Research Article

Ignacio Muga*, Matthew J. W. Tyler and Kristoffer G. van der Zee*

The Discrete-Dual Minimal-Residual Method (DDMRes) for Weak Advection-Reaction Problems in Banach Spaces

https://doi.org/10.1515/cmam-2018-0199
Received August 6, 2018; revised June 8, 2019; accepted June 13, 2019

Abstract: We propose and analyze a minimal-residual method in discrete dual norms for approximating the solution of the advection-reaction equation in a weak Banach-space setting. The weak formulation allows for the direct approximation of solutions in the Lebesgue L^p-space, $1 < p < \infty$. The greater generality of this weak setting is natural when dealing with rough data and highly irregular solutions, and when enhanced qualitative features of the approximations are needed. We first present a rigorous analysis of the well-posedness of the underlying continuous weak formulation, under natural assumptions on the advection-reaction coefficients. The main contribution is the study of several discrete subspace pairs guaranteeing the discrete stability of the method and quasi-optimality in L^p, and providing numerical illustrations of these findings, including the elimination of Gibbs phenomena, computation of optimal test spaces, and application to 2-D advection.

Keywords: Residual Minimization, Discrete Dual Norms, DDMRes, Advection-Reaction, Banach Spaces, Fortin Condition, Compatible FE Pairs, Petrov–Galerkin Method

MSC 2010: 41A65, 65J05, 46B20, 65N12, 65N15, 35L02, 35J25

1 Introduction

Residual minimization encapsulates the idea that an approximation to the solution $u \in \mathbb{U}$ of an (infinite-dimensional) operator equation $Bu = f$ can be found by minimizing the norm of the residual $f - Bw_n$ amongst all w_n in some finite-dimensional subspace $\mathbb{U}_n \subset \mathbb{U}$. This powerful idea provides a stable and convergent discretization method under quite general assumptions, i.e., when $B\colon \mathbb{U} \to \mathbb{V}^*$ is any linear continuous bijection from Banach space $\mathbb{U}$ onto the dual $\mathbb{V}^*$ of a Banach space $\mathbb{V}$, $f \in \mathbb{V}^*$, and $\operatorname{dist}(u, \mathbb{U}_n) \to 0$ as $n \to \infty$; see, e.g., Guermond [30, Section 2] for details. Note that this applies to well-posed weak formulations of linear partial differential equations (PDEs), in which case B is induced by the underlying bilinear form (i.e., $\langle Bw, v\rangle = b(w, v)$ for all $w \in \mathbb{U}$ and all $v \in \mathbb{V}$). As such, residual minimization is essentially an ideal methodology for non-coercive and/or nonsymmetric problems.

However, for many weak formulations of PDEs, $\mathbb{V}^*$ is a *negative* space (such as $H^{-m}(\Omega)$, or more generally $W^{-m,p}(\Omega)$, which is the dual of the Sobolev space $W_0^{m,q}(\Omega)$, where $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $m = 1, 2, \dots$, or the dual of a graph space). In that case, this requires the minimization of the residual in the *non-computable* dual norm $\|\cdot\|_{\mathbb{V}^*}$. To make this tractable, one can instead minimize in a *discrete* dual norm. In other words,

***Corresponding author: Ignacio Muga,** Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile, e-mail: ignacio.muga@pucv.cl. http://orcid.org/0000-0003-4430-5167
***Corresponding author: Kristoffer G. van der Zee,** School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG72RD, United Kingdom, e-mail: kg.vanderzee@nottingham.ac.uk. http://orcid.org/0000-0002-6830-8031
Matthew J. W. Tyler, School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG72RD, United Kingdom, e-mail: pmymt7@nottingham.ac.uk

one aims to find an approximation $u_n \in \mathbb{U}_n$ such that $\|f - Bu_n\|_{(\mathbb{V}_m)^*}$ is minimal, where $\mathbb{V}_m$ is some finite-dimensional subspace of $\mathbb{V}$. We refer to this discretization method as residual minimization in discrete dual norms, or simply as the DDMRes method (*Discrete-Dual Minimal-Residual* method).

In this paper, we consider the DDMRes method when applied to a canonical linear first-order PDE in weak Banach-space settings. In particular, we consider the advection-reaction operator $u \mapsto \boldsymbol{\beta} \cdot \nabla u + \mu u$, with $\boldsymbol{\beta} \colon \Omega \to \mathbb{R}^d$ and $\mu \colon \Omega \to \mathbb{R}$ given advection-reaction coefficients, in a functional setting for which the solution space $\mathbb{U}$ is $L^p(\Omega)$, $1 < p < \infty$, and $\mathbb{V}$ is a suitable Banach graph space (see Section 3 for details). This weak setting allows for the direct approximation of *irregular* solutions, while the greater generality of Banach spaces (over more common Hilbert spaces) is useful for example in the extension to nonlinear hyperbolic PDEs [31],[1] as well as in approximating solutions with discontinuities (allowing the elimination of Gibbs phenomena; see further details below).

Undeniably, solving a linear problem by minimizing its residual in a discrete dual *Hilbert-space* norm has been considered before. Most importantly, a DDMRes principle is at the core of any *discontinuous Petrov–Galerkin* (DPG) technology [20], as well as certain adaptive stabilization strategies [14, 15]. In particular, DPG can be seen as the DDMRes method obtained when B corresponds to a mesh-dependent hybrid formulation of the underlying PDE so that $\mathbb{V}$ is a broken Sobolev-type space. While the vast majority of these methods rely on the inversion of a linear Riesz map (which is a Hilbert-space construct), in more general Banach spaces, the DDMRes method is equivalent to certain (inexact) *nonlinear Petrov–Galerkin* methods, or equivalently, mixed methods with monotone nonlinearity, where the nonlinearity originates from the nonlinear *duality map* $J_{\mathbb{V}} \colon \mathbb{V} \to \mathbb{V}^*$; see Muga and Van der Zee [33] for details, including a schematic overview of connections to other methods. Nonlinearity is unavoidable when minimizing residuals in non-Hilbert Banach spaces (cf. Guermond [30]).

The numerical analysis of the DDMRes method has been carried out abstractly by Gopalakrishnan and Qiu [28] in Hilbert spaces (see also [15, Section 3]) and by Muga and Van der Zee [33] in smooth Banach spaces. A key requirement in these analyses is the *Fortin* compatibility condition on the family of discrete subspace pairs $(\mathbb{U}_n, \mathbb{V}_m)$ under consideration, which, once established, implies stability and quasi-optimal convergence of the method. In some sense, the Fortin condition is rather mild since, for a given $\mathbb{U}_n$, there is the expectation that it will be satisfied for a sufficiently large $\mathbb{V}_m$ (thereby making the discrete dual norm $\|\cdot\|_{(\mathbb{V}_m)^*}$ sufficiently close to $\|\cdot\|_{\mathbb{V}^*}$). Of course, whether this can be established depends crucially on the operator B, therefore also on the particular weak formulation of the PDE that is being studied.

The main contribution of this paper consists in the study of several elementary discrete subspace pairs $(\mathbb{U}_n, \mathbb{V}_m)$ for the DDMRes method for weak advection-reaction, including proofs of Fortin compatibility in the above-mentioned Banach-space setting. It thereby provides the first application and corresponding analysis of DDMRes in genuine (non-Hilbert) Banach spaces. In particular, for the given compatible pairs, DDMRes is thus a quasi-optimal method providing a near-best approximation in $L^p(\Omega)$. Note that our results do not cover DPG-type hybrid weak formulations (with a broken graph space $\mathbb{V}$) so that our discrete spaces $\mathbb{V}_m$ are globally conforming. Broken Banach-space settings will be treated in forthcoming work.

We now briefly discuss some details of our results. To be able to carry out the analysis, our results focus on discrete subspace pairs $(\mathbb{U}_n, \mathbb{V}_m)$, where $\mathbb{U}_n$ is a lowest-order finite element space on mesh $\mathcal{T}_n$ in certain specialized settings. We first consider *continuous linear* finite elements in combination with continuous finite elements of degree k, i.e., $\mathbb{U}_n = \mathbb{P}^1_{\text{cont}}(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^k_{\text{cont}}(\mathcal{T}_n)$. The Fortin condition holds when $k \geq 2$, assuming, e.g., incompressible pure advection ($\operatorname{div} \boldsymbol{\beta} = \mu = 0$) in a one-dimensional setting. Interestingly, we demonstrate that the notorious *Gibbs phenomenon* of spurious numerical over- and undershoots, commonly encountered while approximating discontinuous solutions with continuous approximations, can be *eliminated* with the DDMRes method upon $p \to 1^+$ (see Section 5.1), which is in agreement with previous findings on L^1-methods [30, 32].

We then consider $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$, that is, *discontinuous piecewise-constant* approximations on arbitrary partitionings of the domain Ω in $\mathbb{R}^d$, $d \geq 1$. It turns out that it is possible to define an *optimal* test space

1 Confirm [10–12] for nonlinear PDE examples in Hilbert-space settings using a DPG approach.

$\mathbb{S}_n := B^{-*}\mathbb{U}_n$ and subsequently prove Fortin's condition for any $\mathbb{V}_m \supseteq \mathbb{S}_n$. This result essentially hinges on the fact that $\mathbb{U}_n$ is invariant under the L^p duality map (see the proof of Proposition 5.2). Since the optimal test space is however not explicit, it requires, in general, the computation of an explicit basis (see Section 5.2). Such computations may not be feasible in practise, and in those cases, as an alternative, one could resort to sufficiently rich $\mathbb{V}_m$, e.g., continuous linear finite elements on a sufficiently refined submesh of the original mesh (cf. [7]).

Interestingly however, under certain special, yet nontrivial situations, the optimal test space $\mathbb{S}_n$ happens to coincide with a convenient finite element space. For example, in 2-D in the incompressible pure advection case with $\boldsymbol{\beta}$ piecewise constant on some partition, if $\mathcal{T}_n$ is a triangular mesh of Ω (compatible with the partition) and all triangles are flow-aligned, then we prove that $\mathbb{S}_n = \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$, where $\mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ refers to the space of piecewise-linear functions that are *conforming* with respect to the graph space $\mathbb{V}$. Numerical experiments in 2-D indeed confirm in this case the quasi-optimality of DDMRes (see Section 5.3).

In recent years, several similar methods for weak advection-reaction have appeared, all of which were in Hilbert-space settings (i.e., the solution space is $L^2(\Omega)$) and use a broken weak formulation. Indeed, these include some of the initial DPG methods [8, 18, 19], which were proposed before the importance of Fortin's condition was clarified. Recently however, Broersen, Dahmen and Stevenson [7] studied a higher-order pair using standard finite-element spaces for the DPG method of weak advection-reaction. Under mild conditions on $\boldsymbol{\beta}$, they proved Fortin's condition when $\mathbb{U}_n$ consists of piecewise polynomials of degree k, and $\mathbb{V}_m$ consists of piecewise polynomials of higher degree over a sufficiently deep refinement of the trial mesh. The extension of their proof, based on approximating the optimal test space, to any Banach-space setting seems nontrivial since, currently, the concept of an optimal test space is in general absent in DDMRes in Banach spaces (cf. [33]), exceptions notwithstanding (such as the lowest-order piecewise-constant case discussed above).

Let us finally point out that methods for weak advection-reaction are quite distinct from methods for *strong* advection-reaction (which has its residual in $L^p(\Omega)$ and, a priori, demands more regularity on its solution). Indeed, there is a plethora of methods in the strong case; see, e.g., Ern and Guermond [22, Chapter 5] and Guermond [30], all of which typically exhibit *suboptimal* convergence behavior when measured in $L^p(\Omega)$. In the context of *strong* advection-reaction, the results by Guermond [29] are noteworthy, who proved the Fortin condition for several pairs, consisting of a low-order finite element space and its enrichment with bubbles. These results however do not apply to weak advection-reaction. Similarly for the stability result by Chan, Evans and Qiu [13].

The remainder of this paper is arranged as follows. In Section 2, we first present preliminaries for the advection-reaction equation, allowing us to recall in Section 3 the specifics of the well-posed Banach-space setting (cf. Cantin [9]). In particular, we provide a self-contained proof of the continuous inf-sup conditions using various properties of the L^p duality map. Then, in Section 4, we consider the discrete problem corresponding to the DDMRes method in the equivalent form of the monotone mixed method and establish stability and quasi-optimality of the method, provided the Fortin condition holds. In Section 5, we consider particular discrete subspace pairs $(\mathbb{U}_n, \mathbb{V}_m)$. This section contains several proofs of Fortin conditions, as well as some illustrative numerical examples pertaining to the Gibbs phenomena (Section 5.1), optimal test space basis (Section 5.2), and quasi-optimal convergence for 2-D advection (Section 5.3).

2 Advection-Reaction Preliminaries

For any dimension $d \geq 1$, let $\Omega \subset \mathbb{R}^d$ be an open bounded domain with Lipschitz boundary $\partial\Omega$ oriented by a unit outward normal vector $\boldsymbol{n}$. Let $\boldsymbol{\beta} \in L^\infty(\Omega)$ be an advection-field such that $\operatorname{div}\boldsymbol{\beta} \in L^\infty(\Omega)$, and let $\mu \in L^\infty(\Omega)$ be a (space-dependent) reaction coefficient. The advection field splits the boundary $\partial\Omega$ into an *inflow*, *outflow* and *characteristic* part, which for continuous $\boldsymbol{\beta}$ corresponds to

$$\begin{aligned}\partial\Omega_- &:= \{x \in \partial\Omega : \boldsymbol{\beta}(x)\cdot\boldsymbol{n}(x) < 0\},\\ \partial\Omega_+ &:= \{x \in \partial\Omega : \boldsymbol{\beta}(x)\cdot\boldsymbol{n}(x) > 0\},\\ \partial\Omega_0 &:= \{x \in \partial\Omega : \boldsymbol{\beta}(x)\cdot\boldsymbol{n}(x) = 0\},\end{aligned}$$

respectively; see [7, Section 2] for the definition of the parts in the more general case $\boldsymbol{\beta}$, $\operatorname{div}\boldsymbol{\beta} \in L^\infty(\Omega)$ (which is based on the integration-by-parts formula (2.8)).

