\perp(K) = \{0\}$.

(ii) If $k = 1$, then $u$ is affine and vanishes on at least two facets by Lemma 3.2, so $u$ must vanish (since an affine function vanishing on a facet $F_i$ is a multiple of $\lambda_i$).

In either case, $u \equiv 0$. □

**Lemma 3.8** (Structure of $V(K)$)**.** *Suppose Assumption 3.1 holds, and that either Assumption 3.6 holds or $k = 1$. Then the sum in (19) with $s = s(\beta_K)$ is direct and*

$$\dim V(K) = \dim \gamma(P_k(K)) + \dim P_{k-2}(K).$$

*Proof.* Since $b_s$ and $b$ are nontrivial polynomials (see (18)), the multiplication maps $u \mapsto b_s u$ and $z \mapsto bz$ are injective. Hence $\dim(bP_{k-2}(K)) = \dim P_{k-2}(K)$, and using also Lemma 3.4, $\dim b_s P_k^\perp(K) = \dim P_k^\perp(K) = \dim \gamma(P_k(K))$.

It only remains to show the intersection of the two summands is trivial. Suppose $b_s u = bz$ with $u \in P_k^\perp(K)$, $z \in P_{k-2}(K)$. Restricting to $\partial K$ and using $b|_{\partial K} = 0$ gives $b_s u = 0$ on $\partial K$. Since $b_s|_{F_i} = s_i b_i \neq 0$ on $\operatorname{int} F_i$ for every non-characteristic $F_i$, the trace of $u$ vanishes on all non-characteristic facets. By Lemma 3.7, $u = 0$, so $bz = 0$, which implies $z = 0$. □

Lemma 3.8 shows that $V(K)$ is *minimal* in the sense that $\dim V(K)$ equals the number of Fortin conditions (24) of the next section, with (24) being the smallest set of conditions that forces the Fortin property, since the trial degrees of freedom active on $K$ enter (11) only through $\gamma(P_k(K))$ and through $\beta \cdot \nabla P_{k-1}(K) = P_{k-2}(K)$.

## 4. The Fortin operator

This section establishes the unisolvency of equations defining our first Fortin operator, deferring all quantitative bounds to Section 6, which draws on the scaling estimates of Section 5.

**Lemma 4.1** (Definition of $\Pi_K$)**.** *Suppose Assumption 3.1 holds, and that either Assumption 3.6 holds or $k = 1$. For every $v \in W_q(\partial_\beta, K)$ there is a unique $\Pi_K v \in V(K)$ such that*

$$\langle D_K w, \Pi_K v - v\rangle = 0 \qquad \text{for all } w \in P_k^\perp(K), \tag{24a}$$

$$(z, \Pi_K v - v)_K = 0 \qquad \text{for all } z \in P_{k-2}(K). \tag{24b}$$

*Moreover, (24a) continues to hold for all $w \in P_k(K)$. (For $k = 1$, condition (24b) is vacuous.)*

*Proof.* The terms $\langle D_K w, v\rangle$ and $(z, v)_K$ are well defined when $v \in W_q(\partial_\beta, K)$ (and when $z, w$ are polynomials) due to (6). By Lemmas 3.4 and 3.8, the number of (independent) conditions in (24) equals $\dim P_k^\perp(K) + \dim P_{k-2}(K) = \dim V(K)$, so (24) is a square linear system for $\Pi_K v$ and it suffices to prove uniqueness. Accordingly, we need to prove that any $\varphi = \Pi_K v \in V(K)$ satisfying (24) with $v = 0$ must vanish.

By Lemma 3.8 we may decompose

$$\varphi = b_s \vartheta + b\,\psi, \qquad \vartheta \in P_k^\perp(K), \quad \psi \in P_{k-2}(K),$$

uniquely. By (24a), (7) and (17), since $b$ vanishes on $\partial K$,

$$0 = \langle D_K w, \varphi\rangle = \int_{\partial K} (\beta \cdot n)\, w\, b_s\, \vartheta\, ds = \sum_{i=0}^{N} \int_{F_i} |\beta \cdot n|\; b_i\, w\, \vartheta\, ds,$$

for any $w \in P_k^{\perp}(K)$. Choose $w = \vartheta$. Since every summand is then nonnegative and $b_i > 0$ on $\operatorname{int} F_i$, the trace of $\vartheta$ must vanish on every non-characteristic facet. By Lemma 3.7, $\vartheta = 0$. Thus $\varphi = b\psi$, and choosing the test function $z$ in (24b) to be $\psi$ gives $(b\psi, \psi)_K = 0$, which implies that $\psi = 0$. Thus $\varphi = 0$.

The final claim that (24a) continues to hold for all $w \in P_k(K)$ is seen by splitting $w \in P_k(K)$ into $w = \mathring{w} + w^{\perp}$ with $\mathring{w} \in \mathring{P}_k(K)$, $w^{\perp} \in P_k^{\perp}(K)$, and observing that

$$\langle D_K \mathring{w}, \varPi_K v - v\rangle = 0,$$

because $\mathring{w}$ vanishes on $\partial K$. □

*Remark* 4.2 (Failure at an exactly characteristic facet when $k = 2$). Suppose $N = 2$, $k = 2$, and suppose $\beta_K \cdot n_0 = 0$ for the facet $F_0$ of a triangle $K$ (so $s_0 = 0$), while $F_1, F_2$ are non-characteristic with nonzero signs $s_1, s_2$. Then clearly Assumption 3.6 does not hold. We show that the result of Lemma 4.1 cannot hold. Note that $b_s = s_1 b_1 + s_2 b_2 = \lambda_0(s_1\lambda_2 + s_2\lambda_1)$, since the $b_0 = \lambda_1\lambda_2$ term is absent. Let $u := \lambda_1\lambda_2$. Clearly $u$ is in $P_2^{\perp}(K) = P_2(K)$ as $k = 2 = N$ and $b_s\, u = \lambda_0\lambda_1\lambda_2\,(s_1\lambda_2 + s_2\lambda_1) = b\,(s_1\lambda_2 + s_2\lambda_1)$. Because $b$ vanishes identically on $\partial K$, so does $b_s\, u$. Hence $\langle D_K w, b_s\, u\rangle = 0$ for every $w \in P_2^{\perp}(K)$, i.e., (24a) holds trivially. Set $z_0 = (b_s\, u, 1)_K/(b, 1)_K$. Then $\varphi := b_s\, u - b\, z_0 = b\,(s_1\lambda_2 + s_2\lambda_1 - z_0)$ has zero mean and, since $s_1\lambda_2 + s_2\lambda_1$ is linear but not constant, $\varphi$ is a nonzero function in $V(K)$. Since $\varphi$ vanishes on $\partial K$, (24a) holds and since $\varphi$ has zero mean, (24b) holds too, showing that $\varPi_K$ cannot be uniquely defined by (24).

**Definition 4.3.** Define $\varPi_h : W_q(\partial_\beta, \varOmega_h) \to V_h$ element-wise by $(\varPi_h v)|_K := \varPi_K(v|_K)$.

**Lemma 4.4** (Fortin property)**.** *Suppose Assumption 3.1 holds, and that either Assumption 3.6 holds or $k = 1$. For all $v \in W_q(\partial_\beta, \varOmega_h)$ and all $(w, \mu) \in P_{k-1}(\varOmega_h) \times \hat{X}_h^k$,*

$$\mathscr{b}\big((w, \mu),\, \varPi_h v - v\big) = 0.$$

*Proof.* By (9), $\mu = D_h z_h$ for some $z_h \in S_h^k$. Let $\delta := \varPi_h v - v$ and fix $K \in \varOmega_h$. Since $w|_K \in P_{k-1}(K) \subset P_k(K)$, the definition (6) gives

$$-(w, \beta \cdot \nabla\delta)_K = (\beta \cdot \nabla w, \delta)_K - \langle D_K w, \delta\rangle.$$

Here $\beta \cdot \nabla w \in P_{k-2}(K)$ because $\beta_K$ is constant, so the first term vanishes by (24b). Combining this identity with (8), we obtain

$$-(w, \beta \cdot \nabla\delta)_K + \langle D_K(z_h|_K), \delta\rangle = \langle D_K(z_h|_K - w|_K), \delta\rangle.$$

The last term vanishes by Lemma 4.1, because $z_h|_K - w|_K \in P_k(K)$. Summing over $K$ completes the proof. □

## 5. Scaling estimates

This section develops the element-wise estimates needed to bound $\varPi_K$. The strategy is to prove equivalences on the reference simplex $\hat{K}$ by homogeneity and compactness, treating the advection vector as a variable, and then to transfer them to a physical element $K$ by the affine map (2), tracking how every quantity scales.

5.1. **Pullback conventions.** For $v$ defined on $K$ write $\hat v := v \circ \Phi_K$. The chain rule gives $\hat\nabla \hat v = M_K^t (\nabla v) \circ \Phi_K$, whence for any constant vector $\beta \in \mathbb{R}^N$,

$$(\beta \cdot \nabla v) \circ \Phi_K = \hat\beta \cdot \hat\nabla \hat v, \qquad \hat\beta := M_K^{-1} \beta. \tag{25}$$

Since $\Phi_K$ maps vertices to vertices, $\lambda_i \circ \Phi_K = \hat\lambda_i$, so the bubbles satisfy $b_i \circ \Phi_K = \hat b_i$ and $b \circ \Phi_K = \hat b$, where $\hat b_i$ and $\hat b$ are defined by (14) using $\hat\lambda_i$ in place of $\lambda_i$. Moreover, because the Jacobian of $\Phi_K$ is constant, pullback preserves $L_2$ orthogonality, so

$$\{\hat u = u \circ \Phi_K : u \in \mathring{P}_k(K)\} = \mathring{P}_k(\hat K), \qquad \{\hat u = u \circ \Phi_K : u \in P_k^\perp(K)\} = P_k^\perp(\hat K). \tag{26}$$

**Lemma 5.1** (Sign pattern and non-characteristicity are affine invariants)**.** *Let $\hat n_i$ and $n_i$ denote the outward unit normals of $\hat K$ and $K$ on corresponding facets, and let $\beta \in \mathbb{R}^N \setminus \{0\}$ and $\hat\beta = M_K^{-1}\beta$. Then the sign vector of (13) is mapped exactly,*

$$\operatorname{sgn}(\hat\beta \cdot \hat n_i) = \operatorname{sgn}(\beta \cdot n_i), \qquad i.e., \qquad s(\hat\beta) = s(\beta),$$

*and if $|\beta \cdot n_i| \ge \theta_0 |\beta|$ for all $i$, then*

$$|\hat\beta \cdot \hat n_i| \;\ge\; \frac{\theta_0}{\kappa_K}\, |\hat\beta|$$

*for all $i$, where $\kappa_K$ is as in (4).*

*Proof.* Since $\lambda_i$ is affine, vanishes on $F_i$, and equals 1 at $a_i$, its (constant) gradient $\nabla\lambda_i$ points from $F_i$ toward $a_i$, so the outward unit normal is $n_i = -\nabla\lambda_i/|\nabla\lambda_i|$. Likewise $\hat n_i = -\hat\nabla\hat\lambda_i/|\hat\nabla\hat\lambda_i|$ on $\hat K$. Applying (25) to the affine function $v = \lambda_i$ (whose gradient is constant, so composing with $\Phi_K$ has no effect) gives the equality of constants $\beta \cdot \nabla\lambda_i = \hat\beta \cdot \hat\nabla\hat\lambda_i$. Hence

$$\hat\beta \cdot \hat n_i \;=\; -\frac{\hat\beta \cdot \hat\nabla\hat\lambda_i}{|\hat\nabla\hat\lambda_i|} \;=\; -\frac{\beta \cdot \nabla\lambda_i}{|\hat\nabla\hat\lambda_i|} \;=\; \frac{|\nabla\lambda_i|}{|\hat\nabla\hat\lambda_i|}\,(\beta \cdot n_i), \tag{27}$$

a positive multiple of $\beta \cdot n_i$, which proves the sign assertion. For the second claim, the chain rule $\hat\nabla\hat\lambda_i = M_K^t \nabla\lambda_i$ gives $|\hat\nabla\hat\lambda_i| \le |M_K|\,|\nabla\lambda_i|$, so (27) yields

$$|\hat\beta \cdot \hat n_i| \ge \frac{|\beta \cdot n_i|}{|M_K|} \ge \frac{\theta_0|\beta|}{|M_K|} \ge \frac{\theta_0\,|\hat\beta|}{|M_K||M_K^{-1}|} = \frac{\theta_0}{\kappa_K}|\hat\beta|,$$

using $|\hat\beta| \le |M_K^{-1}||\beta|$. □

5.2. **Parametrization of bubbles.** Throughout this subsection, let $s$ be a sign vector in $\mathscr{S}$ (see Definition 3.5), and let $b_s$ and $\hat b_s$ be as in (15) on $K$ and $\hat K$, respectively. Fix once and for all a basis $\hat\psi_1, \ldots, \hat\psi_m$ of $P_k^\perp(\hat K)$, where

$$m := \dim P_k^\perp(\hat K) = \dim \gamma(P_k(\hat K)) = \binom{N+k}{N} - \binom{k-1}{N},$$

and let $\psi_i := \hat\psi_i \circ \Phi_K^{-1} \in P_k^\perp(K)$ (see (26)). For $c \in \mathbb{R}^m$ define

$$w_K(c) := \sum_{i=1}^m c_i \psi_i \in P_k^\perp(K), \qquad \varphi_K(c, s) := b_s\, w_K(c) \in b_s P_k^\perp(K) \subset V_s(K), \tag{28}$$

and analogously $w_{\hat K}(c)$, $\varphi_{\hat K}(c,s)$ on $\hat K$, so that $w_K(c)\circ\Phi_K = w_{\hat K}(c)$ and $\varphi_K(c,s)\circ\Phi_K = \varphi_{\hat K}(c,s)$. For $1\le p\le\infty$ and a constant vector $\beta\neq 0$ with $s(\beta)=s$ (see (13)), define

$$S_{p,K}(\beta,c,s) := \|\beta\cdot\nabla\varphi_K(c,s)\|_{L_p(K)}, \qquad \tilde S_{p,K}(\beta,c) := \|\beta\cdot\nabla w_K(c)\|_{L_p(K)}, \tag{29a}$$

$$f_{p,K}(c,s) := \|\varphi_K(c,s)\|_{L_p(K)}, \qquad \tilde f_{p,K}(c) := \|w_K(c)\|_{L_p(K)}, \tag{29b}$$

$$d_K(\beta,c) := \sum_{i=0}^{N}\int_{F_i} |\beta\cdot n|\, b_i\, w_K(c)^2\, ds. \tag{29c}$$