Given a possibly *rough* source $f_\circ$ and inflow data g, the advection-reaction model is

$$\begin{aligned} \boldsymbol{\beta}\cdot\nabla u + \mu u &= f_\circ \quad \text{in } \Omega, \\ u &= g \quad \text{on } \partial\Omega_-. \end{aligned} \tag{2.1}$$

Before we give a weak formulation for this model and discuss its well-posedness, we first introduce relevant assumptions and function spaces. We have in mind a weak setting where $u \in L^p(\Omega)$ for any p in $(1,\infty)$. Therefore, throughout this section, let $1 < p < \infty$, and let $q \in (1,\infty)$ denote the *conjugate* exponent of p, satisfying the relation $p^{-1} + q^{-1} = 1$.

The following assumptions are natural extensions of the classical ones in the Hilbert case.

Assumption 2.1 (Friedrich's Positivity Assumption). There exists a constant $\mu_0 > 0$ for which

$$\mu(x) - \frac{1}{p}\operatorname{div}\boldsymbol{\beta}(x) \geq \mu_0, \quad \text{a.e. } x \text{ in } \Omega. \tag{2.2}$$

Assumption 2.2 (Well-Separated In- and Outflow). The in- and outflow boundaries are *well-separated*, i.e., $\overline{\partial\Omega_-} \cap \overline{\partial\Omega_+} = \emptyset$ and, by partition of unity, there exists a function

$$\begin{cases} \phi \in C^\infty(\overline{\Omega}) \quad \text{such that} \\ \qquad \phi(x) = 1 \quad \text{for all } x \in \partial\Omega_-, \\ \qquad \phi(x) = 0 \quad \text{for all } x \in \partial\Omega_+. \end{cases} \tag{2.3}$$

For brevity, we use the following notation for norms and duality pairings:

$$\begin{aligned} \|\cdot\|_\infty &= \operatorname*{ess\,sup}_{x\in\Omega} |\cdot(x)|, \\ \|\cdot\|_\rho &= \Big(\int_\Omega |\cdot|^\rho\Big)^{1/\rho} && \text{for } 1 \leq \rho < \infty, \\ \langle\cdot,\cdot\rangle_{\rho,\sigma} &= \langle\cdot,\cdot\rangle_{L^\rho(\Omega),L^\sigma(\Omega)} && \text{for } 1 < \rho < \infty, \quad \sigma = \frac{\rho}{\rho-1}. \end{aligned} \tag{2.4}$$

Definition 2.3 (Graph Space). For $1 \leq \rho \leq \infty$, the *graph space* is defined by

$$W^\rho(\boldsymbol{\beta};\Omega) := \{w \in L^\rho(\Omega) : \boldsymbol{\beta}\cdot\nabla w \in L^\rho(\Omega)\}.$$

Motivated by the forthcoming introduction of the *duality maps* (see Definition 2.4), the space $W^\rho(\boldsymbol{\beta};\Omega)$ is endowed with the norm

$$\|w\|^2_{\rho,\boldsymbol{\beta}} := \|w\|^2_\rho + \|\boldsymbol{\beta}\cdot\nabla w\|^2_\rho.$$

The "adjoint" norm is defined by

$$|||w|||^2_{\rho,\boldsymbol{\beta}} := \|w\|^2_\rho + \|\operatorname{div}(\boldsymbol{\beta}w)\|^2_\rho. \tag{2.5}$$

These norms are equivalent, which can be shown by means of the identity

$$\operatorname{div}(\boldsymbol{\beta}w) = \operatorname{div}(\boldsymbol{\beta})w + \boldsymbol{\beta}\cdot\nabla w. \tag{2.6}$$

Definition 2.4 (Duality Maps). For $1 < \rho < \infty$, the *duality map of the graph space* $W^\rho(\boldsymbol{\beta};\Omega)$ (=: W^ρ) endowed with the adjoint norm (2.5) is defined by the operator $J_{W^\rho} : W^\rho(\boldsymbol{\beta};\Omega) \to (W^\rho(\boldsymbol{\beta};\Omega))^*$ such that, for all $w, v \in W^\rho(\boldsymbol{\beta};\Omega)$,

$$\langle J_{W^\rho}(w), v\rangle_{(W^\rho)^*,W^\rho} = \langle J_\rho(w), v\rangle_{\sigma,\rho} + \langle J_\rho(\operatorname{div}(\boldsymbol{\beta}w)), \operatorname{div}(\boldsymbol{\beta}v)\rangle_{\sigma,\rho}, \tag{2.7}$$

where $J_\rho : L^\rho(\Omega) \to L^\sigma(\Omega)$ with $\rho^{-1} + \sigma^{-1} = 1$ is the *duality map of* $L^\rho(\Omega)$ defined by

$$J_\rho(w) := \|w\|_\rho^{2-\rho}|w|^{\rho-1}\operatorname{sign}(w) \in L^\sigma(\Omega).$$

The duality maps J_ρ and J_{W^ρ} are *bijective isometries* (i.e., invertible and norm-preserving) and satisfy

$$\begin{aligned} \langle J_\rho(w), w\rangle_{\sigma,\rho} &= \|w\|^2_\rho && \text{for all } w \in L^\rho(\Omega), \\ \langle J_{W^\rho}(w), w\rangle_{(W^\rho)^*,W^\rho} &= |||w|||^2_{\rho,\boldsymbol{\beta}} && \text{for all } w \in W^\rho(\boldsymbol{\beta};\Omega). \end{aligned}$$

They are natural (nonlinear) extensions to $\rho \neq 2$ of the corresponding Riesz maps in the Hilbert case (i.e., J_2 and J_{W^2}, respectively). These definitions of duality maps are consistent with the general definition of the (normalized) duality map $J_{\mathbb{V}} : \mathbb{V} \to \mathbb{V}^*$ in a reflexive Banach space $\mathbb{V}$ for which $\mathbb{V}$ and $\mathbb{V}^*$ are strictly convex (which covers $W^\rho(\boldsymbol{\beta};\Omega)$ and $L^\rho(\Omega)$, $1 < \rho < \infty$). We refer to [33] for a comprehensive development on this topic.

Remark 2.5 (Graph Spaces and Traces). As a consequence of Assumption 2.2, traces of functions in $W^\rho(\boldsymbol{\beta};\Omega)$ are well-defined (by means of a linear and continuous trace operator) as functions in the space

$$L^\rho(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega) := \left\{ w \text{ measurable in } \partial\Omega : \int_{\partial\Omega} |\boldsymbol{\beta}\cdot\boldsymbol{n}||w|^\rho < +\infty \right\},$$

and, moreover, for all $w \in W^\rho(\boldsymbol{\beta};\Omega)$ and all $v \in W^\sigma(\boldsymbol{\beta};\Omega)$, the following integration-by-parts formula holds:

$$\int_\Omega ((\boldsymbol{\beta}\cdot\nabla w)v + (\boldsymbol{\beta}\cdot\nabla v)w + \operatorname{div}(\boldsymbol{\beta})wv) = \int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n})wv. \tag{2.8}$$

The proof of these results is a straightforward extension of the Hilbert-space case given by, e.g., Di Pietro and Ern [21, Section 2.1.5]; see also Dautray and Lions [17, Chapter XXI, §2, Section 2.2] for transport equations, Ern and Guermond [23, Section 3.1] and Cantin [9, Lemma 2.2] for advection-reaction. We can thus define the following two closed subspaces, which are relevant for prescribing boundary conditions at $\partial\Omega_+$ or $\partial\Omega_-$:

$$W^\rho_{0,\pm}(\boldsymbol{\beta};\Omega) := \{ w \in W^\rho(\boldsymbol{\beta};\Omega) : w|_{\partial\Omega_\pm} = 0 \}.$$

The following lemma establishes the surjectivity of the trace operator from $W^\rho(\boldsymbol{\beta};\Omega)$ onto $L^\rho(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$.

Lemma 2.6 (Surjectivity of Graph-Space Trace). *Let $1 < \rho < \infty$. For each $g \in L^\rho(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$, there is*

$$w_g \in W^\rho(\boldsymbol{\beta};\Omega) \quad \textit{such that} \quad w_g = g \quad \textit{a.e. in } \partial\Omega_+ \cup \partial\Omega_-.$$

Proof. By means of the duality map of the graph space (Definition 2.4), the Hilbert-space proof given by Di Pietro and Ern [21, Lemma 2.11] can be extended to the Banach-space setting; see Section A.1 for details. □

Remark 2.7 (Non-separated In- and Outflow). The requirement of separated in- and outflow can be removed, but different trace operators have to be introduced [27].

The case when $\mu \equiv 0$ and $\operatorname{div}\boldsymbol{\beta} \equiv 0$ is special since Assumption 2.1 is not satisfied. An important tool for the analysis of this case is the so-called *curved Poincaré–Friedrichs inequality*; see Lemma 2.10 below. Its proof relies on the following assumption (cf. [1, 2]).

Assumption 2.8 (Ω-Filling Advection). Let $1 < \rho < \infty$. If $\mu \equiv 0$ and $\operatorname{div}\boldsymbol{\beta} \equiv 0$, the advection-field $\boldsymbol{\beta}$ is Ω-*filling*, by which we mean that there exist $z_+, z_- \in W^\infty(\boldsymbol{\beta};\Omega)$ with $\|z_+\|_\infty, \|z_-\|_\infty > 0$ such that

$$\begin{cases} -\boldsymbol{\beta}\cdot\nabla z_\pm = \rho & \text{in } \Omega, \\ z_\pm = 0 & \text{on } \partial\Omega_\pm. \end{cases} \tag{2.9}$$

Remark 2.9 (Method of Characteristics). Assumption 2.8 holds, for example, if $\boldsymbol{\beta}$ is regular enough so that the method of characteristics can be employed to solve for z (cf. Dahmen et al. [15, Remark 2.2]).

Lemma 2.10 (Curved Poincaré–Friedrichs Inequality). *Let $1 < \rho < \infty$. Then, under the hypothesis that Assumption 2.8 holds true, there exists a constant $C_{\mathrm{PF}} > 0$ such that*

$$\|w\|_\rho \le C_{\mathrm{PF}} \|\boldsymbol{\beta}\cdot\nabla w\|_\rho \quad \textit{for all } w \in W^\rho_{0,\pm}(\boldsymbol{\beta};\Omega).$$

Proof. For the Hilbert-space case ($\rho = 2$), the proof can be found in [2]. For completeness, we reproduce here the general ρ-version.

Without loss of generality, take $w \in W^\rho_{0,-}(\boldsymbol{\beta};\Omega)$, and let $z = z_+ \in W^\infty_{0,+}(\boldsymbol{\beta};\Omega)$ as in Assumption 2.8. Notice the important identity

$$z\boldsymbol{\beta}\cdot\nabla(|w|^\rho) = \operatorname{div}(\boldsymbol{\beta}z|w|^\rho) - |w|^\rho\boldsymbol{\beta}\cdot\nabla z. \tag{2.10}$$

Let $\sigma = \frac{\rho}{\rho-1}$. Take $\phi_w = \rho z|w|^{\rho-1}\operatorname{sign}(w) \in L^\sigma(\Omega)$, which satisfies

$$\|\phi_w\|_\sigma \le \rho\|z\|_\infty \|w\|_\rho^{\rho-1}. \tag{2.11}$$

Thus

$$\begin{aligned}
\|\boldsymbol{\beta}\cdot\nabla w\|_\rho &= \sup_{0\neq\phi\in L^\sigma(\Omega)} \frac{\langle \boldsymbol{\beta}\cdot\nabla w, \phi\rangle_{\rho,\sigma}}{\|\phi\|_\sigma} && \text{(by duality)}\\
&\ge \frac{\langle \boldsymbol{\beta}\cdot\nabla w, \phi_w\rangle_{\rho,\sigma}}{\|\phi_w\|_\sigma} && \text{(since } \phi_w \in L^\sigma(\Omega)\text{)}\\
&= \frac{-\int_\Omega |w|^\rho \boldsymbol{\beta}\cdot\nabla z}{\|\phi_w\|_\sigma} && \text{(by (2.10) and } \int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n}) z|w|^\rho = 0\text{)}\\
&\ge \frac{\|w\|_\rho}{\|z\|_\infty} && \text{(by Assumption 2.8 and (2.11)).}
\end{aligned}$$

Hence $C_{\mathrm{PF}} = \|z\|_\infty$. If $w \in W^\rho_{0,+}(\boldsymbol{\beta};\Omega)$, take $z = z_- \in W^\infty_{0,-}(\boldsymbol{\beta};\Omega)$. □

The proof of Lemma 2.10 shows that $C_{\mathrm{PF}} = \|z_\pm\|_\infty$, with $z_\pm$ defined in (2.9); hence C_{PF} depends on Ω, $\boldsymbol{\beta}$ and ρ.

Remark 2.11 (Weaker Condition). Under a weaker condition than Assumptions 2.1 or 2.8, Lemma 2.10 can be generalized to the following situations:

$$\begin{aligned}
\|w\|_p &\lesssim \|\mu w + \boldsymbol{\beta}\cdot\nabla w\|_p \quad \text{for all } w \in W^p_{0,-}(\boldsymbol{\beta};\Omega),\\
\|v\|_q &\lesssim \|\mu v - \operatorname{div}(\boldsymbol{\beta} v)\|_q \quad \text{for all } v \in W^q_{0,+}(\boldsymbol{\beta};\Omega).
\end{aligned}$$

Indeed, it is enough to verify the existence of a constant $\mu_0^* > 0$ and a Lipschitz continuous function $\zeta(x)$ such that

$$\mu(x) - \frac{1}{p}\operatorname{div}\boldsymbol{\beta}(x) - \frac{1}{p}\boldsymbol{\beta}(x)\cdot\nabla\zeta(x) \ge \mu_0^*, \quad \text{a.e. } x \text{ in } \Omega. \tag{2.12}$$

This can be inferred from the recent work by Cantin [9]. Notice that if Assumption 2.1 is satisfied, then (2.12) holds with $\zeta(x) \equiv 0$ and $\mu_0^* = \mu_0$, while, for the case $\operatorname{div}\boldsymbol{\beta} = \mu = 0$, it holds with $\zeta = z_+$ (from Assumption 2.8) and $\mu_0^* = 1$.

3 A Weak Setting for Advection-Reaction

The weak setting for the advection-reaction problem (2.1) considers a trial space $\mathbb{U} := L^p(\Omega)$ endowed with the $\|\cdot\|_p$-norm (see (2.4)) and a test space $\mathbb{V} := W^q_{0,+}(\boldsymbol{\beta};\Omega)$ endowed with the norm $|||\cdot|||_{q,\boldsymbol{\beta}}$ (see (2.5)). The weak-formulation reads as follows:

$$\text{find} \quad u \in \mathbb{U} = L^p(\Omega) \quad \text{such that} \quad \langle Bu, v\rangle_{\mathbb{V}^*,\mathbb{V}} = \langle f, v\rangle_{\mathbb{V}^*,\mathbb{V}} \quad \text{for all } v \in \mathbb{V} = W^q_{0,+}(\boldsymbol{\beta};\Omega), \tag{3.1}$$

where $B\colon \mathbb{U} \to \mathbb{V}^*$ is defined by

$$\langle Bw, v\rangle_{\mathbb{V}^*,\mathbb{V}} := \int_\Omega w(\mu v - \operatorname{div}(\boldsymbol{\beta} v)) \quad \text{for all } w \in \mathbb{U} \text{ and all } v \in \mathbb{V}, \tag{3.2}$$

and the right-hand side f is related to the original PDE data $(f_\circ, g)$ via

$$\langle f, v\rangle_{\mathbb{V}^*,\mathbb{V}} = \int_\Omega f_\circ v + \int_{\partial\Omega_-} |\boldsymbol{\beta}\cdot\boldsymbol{n}| g v,$$

where $f_\circ$ is given in (for example) $L^p(\Omega)$ and g is given in $L^p(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$. More rough $f_\circ$ is allowed as long as $f \in [W^q_{0,+}(\boldsymbol{\beta};\Omega)]^*$.

Remark 3.1 (Non-homogeneous Dirichlet Data). Observe that, for any $g \in L^p(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$, the action

$$v \mapsto \int_{\partial\Omega_-} |\boldsymbol{\beta}\cdot\boldsymbol{n}| g v$$

defines a continuous linear functional on $W^q_{0,+}(\boldsymbol{\beta};\Omega)$. Indeed, owing to Remark 2.5, there is a constant $C_{\boldsymbol{\beta}} > 0$ such that

$$\left| \int_{\partial\Omega_-} |\boldsymbol{\beta}\cdot\boldsymbol{n}| g v \right| \leq \|g\|_{L^p(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)} \|v\|_{L^q(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)} \leq C_{\boldsymbol{\beta}} \|g\|_{L^p(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)} \|v\|_{q,\boldsymbol{\beta}}.$$

Moreover, a sufficient condition to guarantee that $g \in L^p(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$ is indeed the trace over $\partial\Omega_-$ of the solution $u \in \mathbb{U}$ of problem (3.1) is $f_\circ \in L^p(\Omega)$, in which case it is straightforward to verify that $u \in W^p(\boldsymbol{\beta};\Omega)$ and $u|_{\partial\Omega_-} = g$.

Remark 3.2 (Boundedness). The bilinear form in (3.1) is bounded with constant $M_\mu := \sqrt{1 + \|\mu\|^2_\infty}$. Indeed,

$$\left| \int_\Omega u(\mu v - \operatorname{div}(\boldsymbol{\beta} v)) \right| \leq \|u\|_p \|\mu v - \operatorname{div}(\boldsymbol{\beta} v)\|_q \leq M_\mu \|u\|_p \|v\|_{q,\boldsymbol{\beta}}.$$

The following result states the well-posedness of problem (3.1). Although this result can be inferred from the recent result by Cantin [9], we provide a slightly alternative proof based on establishing the adjoint inf-sup conditions using properties of the L^p duality map. For a classical proof of well-posedness in a similar Banach-space setting, we refer to Beirão da Veiga [4, 5] (cf. [3] and [17, Chapter XXI]).

Theorem A (Weak Advection-Reaction: Well-Posedness). *Let $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$. Let $\Omega \subset \mathbb{R}^d$ be an open bounded domain with Lipschitz boundary. Let $\boldsymbol{\beta}\colon \Omega \to \mathbb{R}$ and $\mu\colon \Omega \to \mathbb{R}$ be advection and reaction coefficients (respectively) satisfying either Friedrich's positivity Assumption 2.1 or the Ω-filling Assumption 2.8. Assume further that in- and outflow boundary are well separated (Assumption 2.2).*

(i) *For any $f \in \mathbb{V}^* = [W^q_{0,+}(\boldsymbol{\beta};\Omega)]^*$, there exists a unique solution $u \in L^p(\Omega)$ to the weak advection-reaction problem* (3.1).

(ii) *In the case that Assumption 2.1 holds true, we have the a priori bound*

$$\|u\|_p \leq \frac{1}{\gamma_B}\|f\|_{\mathbb{V}^*} \quad \text{with} \quad \gamma_B = \sqrt{\frac{\mu_0^2}{1 + (\mu_0 + \|\mu\|_\infty)^2}} \tag{3.3}$$

and $\mu_0 > 0$ being the constant in Assumption 2.1.

(ii*) *On the other hand, in the case where Assumption 2.8 holds true, we also have the a priori bound* (3.3), *but γ_B in* (3.3) *must be replaced by the constant $1/(1 + C_{\mathrm{PF}})$, where $C_{\mathrm{PF}} > 0$ is the Poincaré–Friedrichs constant in Lemma 2.10.*

Proof. See Section A.2. □

4 The General Discrete Problem

We now consider the approximate solution of (3.1) given by the DDMRes method. We first present the discrete scheme (and discuss its numerical solution) and then consider its well-posedness and stability.

Given finite-dimensional subspaces $\mathbb{U}_n \subset \mathbb{U} = L^p(\Omega)$ and $\mathbb{V}_m \subset \mathbb{V} = W^q_{0,+}(\boldsymbol{\beta};\Omega)$, we aim to find $u_n \in \mathbb{U}_n$ such that

$$u_n = \arg\min_{w_n \in \mathbb{U}_n} \|f - Bw_n\|_{(\mathbb{V}_m)^*}, \tag{4.1}$$

where the discrete dual norm is given by

$$\|\cdot\|_{(\mathbb{V}_m)^*} = \sup_{v_m \in \mathbb{V}_m} \frac{\langle \cdot, v_m\rangle_{(\mathbb{V}_m)^*,\mathbb{V}_m}}{\|v_m\|_{\mathbb{V}}}.$$

As proven in [33, Theorem 4.A], the minimization problem (4.1) is equivalent to following monotone mixed method: find $(r_m, u_n) \in \mathbb{V}_m \times \mathbb{U}_n$ such that

$$\langle J_{\mathbb{V}}(r_m), v_m\rangle_{\mathbb{V}^*,\mathbb{V}} + \langle Bu_n, v_m\rangle_{\mathbb{V}^*,\mathbb{V}} = \langle f, v_m\rangle_{\mathbb{V}^*,\mathbb{V}} \quad \text{for all } v_m \in \mathbb{V}_m, \tag{4.2a}$$

$$\langle Bw_n, r_m\rangle_{\mathbb{V}^*,\mathbb{V}} = 0 \quad \text{for all } w_n \in \mathbb{U}_n. \tag{4.2b}$$

which, by Definition 2.4 and equation (3.2), reduces to the system

$$\begin{aligned}\langle J_{W^q}(r_m), v_m\rangle_{(W^q)^*,W^q} + \langle u_n, \mu v_m - \operatorname{div}(\boldsymbol{\beta} v_m)\rangle_{p,q} &= \langle f, v_m\rangle_{\mathbb{V}^*,\mathbb{V}} \quad \text{for all } v_m \in \mathbb{V}_m,\\ \langle w_n, \mu r_m - \operatorname{div}(\boldsymbol{\beta} r_m)\rangle_{p,q} &= 0 \qquad\qquad\quad \text{for all } u_n \in \mathbb{U}_n\end{aligned}$$

with $J_{W^q}(\cdot)$ defined in (2.7).

The solution $u_n \in \mathbb{U}_n$ of (4.2) is exactly the residual minimizer of (4.1), while $r_m \in \mathbb{V}_m$ is a *representative of the discrete residual*, i.e., $J_{\mathbb{V}}(r_m) = f - Bu_n$ in $(\mathbb{V}_m)^*$.

Remark 4.1 (Solving the Quasi-Linear System). The DDMRes method leads to a discrete quasi-linear mixed system (4.2) with q-structure monotone nonlinearity, which, for p (and q) moderately close to 2, can be solved effectively with, e.g., Newton's or Picard's method. For p close to 1 (hence q much larger than 2) or for p much larger than 2 (hence q close to 1), the nonlinear problem becomes more tedious to solve, and we have resorted to continuation techniques (with respect to p) or a descent method for the equivalent constrained-minimization formulation

$$r_m = \operatorname*{arg\,min}_{v_m \in \mathbb{V}_m \cap (B\mathbb{U}_n)^\perp} \Big(\frac{1}{2}\|v_m\|_{\mathbb{V}}^2 - \langle f, v_m\rangle_{\mathbb{V}^*,\mathbb{V}} \Big), \tag{4.3}$$

where $u_n \in \mathbb{U}_n$ is the Lagrange multiplier of (4.3). We refer to [33] for the equivalences between (4.1), (4.2) and (4.3).

The well-posedness of the discrete method (4.2) relies on the well-posedness of the continuous problem (3.1) (see Theorem A), together with the following *Fortin* assumption.

Assumption 4.2 (Fortin Condition). Let $B\colon \mathbb{U} \to \mathbb{V}^*$ be a bounded linear operator, and let $\{(\mathbb{U}_n, \mathbb{V}_m)\}$ be a family of *discrete* subspace pairs, where $\mathbb{U}_n \subset \mathbb{U}$ and $\mathbb{V}_m \subset \mathbb{V}$. For each pair $(\mathbb{U}_n, \mathbb{V}_m)$ in this family, there exists an operator $\Pi_{n,m}\colon \mathbb{V} \to \mathbb{V}_m$ and constants $C_\Pi > 0$ (independent of n and m) such that the following conditions are satisfied:

$$\|\Pi_{n,m} v\|_{\mathbb{V}} \le C_\Pi \|v\|_{\mathbb{V}} \quad \text{for all } v \in \mathbb{V}, \tag{4.4a}$$

$$\langle Bw_n, v - \Pi_{n,m} v\rangle_{\mathbb{V}^*,\mathbb{V}} = 0 \quad \text{for all } w_n \in \mathbb{U}_n \text{ and all } v \in \mathbb{V}. \tag{4.4b}$$

For simplicity, we write Π instead of $\Pi_{n,m}$.

Theorem B (Weak Advection-Reaction: DDMRes Method). *Under the conditions of Theorem A, let the pair* $(\mathbb{U}_n, \mathbb{V}_m)$ *satisfy the (Fortin) Assumption 4.2 with operator B given by* (3.2).

(i) *There exists a unique solution* (r_m, u_n) *to* (4.2)*, which satisfies the a priori bounds*

$$|||r_m|||_{q,\boldsymbol{\beta}} \le \|f\|_{\mathbb{V}^*} \quad \textit{and} \quad \|u_n\|_p \le \widetilde{C}\|f\|_{\mathbb{V}^*}$$

with $\widetilde{C} := C_\Pi(1 + C_{\mathrm{AO}}(\mathbb{V}))/\gamma_B$.

(ii) *Moreover, we have the a priori error estimate*

$$\|u - u_n\|_p \le C \inf_{w_n \in \mathbb{U}_n} \|u - w_n\|_p,$$

where $C = \min\{2^{\left|\frac{2}{p}-1\right|} M_\mu \widetilde{C}, 1 + M_\mu \widetilde{C}\}$.

The constants involved are C_Π*, which is given in Assumption 4.2, the boundedness constant* M_μ *given in Remark 3.2, the stability constant* γ_B *given in* (3.3) *(see also statement* (ii*) *in Theorem A) and the geometrical constant* $C_{\mathrm{AO}}(\mathbb{V})$ *(for* $\mathbb{V} = W^q_{0,+}(\boldsymbol{\beta}; \Omega)$*) defined in* [33, Definition 2.14].