By (17) and (7), using $\varphi_K(c,s)$ with $s=s(\beta)$,

$$d_K(\beta,c) = \int_{\partial K} (\beta\cdot n)\, w_K(c)\,\varphi_K(c,s(\beta))\, ds = \langle D_K w_K(c), \varphi_K(c,s(\beta))\rangle. \tag{30}$$

**Lemma 5.2** (Reference element estimates)**.** *Let $1\le p\le\infty$. There exist positive constants $C_0, C_1, C_2$ depending only on $N,k,p$, and a positive constant $C_0^\theta$ depending on $N,k,p$ and additionally on $\theta$, all independent of $s$, such that for all $c\in\mathbb{R}^m$ and all $\beta\in\mathbb{R}^N\setminus\{0\}$ with $s(\beta)=s$, the following statements hold:*

*(1) On the reference element $\hat K$,*

$$S_{p,\hat K}(\beta,c,s) + \tilde S_{p,\hat K}(\beta,c) \le C_2\,|\beta|\,|c|, \tag{31a}$$

$$C_1|c| \,\le\, f_{p,\hat K}(c,s) \le C_2|c|, \qquad C_1|c| \,\le\, \tilde f_{p,\hat K}(c) \le C_2|c|, \tag{31b}$$

$$d_{\hat K}(\beta,c) \le C_2\,|\beta|\,|c|^2. \tag{31c}$$

*(2) If $k=1$, then*

$$C_0\,|\beta|\,|c|^2 \,\le\, d_{\hat K}(\beta,c). \tag{32}$$

*(3) If there is a $\theta\in(0,1]$ such that $|\beta\cdot\hat n_i|\ge\theta|\beta|$ for all $i$, then*

$$C_0^\theta\,|\beta|\,|c|^2 \,\le\, d_{\hat K}(\beta,c). \tag{33}$$

*Proof.* It suffices to prove (31) with the constants $C_1(s)$ and $C_2(s)$ depending on $s$ for the $s$-dependent quantities. Since $s$ ranges over the finite subset $\mathcal{S}$ (of Definition 3.5) the same inequalities then continue to hold with $C_2 = \max_{s\in\mathcal{S}} C_2(s)$ in place of $C_2(s)$ and $C_1 = \min_{s\in\mathcal{S}} C_1(s)$, respectively. (The constants in (32) and (33) are obviously $s$-independent since those inequalities govern $s$-independent quantities.)

All quantities are positively homogeneous in $|\beta|$ and $|c|$: indeed $c\mapsto w_{\hat K}(c)$ and $c\mapsto \varphi_{\hat K}(c,s)$ are linear, and (29c) is homogeneous of degree 1 in $|\beta|$ (as seen by writing $|\beta\cdot\hat n| = |\beta|\,|(\beta/|\beta|)\cdot\hat n|$) and of degree 2 in $c$. Hence it suffices to prove all bounds for $|\beta|=|c|=1$. The upper bounds follow because each quantity is a continuous function of $(\beta,c)$ on the compact set $\{|\beta|=1\}\times\{|c|=1\}$.

For the lower bounds, start with (31b). Observe that $f_{p,\hat K}$ and $\tilde f_{p,\hat K}$ are norms of linear images of $c$, so $\tilde f_{p,\hat K}(c)$ vanishes only if $w_{\hat K}(c)=0$, i.e. only if $c=0$, since the $\hat\psi_i$ form a basis. Similarly, $f_{p,\hat K}(c,s)$ vanishes only if $\hat b_s\, w_{\hat K}(c)=0$, which forces $w_{\hat K}(c)=0$ and $c=0$. Both $f_{p,\hat K}(c,s)$ and $\tilde f_{p,\hat K}(c)$ are therefore norms on $\mathbb{R}^m$, equivalent to $|c|$, with constants depending only on $(N,k,p)$ and the fixed basis.

For (33), consider the set

$$A_\theta := \big\{(\beta,c) : |\beta|=1,\ |c|=1,\ |\beta\cdot\hat n_i|\ge\theta \ \text{ for } i=0,\dots,N\big\},$$

which is compact. The function $d_{\hat K}$ is obviously continuous on $A_\theta$. It is also positive there. Indeed, each summand of (29c) defining $d_{\hat K}$ is nonnegative, so if all of them vanish, then, since $|\beta \cdot \hat n| \ge \theta > 0$ on each facet and $\hat b_i > 0$ on int $\hat F_i$, the trace of $w_{\hat K}(c)$ vanishes on every facet, thus making $w_{\hat K}(c) \in \mathring{P}_k(\hat K) \cap P_k^\perp(\hat K) = \{0\}$, contradicting $|c| = 1$. The minimum of $d_{\hat K}$ over $A_\theta$ is attained and positive; call it $C_0^\theta$.

For (32), we apply the same argument noting that when $k = 1$ we may use the larger compact set $A := \{(\beta, c) : |\beta| = 1,\ |c| = 1\}$. If $d_{\hat K}(\beta, c) = 0$, then the trace of the affine function $w_{\hat K}(c)$ vanishes on every facet with $\beta \cdot \hat n_i \ne 0$ (of which there are at least two, by Lemma 3.2) and an affine function vanishing on two distinct facets is identically zero, contradicting $|c| = 1$. Hence for $k = 1$ the minimum of $d_{\hat K}$ over all of $A$, named $C_0$, is positive (and $C_0$ is $\theta$-free). ☐

*Remark* 5.3. No lower bound for $S_{p,\hat K}$ of the form $C|\beta||c| \le S_{p,\hat K}(\beta, c, s)$ is claimed in Lemma 5.2. For $k \ge 2$ such a bound is false in general, since $\beta \cdot \nabla \varphi_{\hat K}(c, s)$ can vanish for $c \ne 0$ when $\varphi_{\hat K}(c, s)$ is constant along streamlines. It is also never needed in our analysis.

**Lemma 5.4** (Scaling identities)**.** *Let $1 \le p \le \infty$ (with the convention $1/\infty = 0$), $c \in \mathbb{R}^m$, and let $\beta \ne 0$ satisfy $s(\beta) = s$ on $K$. Then, with $\hat\beta = M_K^{-1}\beta$,*

$$S_{p,K}(\beta, c, s) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/p} S_{p,\hat K}(\hat\beta, c, s), \qquad \tilde S_{p,K}(\beta, c) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/p} \tilde S_{p,\hat K}(\hat\beta, c), \tag{34a}$$

$$f_{p,K}(c, s) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/p} f_{p,\hat K}(c, s), \qquad \tilde f_{p,K}(c) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/p} \tilde f_{p,\hat K}(c), \tag{34b}$$

$$d_K(\beta, c) = \frac{|K|}{|\hat K|}\, d_{\hat K}(\hat\beta, c). \tag{34c}$$

*Proof.* Since $\varphi_K(c, s) \circ \Phi_K = \varphi_{\hat K}(c, s)$ and $w_K(c) \circ \Phi_K = w_{\hat K}(c)$, the identities (34a)–(34b) follow from (25) and the change of variables formula (for $p = \infty$, suprema are invariant under composition with $\Phi_K$). For (34c) we use the representation (30) with $s = s(\beta)$ and integrate by parts using (6):

$$\begin{aligned} d_K(\beta, c) &= (\beta \cdot \nabla w_K(c), \varphi_K(c, s))_K + (w_K(c), \beta \cdot \nabla \varphi_K(c, s))_K \\ &= \frac{|K|}{|\hat K|}\Big[(\hat\beta \cdot \hat\nabla w_{\hat K}(c), \varphi_{\hat K}(c, s))_{\hat K} + (w_{\hat K}(c), \hat\beta \cdot \hat\nabla \varphi_{\hat K}(c, s))_{\hat K}\Big] = \frac{|K|}{|\hat K|}\, d_{\hat K}(\hat\beta, c), \end{aligned}$$

where the last equality is (30) on $\hat K$, valid because $s(\hat\beta) = s(\beta) = s$ on $\hat K$ by Lemma 5.1. ☐

**Lemma 5.5** (Interior bubble estimate)**.** *Let $j \ge 0$, let $1 \le r < \infty$, and let $r'$ be the conjugate exponent. There is a $C = C(N, j, r)$ such that for every simplex $K$ and every $z \in P_j(K)$,*

$$\|bz\|_{L_r(K)}\, \|z\|_{L_{r'}(K)} \le C(z, bz)_K.$$

*Proof.* On $\hat K$, the three maps $z \mapsto \|\hat b z\|_{L_r(\hat K)}$, $z \mapsto \|z\|_{L_{r'}(\hat K)}$ and $z \mapsto (z, \hat b z)_{\hat K}^{1/2}$ are norms on the finite dimensional space $P_j(\hat K)$ (the first because $\hat b \ne 0$ and polynomials form an integral domain, the third because $\hat b > 0$ in the interior of $\hat K$), hence pairwise equivalent with constants depending only on $(N, j, r)$. This proves the claim on $\hat K$. Mapping to $K$, the left-hand side scales by $(|K|/|\hat K|)^{1/r}(|K|/|\hat K|)^{1/r'} = |K|/|\hat K|$ and the right-hand side by $|K|/|\hat K|$ as well, so the inequality transfers with the same constant. ☐

From now on, $A \lesssim B$ means that $A \le CB$ with a positive constant $C$ depending on $N, k, p, \kappa_0$, and generally on $\theta_0$, but independent of $\theta_0$ when $k = 1$.

**Lemma 5.6** (Element estimates). *Suppose Assumption 3.1 and (4) hold, and that either Assumption 3.6 holds or $k = 1$. Let $1 \le p < \infty$, let $q$ be as in (5), let $K \in \Omega_h$, $\beta = \beta_K$, and let $s := s(\beta)$ be its sign vector (13). Then, for all $c \in \mathbb{R}^m$,*

$$f_{p,K}(c,s)^2 \lesssim |K|^{\frac{1}{p}-\frac{1}{q}} \frac{|M_K|}{|\beta|} d_K(\beta, c), \tag{35a}$$

$$\tilde{f}_{q,K}(c) \lesssim |K|^{\frac{1}{q}-\frac{1}{p}} f_{p,K}(c,s), \tag{35b}$$

$$\tilde{S}_{q,K}(\beta, c) \lesssim |K|^{\frac{1}{q}-\frac{1}{p}} |M_K^{-1}|\, |\beta|\, f_{p,K}(c,s). \tag{35c}$$

*Proof.* Throughout, $\hat\beta = M_K^{-1}\beta$, which by Lemma 5.1 satisfies $s(\hat\beta) = s(\beta) = s$ on $\hat K$ and, when $k \ge 2$, satisfies $|\hat\beta \cdot \hat n_i| \ge (\theta_0/\kappa_0)|\hat\beta|$ for all $i$; thus Lemma 5.2 applies on $\hat K$, giving (33) with $\theta = \theta_0/\kappa_0$ when $k \ge 2$, and (32) (with no $\theta$-restriction) when $k = 1$. Also note $1/|\hat\beta| \le |M_K|/|\beta|$, since $|\beta| = |M_K\hat\beta| \le |M_K||\hat\beta|$. Factors of $|\hat K|$ are absorbed into constants.

Proof of (35a):

$$\begin{aligned}
f_{p,K}(c,s)^2 &= \left(\tfrac{|K|}{|\hat K|}\right)^{2/p} f_{p,\hat K}(c,s)^2 && \text{by (34b)}\\
&\lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{2/p} |c|^2 \lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{2/p} \frac{d_{\hat K}(\hat\beta, c)}{|\hat\beta|} && \text{by (31b), (32)/(33)}\\
&= \left(\tfrac{|K|}{|\hat K|}\right)^{2/p-1} \frac{d_K(\beta,c)}{|\hat\beta|} \le \left(\tfrac{|K|}{|\hat K|}\right)^{2/p-1} \frac{|M_K|}{|\beta|} d_K(\beta,c) && \text{by (34c),}
\end{aligned}$$

and $2/p - 1 = 1/p - 1/q$.

Proof of (35b):

$$\tilde f_{q,K}(c) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/q} \tilde f_{q,\hat K}(c) \lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{1/q} |c| \lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{1/q} f_{p,\hat K}(c,s) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/q-1/p} f_{p,K}(c,s),$$

by (34b) (twice, with exponents $q$ and $p$) and (31b) (upper bound at exponent $q$, lower bound at exponent $p$).

Proof of (35c): Similarly,

$$\tilde S_{q,K}(\beta,c) = \left(\tfrac{|K|}{|\hat K|}\right)^{1/q} \tilde S_{q,\hat K}(\hat\beta, c) \lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{1/q} |\hat\beta|\,|c| \lesssim \left(\tfrac{|K|}{|\hat K|}\right)^{1/q-1/p} |M_K^{-1}|\,|\beta|\, f_{p,K}(c,s),$$

by (34a), (31a), $|\hat\beta| \le |M_K^{-1}||\beta|$, and (31b) with (34b) as before. □

## 6. Continuity of Fortin and augmented Fortin operators

### 6.1. Bounding the Fortin operator.

For $1 < p < \infty$, $K \in \Omega_h$ and $v \in W_q(\partial_\beta, K)$, define

$$|||v|||_{q,K} := \|v\|_{L_q(K)} + \frac{h_K}{|\beta_K|} \|\beta \cdot \nabla v\|_{L_q(K)}, \qquad |||v|||^q_{q,\Omega_h} := \sum_{K \in \Omega_h} |||v|||^q_{q,K}. \tag{36}$$

On each fixed $K$ this norm generates the same topology as the canonical norm $\|v\|_{L_q(K)} + \|\beta \cdot \nabla v\|_{L_q(K)}$ of $W_q(\partial_\beta, K)$, but with equivalence constants depending on $h_K/|\beta_K|$. The weighting in (36) is chosen so that the constant $C$ of Theorem 6.1 below depends on neither $h_K$ nor $|\beta_K|$, and is therefore uniform over the shape-regular mesh family and invariant under rescaling of $\beta$. In the proof below, we will also use the standard polynomial inverse estimate: for any $j \geq 0$ and $1 \leq r \leq \infty$,

$$\|\nabla g\|_{L_r(K)} \lesssim |M_K^{-1}| \, \|g\|_{L_r(K)} \qquad \text{for all } g \in P_j(K). \tag{37}$$

**Theorem 6.1** (Boundedness of $\Pi_h$)**.** *Suppose Assumption 3.1 and the shape regularity bound (4) hold, and that either Assumption 3.6 holds or $k = 1$. Let $1 < p < \infty$. There is a constant $C = C(N, k, q, \theta_0, \kappa_0)$, independent of $\theta_0$ in the lowest-order case $k = 1$, such that for all $K \in \Omega_h$ and $v \in W_q(\partial_\beta, K)$,*

$$|||\Pi_K v|||_{q,K} \leq C \, |||v|||_{q,K}. \tag{38}$$

*Consequently $|||\Pi_h v|||_{q,\Omega_h} \leq C|||v|||_{q,\Omega_h}$ for all $v \in W_q(\partial_\beta, \Omega_h)$. Moreover, if $v$ is such that*

$$(z, v)_K = 0 \quad \text{for all } z \in P_{k-2}(K), \qquad (\beta \cdot \nabla w, v)_K = 0 \quad \text{for all } w \in P_k^{\perp}(K), \tag{39}$$

*then the sharper bounds*

$$\|\Pi_K v\|_{L_q(K)} \leq C \, a_K \, \|\beta \cdot \nabla v\|_{L_q(K)}, \qquad \|\beta \cdot \nabla \Pi_K v\|_{L_q(K)} \leq C \, \|\beta \cdot \nabla v\|_{L_q(K)} \tag{40}$$

*hold, where $a_K := h_K/|\beta_K|$.*

*Proof.* Write $\beta = \beta_K$ and $s := s(\beta)$, fix $v \in W_q(\partial_\beta, K)$, and let $\varphi := \Pi_K v$. By Lemma 3.8 and the parametrization (28), there are unique $c \in \mathbb{R}^m$ and $z \in P_{k-2}(K)$ such that

$$\varphi = \varphi^{\perp} + \mathring{\varphi}, \qquad \varphi^{\perp} = \varphi_K(c, s) = b_s \, w, \quad w := w_K(c), \qquad \mathring{\varphi} = bz. \tag{41}$$

*Step 1.* Let us bound $\|\varphi^{\perp}\|_{L_q(K)} = f_{q,K}(c, s)$. Since $\mathring{\varphi}$ is a polynomial vanishing on $\partial K$, equation (7) gives $\langle D_K w, \mathring{\varphi} \rangle = 0$, so by (30) and the Fortin condition (24a),

$$d_K(\beta, c) = \langle D_K w, \varphi^{\perp} \rangle = \langle D_K w, \varphi \rangle = \langle D_K w, v \rangle = (\beta \cdot \nabla w, v)_K + (w, \beta \cdot \nabla v)_K.$$

Hölder's inequality with the conjugate exponents (5), followed by (35c) and (35b) applied with $q$ in place of their generic exponent $p$, yields

$$\begin{aligned} d_K(\beta, c) &\leq \tilde{S}_{p,K}(\beta, c) \, \|v\|_{L_q(K)} + \tilde{f}_{p,K}(c) \, \|\beta \cdot \nabla v\|_{L_q(K)} \\ &\lesssim |K|^{\frac{1}{p} - \frac{1}{q}} \Big( |M_K^{-1}| \, |\beta| \, \|v\|_{L_q(K)} + \|\beta \cdot \nabla v\|_{L_q(K)} \Big) \, f_{q,K}(c, s). \end{aligned}$$

Combining with (35a), again applied with $q$ in place of its generic exponent $p$, and noting that the powers of $|K|$ cancel,

$$f_{q,K}(c, s)^2 \lesssim \frac{|M_K|}{|\beta|} \Big( |M_K^{-1}| \, |\beta| \, \|v\|_{L_q(K)} + \|\beta \cdot \nabla v\|_{L_q(K)} \Big) \, f_{q,K}(c, s).$$

Canceling one factor of $f_{q,K}(c, s)$ and using (3) and (4),

$$\|\varphi^{\perp}\|_{L_q(K)} \lesssim \|v\|_{L_q(K)} + \frac{h_K}{|\beta|} \, \|\beta \cdot \nabla v\|_{L_q(K)} \lesssim |||v|||_{q,K}. \tag{42}$$

Under the additional hypothesis (39), the term $\|v\|_{L_q(K)}$ that entered through $(\beta \cdot \nabla w, v)_K$ is absent, so the above simplifies to

$$\|\varphi^{\perp}\|_{L_q(K)} \lesssim a_K \, \|\beta \cdot \nabla v\|_{L_q(K)}. \tag{43}$$

*Step 2.* Next we bound the component $\mathring{\varphi}$. If $k = 1$ then $\mathring{\varphi} = 0$. So suppose $k \geq 2$ and $z \not\equiv 0$ in (41) so $\mathring{\varphi} = bz \not\equiv 0$. The Fortin condition (24b) with test function $z$ gives

$$(z, bz)_K = (z, v - \varphi^{\perp})_K \leq \|z\|_{L_p(K)} \big( \|v\|_{L_q(K)} + \|\varphi^{\perp}\|_{L_q(K)} \big).$$

By Lemma 5.5, applied with $j = k-2$, $r = q$, and $r' = p$, $\|bz\|_{L_q(K)}\|z\|_{L_p(K)} \le C(z, bz)_K$, so dividing by $\|z\|_{L_p(K)}$ and using (42),

$$\|\mathring{\varphi}\|_{L_q(K)} = \|bz\|_{L_q(K)} \lesssim \|v\|_{L_q(K)} + \|\varphi^\perp\|_{L_q(K)} \lesssim \|\!|v|\!\|_{q,K}. \tag{44}$$

Under the additional hypothesis (39), the term $\|v\|_{L_q(K)}$ that arose from $(z, v)_K$ vanishes, so $(z, bz)_K = -(z, \varphi^\perp)_K$ and, by (43),

$$\|\mathring{\varphi}\|_{L_q(K)} \lesssim \|\varphi^\perp\|_{L_q(K)} \lesssim a_K \, \|\beta \cdot \nabla v\|_{L_q(K)}. \tag{45}$$

*Step 3.* By (42) and (44), $\|\varphi\|_{L_q(K)} \le C\|\!|v|\!\|_{q,K}$. For the streamline derivative, since $\varphi \in P_{k+N}(K)$, the inverse estimate (37) together with (3) and (4) gives

$$\frac{h_K}{|\beta|}\|\beta \cdot \nabla\varphi\|_{L_q(K)} \le h_K\|\nabla\varphi\|_{L_q(K)} \le C\, h_K|M_K^{-1}|\, \|\varphi\|_{L_q(K)} \le C\kappa_0\|\varphi\|_{L_q(K)} \le C\, \|\!|v|\!\|_{q,K},$$

so the local bound (38) follows. Raising it to the $q$-th power and summing over $K \in \Omega_h$ proves the global one. Finally, under (39), adding (43) and (45) gives the first bound of (40), and applying to it the same argument using the inverse inequality (37), gives the second. ☐

6.2. **Augmented Fortin operators.** The derivative bound implicit in Theorem 6.1 carries the weight $a_K$, i.e., on each element, for any $v \in W_q(\partial_\beta, K)$,

$$\begin{aligned} \|\Pi_K v\|_{L_q(K)} &\le C\big(\|v\|_{L_q(K)} + a_K\|\beta \cdot \nabla v\|_{L_q(K)}\big), \\ \|\beta \cdot \nabla \Pi_K v\|_{L_q(K)} &\le C a_K^{-1}\big(\|v\|_{L_q(K)} + a_K\|\beta \cdot \nabla v\|_{L_q(K)}\big), \end{aligned} \tag{46}$$

and the term $a_K^{-1}\|v\|_{L_q(K)}$ reflects the mapping of constants by $\Pi_K$ to nonconstant functions, whose streamline derivatives scale like $a_K^{-1}$ times their size. Adapting a correction trick from the proof of [14, Lemma 3.2] (see also [9, Theorems 5.4 and 5.7] and [11]), we now improve (46) to (50) below, at the price of augmenting the test space by one constant per element.

To show the idea in general, for any linear operator $R_K : W_q(\partial_\beta, K) \to W_q(\partial_\beta, K)$, define

$$\Pi_K^+ v \;:=\; R_K v + \Pi_K(v - R_K v), \qquad (\Pi_h^+ v)|_K := \Pi_K^+(v|_K). \tag{47}$$

Two properties are immediate from the identity $\Pi_K^+ v - v = (\Pi_K - I)(v - R_K v)$. First, $\Pi_K^+$ satisfies the same Fortin conditions (24) as $\Pi_K$ and the Fortin property of Lemma 4.4 holds verbatim for $\Pi_h^+$. Second, $\Pi_K^+$ maps into the augmented space $V(K) + \operatorname{range}(R_K)$ and, if $R_K$ is a projection,

$$\Pi_K^+\psi = \psi \qquad \text{for every } \psi \in \operatorname{range}(R_K). \tag{48}$$

Now, for $k = 1$ the conditions (39) on $v$ say exactly that $v$ has mean zero: the space $P_{k-2}(K)$ is trivial, and by definition (16), $\beta \cdot \nabla P_1^\perp(K) = \beta \cdot \nabla P_1(K)$ consists of constants. This motivates the choice of $R_K$ in the next result.