Proof. Statement (i) directly follows from [33, Theorem 4.B] applied to the current situation, while statement (ii) follows from [33, Theorem 4.D], which can be applied since the spaces $\mathbb{U} = L^p(\Omega)$ and $\mathbb{V} = W^q_{0,+}(\boldsymbol{\beta}; \Omega)$ are strictly convex and reflexive for $1 < p, q < +\infty$, as well as the dual spaces $\mathbb{U}^*$ and $\mathbb{V}^*$. The factor $2^{\left|\frac{2}{p}-1\right|}$ is the value of the Banach–Mazur constant $C_{\mathrm{BM}}(\mathbb{U})$ (appearing in [33, Theorem 4.D]) for $\mathbb{U} = L^p(\Omega)$; see [36, Section 5]. □

Remark 4.3 (Finite Elements). Theorem B implies that the convergence of the method is *quasi-optimal* in $L^p(\Omega)$ for finite element subspaces $\mathbb{U}_n \equiv \mathbb{U}_h$, provided C_Π is uniformly bounded. For example, on a sequence

of approximation subspaces $\{\mathbb{P}^k(\mathcal{T}_h)\}_{h>0}$ of piecewise polynomials of fixed degree k on quasi-uniform shape-regular meshes $\mathcal{T}_h$ with mesh-size parameter h, well-known best-approximation estimates (see, e.g., [6, Section 4.4], [22, Section 1.5] and [25]) imply

$$\|u - u_n\|_p \lesssim \inf_{w_h \in \mathbb{U}_h} \|u - w_h\|_p \lesssim h^s |u|_{W^{s,p}(\Omega)} \quad \text{for } 0 \le s \le k+1,$$

where $|\cdot|_{W^{s,p}(\Omega)}$ denotes a standard semi-norm of $W^{s,p}(\Omega)$ (e.g., of Sobolev–Slobodeckij type). For a relevant regularity result in $W^{1,\rho}(\Omega)$, with $\rho \ge 2$, see Girault and Tartar [26] (see also [34]).

5 Applications

In this section, we apply the general discrete method (4.2) to particular choices of discrete subspace pairs $(\mathbb{U}_n, \mathbb{V}_m)$ involving low-order finite-element spaces.

For simplicity, throughout this section, $\Omega \subset \mathbb{R}^d$ will be a *polyhedral* domain, and $\mathcal{T}_n$ will denote a finite partition of Ω, i.e., $\mathcal{T}_n = \{T\}$ consists of a finite number of non-overlapping elements T for which $\overline{\Omega} = \bigcup_{T\in\mathcal{T}_n} \overline{T}$.

5.1 The Pair $\mathbb{P}^1_{\text{cont}}(\mathcal{T}_n)$, $\mathbb{P}^k_{\text{cont}}(\mathcal{T}_n)$: Eliminating the Gibbs Phenomena

By first considering *continuous* finite elements for $\mathbb{U}_n$, we briefly illustrate how the discrete method (4.2) eliminates the well-known *Gibbs phenomena* when approaching discontinuous solutions. For simplicity, consider the advection-reaction problem (2.1) with $\Omega \equiv (-1, 1) \subset \mathbb{R}$, $\beta = 1$, $\mu = 0$, $g = -1$, and let the source $f_\circ$ be $2\delta_0$, where δ_0 is the *Dirac delta* at $x = 0$, i.e.,

$$\begin{aligned} u'(x) &= 2\delta_0(x) \quad \text{for all } x \in (-1, 1), \\ u(-1) &= -1. \end{aligned} \tag{5.1}$$

Notice that the exact solution of (5.1) corresponds to the sign of x,

$$u(x) = \operatorname{sign}(x) := \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x > 0. \end{cases}$$

We endow $\mathbb{V} = W^q_{0,+}(\boldsymbol{\beta}; \Omega) = W^{1,q}_{0,\{1\}}(\Omega)$ with the norm $\|\cdot\|_{\mathbb{V}} = \|(\cdot)'\|_q$, which simplifies the duality map (2.7) to a normalized *q-Laplace operator*

$$\langle J_{\mathbb{V}}(r), v\rangle_{\mathbb{V}^*,\mathbb{V}} = \langle J_q(r'), v'\rangle_{p,q} = \|r'\|_q^{2-q} \langle |r'|^{q-1} \operatorname{sign}(r'), v'\rangle_{p,q}.$$

Moreover, in this setting, it is not difficult to show that residual minimization in $[W^{1,q}_{0,\{1\}}(\Omega)]^*$ now coincides with finding the *best L^p-approximation* to $\operatorname{sign}(x)$. The Gibbs phenomena for best L^p-approximation was studied analytically by Saff and Tashev [35] when using continuous piecewise-linear approximations on n equal-sized elements. They clarified that the overshoot next to a discontinuity *remains* as $n \to \infty$ whenever $p > 1$; however remarkably, the *overshoot tends to zero as* $p \to 1^+$.

To illustrate these findings, we plot in Figure 1 the best L^2-approximation using continuous piecewise-linears for various mesh-size parameters h. Clearly, the overshoots remain present, signifying the Gibbs phenomenon. Next, in Figure 2, we plot the solution to ideal residual minimization (i.e., in the so-called *ideal* case where $\mathbb{V}_m = \mathbb{V}$) on a fixed mesh consisting of nine elements, for different values of $p > 1$.[2] In Figure 2, we also plot the corresponding ideal residual $r'(x)$ as defined by the mixed formulation (4.2) in the case where $\mathbb{V}_m = \mathbb{V}$. It can be shown that, in this ideal 1-D situation, $r' = \|u_n - u\|_p^{2-p} |u_n - u|^{p-1} \operatorname{sign}(u_n - u)$. The plots in Figure 2 clearly illustrate the elimination of the Gibbs phenomenon as $p \to 1^+$.

2 These plots were obtained by using the analytical results by Saff and Tashev for the L^p-approximation of $\operatorname{sign}(x)$.

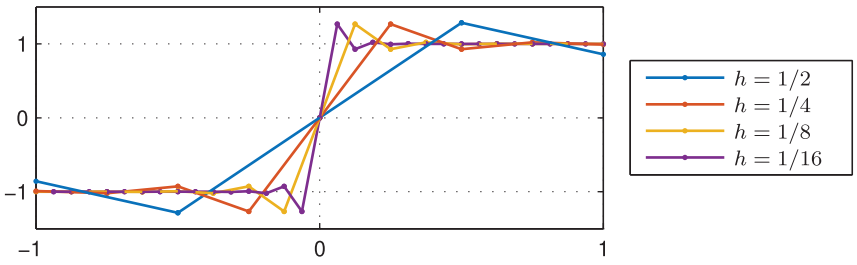
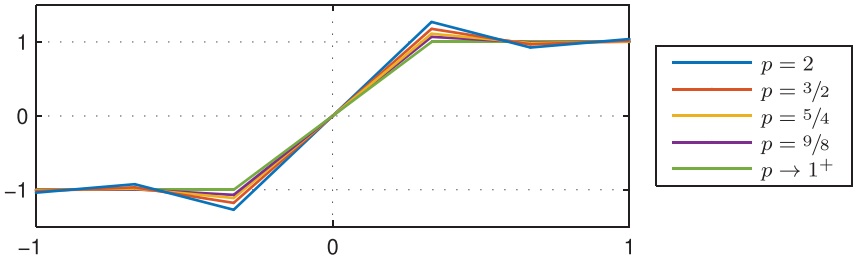
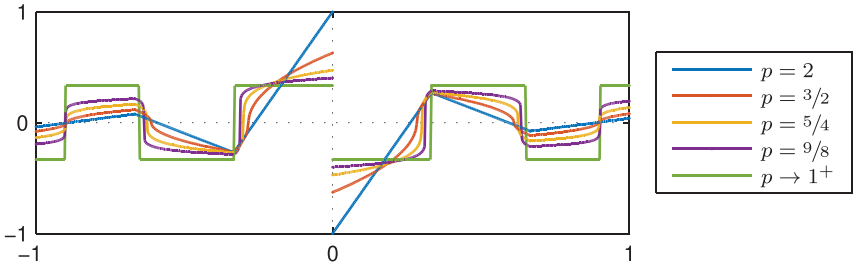


Figure 1: The best L^2-approximation of $u(x) = \text{sign}(x)$ displays the Gibbs phenomenon: the overshoot next to the discontinuity persists on any mesh.



(a) $u_n(x)$



(b) $r'(x)$

Figure 2: Vanishing Gibbs phenomena as $p \to 1^+$ for approximations to the discontinuous solution $u(x) = \text{sign}(x)$ given by ideal residual-minimization of weak advection.

We note that the elimination of the Gibbs phenomena was also observed for residual minimization of the *strong form* of the advection-reaction residual in $L^1(\Omega)$ (see [30]), the explanation of which remains somewhat elusive.

Next, consider DDMRes method (4.2) with the discrete space pair $(\mathbb{U}_n, \mathbb{V}_m)$ defined by

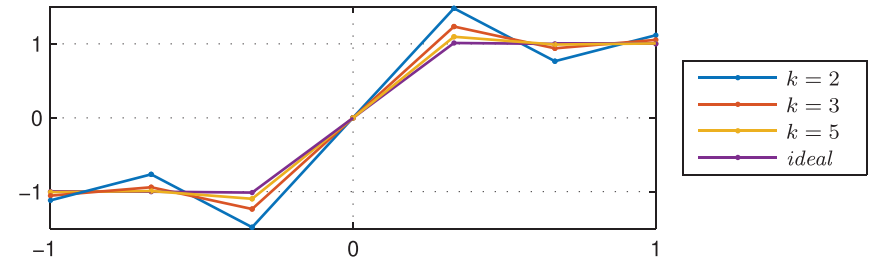
$$\mathbb{U}_n \subseteq \mathbb{P}^1_{\text{cont}}(\mathcal{T}_n) := \{w_n \in \mathcal{C}^0[-1,1] : w_n|_T \in \mathbb{P}^1(T) \text{ for all } T \in \mathcal{T}_n\},$$
$$\mathbb{V}_m \supseteq \mathbb{P}^k_{\text{cont},0,\{1\}}(\mathcal{T}_n) := \{\phi_m \in \mathcal{C}^0[-1,1] : \phi_m|_T \in \mathbb{P}^k(T) \text{ for all } T \in \mathcal{T}_n \text{ and } \phi_m(1) = 0\},$$

where k is the polynomial degree of the test space and $\mathcal{T}_n$ is any partition of the interval $(-1, 1)$.

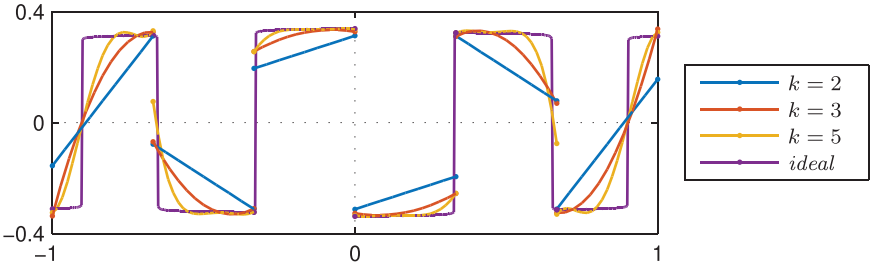
Proposition 5.1 (1-D Advection: Compatible Pair). *Let* $\mathbb{U} = L^p(-1,1)$, $\mathbb{V} = W^{1,q}_{0,\{1\}}(-1,1)$, *and let* $(\mathbb{U}_n, \mathbb{V}_m)$ *be defined as above. If* $k \geq 2$, *then the Fortin condition (Assumption 4.2) holds for the operator* $B : \mathbb{U} \to \mathbb{V}^*$ *defined by* $\langle Bw, v\rangle_{\mathbb{V}^*,\mathbb{V}} = -\int_{-1}^{1} wv'$.

Proof. See Section A.3. □

Figure 3 displays numerical results obtained using DDMRes method (4.2) with the above discrete spaces and for $p = 1.01$ (hence $q = 101$). The nonlinear system was solved as explained in Remark 4.1. We plot u_n and r'_m for various test-space degrees $k \geq 2$. While the method is stable for any $k \geq 2$ (owing to Proposition 5.1), there is no reason for the DDMRes method to directly inherit any qualitative feature of the exact residual



(a) $u_n(x)$



(b) $r'_m(x)$

Figure 3: Approximations to $u(x) = \mathrm{sign}(x)$ given by the DDMRes method for weak advection with trial space $\mathbb{U} = L^p(\Omega)$ and $p = 1.01$. The discrete space $\mathbb{U}_n$ consists of continuous piecewise-linears and $\mathbb{V}_m$ of continuous piecewise-polynomials of degree k. The line *ideal* corresponds to the case that $r \in \mathbb{V}_m$ or $\mathbb{V}_m = \mathbb{V}$ (or $k \to \infty$).

minimization (i.e., $\mathbb{V}_m = \mathbb{V}$). Indeed, overshoot is present for small k. However, results are qualitatively converging once the test space $\mathbb{V}_m$ starts resolving the ideal r. This is expected by [33, Proposition 4.2], which states that the ideal u_n is obtained if the ideal r happens to be in $\mathbb{V}_m$. The lines indicated by "ideal" in Figure 3 correspond to the case that r is fully resolved.