**Theorem 6.2** (Constant-preserving Fortin operator)**.** *Adopt the hypotheses of Theorem 6.1. On each element let $R_K v := \bar{v}$, the mean value of $v$ over $K$, and define $\Pi_K^+$ by (47). Then the following statements hold.*

*(a) For every $k \ge 1$, the operator $\Pi_K^+$ maps $W_q(\partial_\beta, K)$ into*

$$V^+(K) := V(K) + P_0(K), \tag{49}$$

*satisfies the Fortin conditions (24), and reproduces constants, $\Pi_K^+ c = c$ for all $c \in P_0(K)$. The sum in (49) is direct whenever $N \ge 2$.*

(b) *If $k = 1$ (in which case Assumption 3.6 is* not *needed), then there is a $C = C(N, q, \kappa_0)$ such that for all $K \in \Omega_h$ and $v \in W_q(\partial_\beta, K)$,*

$$\begin{aligned} \|\Pi_K^+ v\|_{L_q(K)} &\le C\big(\|v\|_{L_q(K)} + a_K\|\beta\cdot\nabla v\|_{L_q(K)}\big),\\ \|\beta\cdot\nabla\Pi_K^+ v\|_{L_q(K)} &\le C\|\beta\cdot\nabla v\|_{L_q(K)}. \end{aligned} \tag{50}$$

*Consequently, for every choice of per-element weights $\sigma_K \ge a_K$,*

$$\sigma_K^{-1}\|\Pi_K^+ v\|_{L_q(K)} + \|\beta\cdot\nabla\Pi_K^+ v\|_{L_q(K)} \;\le\; C\big(\sigma_K^{-1}\|v\|_{L_q(K)} + \|\beta\cdot\nabla v\|_{L_q(K)}\big). \tag{51}$$

(c) *For every $k \ge 1$, there is a $C = C(N, k, q, \theta_0, \kappa_0)$, independent of $\theta_0$ when $k = 1$, such that*

$$|\!|\!|\Pi_K^+ v|\!|\!|_{q,K} \le C|\!|\!|v|\!|\!|_{q,K}, \qquad v \in W_q(\partial_\beta, K). \tag{52}$$

*Consequently, $|\!|\!|\Pi_h^+ v|\!|\!|_{q,\Omega_h} \le C|\!|\!|v|\!|\!|_{q,\Omega_h}$ for all $v \in W_q(\partial_\beta, \Omega_h)$.*

*Proof. Proof of (a).* The first statement of (a) is contained in the two properties noted after (47), $R_K$ being a projection onto $P_0(K)$. For the directness of the sum (49), suppose $b_s\, u + b\, z$ is a nonzero constant $c$, with $u \in P_k^\perp(K)$ and $z \in P_{k-2}(K)$. Restricting to $\partial K$, where $b$ vanishes, gives $b_s\, u = c$ there. Choose a non-characteristic facet $F_i$, which exists by Lemma 3.2. On $F_i$ we have $b_s|_{F_i} = s_i b_i$, and $b_i$ vanishes on the relative boundary $\partial F_i$, which is nonempty because $N \ge 2$. Hence $c = 0$, a contradiction.

*Proof of (b).* For (50), set $g := v - \bar v$. Then $(\beta\cdot\nabla w, g)_K = 0$ for all $w \in P_1^\perp(K) = P_1(K)$. The $P_{-1}(K)$-condition in (39) is vacuous. So the refined bounds (40) of Theorem 6.1 apply to $g$ and give, using $\beta\cdot\nabla g = \beta\cdot\nabla v$,

$$\|\Pi_K(v-\bar v)\|_{L_q(K)} \le C a_K\|\beta\cdot\nabla v\|_{L_q(K)}, \qquad \|\beta\cdot\nabla\Pi_K(v-\bar v)\|_{L_q(K)} \le C\|\beta\cdot\nabla v\|_{L_q(K)}.$$

Since $\|\bar v\|_{L_q(K)} \le \|v\|_{L_q(K)}$ by Jensen's inequality and $\beta\cdot\nabla\bar v = 0$, the bounds (50) follow from (47). Multiplying the first bound in (50) by $\sigma_K^{-1}$ and using $a_K/\sigma_K \le 1$ gives (51).

*Proof of (c).* From (47) and Theorem 6.1,

$$|\!|\!|\Pi_K^+ v|\!|\!|_{q,K} \le \|\bar v\|_{L_q(K)} + C|\!|\!|v - \bar v|\!|\!|_{q,K} \le C|\!|\!|v|\!|\!|_{q,K},$$

where the last inequality follows from Jensen's inequality and $\beta\cdot\nabla\bar v = 0$. Raising this estimate to the $q$-th power and summing over the mesh proves the global bound. ☐

*Remark* 6.3 (Benign one-dimensional exclusion). The restriction to $N \ge 2$ in the directness assertion of Theorem 6.2(a) is not an artifact: for $N = 1$ and $k = 2$, with $K = (0,1)$ and $\beta > 0$, observe that $b_s = 2x - 1$ and $b_s\cdot(2x-1) + 4\,b = (2x-1)^2 + 4x(1-x) = 1$, where $2x - 1 \in P_2^\perp(K)$. So $V(K)$ already contains the constants and $V^+(K) = V(K)$.

One further augmentation will be helpful. Define

$$V^{++}(K) := V^+(K) + \operatorname{span}\{b\} \subset P_{k+N}(K), \qquad V_h^{++} := \bigtimes_{K\in\Omega_h} V^{++}(K), \tag{53}$$

$$\Pi_K^{++} v := \Pi_K^+ v + \frac{(v - \Pi_K^+ v, 1)_K}{(b,1)_K}\, b, \qquad (\Pi_h^{++} v)|_K = \Pi_K^{++}(v|_K). \tag{54}$$

**Theorem 6.4** (Fortin operator that preserves both constants and means)**.** *Adopt the hypotheses and notation of Theorem 6.2. Then the following statements hold.*

(a) *For every $k \ge 1$, the operator $\Pi_K^{++}$ maps $W_q(\partial_\beta, K)$ into $V^{++}(K)$, satisfies the Fortin conditions (24), reproduces constants, and preserves mean: for $c \in P_0(K)$,*

$$\Pi_K^{++} c = c, \qquad (c, \Pi_K^{++} v - v)_K = 0. \tag{55}$$

*The global Fortin property of Lemma 4.4 holds with* $\mathit{\Pi}_h^{++}$. *Moreover, if* $k \geq 2$, *then*

$$\mathit{\Pi}_K^{++} = \mathit{\Pi}_K^{+}, \qquad V^{++}(K) = V^{+}(K),$$

*while if* $k = 1$ *and* $N \geq 2$, *then* $V^{++}(K) = V(K) \oplus P_0(K) \oplus \operatorname{span}\{b\}$.

(b) *If* $k = 1$, *the estimates* (50) *and* (51) *remain valid with* $\mathit{\Pi}_K^{++}$ *in place of* $\mathit{\Pi}_K^{+}$.

(c) *For* $k \geq 1$, *there is a* $C = C(N, k, q, \theta_0, \kappa_0)$, *independent of* $\theta_0$ *when* $k = 1$, *such that*

$$\|\!|\mathit{\Pi}_K^{++} v|\!\|_{q,K} \leq C \|\!|v|\!\|_{q,K}, \qquad v \in W_q(\partial_\beta, K). \tag{56}$$

*Consequently,* $\|\!|\mathit{\Pi}_h^{++} v|\!\|_{q,\Omega_h} \leq C \|\!|v|\!\|_{q,\Omega_h}$ *for all* $v \in W_q(\partial_\beta, \Omega_h)$.

*Proof. Proof of (a).* Since $b$ vanishes on $\partial K$, $\langle D_K w, b\rangle = 0$ for all $w$ in $P_k(K)$. Thus the bubble correction in (54) does not change (24a). If $k = 1$, condition (24b) is vacuous. If $k \geq 2$, then $1 \in P_{k-2}(K)$, and (24b) for $\mathit{\Pi}_K^{+}$ shows that the correction coefficient vanishes. Hence $\mathit{\Pi}_K^{++} = \mathit{\Pi}_K^{+}$, and also $V^{++}(K) = V^{+}(K)$ because $b \in bP_{k-2}(K) \subset V(K)$. The definition (54) gives the moment condition in (55). If $v = c \in P_0(K)$, then $\mathit{\Pi}_K^{+} c = c$ by Theorem 6.2(a), so the correction vanishes and $\mathit{\Pi}_K^{++} c = c$. The element-wise Fortin conditions (24) obviously imply the global Fortin property. It remains to verify the direct sum asserted when $k = 1$ and $N \geq 2$. Suppose $b_s u + c + bd = 0$, for some $u \in P_1(K)$ and $c, d \in \mathbb{R}$. Restricting first to a non-characteristic facet and then to its relative boundary gives $c = 0$. It follows that $b_s u$ vanishes on $\partial K$, so the boundary argument in Lemma 4.1 gives $u = 0$. Finally, $bd = 0$ gives $d = 0$, thus proving that the sum is direct.

*Proof of (b).* Let $k = 1$, set $g := v - \bar{v}$, and denote the bubble correction by $t_K := \mathit{\Pi}_K^{++} v - \mathit{\Pi}_K^{+} v$. Since $(g, 1)_K = 0$, (47) gives $t_K = -(\mathit{\Pi}_K g, 1)_K \, b/(b, 1)_K$. Lemma 5.5, applied with $j = 0$, $r = q$, $r' = p$, and $z = 1$, gives

$$\|b\|_{L_q(K)} \|1\|_{L_p(K)} \lesssim (b, 1)_K.$$

Therefore, Hölder's inequality and (40) for the mean-zero function $g$ yield

$$\|t_K\|_{L_q(K)} \lesssim \|\mathit{\Pi}_K g\|_{L_q(K)} \lesssim a_K \|\beta \cdot \nabla v\|_{L_q(K)}.$$

Since $t_K$ is a polynomial of fixed degree, the inverse estimate (37) and shape regularity give

$$\|\beta \cdot \nabla t_K\|_{L_q(K)} \lesssim a_K^{-1} \|t_K\|_{L_q(K)} \lesssim \|\beta \cdot \nabla v\|_{L_q(K)}.$$

Adding these two estimates to (50) proves that bound for $\mathit{\Pi}_K^{++}$. If $\sigma_K \geq a_K$, the same estimates, with $a_K/\sigma_K \leq 1$, also give (51).

*Proof of (c).* For $k = 1$, the assertion follows from part (b). For $k \geq 2$, part (a) gives $\mathit{\Pi}_K^{++} = \mathit{\Pi}_K^{+}$, so both the local and global assertions follow from Theorem 6.2(c). ☐

## 7. Quasioptimality of the practical DPG method

Section 2 introduced the practical method (12), computed with any finite dimensional test space $Y_h$ satisfying (22). In this section we place it in a functional-analytic setting adapted to the continuous problem (1), use the Fortin operator to prove that it is quasioptimal, and derive convergence rates in $h$. Throughout this section we fix a Lebesgue exponent $1 < p < \infty$ for the *trial* side, so that its conjugate $q$ of (5) is the exponent of the *test* side.

7.1. **The undiscretized problem.** Since $\operatorname{div}\beta = 0$, no coercivity is available from a reaction term, so the solvability of the non-coercive transport problem (1), as well as the trace theory needed to give its boundary condition meaning, is secured by conditions on the flow. We assume the standard ones (see, e.g., [1, 7, 17]). Here $\Gamma_{\mathrm{out}} := \{x \in \partial\Omega : \beta\cdot n > 0\}$ denotes the outflow boundary, complementing the inflow boundary $\Gamma_{\mathrm{in}}$ of (1).

**Assumption 7.1** (Flow conditions on $\Omega$)**.** (i) Separated inflow and outflow: *The sets* $\overline{\Gamma}_{\mathrm{in}}$ *and* $\overline{\Gamma}_{\mathrm{out}}$ *are disjoint.* (ii) $\Omega$-filling flow: *There exist* $z_+, z_- \in L_\infty(\Omega)$ *with* $\beta\cdot\nabla z_\pm \in L_\infty(\Omega)$*, and a constant* $c_0 > 0$*, such that*

$$-\beta\cdot\nabla z_\pm = c_0 \quad \text{in } \Omega, \qquad z_+ = 0 \quad \text{on } \Gamma_{\mathrm{out}}, \qquad z_- = 0 \quad \text{on } \Gamma_{\mathrm{in}}. \tag{57}$$

Neither part is new. Part (i) is the well-separation condition of [2, eq. (9)] (see also [17, Assumption 2.2]). Few results are known without it (an exception being [12, Lemma 4.1]). Part (ii) is the $\Omega$-filling condition of [17, Assumption 2.8]. It states informally that the flow of $\beta$ has no closed streamlines and that every streamline joins $\Gamma_{\mathrm{in}}$ to $\Gamma_{\mathrm{out}}$ in finite time (indeed $z_+/c_0$ is the time needed to reach $\Gamma_{\mathrm{out}}$) and it holds, e.g., whenever the method of characteristics applies (see [17, Remark 2.9] and [6]). The constant $c_0$ in (57) is immaterial: rescaling $z_\pm$ changes it at will. Another condition in the literature, perhaps easier to verify, is the potential condition of [2, Section 2.4]. In our case the reaction coefficient $\mu$, the divergence $\operatorname{div}\beta$, and hence the Friedrichs coefficient $\sigma_{\beta,\mu;p} = \mu - \frac{1}{p}\operatorname{div}\beta$ all vanish, and that condition reduces to the existence of a Lipschitz $\psi = -\zeta$ with $\beta\cdot\nabla\psi \geq 2\mu_0 > 0$, from which (57) follows by taking $z_+$ to be the transit time to $\Gamma_{\mathrm{out}}$. Note that Assumption 7.1 rules out $\beta$ with closed streamlines in $\Omega$ that never meet $\Gamma_{\mathrm{in}}$, i.e., periodic orbits of $x'(t) = \beta(x(t))$ with period $T > 0$: along it, $dz_+(x(t))/dt = \beta\cdot\nabla z_+ = -c_0$, so $z_+$ strictly decreases at a nonzero constant rate, yet $z_+(x(T)) = z_+(x(0))$ by periodicity, which is impossible. Note that no quantitative stability constant of the continuous problem will enter Theorem 7.5 or Corollary 7.9 below; the constant $C_{\mathrm{PF}}(q,\beta)$ is used only in Corollary 7.10, where we convert energy control into mesh-independent $L_p(\Omega)$ control.