Interestingly, the results seem to indicate that different values of k in each element would be needed for efficiently addressing Gibbs phenomena. This is reminiscent of the idea of *adaptive stabilization* [14], in which, for a given $\mathbb{U}_n$, a sufficiently large test space $\mathbb{V}_m$ is found in an adaptive manner so as to achieve stability.

5.2 The Pair $\mathbb{P}^0(\mathcal{T}_n)$, $B^{-*}(\mathbb{P}^0(\mathcal{T}_n))$: An Optimal Compatible Pair

In the remainder of Section 5, we consider for $\mathbb{U}_n$ piecewise-constant functions on mesh partitions $\mathcal{T}_n$ of $\Omega \subset \mathbb{R}^d$, i.e.,

$$\mathbb{U}_n \subseteq \mathbb{P}^0(\mathcal{T}_n) := \{w_n \in L^\infty(\Omega) : w_n|_T \in \mathbb{P}^0(T) \text{ for all } T \in \mathcal{T}_n\} \subset \mathbb{U}. \tag{5.2a}$$

For the discrete test space $\mathbb{V}_m \subset \mathbb{V}$, we assume that it includes the *optimal space* $\mathbb{S}_n := B^{-*}(\mathbb{P}^0(\mathcal{T}_n)) \subset \mathbb{V}$, i.e.,

$$\mathbb{V}_m \supseteq \mathbb{S}_n := B^{-*}(\mathbb{P}^0(\mathcal{T}_n)) = \{\phi_n \in \mathbb{V} : B^*\phi_n = \chi_n \text{ for some } \chi_n \in \mathbb{P}^0(\mathcal{T}_n)\}. \tag{5.2b}$$

Note that $\dim \mathbb{U}_n \le \dim \mathbb{P}^0(\mathcal{T}_n) = \dim \mathbb{S}_n \le \dim \mathbb{V}_m$.

Without any further assumptions, the following striking result show that this pair satisfies the Fortin condition. Its proof hinges on the fact that the L^p duality map of any $w_n \in \mathbb{P}^0(\mathcal{T}_n)$ is also in $\mathbb{P}^0(\mathcal{T}_n)$.

Proposition 5.2 (Weak Advection-Reaction: Compatible Pair). *Let $\mathbb{U} = L^p(\Omega)$, $\mathbb{V} = W^q_{0,+}(\boldsymbol{\beta}; \Omega)$, and let the discrete pair $(\mathbb{U}_n, \mathbb{V}_m)$ be defined as in* (5.2a) *and* (5.2b). *Then the Fortin condition (Assumption 4.2) holds for B defined in* (3.2), *with $C_\Pi = M_\mu/\gamma_B$, where M_μ is the continuity constant of B (see Remark 3.2) and γ_B the bounded-below constant of B (see Theorem A).*

Proof. In view of the equivalence between the discrete inf-sup condition and the Fortin condition (see Ern and Guermond [24]), we prove this proposition by directly establishing the discrete inf-sup condition.

Let $w_n \in \mathbb{U}_n \subseteq \mathbb{P}^0(\mathcal{T}_n)$. Then

$$\sup_{v_m \in \mathbb{V}_m} \frac{\langle Bw_n, v_m\rangle_{\mathbb{V}^*,\mathbb{V}}}{\|v_m\|_{\mathbb{V}}} \geq \sup_{\phi_n \in \mathbb{S}_n} \frac{\langle Bw_n, \phi_n\rangle_{\mathbb{V}^*,\mathbb{V}}}{\|\phi_n\|_{\mathbb{V}}} = \sup_{\chi_n \in \mathbb{P}^0(\mathcal{T}_n)} \frac{\langle w_n, \chi_n\rangle_{p,q}}{\|B^{-*}\chi_n\|_{\mathbb{V}}}.$$

Let $J_p(w_n) := \|w_n\|_p^{2-p}|w_n|^{p-1}\operatorname{sign}(w_n)$ denote the L^p duality map of w_n, and notice that it is also in $\mathbb{P}^0(\mathcal{T}_n)$. Furthermore, we have the duality-map property $\langle w_n, J_p(w_n)\rangle_{p,q} = \|w_n\|_p\|J_p(w_n)\|_q$. Therefore,

$$\sup_{\chi_n \in \mathbb{P}^0(\mathcal{T}_n)} \frac{\langle w_n, \chi_n\rangle_{p,q}}{\|B^{-*}\chi_n\|_{\mathbb{V}}} \geq \frac{\langle w_n, J_p(w_n)\rangle_{p,q}}{\|B^{-*}J_p(w_n)\|_{\mathbb{V}}} = \frac{\|w_n\|_p\|J_p(w_n)\|_q}{\|B^{-*}J_p(w_n)\|_{\mathbb{V}}} \geq \gamma_B\|w_n\|_p,$$

where, in the last step, we used that $\|B^{-*}\chi\|_{\mathbb{V}} \leq \gamma_B^{-1}\|\chi\|_q$ for all $\chi \in L^q(\Omega)$ (this is nothing but the dual counterpart of Theorem A). Finally, [24, Theorem 1] implies the existence of a Fortin operator $\Pi\colon \mathbb{V} \to \mathbb{V}_m$ with $C_\Pi = M_\mu/\gamma_B$. □

Remark 5.3 (Petrov–Galerkin Method). If $(\mathbb{U}_n, \mathbb{V}_m) \equiv (\mathbb{P}^0(\mathcal{T}_n), \mathbb{S}_n)$, then $\dim \mathbb{U}_n = \dim \mathbb{V}_m$. Then Proposition 5.2 together with (4.2b) implies that $r_m = 0$. Thus we obtain from (4.2a) that the approximation u_n satisfies the Petrov–Galerkin statement (cf. [33, Section 5])

$$\langle Bu_n, v_n\rangle_{\mathbb{V}^*,\mathbb{V}} = \langle f, v_n\rangle_{\mathbb{V}^*,\mathbb{V}} \quad \text{for all } v_n \in \mathbb{S}_n. \tag{5.2}$$

Remark 5.4 (Cell Average). If $(\mathbb{U}_n, \mathbb{V}_m) \equiv (\mathbb{P}^0(\mathcal{T}_n), \mathbb{S}_n)$, the approximation u_n is in fact the element average of the exact solution u, i.e.,

$$u_n|_T = |T|^{-1}\int_T u \quad \text{for all } T \in \mathcal{T}_n. \tag{5.3}$$

To prove (5.3), note that (5.2) can be written as

$$\langle u_n, B^*v_n\rangle = \langle u, B^*v_n\rangle \quad \text{for all } v_n \in \mathbb{S}_n.$$

Let χ_T be the characteristic function of the element T. Then the test function $v_T = B^{-*}\chi_T$ determines $u_n|_T$. Indeed,

$$|T|u_n|_T = \langle u_n, \chi_T\rangle_{p,q} = \langle u_n, B^*v_T\rangle_{p,q} = \langle u, B^*v_T\rangle_{p,q} = \langle u, \chi_T\rangle_{p,q} = \int_T u.$$

Remark 5.5 (Quasi-Uniform Meshes). In the case that $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ where the partitions $\{\mathcal{T}_n\}$ are quasi-uniform shape-regular meshes with mesh-size parameter h, the following a priori error estimate is immediate (apply Remark 4.3 with $k = 0$), provided that $u \in W^{s,p}(\Omega)$ for $0 \leq s \leq 1$:

$$\|u - u_n\|_p \lesssim h^s|u|_{W^{s,p}(\Omega)}.$$

Example 5.6 (Quasi-Optimality: Solution with Jump Discontinuity). To illustrate the convergence of approximations given by the compatible pair $(\mathbb{U}_n, \mathbb{V}_m) \equiv (\mathbb{P}^0(\mathcal{T}_n), \mathbb{S}_n)$ for $\Omega \equiv (0, 1)$ on uniform meshes using $n = 2, 4, 8, \ldots$ elements of size $h = \frac{1}{n}$, consider the following exact solution with jump discontinuity (never aligned with the mesh):[3]

$$u(x) = \operatorname{sign}\Big(x - \frac{\sqrt{2}}{2}\Big) \quad \text{for } x \in (0, 1).$$

It can be shown (e.g., by computing the Sobolev–Slobodeckij norm) that $u \in W^{s,p}(0, 1)$ for any $0 < s < \frac{1}{p}$, but not $s = \frac{1}{p}$. While approximations u_n can be computed using (4.2) or (5.3), in this case, u_n can be obtained simply using the cell average (5.3) since the exact solution is known. Figure 4 shows the convergence of $\|u - u_n\|_p$ with respect to h for various p. The observed convergence behavior, as anticipated in Remark 5.5, is indeed close to $O(h^{1/p})$.

Example 5.7 (Basis for Optimal Test Space). Let us illustrate the discrete test space $\mathbb{S}_n$ in 1-D for the particular case where the (scalar-valued) advection $\beta(x)$ is space-dependent and $\mu \equiv 0$. Let $\Omega = (0, 1)$, and let β be a strictly decreasing and positive function such that $\beta'(x)$ is bounded away from zero (hence Assumption 2.1

3 The approximations are given by (5.3) or can be obtained by solving the nonlinear discrete problem (see Remark 4.1).

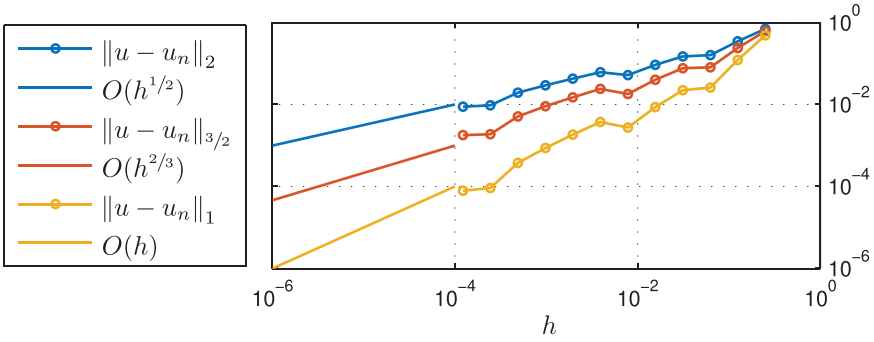
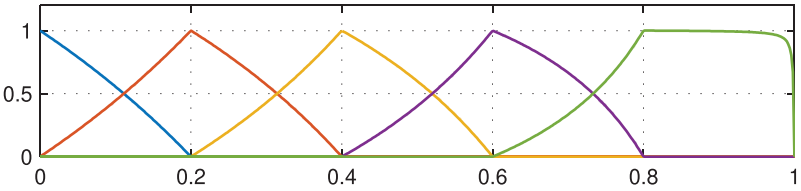
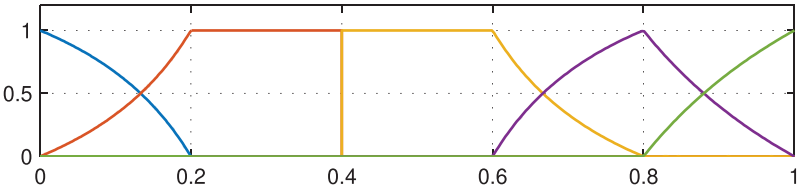


Figure 4: Approximating an advection-reaction problem with a discontinuous solution using the optimal pair $(\mathbb{P}^0(\mathcal{T}_n), \mathbb{S}_n)$: the convergence in $\|u-u_n\|_p$ is close to $O(h^{1/p})$, which is optimal for near-best approximations.



(a) $\beta(x) = 1.001 - x$



(b) $\beta(x) = \frac{2}{5} - x$

Figure 5: Basis for the optimal test space $\mathbb{S}_n$ that is compatible with $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$, in the case of two different space-dependent advection fields $\beta(x)$ corresponding to (a) left-sided inflow and (b) two-sided inflow, respectively.

is valid). The space $\mathbb{V}$ is given by

$$\mathbb{V} = \{v \in L^q(0,1) : (\beta v)' \in L^q(0,1) \text{ and } v(1) = 0\}.$$

Let $0 = x_1 < x_1 < \cdots < x_{n+1} = 1$ be a partition of Ω, and define $\mathcal{T}_n = \{T_j\}$, where $T_j = (x_{j-1}, x_j)$. Let χ_{T_j} be the characteristic function of T_j and $h_j = |T_j|$. The discrete test space $\mathbb{S}_n$ is defined as the span of the functions $v_j \in \mathbb{V}$ such that $-(\beta v_j)' = \chi_{T_j}$, which, upon integrating over the interval $[x, 1]$, gives

$$v_j(x) = \begin{cases} h_j/\beta(x) & \text{if } x \le x_{j-1}, \\ (x_j - x)/\beta(x) & \text{if } x \in T_j, \\ 0 & \text{if } x \ge x_j. \end{cases}$$

Moreover, we can combine them in order to produce the local, nodal basis functions

$$\tilde{v}_1(x) = \beta(x_1)\frac{v_1(x)}{h_1} \quad \text{and} \quad \tilde{v}_j(x) = \beta(x_j)\Big(\frac{v_j(x)}{h_j} - \frac{v_{j-1}(x)}{h_{j-1}}\Big), \quad j \ge 2.$$

See Figure 5 (a) for an illustration of these basis functions with $h_j = 0.2$ for all j and $\beta(x) = 1.001 - x$.

Another interesting example is when we have two inflows, each on one side of the interval $\Omega = (0,1)$. This is possible by means of a strictly decreasing $\beta(x)$ such that $\beta(0) > 0$ and $\beta(1) < 0$, and such that $\beta'(x)$ is bounded away from zero. The solution $u \in L^p(\Omega)$ of problem (3.1) may be singular at the point $\tilde{x} \in \Omega$ for which $\beta(\tilde{x}) = 0$, even for smooth right-hand sides. The test functions computed by solving $-(\beta v_j)' = \chi_{T_j}$ may be discontinuous when $\tilde{x}$ matches one of the mesh points. This is illustrated in Figure 5 (b) for $\beta(x) = \frac{2}{5} - x$.

Example 5.8 (A Practical Alternative to $\mathbb{S}_n$). In practise, it may not be feasible to explicitly compute a basis for $\mathbb{S}_n$. Practical alternatives consist of, for example, continuous piecewise polynomials of sufficiently high degree k on $\mathcal{T}_n$, or continuous piecewise linear polynomials on $\mathsf{Refine}_\ell(\mathcal{T}_n)$, which is the submesh obtained from the original mesh $\mathcal{T}_n$ by performing ℓ uniform refinements of all elements (see [7] for a similar alternative in a DPG setting).

To illustrate the latter alternative for the DDMRes method, consider the domain $\Omega = (0, 1)$, coefficients $\beta(x) = 1 - 12x$ and $\mu(x) = -4$, source $f_\circ(x) = 0$ and inflow data g such that the exact solution is $u(x) = |1 - 12x|^{-\frac{1}{3}}$ for all $x \in \Omega \setminus \{\frac{1}{12}\}$. Note that u has a singularity and that $u \in L^r(\Omega)$ for any $1 \le r < 3$, but not for $r \ge 3$.

In method (4.2), we take $p = q = 2$ (the Hilbert-space case), $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^1_{\text{cont}}(\mathsf{Refine}_\ell(\mathcal{T}_n))$, where $\mathcal{T}_n$ is a mesh of uniform elements of size $h = \frac{1}{n}$ and $\mathsf{Refine}_\ell(\mathcal{T}_n)$ is an ℓ-refined submesh with uniform elements of size $h_\ell = h/(2^\ell)$.

Figure 6 plots the convergence of the $\|u - u_n\|_2$ versus h for $\ell = 1$, 2 and 4 (error plots are actually similar for all $\ell \ge 1$). We note that $\ell = 0$ is, in general, not sufficiently rich, as it leads to a singular matrix for $h = 1/2$, while the results for $\ell \ge 1$ did not show any instabilities. To anticipate the rate of convergence, note the Sobolev embedding result $W^{s,2}(\Omega) \subset L^r(\Omega)$ for $s \ge \frac{1}{2} - \frac{1}{r}$ and $r \ge 2$. Therefore, one expects a convergence of $O(h^s)$ with $s = \frac{1}{6}$, which is indeed consistent with the numerical observation in Figure 6. The oscillations are caused by the singularity location ($x = \frac{1}{12}$) being closer to the left or right element edge depending on h.

To investigate for a fixed mesh with $h = \frac{1}{16}$ the convergence of the obtained approximations u_n with respect to ℓ, we consider $\beta(x) = 2 - x$, $\mu(x) = 0$ and exact solution $u(x) = 1 + 2x$ for $x \in \Omega$. Figure 7 plots the error $\|u_{n|\infty} - u_{n|\ell}\|$ with respect to $h_\ell = h/(2^\ell)$, where $u_{n|\infty}$ denotes the ideal approximation ($\mathbb{V}_m = \mathbb{V}$). For this error, we observe a rate of convergence $O(h_\ell^2)$.

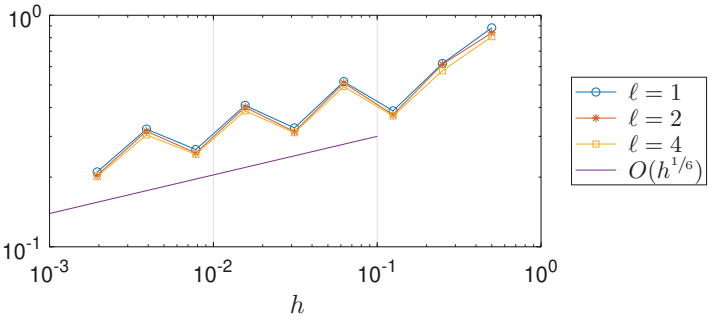


Figure 6: Approximating a singular solution $u(x) = |1 - 12x|^{-\frac{1}{3}}$ for $x \in (0, 1) \setminus \{\frac{1}{12}\}$ to the advection-reaction problem with the DDMRes method using $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^1_{\text{cont}}(\mathsf{Refine}_\ell(\mathcal{T}_n))$. The convergence is close to $O(h^{\frac{1}{6}})$, which is optimal for near-best approximations.

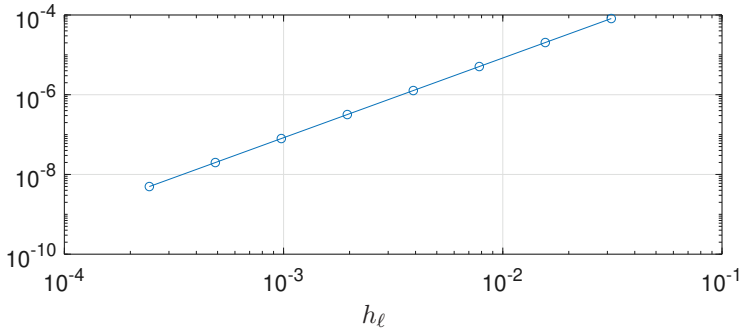


Figure 7: Convergence of approximation $u_{n|\ell}$ toward $u_{n|\infty}$ on a fixed mesh $\mathcal{T}_n$ for a smooth exact solution, where $u_{n|\ell}$ denotes the approximation obtained by the DDMRes method using $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^1_{\text{cont}}(\mathsf{Refine}_\ell(\mathcal{T}_n))$. The observed convergence is $O(h_\ell^2)$.

5.3 The Pair $\mathbb{P}^0(\mathcal{T}_n)$, $\mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$: An Optimal Pair in Special Situations

As a last application, we consider a special multi-dimensional situation such that the optimal test space $\mathbb{S}_n$ defined in (5.2b) reduces to a convenient finite element space. We focus on a 2-D setting and assume $\Omega \subset \mathbb{R}^2$ is polygonal and $\mathcal{T}_n$ is a simplicial mesh (triangulation) of Ω. Let $\mathcal{F}_n = \{F\}$ denote all mesh interior faces.[4] Assume that $\mu \equiv 0$, $\operatorname{div} \boldsymbol{\beta} \equiv 0$ and that the hypothesis of Assumption 2.8 is fulfilled. Assume additionally that $\boldsymbol{\beta}$ is piecewise constant on some partition of Ω, and let the mesh $\mathcal{T}_n$ be compatible with this partition, i.e.,

$$\begin{aligned} &\boldsymbol{\beta}|_T \in \mathbb{P}^0(T) \times \mathbb{P}^0(T) \quad \text{for all } T \in \mathcal{T}_n, \\ &[\![\boldsymbol{\beta} \cdot \boldsymbol{n}_F]\!]_F = 0 \qquad\qquad \text{for all } F \in \mathcal{F}_n. \end{aligned}$$

where $[\![\cdot]\!] = (\cdot)_+ - (\cdot)_-$ denotes the jump. Finally, assume that the mesh is *flow-aligned* in the sense that each triangle $T \in \mathcal{T}_n$ has exactly one tangential-flow face $F \subset \partial T$ for which $\boldsymbol{\beta} \cdot \boldsymbol{n}_T = 0$ on F. Necessarily, the other two faces of T correspond to in- and out-flow on which $\boldsymbol{\beta}|_T \cdot \boldsymbol{n}_T < 0$ and $\boldsymbol{\beta}|_T \cdot \boldsymbol{n}_T > 0$, respectively.

The main result for this special situation is the following characterization of $\mathbb{S}_n$.

Proposition 5.9 (Optimal Space $\mathbb{S}_n$: Flow-Aligned Case). *Under the above assumptions,*

$$\mathbb{S}_n = \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n) := \{\phi_n \in \mathbb{V} = W^q_{0,+}(\boldsymbol{\beta}; \Omega) : \phi_n|_T \in \mathbb{P}^1(T) \text{ for all } T \in \mathcal{T}_n\}.$$

Note that $\mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ consists of $W^q_{0,+}(\boldsymbol{\beta}; \Omega)$-conforming, piecewise-linear functions, which can be discontinuous across tangential-flow faces, but must be continuous across the other faces. Furthermore, they are zero on $\partial\Omega_+$.

Proof. The proof follows upon demonstrating that $B^* \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n) = \mathbb{P}^0(\mathcal{T}_n)$. First note (under the above assumptions) that $B^* = -\boldsymbol{\beta} \cdot \nabla_n$, where ∇_n is the element-wise (or broken) gradient, i.e., $(\nabla_n \phi)|_T = \nabla(\phi|_T)$ for all $T \in \mathcal{T}_n$. Since functions in $\mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ are element-wise linear, we thus have $B^* \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n) \subset \mathbb{P}^0(\mathcal{T}_n)$.

We next show that $\mathbb{P}^0(\mathcal{T}_n) \subset B^* \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$. Note that $\mathbb{P}^0(\mathcal{T}_n) = \operatorname{Span}\{\chi_T, T \in \mathcal{T}_n\}$, where χ_T is the characteristic function for T. Let ϕ_T be the unique solution in $\mathbb{V}$ such that $B^* \phi_T = \chi_T$. The Ω-filling assumption (see Assumption 2.8) guarantees that $\boldsymbol{\beta} \neq \boldsymbol{0}$ a.e. in Ω (otherwise, we would have $-\boldsymbol{\beta} \cdot \nabla z_\pm = 0$ in some element, contradicting (2.9)). Thus, for a.e. $x \in \Omega$, consider the polygonal path $\Gamma(x) \subset \overline{\Omega}$ that starts from x and moves along the advection field $\boldsymbol{\beta}$. By the Ω-filling assumption, the path $\Gamma(x)$ has to end in some point on the outflow boundary $\partial\Omega_+$ (otherwise, it will stay forever within Ω, contradicting the existence of a bounded function $z_\pm \in W^\infty(\boldsymbol{\beta}; \Omega)$ whose absolute value grows linearly along $\Gamma(x)$). Hence we can construct ϕ_T integrating χ_T over the polygonal path $\Gamma(x)$ from $\partial\Omega_+$ to x. By construction, ϕ_T is a piecewise linear polynomial, which can be discontinuous only across $\{F \in \mathcal{F}_n : \boldsymbol{\beta} \cdot \boldsymbol{n}_F = 0\}$. Besides, ϕ_T satisfies the homogeneous boundary condition over $\partial\Omega_+$. Hence $\phi_T \in \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ and $\chi_T \in B^* \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$. □

Remark 5.10 (Petrov–Galerkin Method). In the above situation, the DDMRes with the FE pair

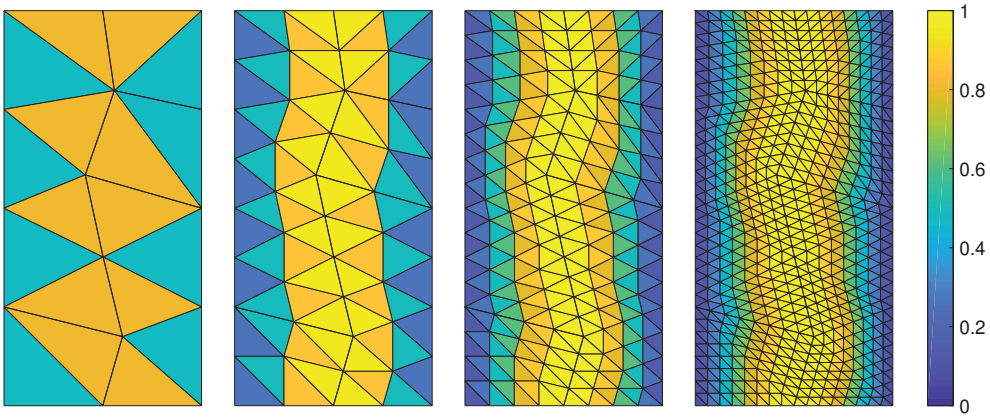
$$(\mathbb{U}_n, \mathbb{V}_m) = (\mathbb{P}^0(\mathcal{T}_n), \mathbb{S}_n) = (\mathbb{P}^0(\mathcal{T}_n), \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n))$$

gives the Petrov-Galerkin method (5.2), which can be written explicitly as (after performing an element-wise integration by parts): find $u_n \in \mathbb{P}^0(\mathcal{T}_n)$ such that

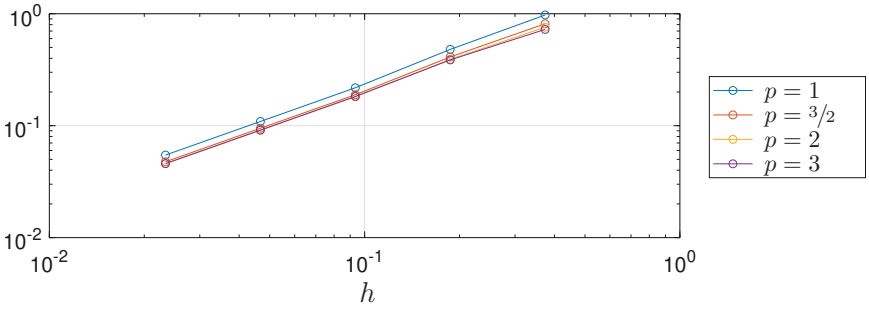
$$\langle B u_n, v_n \rangle_{\mathbb{V}^*, \mathbb{V}} = \int_{\partial\Omega_-} |\boldsymbol{\beta} \cdot \boldsymbol{n}| u_n v_n - \sum_{F \in \mathcal{F}_n} \int_{\partial F} (\boldsymbol{\beta} \cdot \boldsymbol{n}) [\![u_n]\!] v_n = \langle f, v_n \rangle_{\mathbb{V}^*, \mathbb{V}} \tag{5.4}$$

for all $v_n \in \mathbb{S}_n = \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$. While this discrete formulation (5.4) is reminiscent of a lowest-order discontinuous Galerkin (DG) scheme (with, e.g., centered fluxes [21, Section 2.2]), it is however a truly distinct scheme since v_n comes from the conforming *non-broken* space $\mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ (and not from the *broken* space $\mathbb{P}^0(\mathcal{T}_n)$).