We need the unbroken graph spaces on $\Omega$: for $1 < r < \infty$,

$$W_r(\partial_\beta, \Omega) = \{v \in L_r(\Omega) : \beta\cdot\nabla v \in L_r(\Omega)\}.$$

Under Assumption 7.1(i), functions in $W_r(\partial_\beta,\Omega)$ have traces on $\partial\Omega$ in the $|\beta\cdot n|$-weighted $L_r$ sense, smooth functions are dense in $W_r(\partial_\beta,\Omega)$, and the integration by parts formula extends to

$$\int_\Omega \beta\cdot\nabla v\, w + \int_\Omega v\beta\cdot\nabla w = \int_{\partial\Omega} \beta\cdot n\, v\, w, \qquad v \in W_p(\partial_\beta,\Omega),\ w \in W_q(\partial_\beta,\Omega) \tag{58}$$

(see e.g., [2, Lemmas 2.1–2.2], [17, Remark 2.5, Lemma 2.6] and [1, 7, 16]). Let

$$W_{r,0}(\partial_\beta,\Omega) := \{v \in W_r(\partial_\beta,\Omega) : v|_{\Gamma_{\mathrm{in}}} = 0\}$$

denote the subspace incorporating the inflow boundary condition of (1) in that trace sense.

**Lemma 7.2** (Wellposedness and curved Poincaré–Friedrichs inequality)**.** *Suppose* $1 < r < \infty$*, Assumptions 3.1 and 7.1 hold, and* $C_{\mathrm{PF}}(r,\beta) := r\max(\|z_+\|_{L_\infty(\Omega)}, \|z_-\|_{L_\infty(\Omega)})/c_0$*. Then*

$$\|v\|_{L_r(\Omega)} \leq C_{\mathrm{PF}}(r,\beta)\|\beta\cdot\nabla v\|_{L_r(\Omega)} \qquad \text{for all } v \in W_r(\partial_\beta,\Omega) \text{ vanishing on } \Gamma_{\mathrm{in}} \text{ or } \Gamma_{\mathrm{out}}, \tag{59}$$

*and for every* $f \in L_r(\Omega)$ *there is a unique* $u \in W_{r,0}(\partial_\beta,\Omega)$ *with* $\beta\cdot\nabla u = f$*.*

*Proof.* The inequality (59) follows from [17, Lemma 2.10] and unique solvability from [17, Theorem A and Remark 3.1]. □

We first complete the prescription of the method (12) by specifying the norm whose dual appears there. As in (8), let $Y = W_q(\partial_\beta, \Omega_h)$. Observe that the spaces of (22) form the chain $V_h \subseteq Y_h \subset Y$. Equip $Y$ with the $\beta$-weighted test norm obtained by rescaling the norm (36) element by element. Recall that $a_K = h_K/|\beta_K|$. Then

$$\|v\|_{Y(K)}^q := a_K^{-q}\, \|v\|_{L_q(K)}^q + \|\beta \cdot \nabla v\|_{L_q(K)}^q, \qquad \|v\|_Y^q := \sum_{K \in \Omega_h} \|v\|_{Y(K)}^q. \tag{60}$$

Linear functionals $g$ on $Y_h$ are normed by

$$\|g\|_{Y_h'} := \sup_{0 \neq v \in Y_h} \frac{g(v)}{\|v\|_Y}, \tag{61}$$

which is the norm used in the specification of the method (12).

Next we set the undiscretized trial space, realizing interface variables as functionals on $Y$, in the spirit of [4]. Formula (8) defines $D_h : W_p(\partial_\beta, \Omega) \to Y'$ by the same element-wise sum (where the right hand side is well defined even when $z$ is taken to be in $W_p(\partial_\beta, \Omega)$ by virtue of (6)). Set

$$\hat{X} := D_h\big(W_p(\partial_\beta, \Omega)\big), \qquad \hat{X}_0 := D_h\big(W_{p,0}(\partial_\beta, \Omega)\big), \qquad X := L_p(\Omega) \times \hat{X}_0. \tag{62}$$

The definitions (9) now give the inclusions $\hat{X}_h^k \subset \hat{X}$, $\hat{X}_{h,0}^k \subset \hat{X}_0$, and hence $X_h \subset X$. The DPG bilinear form (11) extends to $X \times Y$ as

$$\mathscr{b}\big((w, \mu), v\big) = \langle \mu, v \rangle - \sum_{K \in \Omega_h} (w,\, \beta \cdot \nabla v)_K, \qquad \text{for any } \mathscr{w} = (w, \mu) \in X,\ v \in Y. \tag{63}$$

**Lemma 7.3** (Injectivity)**.** *Suppose Assumptions 3.1 and 7.1 hold. If $\mathscr{w} = (w, \mu) \in X$ satisfies $\mathscr{b}(\mathscr{w}, v) = 0$ for all $v \in Y$, then $\mathscr{w} = 0$.*

*Proof.* We use the space of (smooth compactly supported) Schwartz test functions on $K$ and $\Omega$, denoted by $C_c^\infty(K)$ and $C_c^\infty(\Omega)$, respectively. By (62), $\mu = D_h z$ for some $z \in W_{p,0}(\partial_\beta, \Omega)$. Fix $K \in \Omega_h$ and let $v \in C_c^\infty(K)$, extended by zero to an element of $Y$. Then, since $\langle D_K z, v\rangle = 0$, we have $0 = \mathscr{b}(\mathscr{w}, v) = -(w, \beta \cdot \nabla v)_K$ for all $v \in C_c^\infty(K)$. Since $\beta_K$ is constant, this implies that the distributional derivative $\beta \cdot \nabla w$ vanishes on $K$. In particular $w|_K \in W_p(\partial_\beta, K)$ with

$$\beta \cdot \nabla w|_K = 0, \qquad \text{for every } K \in \Omega_h. \tag{64}$$

Using this and (6), $-(w, \beta \cdot \nabla v)_K = (\beta \cdot \nabla w, v)_K - \langle D_K w, v\rangle = -\langle D_K w, v\rangle$, so

$$0 = \mathscr{b}(\mathscr{w}, v) = \langle D_h(z - w), v\rangle, \qquad \text{for all } v \in Y. \tag{65}$$

Let $\zeta := z - w \in W_p(\partial_\beta, \Omega_h)$. The distributional derivative $\beta \cdot \nabla \zeta$ coincides with the element-wise one: indeed, for any $v \in C_c^\infty(\Omega)$, we have $(\beta \cdot \nabla \zeta)(v) = -(\zeta, \beta \cdot \nabla v)_\Omega = (\beta \cdot \nabla \zeta, v)_\Omega$ using (6) and (65). By (58) and (65), now taken for $v \in W_q(\partial_\beta, \Omega) \subset Y$, $0 = (\beta \cdot \nabla \zeta, v)_\Omega + (\zeta, \beta \cdot \nabla v)_\Omega = \int_{\partial\Omega} (\beta \cdot n)\, \zeta\, v$, so the trace of $\zeta$ must vanish on $\Gamma_{\text{in}}$.

Consequently $w = z - \zeta \in W_{p,0}(\partial_\beta, \Omega)$ (since the trace of both $z$ and $\zeta$ on $\Gamma_{\text{in}}$ vanishes) and $\beta \cdot \nabla w = 0$. The uniqueness part of Lemma 7.2 now implies $w = 0$. Returning to (65) with $w = 0$, we find $\langle \mu, v\rangle = \langle D_h z, v\rangle = 0$ for all $v \in Y$, i.e., $\mu = 0$. $\square$

7.2. **Stability of the practical DPG method.** Since $X_h = P_{k-1}(\Omega_h) \times \hat{X}^k_{h,0} \subset P_{k-1}(\Omega_h) \times \hat{X}^k_h$, Lemma 4.4 gives

$$\mathscr{b}(\mathscr{w}_h, v - \Pi_h v) = 0, \qquad \mathscr{w}_h \in X_h, \quad v \in Y, \tag{66}$$

the key property that will get us stability. Define the "energy norm" and a discrete counterpart, both for any $\mathscr{w} \in X$, by

$$\|\mathscr{w}\|_{\mathscr{E}} := \sup_{0 \neq v \in Y} \frac{\mathscr{b}(\mathscr{w}, v)}{\|v\|_Y}, \qquad \|\mathscr{w}\|_{\mathscr{E},h} := \sup_{0 \neq v_h \in Y_h} \frac{\mathscr{b}(\mathscr{w}, v_h)}{\|v_h\|_Y}. \tag{67}$$

Since $Y_h \subset Y$,

$$\|\mathscr{w}\|_{\mathscr{E},h} \leq \|\mathscr{w}\|_{\mathscr{E}}. \tag{68}$$

By Lemma 7.3, $\|\cdot\|_{\mathscr{E}}$ is a norm on $X$. That $\|\cdot\|_{\mathscr{E},h}$ is a norm on $X_h$ follows by the Fortin property: if $\mathscr{x}_h \in X_h$ satisfies $\|\mathscr{x}_h\|_{\mathscr{E},h} = 0$, then $0 = \mathscr{b}(\mathscr{x}_h, \Pi_h v) = \mathscr{b}(\mathscr{x}_h, v)$ for all $v \in Y$ by (66), so $\mathscr{x}_h = 0$ by Lemma 7.3. We begin with a quick observation that the Fortin operator is bounded also in the rescaled test norm (60).

**Corollary 7.4.** *In the setting of Theorem 6.1, there is a $C_\Pi = C_\Pi(N, k, q, \theta_0, \kappa_0)$, independent of $\theta_0$ when $k = 1$, such that $\|\Pi_h v\|_Y \leq C_\Pi \|v\|_Y$ for all $v \in Y$. The same conclusion holds for the mean-preserving operator: there is a constant $C_{\Pi^{++}}$ with the same dependencies such that*

$$\|\Pi_h^{++} v\|_Y \leq C_{\Pi^{++}} \|v\|_Y, \qquad v \in Y. \tag{69}$$

*Proof.* Multiplying the local bound $\|\!|\Pi_K v|\!\|_{q,K} \leq C \|\!|v|\!\|_{q,K}$ of Theorem 6.1 by $a_K^{-1}$ gives

$$a_K^{-1} \|\Pi_K v\|_{L_q(K)} + \|\beta \cdot \nabla \Pi_K v\|_{L_q(K)} \;\leq\; C\Big(a_K^{-1} \|v\|_{L_q(K)} + \|\beta \cdot \nabla v\|_{L_q(K)}\Big),$$

so $\|\Pi_K v\|_{Y(K)} \leq 2^{1/p}\, C\, \|v\|_{Y(K)}$ by the equivalence $(a^q + b^q)^{1/q} \leq a + b \leq 2^{1-1/q}(a^q + b^q)^{1/q}$ of the $\ell_1$ and $\ell_q$ norms on $\mathbb{R}^2$. Raise to the $q$-th power, sum over $K$, and set $C_\Pi := 2^{1/p}\, C$. Applying the same argument to (56) proves (69). □

Let $u \in W_{p,0}(\partial_\beta, \Omega)$ solve (1) (Lemma 7.2) for the data $f$ of $\ell$ in (12). The exact pair $\mathscr{u} := (u, D_h u) \in X$ is consistent: by (6),

$$\mathscr{b}(\mathscr{u}, v) = \sum_{K \in \Omega_h} \big( -(u, \beta \cdot \nabla v)_K + (\beta \cdot \nabla u, v)_K + (u, \beta \cdot \nabla v)_K \big) = (f, v)_\Omega = \ell(v) \tag{70}$$

for all $v \in Y$. By (70), $\|\ell - \mathscr{b}(\mathscr{w}_h, \cdot)\|_{Y_h'} = \|\mathscr{u} - \mathscr{w}_h\|_{\mathscr{E},h}$, so (12) is equivalent to

$$\|\mathscr{u} - \mathscr{u}_h\|_{\mathscr{E},h} = \min_{\mathscr{w}_h \in X_h} \|\mathscr{u} - \mathscr{w}_h\|_{\mathscr{E},h}. \tag{71}$$

The existence of this minimizer and its quasioptimality in $\|\cdot\|_{\mathscr{E}}$ is proved next.

**Theorem 7.5** (Quasioptimality)**.** *Let $1 < p < \infty$. Suppose Assumptions 3.1 and 7.1 hold, $Y_h$ satisfy (22), the shape regularity bound (4) holds, and that either Assumption 3.6 holds or $k = 1$. Let $f \in L_p(\Omega)$ and let $C_\Pi$ be the constant of Corollary 7.4. Then the discrete solution $\mathscr{u}_h$ of (12) exists, is unique, and satisfies*

$$\|\mathscr{u} - \mathscr{u}_h\|_{\mathscr{E}} \;\leq\; (1 + 2C_\Pi) \inf_{\mathscr{w}_h \in X_h} \|\mathscr{u} - \mathscr{w}_h\|_{\mathscr{E}}. \tag{72}$$

*Proof.* The map $w_h \mapsto \|\ell - \mathscr{b}(w_h,\cdot)\|_{Y_h'} = \|u - w_h\|_{\mathscr{E},h}$ is convex, continuous and coercive on $X_h$, so minimizers of (12) exist. If $u_1, u_2$ are two minimizers, their midpoint is also a minimizer, so the triangle inequality holds with equality for the two residuals $\|\ell - \mathscr{b}(u_i,\cdot)\|_{Y_h'}$. Strict convexity of $\|\cdot\|_{Y_h'}$ (since $\|\cdot\|_Y$ restricted to $Y_h$ is smooth, being isometric to a subspace of an $L_q$ space with $1 < q < \infty$) then implies that the two residuals must coincide as functionals on $Y_h$, i.e. $\|u_1 - u_2\|_{\mathscr{E},h} = 0$, so $u_1 = u_2$. Thus the discrete solution $u_h$ of (12) exists and is unique.