4 That is, $\text{length}(F) > 0$, and $F = \partial T_1 \cap \partial T_2$ for distinct T_1 and T_2 in $\mathcal{T}_n$.



(a) Approximations u_n on different (nested) meshes.



(b) The convergence in $\|u - u_n\|_p$ is $O(h)$ (optimal).

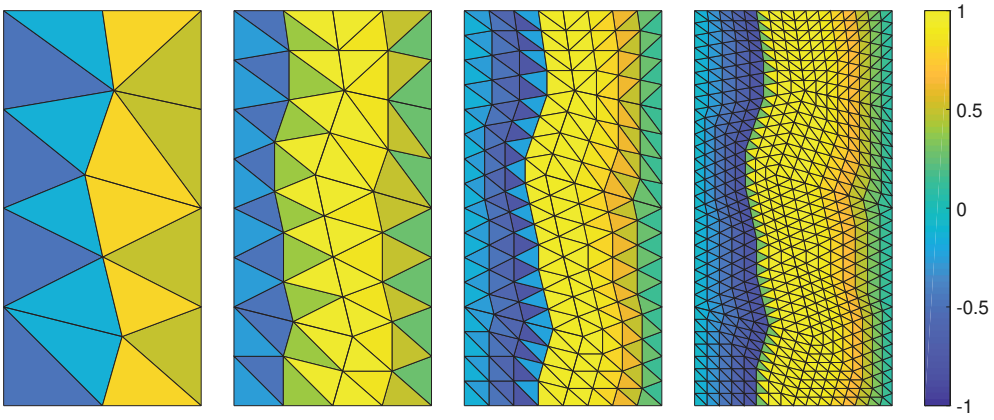
Figure 8: DDMRes approximations using $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ for an incompressible advection problem with special piecewise constant $\boldsymbol{\beta}$ and a *smooth* inflow boundary condition g.

Example 5.11 (2-D Numerical Illustration). To illustrate the above setting with a numerical example, let $\Omega = (0, 1) \times (0, 2) \subset \mathbb{R}^2$, $f_\circ = 0$, and let g be nonzero on the inflow boundary $\partial\Omega_- = \{(x, 0),\ x \in (0, 1)\}$. Let an initial triangulation of the domain be as in Figure 8 (top-left mesh). The advection $\boldsymbol{\beta}$ is such that, for the bottom, left, right and top boundary, we have that $\boldsymbol{\beta} \cdot \boldsymbol{n}$ is -1, 0, 0 and 1, respectively. Next, within each triangle, $\boldsymbol{\beta}$ is some constant vector with a positive vertical component, while satisfying the above requirements (i.e., $[\![\boldsymbol{\beta} \cdot \boldsymbol{n}_F]\!]_F = 0$ on each interior face F, and each triangle has a tangential-flow, in-flow and out-flow face).[5] We computed our approximations by implementing (5.4), using a large number of quadrature points to evaluate the right-hand side $\langle f, v_n\rangle_{\mathbb{V}^*,\mathbb{V}} = \int_{\partial\Omega_-} |\boldsymbol{\beta} \cdot \boldsymbol{n}| g v_n \, \mathrm{d}s$.

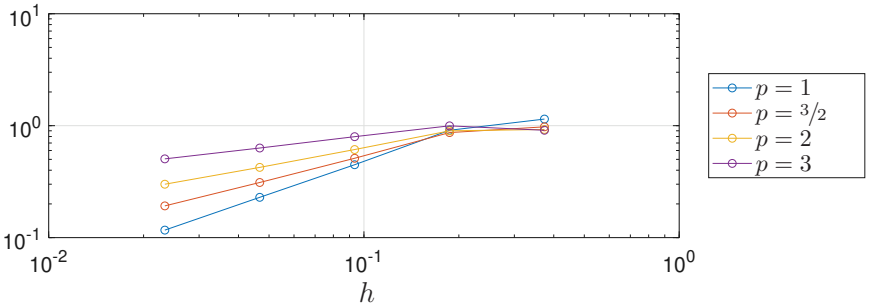
We first consider the smooth inflow boundary condition $g(x, 0) = \sin(\pi x)$ for $x \in (0, 1)$. Figure 8 (a) shows the approximations for u_n obtained on the initial triangulation and three finer meshes. The finer meshes were obtained by uniform refinements of the initial triangulation using so-called *red*-refinement [37, Section 2.1.2] (splitting each triangle into four similar triangles), which preserves the above flow-aligned mesh requirement. The approximations nicely illustrate the cell-average property mentioned in Remark 5.4 (the exact solution is simply found by traversing g along the characteristics). In Figure 8 (b), the convergence of $\|u - u_n\|_p$ is shown to be optimal (rate is $O(h)$) for various values of p.

Figure 9 shows the same results as before, but now for a discontinuous inflow boundary condition $g(x, 0) = \sin(\pi x)\,\mathrm{sign}(x - \frac{1}{3})$ for $x \in (0, 1)$. Again, the method provides a near-best approximation; as anticipated, the observed rate of convergence is $O(h^{1/p})$ (cf. discussion in Example 5.6).

5 For a given mesh, such a $\boldsymbol{\beta}$ can be constructed by traversing through the mesh in an element-by-element fashion, starting at the inflow boundary, and assigning $\boldsymbol{\beta}$ in each element so as to satisfy the requirements.



(a) Approximations u_n on different (nested) meshes.



(b) The convergence in $\|u - u_n\|_p$ is $O(h^{1/p})$, which is optimal for near-best approximations to the discontinuous solution u.

Figure 9: DDMRes approximations using $\mathbb{U}_n = \mathbb{P}^0(\mathcal{T}_n)$ and $\mathbb{V}_m = \mathbb{P}^1_{\text{conf}}(\mathcal{T}_n)$ for an incompressible advection problem with special piecewise constant $\boldsymbol{\beta}$ and a *discontinuous* inflow boundary condition g.

A Proofs of the Main Results

A.1 Proof of Lemma 2.6

Recall that $\sigma = \frac{\rho}{\rho-1}$. Given $g \in L^\rho(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$, we define the linear functional $F_g\colon W^\sigma(\boldsymbol{\beta};\Omega) \to \mathbb{R}$ by

$$\langle F_g, v\rangle_{(W^\sigma)^*,W^\sigma} = \int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n})gv \quad \text{for all } v \in W^\sigma(\boldsymbol{\beta};\Omega).$$

The fact that $F_g \in (W^\sigma(\boldsymbol{\beta};\Omega))^*$ is well-defined is because traces of $W^\sigma(\boldsymbol{\beta};\Omega)$ are well-defined as a subspace of $L^\sigma(|\boldsymbol{\beta}\cdot\boldsymbol{n}|;\partial\Omega)$ (see Remark 2.5). Since $\sigma \in (1,+\infty)$, by surjectivity of the duality map

$$J_{W^\sigma}\colon W^\sigma(\boldsymbol{\beta};\Omega) \to (W^\sigma(\boldsymbol{\beta};\Omega))^*$$

(see Definition 2.4), there is a $v_g \in W^\sigma(\boldsymbol{\beta};\Omega)$ such that

$$J_{W^\sigma}(v_g) = F_g,$$

i.e.,

$$\langle J_\sigma(v_g), v\rangle_{\rho,\sigma} + \langle J_\sigma(\operatorname{div}(\boldsymbol{\beta}v_g)), \operatorname{div}(\boldsymbol{\beta}v)\rangle_{\rho,\sigma} = \int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n})gv \quad \text{for all } v \in W^\sigma(\boldsymbol{\beta};\Omega).$$

In particular, testing with $v \in C_0^\infty(\Omega)$, we conclude that $w_g := J_\sigma(\operatorname{div}(\boldsymbol{\beta}v_g)) \in L^\rho(\Omega)$ has a *weak derivative* $\boldsymbol{\beta}\cdot\nabla w_g = J_\sigma(v_g) \in L^\rho(\Omega)$. Hence $w_g \in W^\rho(\boldsymbol{\beta};\Omega)$. We claim now that $w_g = g$ a.e. in $\partial\Omega_+ \cup \partial\Omega_-$. Indeed, using

the integration by parts formula (2.8) and the identity (2.6), we have

$$\int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n}) w_g v = \int_\Omega (\boldsymbol{\beta}\cdot\nabla w_g) v + \int_\Omega w_g \operatorname{div}(\boldsymbol{\beta} v)$$
$$= \langle J_\sigma(v_g), v\rangle_{\rho,\sigma} + \langle J_\sigma(\operatorname{div}(\boldsymbol{\beta} v_g)), \operatorname{div}(\boldsymbol{\beta} v)\rangle_{\rho,\sigma} = \int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n}) g v.$$

Thus $\int_{\partial\Omega}(\boldsymbol{\beta}\cdot\boldsymbol{n})(w_g - g)v = 0$ for all $v \in W^\sigma(\boldsymbol{\beta};\Omega)$. We conclude the result owing to the density of $C^\infty(\overline{\Omega})$ in $W^\sigma(\boldsymbol{\beta};\Omega)$. □

A.2 Proof of Theorem A

In this section, we give the proof of Theorem A by means of the *Banach–Nečas–Babuška inf-sup conditions* (see, e.g., [22, Theorem 2.6]):

$$\|w\|_{\mathbb{U}} \lesssim \sup_{0\neq v\in\mathbb{V}} \frac{|b(w,v)|}{\|v\|_{\mathbb{V}}} \quad \text{for all } w \in \mathbb{U}, \tag{BNB1}$$

$$\{v \in \mathbb{V} : b(w,v) = 0 \text{ for all } w \in \mathbb{U}\} = \{0\}. \tag{BNB2}$$

In fact, we prove (BNB1), (BNB2) on the adjoint bilinear form. Recall that the primal operator is a continuous bijection if and only if the adjoint operator is a continuous bijection, in which case both inf-sup constants are the same.

We start by giving some properties that we need for the Banach-space setting. Recall from Definition 2.4 that $J_q(v) = \|v\|_q^{2-q}|v|^{q-1}\operatorname{sign}(v) \in L^p(\Omega) = \mathbb{U}$ denotes the duality map of $L^q(\Omega)$, i.e.,

$$\langle J_q(v), v\rangle_{p,q} = \|v\|_q^2 \quad \text{and} \quad \|J_q(v)\|_p = \|v\|_q \quad \text{for all } v \in L^q(\Omega). \tag{A.1}$$

Additionally, for any $v \in \mathbb{V} = W^q_{0,+}(\boldsymbol{\beta};\Omega) \subset L^q(\Omega)$, notice the identity

$$\boldsymbol{\beta}\cdot\nabla v|v|^{q-1}\operatorname{sign}(v) = \frac{1}{q}\operatorname{div}(\boldsymbol{\beta}|v|^q) - \frac{1}{q}\operatorname{div}(\boldsymbol{\beta})|v|^q \quad \text{for all } v \in \mathbb{V}. \tag{A.2}$$

We will use these definitions and properties also for their analogous "p" version, i.e., obtained by replacing q by p.

Remark A.1 (Differences to Proof in [9]). Our technique to proving Theorem A is similar to the one used by Cantin [9], which is also analogous to the proof in Hilbert spaces given by Di Pietro and Ern [21, Section 2.1]. Minor differences to Cantin are that we work with the operator $-\operatorname{div}(\boldsymbol{\beta} v) + \mu v$ (instead of $\boldsymbol{\beta}\cdot\nabla v + \mu v$) and that we use a graph-space norm (2.5) defined using squares (instead of the exponent ρ).

However, the main difference with Cantin's proof is in proving (BNB2), in particular, in concluding that a function $w \in W^p(\boldsymbol{\beta};\Omega)$ satisfies $w|_{\partial\Omega_\pm} = 0$ whenever

$$\int_{\partial\Omega} (\boldsymbol{\beta}\cdot\boldsymbol{n}) w v = 0 \quad \text{for all } v \in W^q_{0,\mp}(\boldsymbol{\beta};\Omega).$$

The approach by Cantin uses a density argument and an explicit characterization of the traces of the subspace of $W^q_{0,\mp}(\boldsymbol{\beta};\Omega)$ consisting of Lipschitz continuous functions (see [9, Proposition 2.4]). In our proof, we either use a regularity result for $w \in W^p(\boldsymbol{\beta};\Omega)$ (knowing that $\boldsymbol{\beta}\cdot\nabla w + \mu w = 0$) that allows us to pick $v = \phi|w|^{p-1}\operatorname{sign}(w)$ (with ϕ defined in Assumption 2.2), or we employ the surjectivity Lemma 2.6, to construct a particular function $v \in W^q_{0,\mp}(\boldsymbol{\beta};\Omega)$ such that $v|_{\partial\Omega_\pm} = |w|^{p-1}\operatorname{sign}(w)|_{\partial\Omega_\pm}$ (see Section A.2.2 below).