To prove quasioptimality, let $\delta := u - u_h$ and fix a $w_h \in X_h$. For any $v \in Y$, we have $\Pi_h v \in V_h \subseteq Y_h$ by (22). Insert $\pm w_h$ and $\pm \Pi_h v$ and use (66) on $w_h - u_h \in X_h$ to get

$$\begin{aligned}\mathscr{b}(\delta, v) &= \mathscr{b}(u - w_h,\, v - \Pi_h v) + \mathscr{b}(w_h - u_h,\, v - \Pi_h v) + \mathscr{b}(\delta,\, \Pi_h v)\\ &= \mathscr{b}(u - w_h,\, v - \Pi_h v) + \mathscr{b}(\delta,\, \Pi_h v).\end{aligned}$$

By the definitions (67) and Corollary 7.4,

$$\begin{aligned}\mathscr{b}(u - w_h, v - \Pi_h v) &\le \|u - w_h\|_{\mathscr{E}}\, \|v - \Pi_h v\|_Y \le (1 + C_\Pi)\, \|u - w_h\|_{\mathscr{E}}\, \|v\|_Y,\\ \mathscr{b}(\delta, \Pi_h v) &\le \|\delta\|_{\mathscr{E},h}\, \|\Pi_h v\|_Y \le C_\Pi\, \|\delta\|_{\mathscr{E},h}\, \|v\|_Y.\end{aligned}$$

Dividing by $\|v\|_Y$ and taking the supremum over $v \in Y$,

$$\|\delta\|_{\mathscr{E}} \le (1 + C_\Pi)\|u - w_h\|_{\mathscr{E}} + C_\Pi \|\delta\|_{\mathscr{E},h}.$$

Finally, by (71) and (68), the last term above is bounded by $\|\delta\|_{\mathscr{E},h} \le \|u - w_h\|_{\mathscr{E},h} \le \|u - w_h\|_{\mathscr{E}}$, so $\|\delta\|_{\mathscr{E}} \le (1 + 2C_\Pi)\|u - w_h\|_{\mathscr{E}}$, thus proving (72). □

*Remark* 7.6. The argument of Theorem 7.5 is general and can be applied beyond the advection problem. Another general approach is the theory of [18], which we did not apply directly for brevity because it requires us to first prove that the infinite-dimensional broken Banach formulation is wellposed. If we had applied [18, Theorem 4.14], then we would have obtained the quasioptimality constant $1 + C_\Pi(1 + C_{\mathrm{AO}}(Y))$, where $C_{\mathrm{AO}}(Y) \in [0, 1]$ is their asymmetric-orthogonality constant of $Y$, instead of $1 + 2C_\Pi$.

*Remark* 7.7. In the Hilbert space case $p = q = 2$, by the standard route of [14], we obtain (72) with $C_\Pi$ instead of $1 + 2C_\Pi$.

7.3. **Convergence rates.** To obtain convergence rates from Theorem 7.5, we measure the best approximation error in a convenient norm on $X$. Viewing $\hat{X}$ as a subspace of $Y'$, define

$$\|(w, \mu)\|_X^p := \|w\|_{L_p(\Omega)}^p + \|\mu\|_{Y'}^p. \tag{73}$$

**Lemma 7.8.** *Suppose Assumption 3.1 holds. For all $z \in W_p(\partial_\beta, \Omega)$, $(w, \mu) \in X$, $v \in Y$,*

$$\|D_h z\|_{Y'}^p \le \sum_{K \in \Omega_h} \Big( \|z\|_{L_p(K)}^p + a_K^p \|\beta \cdot \nabla z\|_{L_p(K)}^p \Big), \tag{74}$$

$$\big|\mathscr{b}\big((w, \mu), v\big)\big| \le 2^{1/q}\, \|(w, \mu)\|_X\, \|v\|_Y. \tag{75}$$

*Proof.* By (8), (6), and the Hölder inequality,

$$\begin{aligned}\langle D_h z, v\rangle &= \sum_K \Big[(\beta\cdot\nabla z, v)_K + (z, \beta\cdot\nabla v)_K\Big]\\ &\le \sum_K \Big[\Big(a_K\|\beta\cdot\nabla z\|_{L_p(K)}\Big)\Big(a_K^{-1}\|v\|_{L_q(K)}\Big) + \|z\|_{L_p(K)}\,\|\beta\cdot\nabla v\|_{L_q(K)}\Big]\\ &\le \bigg(\sum_K \Big[\|z\|^p_{L_p(K)} + a_K^p\|\beta\cdot\nabla z\|^p_{L_p(K)}\Big]\bigg)^{1/p}\|v\|_Y.\end{aligned}$$

Continuity of $\mathscr{b}$ is now immediate: bounding the second (volume) term of (63) by Hölder inequality on each element and over the mesh, and the first (interface) term by $\|\mu\|_{Y'}\|v\|_Y$, the inequality (75) follows. ◻

In particular, inequality (75) implies $\|\boldsymbol{w}\|_{\mathscr{E}} \le 2^{1/q}\|\boldsymbol{w}\|_X$, so (72) of Theorem 7.5 also gives

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathscr{E}} \;\le\; 2^{1/q}\,(1+2C_\Pi)\inf_{\boldsymbol{w}_h\in X_h}\|\boldsymbol{u}-\boldsymbol{w}_h\|_X. \tag{76}$$

Our convergence rates are in terms of $h := \max_{K\in\Omega_h} h_K$. We will use the element-wise $L_2$-orthogonal projection onto $P_j(\Omega_h)$ extended to $L_1$, namely let $Q_j : L_1(K)\to P_j(K)$ be defined by these moment conditions on each $K\in\Omega_h$ for any $v\in L_1(K)$:

$$(Q_j v, r)_K = (v, r)_K, \qquad r\in P_j(K). \tag{77}$$

**Corollary 7.9** (Convergence rates)**.** *In addition to the hypotheses of Theorem 7.5, suppose the solution $u$ of (1) satisfies $u\in W_p^s(\Omega)$ for an integer $1\le s\le k+1$. Then,*

$$\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathscr{E}} \;\le\; C\,h^{\min(s,k)}\,\|u\|_{W_p^s(\Omega)},$$

*with $C = C(N,k,s,p,\theta_0,\kappa_0)$, independent of $\theta_0$ when $k=1$.*

*Proof.* By (76) it suffices to exhibit one $\boldsymbol{w}_h\in X_h$ with $\|\boldsymbol{u}-\boldsymbol{w}_h\|_X \le Ch^{\min(s,k)}\|u\|_{W_p^s(\Omega)}$. Take $\boldsymbol{w}_h := \big(Q_{k-1}u,\ D_h(I_h u)\big)$, where $Q_{k-1}$ is as in (77) and $I_h$ is a Scott–Zhang interpolation operator [20] of degree $k$, chosen so as to preserve the homogeneous boundary values on $\Gamma_{\text{in}}$. Write $\varsigma_0 := \min(s,k)$ and $\varsigma_1 := \min(s,k+1)$. For the interior variable, the Bramble–Hilbert lemma immediately gives

$$\|u - Q_{k-1}u\|_{L_p(\Omega)} \lesssim h^{\varsigma_0}|u|_{W_p^{\varsigma_0}(\Omega)}, \tag{78}$$

since $P_{k-1}(K)\supseteq P_{\varsigma_0-1}(K)$.

For the trace component, note that $z := u - I_h u \in W_{p,0}(\partial_\beta,\Omega)$, because $u\in W_{p,0}(\partial_\beta,\Omega)$ and $I_h u$ is a continuous piecewise polynomial vanishing on $\Gamma_{\text{in}}$. By (74) of Lemma 7.8 and $\|\beta\cdot\nabla z\|_{L_p(K)} \le |\beta_K|\,|z|_{W_p^1(K)}$,

$$\|D_h z\|^p_{Y'} \le \sum_K \Big(\|z\|^p_{L_p(K)} + h_K^p|z|^p_{W_p^1(K)}\Big).$$

The simultaneous approximation and stability properties [20] of the Scott–Zhang operator in $W_p^s$ yield, for $1\le\varsigma_1\le k+1$, $\|z\|_{L_p(K)} + h_K|z|_{W_p^1(K)} \le C\,h_K^{\varsigma_1}|u|_{W_p^{\varsigma_1}(\omega_K)}$, where $\omega_K$ is the patch of elements meeting $K$; summing over $K$ and using the finite overlap of the patches (by shape regularity),

$$\|D_h u - D_h(I_h u)\|_{Y'} \le C\,h^{\varsigma_1}\,|u|_{W_p^{\varsigma_1}(\Omega)}. \tag{79}$$

Combining (78) and (79) with $\varsigma_0\le\varsigma_1$ completes the proof. ◻

The energy norm above is mesh-dependent. We now convert this into a convergence rate in the mesh-independent $L_p(\Omega)$ norm. The conversion costs one power of $h$. Let

$$\bar\beta := \max_K |\beta_K|. \tag{80}$$

Some results use quasi-uniform meshes where there is a $c_{\mathrm{qu}} > 0$ such that $h_K \ge c_{\mathrm{qu}} h$ for all $K \in \Omega_h$ with $h := \max_K h_K$.

**Corollary 7.10** (Interior variable error in $L_p$)**.** *Adopt the hypotheses of Corollary 7.9, write $\boldsymbol{u} = (u, D_h u)$ and $\boldsymbol{u}_h = (u_h, D_h z_h)$. Then*

$$\|u - u_h\|_{L_p(\Omega)} \le \Big(1 + C_{\mathrm{PF}}(q,\beta)^q \max_{K\in\Omega_h} a_K^{-q}\Big)^{1/q} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathscr{E}}. \tag{81}$$

*In particular, on quasi-uniform mesh families, with $\bar\beta$ as in (80),*

$$\|u - u_h\|_{L_p(\Omega)} \le C\big(1 + C_{\mathrm{PF}}(q,\beta)\,\bar\beta\, h^{-1}\big)\, h^{\min(s,k)}\, \|u\|_{W_p^s(\Omega)}.$$

*Proof.* We use a duality argument. Let $g \in L_q(\Omega)$ be arbitrary. Applying Lemma 7.2, with exponent $q$, to the reversed advection field $-\beta$ (whose inflow boundary is $\Gamma_{\mathrm{out}}$, and for which Assumption 7.1(ii) holds with the roles of $z_+$ and $z_-$ interchanged), we obtain $z \in W_q(\partial_\beta, \Omega)$ with $-\beta\cdot\nabla z = g$ in $\Omega$ and $z = 0$ on $\Gamma_{\mathrm{out}}$. Then (59) gives $\|z\|_{L_q(\Omega)} \le C_{\mathrm{PF}}(q, -\beta)\|g\|_{L_q(\Omega)} = C_{\mathrm{PF}}(q,\beta)\|g\|_{L_q(\Omega)}$, the last equality because the interchange of $z_+$ and $z_-$ leaves the maximum defining $C_{\mathrm{PF}}$ unchanged. Viewing $z$ as an element of the broken space $Y$,

$$\|z\|_Y^q = \sum_K \Big[a_K^{-q}\|z\|^q_{L_q(K)} + \|\beta\cdot\nabla z\|^q_{L_q(K)}\Big] \le \Big(1 + C_{\mathrm{PF}}(q,\beta)^q \max_K a_K^{-q}\Big)\, \|g\|^q_{L_q(\Omega)}. \tag{82}$$

The trace component of $\boldsymbol{u} - \boldsymbol{u}_h$ is $D_h(u - z_h)$ for a $z_h \in S^k_{h,0}$. Since $u - z_h \in W_{p,0}(\partial_\beta, \Omega)$ vanishes on $\Gamma_{\mathrm{in}}$, while $z \in W_q(\partial_\beta,\Omega)$ vanishes on $\Gamma_{\mathrm{out}}$, and $\beta\cdot n$ vanishes on the remainder of the boundary $\partial\Omega$, the integration by parts formula (58) yields

$$\langle D_h(u - z_h), z\rangle = \int_{\partial\Omega} (\beta\cdot n)(u - z_h)\, z\, ds = 0.$$

Hence, by (63),

$$(u - u_h, g)_\Omega = -(u - u_h, \beta\cdot\nabla z)_\Omega = \mathscr{b}(\boldsymbol{u} - \boldsymbol{u}_h,\, z) \le \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathscr{E}}\, \|z\|_Y.$$

Combining with (82) and taking the supremum over $\|g\|_{L_q(\Omega)} = 1$ proves (81) by $L_p$–$L_q$ duality. The second assertion follows from Corollary 7.9. ☐

## 8. A posteriori error control

The quantity minimized in (12) is computable, and is a natural a posteriori error estimator:

$$\eta := \big\|\ell - \mathscr{b}(\boldsymbol{u}_h, \cdot)\big\|_{Y_h'}. \tag{83}$$

In practice $\eta$ is evaluated through the discrete *residual representative* $\varrho_h \in Y_h$ defined by

$$\langle J_Y \varrho_h, v\rangle = \ell(v) - \mathscr{b}(\boldsymbol{u}_h, v), \qquad \text{for all } v \in Y_h, \tag{84}$$

where $J_Y : Y \to Y'$ is the normalized duality map of $(Y, \|\cdot\|_Y)$, which reduces to the linear Riesz map when $p = 2$. The pair $(\varrho_h, \boldsymbol{u}_h)$ is exactly the solution of the mixed system equivalent to (12) (see [17, eq. (4.2)]), so $\varrho_h$ is available at no extra cost. Since normalized duality maps preserve norms, $\eta = \|\varrho_h\|_Y$, and the estimator localizes into element-wise computable indicators,

$$\eta^q = \sum_{K\in\Omega_h} \eta_K^q, \qquad \eta_K := \|\varrho_h\|_{Y(K)}. \tag{85}$$

Set $m_k := \max\{k-2, 0\}$. Reliability is measured against the data oscillation

$$\operatorname{osc}(f) := \Bigg( \sum_{K\in\Omega_h} a_K^p \, \|f - Q_{m_k} f\|_{L_p(K)}^p \Bigg)^{1/p}, \tag{86}$$

where $Q_{m_k}$ is as in (77) and $a_K = h_K/|\beta_K|$ as before. For $p = 2$, estimates of the kind below were deduced from the existence of a Fortin operator in [3] (see also [9]). We now provide them in the present setting for all $1 < p < \infty$, with efficiency constant one.

**Theorem 8.1** (Global reliability and efficiency)**.** *Under the hypotheses of Theorem 7.5, suppose in addition that $V_h^{++} \subseteq Y_h$. Then*

$$\eta \;\le\; \|\mathscr{u} - \mathscr{u}_h\|_{\mathscr{E}} \;\le\; \|\varPi_h^{++}\|\, \eta \;+\; \|I - \varPi_h^{++}\|\, \operatorname{osc}(f)$$

*where the operator norms $\|\varPi_h^{++}\|$ and $\|I - \varPi_h^{++}\|$ are bounded mesh-independently by $C_{\varPi^{++}}$ and $1 + C_{\varPi^{++}}$, respectively. In particular,* $\operatorname{osc}(f)$ *vanishes if $f$ is piecewise constant.*

*Proof.* Efficiency follows trivially by (70) and (68), $\eta = \|\mathscr{u} - \mathscr{u}_h\|_{\mathscr{E},h} \le \|\mathscr{u} - \mathscr{u}_h\|_{\mathscr{E}}$.

To prove reliability, let $\delta := \mathscr{u} - \mathscr{u}_h$ and $v \in Y$, and split $\mathscr{b}(\delta, v) = \mathscr{b}(\delta, \varPi_h^{++} v) + \mathscr{b}(\delta, v - \varPi_h^{++} v)$. Since $\varPi_h^{++} v \in V_h^{++} \subseteq Y_h$, Corollary 7.4 gives

$$\mathscr{b}(\delta, \varPi_h^{++} v) = \ell(\varPi_h^{++} v) - \mathscr{b}(\mathscr{u}_h, \varPi_h^{++} v) \;\le\; \eta \, \|\varPi_h^{++} v\|_Y \;\le\; \eta\, \|\varPi_h^{++}\| \, \|v\|_Y. \tag{87}$$

For the second term, we start by applying the global Fortin property of Theorem 6.4 to $\mathscr{u}_h \in X_h$, followed by consistency:

$$\begin{aligned} \mathscr{b}(\delta, v - \varPi_h^{++} v) &= \mathscr{b}(\mathscr{u}, v - \varPi_h^{++} v) = \ell(v - \varPi_h^{++} v) = \sum_{K\in\Omega_h} (f, v - \varPi_K^{++} v)_K \\ &= \sum_K (f - Q_{m_k} f, v - \varPi_K^{++} v)_K \\ &\le \sum_K \big(a_K \|f - Q_{m_k} f\|_{L_p(K)}\big)\big(a_K^{-1} \|v - \varPi_K^{++} v\|_{L_q(K)}\big) \\ &\le \operatorname{osc}(f)\, \|I - \varPi_h^{++}\| \, \|v\|_Y. \end{aligned} \tag{88}$$

In the second line, we have used (24b) and (55), by which $v - \varPi_K^{++} v$ is $L_2(K)$-orthogonal to $P_{m_k}(K)$, allowing us to replace $f$ by $f - Q_{m_k} f$ in each summand. Afterward, we used Hölder inequalities on each element and over the mesh, together with $a_K^{-1}\|\cdot\|_{L_q(K)} \le \|\cdot\|_{Y(K)}$. Adding the bounds (87) and (88), dividing by $\|v\|_Y$, and taking the supremum over $0 \ne v \in Y$ completes the proof. The bounds with $C_{\varPi^{++}}$ follow from (69). ◻

It is well known that the $\varrho_h$ solving (84) also satisfies $\mathscr{b}(\mathscr{w}_h, \varrho_h) = 0$ for all $\mathscr{w}_h \in X_h$ (which is the second equation of equivalent mixed system of the practical DPG method: see [9, eq. (6.7)] or [17, eq. (4.2b)]), i.e., $\varrho_h$ is in the *residual representation space*

$$Y_h^0 := \{v \in Y_h : \mathscr{b}(\mathscr{w}_h, v) = 0 \text{ for all } \mathscr{w}_h \in X_h\}.$$

If $Y_h^0$ were trivial, then $\eta \equiv 0$ would carry no information, but this is not the case:

**Proposition 8.2.** *Under the hypotheses of Theorem 8.1, when $N \geq 2$, the dimension of the residual representation space $Y_h^0$ is not smaller than the number of mesh elements $N_{\mathrm{el}}$.*

*Proof.* Define $B_h : X_h \to Y_h'$ by $(B_h \mathcal{w}_h)(v_h) := \mathcal{b}(\mathcal{w}_h, v_h)$, for any $\mathcal{w}_h \in X_h, v_h \in Y_h$. This map is injective. Indeed, if $B_h \mathcal{w}_h = 0$, then for every $v \in Y$, the Fortin property (66) gives $\mathcal{b}(\mathcal{w}_h, v) = \mathcal{b}(\mathcal{w}_h, \Pi_h v) = 0$, because $\Pi_h v \in V_h \subseteq V_h^{++} \subseteq Y_h$. Lemma 7.3 therefore gives $\mathcal{w}_h = 0$. By rank-nullity theorem, rank of $B_h$ must therefore equal $\dim X_h$. The transpose $B_h^* : Y_h \to X_h'$ has the same rank as $B_h$, and its kernel is $Y_h^0$, so rank-nullity theorem gives

$$\dim Y_h^0 = \dim Y_h - \dim X_h. \tag{89}$$

The same bilinear form also generates an injective map from $X_h \to V_h'$ by the same Fortin property and by the same argument as above. Hence $\dim X_h$ cannot be larger than $\dim V_h$, so (89) implies

$$\dim Y_h^0 \geq \dim Y_h - \dim V_h. \tag{90}$$

Finally, since $N \geq 2$, Theorem 6.2(a) gives the direct sum $V^+(K) = V(K) \oplus P_0(K)$. Because $V^+(K) \subseteq V^{++}(K)$, it follows that $\dim V_h^{++} \geq \dim V_h + N_{\mathrm{el}}$. This provides a lower bound on the dimension of $Y_h$ since $V_h^{++} \subseteq Y_h$. The inequality (90) then implies $\dim Y_h^0 \geq \dim V_h^{++} - \dim V_h \geq N_{\mathrm{el}}$. □

## 9. Lowest order special case

In this section, we show how a change of the test space metric by a scale $\sigma$ yields optimal first-order rate in $L_2(\Omega)$ for the lowest-order case (improving upon Corollary 7.10's rate). The analysis of this section is the first, to our knowledge, to show the influence of such a scale $\sigma$ on the practical DPG method's error. The key ingredients are the augmented Fortin operator $\Pi_h^{++}$ and a characterization from [9] of how the energy norm changes with $\sigma$. The latter result is only for Hilbert spaces, so for brevity, without attempting to extend it here, we will restrict to $p = q = 2$ in this section.

For $0 < \sigma < \infty$, the modified norms are given by

$$\|v\|_{Y,\sigma}^2 := \sigma^{-2}\|v\|_{L_2(\Omega)}^2 + \sum_{K \in \Omega_h} \|\beta \cdot \nabla v\|_{L_2(K)}^2, \qquad \|\mathcal{w}\|_{\mathscr{E},\sigma} := \sup_{0 \neq v \in Y} \frac{\mathcal{b}(\mathcal{w}, v)}{\|v\|_{Y,\sigma}}, \tag{91}$$

for all $v \in Y := W_2(\partial_\beta, \Omega_h)$ and $\mathcal{w} \in X = L_2(\Omega) \times \hat{X}_0$ where $\hat{X}_0 = D_h(W_{2,0}(\partial_\beta, \Omega))$ as in (62). Clearly, the second norm in (91) is precisely the energy norm (67) when the original test norm is replaced by the new test norm $\|\cdot\|_{Y,\sigma}$. The subspace $Y_h$ is now considered with the norm $\|\cdot\|_{Y,\sigma}$ and discrete dual norm in (61) with $\|\cdot\|_Y$ replaced by $\|\cdot\|_{Y,\sigma}$ is denoted by $\|\cdot\|_{Y_{h,\sigma}'}$. In this section, we study the solution $\mathcal{u}_h^{(\sigma)} = (u_h^{(\sigma)}, \mu_h^{(\sigma)})$ of the residual minimization (12) and the estimator $\eta^{(\sigma)}$ of (83), both obtained using $\|\cdot\|_{Y_{h,\sigma}'}$ in place of $\|\cdot\|_{Y_h'}$. The effect of such a DPG scale was first analyzed in [13] for the Helmholtz equation, and later, more generally in [9], both only in the context of ideal DPG methods. With our augmented Fortin operator, we can now give the first analysis of the practical DPG method with the scale.

For $\mu \in \hat{X}_0$, let $E\mu \in W_{2,0}(\partial_\beta, \Omega)$ minimize $\|\beta \cdot \nabla z\|_{L_2(\Omega)}$ over all $z \in W_{2,0}(\partial_\beta, \Omega)$ satisfying $D_h z = \mu$. The theory of [9, Theorem 7.9 and Propositions 7.10–7.11], whose bijectivity hypothesis follows from Lemma 7.2, shows that for all $\mathcal{w} = (w, \mu)$ in $X$,

$$(1 + k_\sigma)^{-1}\|\mathcal{w}\|_{\mathscr{E},\sigma}^2 \leq \|w\|_{L_2(\Omega)}^2 + \sigma^2\|\beta \cdot \nabla E\mu\|_{L_2(\Omega)}^2 \leq (1 + k_\sigma)\|\mathcal{w}\|_{\mathscr{E},\sigma}^2, \tag{92}$$

with $k_\sigma = \frac{1}{2}(c_\sigma + \sqrt{c_\sigma^2 + 4c_\sigma})$ and $c_\sigma = C_{\mathrm{PF}}(2,\beta)^2/\sigma^2$. In particular, for every $\sigma > 0$,

$$\|w\|_{L_2(\Omega)} \le \sqrt{1+k_\sigma}\,\|(w,\mu)\|_{\mathscr{E},\sigma}. \tag{93}$$

Write $\operatorname{osc}^{(\sigma)}(f) := \sigma\|f - Q_0 f\|_{L_2(\Omega)}$. Recall that $a_K := h_K/|\beta_K|$, that $\bar\beta$ is as in (80), and that $Q_0$ is the element-wise $L_2(\Omega)$-orthogonal projection onto $P_0(\Omega_h)$. In the next theorem, no non-characteristic assumption is needed.

**Theorem 9.1** (A priori and a posteriori estimates)**.** *Let $k = 1$ and $p = 2$, suppose Assumptions 3.1 and 7.1 and (4) hold, and that $V_h^{++} \subseteq Y_h$ with $Y_h$ satisfying (22). Fix any $\sigma > 0$ and consider meshes with $\max_K a_K \le \sigma$. Then, there are constants $C_i = C_i(N,\kappa_0)$, $i = 1,2$, independent of $\beta$, $\sigma$, and the mesh, such that the following statements hold.*

*(a) The estimator $\eta^{(\sigma)}$ is efficient and reliable up to data oscillation,*

$$\eta^{(\sigma)} \le \|\boldsymbol{u} - \boldsymbol{u}_h^{(\sigma)}\|_{\mathscr{E},\sigma} \le C_1\eta^{(\sigma)} + (1 + C_1)\operatorname{osc}^{(\sigma)}(f), \tag{94}$$

*and is also a reliable indicator of the error in the interior variable,*

$$\|u - u_h^{(\sigma)}\|_{L_2(\Omega)} \le \sqrt{1+k_\sigma}\big[C_1\eta^{(\sigma)} + (1 + C_1)\operatorname{osc}^{(\sigma)}(f)\big]. \tag{95}$$

*If $f$ is piecewise constant on the mesh, then $\operatorname{osc}^{(\sigma)}(f) = 0$.*

*(b) If $u \in H^2(\Omega)$, then with $\bar\beta$ as in (80),*

$$\|u - u_h^{(\sigma)}\|_{L_2(\Omega)} \;\le\; C_2(1 + k_\sigma)\big(1 + \sigma\bar\beta\big)\,h\,\|u\|_{H^2(\Omega)}. \tag{96}$$

*Proof.* Theorem 6.4 shows that $\Pi_h^{++}$ maps into $V_h^{++}$, has the Fortin property, and, by (51), with $\sigma_K = \sigma \ge a_K$, satisfies $\|\Pi_h^{++} v\|_{Y,\sigma} \le C_1\|v\|_{Y,\sigma}$, with a constant $C_1$ depending only on $N$ and $\kappa_0$.

*Proof of (a).* Consistency (70) and the inclusion $Y_h \subset Y$ immediately give the efficiency bound $\eta^{(\sigma)} \le \|\boldsymbol{u} - \boldsymbol{u}_h^{(\sigma)}\|_{\mathscr{E},\sigma}$ as in the proof of Theorem 8.1. For reliability, repeating the same splitting in that proof, but now with $\Pi_h^{++}$, its Fortin property eliminates $\mathscr{b}(\boldsymbol{u}_h^{(\sigma)}, v - \Pi_h^{++} v)$, while its mean-preserving property gives, on every element,

$$(f, v - \Pi_K^{++} v)_K = (f - Q_0 f, v - \Pi_K^{++} v)_K.$$

The Cauchy–Schwarz inequality, first element-wise and then over the mesh, yields

$$\big|(f, v - \Pi_h^{++} v)_\Omega\big| \le \operatorname{osc}^{(\sigma)}(f)\,\|(I - \Pi_h^{++})v\|_{Y,\sigma} \le (1 + C_1)\operatorname{osc}^{(\sigma)}(f)\|v\|_{Y,\sigma}.$$

Taking the supremum over $v$ proves (94). Applying (93) to $\boldsymbol{u} - \boldsymbol{u}_h^{(\sigma)}$ gives (95) and its piecewise-constant specialization.

*Proof of (b).* The proof of Theorem 7.5, with the scaled test and energy norms, gives

$$\|\boldsymbol{u} - \boldsymbol{u}_h^{(\sigma)}\|_{\mathscr{E},\sigma} \;\lesssim\; \inf_{\boldsymbol{w}_h \in X_h} \|\boldsymbol{u} - \boldsymbol{w}_h\|_{\mathscr{E},\sigma}. \tag{97}$$

Take $\boldsymbol{w}_h = (Q_0 u, D_h I_h u)$, where $I_h$ is the degree-one Scott–Zhang interpolant preserving the inflow boundary condition. Since $u - I_h u$ is an admissible extension of $D_h(u - I_h u)$, standard approximation estimates and (92) give

$$\begin{aligned}\|\boldsymbol{u} - \boldsymbol{w}_h\|_{\mathscr{E},\sigma} &\le \sqrt{1+k_\sigma}\,\big(\|u - Q_0 u\|_{L_2(\Omega)} + \sigma\|\beta\cdot\nabla(u - I_h u)\|_{L_2(\Omega)}\big)\\ &\le C\sqrt{1+k_\sigma}(1 + \sigma\bar\beta)h\|u\|_{H^2(\Omega)}.\end{aligned}$$

Combining this with (97) and (93) proves (96). ∎

*Remark* 9.2 (Optimal $\sigma$)*.* The two factors in (96) pull in opposite directions: $1+k_\sigma$ decreases to 1 while $1+\sigma\bar{\beta}$ grows linearly as $\sigma \to \infty$. As $\sigma \to 0$, an elementary expansion shows that $k_\sigma = (C_{\mathrm{PF}}(2,\beta)/\sigma)^2 + 1 + O(\sigma^2)$. The product $(1+k_\sigma)\bigl(1+\sigma\bar{\beta}\bigr)$ therefore tends to infinity at both ends so must have a minimum determined by $C_{\mathrm{PF}}(2,\beta)$ and $\bar{\beta}$ alone. Also observe that it is essential to hold $\sigma$ fixed as $h \to 0$ (so that the hypothesis $\max_K a_K \le \sigma$ of Theorem 9.1 is satisfied as $h \to 0$.) The smallest admissible choice $\sigma = \max_K a_K$ destroys the rate (as then $\sigma \eqsim h$, $k_\sigma \eqsim h^{-2}$, and the right hand side of (96) diverges like $h^{-1}$). The optimal rate is thus a consequence of decoupling the test norm scale from the mesh scale.

## 10. Closing remarks

Discrete stability analysis of practical DPG methods is more involved for transport equations than for second-order elliptic problems, for which enriched polynomial test spaces are available [14, 19]. Ever since [7] exhibited nonpolynomial optimal test functions for advection, finding a stable polynomial test space surrogate has remained a challenge. The key difficulty is to obtain estimates uniform in the direction of the flow relative to the mesh. Until now, the only DPG stability result for advection, using standard finite element spaces, that was free of any condition on the flow direction was that of [1]. They traced the difficulty to the instability of optimal test functions as $\beta$ becomes tangent to an element facet, and recovered orientation-uniform stability using polynomial surrogates on a sufficiently deep fixed subgrid [1, Section 4.4 and Theorem 4.8]. Because the required refinement depth is not quantified, their result left open whether an explicit polynomial test space on the original mesh can suffice. We have answered this question affirmatively without any directional restrictions for $k=1$, and under a uniform non-characteristic facet condition for $k \ge 2$.

Our Fortin construction for the ultraweak advection formulation splits into facet conditions involving the signed facet-bubble combination $b_s$, and interior moment conditions involving the interior bubble $b$. Building the local test space out of exactly these two ingredients makes the Fortin system square and unisolvent, yielding a minimal test space on the same mesh (without any refinement). Everything quantitative then follows from the scaling estimates of Section 5, in which the advection vector is carried as a variable through a compactness argument, so that the constants are uniform over admissible flow directions and invariant under rescaling of $\beta$.

At lowest order $k=1$, our results have no directional restrictions and admit meshes containing characteristic facets. In our first step, the minimal test space, without any augmentation, gave the vacuous rate $h^{\min(s,k)-1} = 1$ for the field in the mesh-independent norm. Adding one constant per element and choosing a fixed scale $\sigma \ge \max_K a_K$ recovered the optimal first-order rate in $L_2(\Omega)$, and an additional mean-preserving bubble augmentation yielded an a posteriori bound whose data oscillation vanishes for piecewise constant data. That so small a modification should close the gap is striking. It suggests that the scale $\sigma$ deserves to be treated as a genuine design parameter whose numerical and adaptive selection merits further study. For $k \ge 2$, the result is qualified (needing Assumption 3.6).

Finding a test space with this qualification removed is an open problem. The Fortin conditions themselves are not the obstruction: the correction formula (47) remains Fortin for every linear map $R_K$. The missing step is to choose $R_K$ so that $v - R_K v$ annihilates the data in (39), allowing the refined bounds (40) to apply. The mean projection has this property for $k=1$. For $k=2$, it annihilates the interior data $P_0(K)$, but not the generally nonconstant boundary data $\beta \cdot \nabla P_2^{\perp}(K)$; for $k \ge 3$, it generally annihilates neither. This suggests seeking a bounded linear map $R_K : W_q(\partial_\beta, K) \to Z_{k'}(K) := \{\varphi \in P_{k'}(K) : \beta_K \cdot \nabla\varphi = 0\}$ satisfying $(v - R_K v, \mu)_K = 0$ for all $\mu \in M_k(K) := P_{k-2}(K) + \beta \cdot \nabla P_k^\perp(K) \subseteq P_{k-1}(K)$. Then $v - R_K v$ satisfies (39), and the proof of Theorem 6.2(b) extends with $V(K) + Z_{k'}(K)$ in place of $V(K) + P_0(K)$. (No projection property is required.) The open question is whether such an $R_K$ can be constructed with uniformly bounded degree and norm over the admissible flow directions. The studies in [10] suggest $k' \geq 2k-2$ in two dimensions, while the optimal test space of [7] supplies the corresponding nonpolynomial streamline-constant correction. One other limitation of this work is that $\beta$ is assumed piecewise constant. Extension of the Fortin analysis to variable fields is desirable and is postponed to future work.

Ongoing work aims to apply the results here in a scientific machine learning context, where we approximate the high-dimensional parameter-to-solution map $\beta \mapsto u$ of (1) by neural networks (NNs). Traditional applications of Theorem 8.1 and (94) are in adaptivity, but as explained in [5], they also generate variationally correct loss functions for training NNs. The bound (94) for $\eta^{(\sigma)}$ is particularly interesting in this context: for $k = 1$ it is robust in the element-wise directions of $\beta$ on every shape-regular mesh with $\max_K a_K \leq \sigma$. Moreover, $\sigma$ may be supplied to the NN as an input parameter alongside the element-wise $\beta$. Because of this ongoing numerical work, and ample prior numerical studies on DPG methods for advection [7, 8, 10], no further numerical experiments are reported here.

## Funding

This work was supported in part by NSF Grant 2245077.

## References

[1] D. Broersen, W. Dahmen, and R. P. Stevenson, *On the stability of DPG formulations of transport equations*, Math. Comp., 87 (2018), pp. 1051–1082.

[2] P. Cantin, *Well-posedness of the scalar and the vector advection–reaction problems in Banach graph spaces*, C. R. Acad. Sci. Paris, Ser. I, 355 (2017), pp. 892–902.

[3] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan, *A posteriori error control for DPG methods*, SIAM J. Numer. Anal., 52 (2014), pp. 1335–1353.

[4] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan, *Breaking spaces and forms for the DPG method and applications including Maxwell equations*, Comput. Math. Appl., 72 (2016), pp. 494–522.

[5] P. Cortés Castillo, W. Dahmen, and J. Gopalakrishnan, *DPG loss functions for learning parameter-to-solution maps by neural networks*, ArXiv Preprint, (2025).

[6] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive Petrov-Galerkin methods for first order transport equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445.

[7] L. Demkowicz and J. Gopalakrishnan, *A class of discontinuous Petrov–Galerkin methods. Part I: The transport equation*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 1558–1572.

[8] ———, *A class of discontinuous Petrov–Galerkin methods. II. Optimal test functions*, Numer. Methods Partial Differential Equations, 27 (2011), pp. 70–105.

[9] L. Demkowicz and J. Gopalakrishnan, *The discontinuous Petrov–Galerkin method*, Acta Numerica, 34 (2025), pp. 293–384.

[10] L. Demkowicz, N. V. Roberts, and J. Muñoz-Matute, *The DPG method for the convection-reaction problem, revisited*, Comput. Methods Appl. Math., 23 (2023), pp. 93–125.

[11] T. Führer and N. Heuer, *Robust DPG test spaces and Fortin operators—The $H^1$ and $H(\mathrm{div})$ cases*, SIAM Journal on Numerical Analysis, 62 (2024), pp. 718–748.

[12] J. Gopalakrishnan, P. Monk, and P. Sepúlveda, *A tent pitching scheme motivated by Friedrichs theory*, Computers and Mathematics with Applications, 70 (2015), pp. 1114–1135.

[13] J. Gopalakrishnan, I. Muga, and N. Olivares, *Dispersive and dissipative errors in the DPG method with scaled norms for the Helmholtz equation*, SIAM J. Sci. Comput., 36 (2014), pp. A20–A39.

[14] J. Gopalakrishnan and W. Qiu, *An analysis of the practical DPG method*, Math. Comp., 83 (2014), pp. 537–552.

[15] J.-L. Guermond, *Stabilization of Galerkin approximations of transport equations by subgrid modeling*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 1293–1316.

[16] M. Jensen, *Discontinuous Galerkin Methods for Friedrichs Systems with Irregular Solutions*, PhD thesis, University of Oxford, 2004.

[17] I. Muga, M. J. W. Tyler, and K. G. van der Zee, *The discrete-dual minimal-residual method (DDMRes) for weak advection-reaction problems in Banach spaces*, Comput. Methods Appl. Math., 19 (2019), pp. 557–579.

[18] I. Muga and K. G. van der Zee, *Discretization of linear problems in Banach spaces: residual minimization, nonlinear Petrov–Galerkin, and monotone mixed methods*, SIAM J. Numer. Anal., 58 (2020), pp. 3406–3426.

[19] S. Nagaraj, S. Petrides, and L. Demkowicz, *Construction of DPG Fortin operators for second order problems*, Comput. Math. Appl., 74 (2017), pp. 1964–1980.

[20] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493.

Portland State University, PO Box 751, Portland OR 97201, USA
*Email address*: pcortes@pdx.edu

Portland State University, PO Box 751, Portland OR 97201, USA
*Email address*: gjay@pdx.edu