A.2.1 Proof of inf-sup Condition (BNB1) on the Adjoint

Let $b\colon \mathbb{U}\times\mathbb{V}\to\mathbb{R}$ be the bilinear form corresponding to the weak form in (3.1), i.e.,

$$b(w,v) = \int_\Omega w(\mu v - \operatorname{div}(\boldsymbol{\beta} v)). \tag{A.3}$$

For any $0 \neq v \in \mathbb{V}$, we have

$$\begin{aligned}
\sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p} &\geq \frac{|b(J_q(v),v)|}{\|J_q(v)\|_p} \\
&= \|v\|_q^{1-q}\left|\int_\Omega |v|^{q-1}\operatorname{sign}(v)(\mu v - \operatorname{div}(\boldsymbol{\beta} v))\right| && \text{(by (A.1) and (A.3))} \\
&= \|v\|_q^{1-q}\left|\int_\Omega |v|^{q-1}\operatorname{sign}(v)(\mu v - \operatorname{div}(\boldsymbol{\beta})v - \boldsymbol{\beta}\cdot\nabla v)\right| && \text{(by (2.6))} \\
&= \|v\|_q^{1-q}\left|\int_\Omega |v|^{q}\left(\mu - \frac{1}{p}\operatorname{div}(\boldsymbol{\beta})\right) - \frac{1}{q}\operatorname{div}(\boldsymbol{\beta}|v|^q)\right| && \text{(by (A.2))} \\
&\geq \mu_0\|v\|_q + \|v\|_q^{1-q}\frac{1}{q}\int_{\partial\Omega_-} |\boldsymbol{\beta}\cdot\boldsymbol{n}||v|^q && \text{(by (2.2))} \\
&\geq \mu_0\|v\|_q.
\end{aligned}$$

Hence we obtain control on v in the $\|\cdot\|_q$-norm.[6] To control the entire graph norm $|||\cdot|||_{q,\boldsymbol{\beta}}$, we also need to control the divergence part,

$$\begin{aligned}
\|\operatorname{div}(\boldsymbol{\beta}v)\|_q &= \sup_{0\neq w\in\mathbb{U}} \frac{\langle w, \operatorname{div}(\boldsymbol{\beta}v)\rangle_{p,q}}{\|w\|_p} && \text{(by duality)} \\
&= \sup_{0\neq w\in\mathbb{U}} \frac{\left|b(w,v) - \int_\Omega \mu w v\right|}{\|w\|_p} && \text{(by (A.3))} \\
&\leq \sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p} + \sup_{0\neq w\in\mathbb{U}} \frac{\left|\int_\Omega \mu w v\right|}{\|w\|_p} \\
&\leq \sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p} + \|\mu\|_\infty\|v\|_q && \text{(by the Cauchy–Schwarz inequality)} \\
&\leq \left(1 + \frac{\|\mu\|_\infty}{\mu_0}\right)\sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p} && \text{(using the previous bound).}
\end{aligned}$$

Combining both bounds, we have

$$|||v|||_{q,\boldsymbol{\beta}} \leq \frac{\sqrt{1 + (\mu_0 + \|\mu\|_\infty)^2}}{\mu_0^2}\sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p}.$$

The case when $\mu \equiv 0$ and $\operatorname{div}\boldsymbol{\beta} \equiv 0$ (under Assumption 2.8) is simpler since, by Lemma 2.10, we immediately have

$$\|v\|_q \leq C_{\mathrm{PF}}\|\boldsymbol{\beta}\cdot\nabla v\|_q = C_{\mathrm{PF}}\|\operatorname{div}(\boldsymbol{\beta}v)\|_q = C_{\mathrm{PF}}\sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p}.$$

Hence

$$|||v|||_{q,\boldsymbol{\beta}} \leq (1 + C_{\mathrm{PF}})\sup_{0\neq w\in\mathbb{U}} \frac{|b(w,v)|}{\|w\|_p}. \qquad \square$$

A.2.2 Proof of inf-sup Condition (BNB2) on the Adjoint

Next, we prove (BNB2) for the adjoint, which corresponds to injectivity of the primal operator. In other words, we need to show that $w = 0$ if $w \in L^p(\Omega)$ is such that

$$b(w,v) = 0 \quad \text{for all } v \in \mathbb{V} = W^q_{0,+}(\boldsymbol{\beta};\Omega). \tag{A.4}$$

6 This result is an extension of the 1-D result with constant advection in [16, Chapter XVII A, § 3, Section 3.7].

We first take $v \in C_0^\infty(\Omega)$ to obtain $\boldsymbol{\beta} \cdot \nabla w + \mu w = 0$ in the sense of distributions, and hence

$$\boldsymbol{\beta} \cdot \nabla w = -\mu w \in L^p(\Omega),$$

which implies $w \in W^p(\boldsymbol{\beta}; \Omega)$.

This means that w has sufficient regularity so that traces make sense (see Remark 2.5). Hence, going back to (A.4) and integrating by parts, we have

$$\int_{\partial\Omega_-} \boldsymbol{\beta} \cdot \boldsymbol{n} w v = 0 \quad \text{for all } v \in W^q_{0,+}(\boldsymbol{\beta}; \Omega). \tag{A.5}$$

We give now two different (but similar) proofs to show that $w \in W^p_{0,-}(\boldsymbol{\beta}; \Omega)$. The first proof considers the function $\widetilde{J}_p(w) := |w|^{p-1} \operatorname{sign}(w) \in L^q(\Omega)$. The fact that $\widetilde{J}_p(w)$ is actually in $W^q(\boldsymbol{\beta}; \Omega)$ is proven in Lemma A.2 below. For the function $\phi \in C^\infty(\overline{\Omega})$ defined in (2.3), we then have that $\phi \widetilde{J}_p(w)$ belongs to $W^q_{0,+}(\boldsymbol{\beta}; \Omega)$ (since ϕ vanishes on $\partial\Omega_+$). Using $v = \phi \widetilde{J}_p(w)$ in (A.5), we immediately obtain

$$\int_{\partial\Omega_-} \boldsymbol{\beta} \cdot \boldsymbol{n} |w|^p = 0, \tag{A.6}$$

hence $w \in W^p_{0,-}(\boldsymbol{\beta}; \Omega)$. Alternatively, a second proof (thanks to Lemma 2.6) considers a function $v_w \in W^q(\boldsymbol{\beta}; \Omega)$ such that $v_w|_{\partial\Omega_-} = |w|^{p-1} \operatorname{sign}(w) \in L^q(|\boldsymbol{\beta} \cdot \boldsymbol{n}|; \partial\Omega)$. Using $v = \phi v_w$ in (A.5), we also obtain (A.6).

Finally, we conclude using an energy argument

$$\begin{aligned}
0 &= \int_\Omega (\boldsymbol{\beta} \cdot \nabla w + \mu w) J_p(w) \\
&= \|w\|_p^{2-p} \Big[\int_\Omega |w|^p \Big(\mu - \frac{1}{p} \operatorname{div}(\boldsymbol{\beta}) \Big) + \frac{1}{p} \int_{\partial\Omega_+} \boldsymbol{\beta} \cdot \boldsymbol{n} |w|^p \Big] \quad \text{(by (A.2) and (A.6))} \\
&\geq \mu_0 \|w\|_p^2 + \frac{1}{p} \|w\|_p^{2-p} \int_{\partial\Omega_+} \boldsymbol{\beta} \cdot \boldsymbol{n} |w|^p \qquad \text{(by (2.2))} \\
&\geq \mu_0 \|w\|_p^2.
\end{aligned}$$

Hence $w = 0$.

On the other hand, the case when $\mu \equiv 0$ and $\operatorname{div} \boldsymbol{\beta} \equiv 0$ is straightforward (under Assumption 2.8) since $\boldsymbol{\beta} \cdot \nabla w = 0$ implies

$$0 = \|\boldsymbol{\beta} \cdot \nabla w\|_p \geq \frac{1}{C_{\mathrm{PF}}} \|w\|_p \quad \text{(by Lemma 2.10).}$$ □

We are left with a proof of the statement $\widetilde{J}_p(w) \in W^q(\boldsymbol{\beta}; \Omega)$.

Lemma A.2 (Regularity of $|w|^{p-1} \operatorname{sign}(w)$). *Let $\mu, \boldsymbol{\beta} \in L^\infty(\Omega)$ and $w \in L^p(\Omega)$ satisfy the homogeneous advection-reaction equation*

$$\boldsymbol{\beta} \cdot \nabla w + \mu w = 0 \quad \text{in } L^p(\Omega).$$

Then the function $\widetilde{J}_p(w) := |w|^{p-1} \operatorname{sign}(w) \in L^q(\Omega)$ satisfies

$$\boldsymbol{\beta} \cdot \nabla \widetilde{J}_p(w) \in L^q(\Omega).$$

Proof. First observe that $\widetilde{J}_p(w)$ has a Gâteaux derivative in the direction $\boldsymbol{\beta} \cdot \nabla w$. Indeed,

$$\begin{aligned}
\widetilde{J}_p'(w)[\boldsymbol{\beta} \cdot \nabla w] &= \lim_{t \to 0} \frac{\widetilde{J}_p(w + t\boldsymbol{\beta} \cdot \nabla w) - \widetilde{J}_p(w)}{t} \\
&= \lim_{t \to 0} \frac{\widetilde{J}_p(w - t\mu w) - \widetilde{J}_p(w)}{t} \\
&= \Big(\lim_{t \to 0} \frac{|1 - t\mu|^{p-2}(1 - t\mu) - 1}{t} \Big) |w|^{p-1} \operatorname{sign}(w) \\
&= -(p-1)\mu |w|^{p-1} \operatorname{sign}(w).
\end{aligned}$$

Hence $\widetilde{J}_p'(w)[\boldsymbol{\beta}\cdot\nabla w]\in L^q(\Omega)$. The conclusion of the lemma follows from the identity

$$\boldsymbol{\beta}\cdot\nabla\widetilde{J}_p(w)=\widetilde{J}_p'(w)[\boldsymbol{\beta}\cdot\nabla w]\quad\text{a.e. in }\Omega,$$

which is straightforward to verify. □

A.3 Proof of Proposition 5.1

We construct explicitly a Fortin operator $\Pi\colon\mathbb{V}\to\mathbb{V}_m$ satisfying Assumption 4.2. We note that this 1-D proof is similar to the 1-D version of the proof of [22, Lemma 4.20, p. 190].

Let $-1=x_0<x_1<\cdots<x_n=1$ be the set of nodes defining the partition $\mathcal{T}_n$. Over each element

$$T_j=(x_{j-1},x_j)\in\mathcal{T}_n,$$

we define Π to be the linear interpolant Π_1 plus a quadratic bubble, i.e.,

$$\Pi(v)|_{T_j}=\Pi_1(v)|_{T_j}+\alpha_j Q_j(x)\in\mathbb{P}^2(T_j)\quad\text{for all } v\in\mathbb{V},$$

where

$$\Pi_1(v)|_{T_j}=|T_{j-1}|^{-1}\big(v(x_{j-1})(x_j-x)+v(x_j)(x-x_{j-1})\big)\quad\text{and}\quad Q_j(x)=(x-x_{j-1})(x-x_j).$$

The coefficient α_j multiplying the bubble $Q_j(x)$ is selected in order to fulfill the equation

$$\int_{T_j}\Pi(v)=\int_{T_j}v. \tag{A.7}$$

Observe that $\Pi(v)\in\mathbb{P}^2_{\mathrm{cont},0,\{1\}}(\mathcal{T}_n)\subseteq\mathbb{P}^k_{\mathrm{cont},0,\{1\}}(\mathcal{T}_n)$ since $k\geq 2$, and for all $w_n\in\mathbb{U}_n$, we have

$$\begin{aligned}
b(w_n,\Pi(v))&=\sum_{j=1}^n\int_{T_j}w_n'\Pi(v)-w_n\Pi(v)|_{x_{j-1}}^{x_j} && \text{(by integration by parts)}\\
&=\sum_{j=1}^n w_n'\int_{T_j}\Pi(v)-w_n\Pi(v)|_{x_{j-1}}^{x_j} && \text{(since } w_n\in\mathbb{P}^1(T_j)\text{)}\\
&=\sum_{j=1}^n w_n'\int_{T_j}v-w_nv|_{x_{j-1}}^{x_j} && \text{(by interpolation and (A.7))}\\
&=b(w_n,v) && \text{(by integration by parts).}
\end{aligned}$$

Hence the requirement (4.4b) is satisfied. Now we recall that $\|(\cdot)\|_{\mathbb{V}}:=\|(\cdot)'\|_q$. Therefore, to obtain the requirement (4.4a) (i.e., the boundedness of the operator Π), we note that, on each element,

$$\begin{aligned}
|\alpha_j|&\leq\frac{6}{|T_j|^3}\int_{T_j}|v-\Pi_1(v)|\leq\frac{6}{|T_j|^{3-\frac{1}{p}}}\|v-\Pi_1(v)\|_q\leq\frac{6}{|T_j|^{2-\frac{1}{p}}}\|v'-\Pi_1(v)'\|_q,\\
\|\Pi_1(v)'\|_q&=\frac{|v(x_j)-v(x_{j-1})|}{|T_j|^{1-\frac{1}{q}}}=\frac{1}{|T_j|^{1-\frac{1}{q}}}\left|\int_{T_j}v'\right|\leq\|v'\|_q,\\
\|Q_j'\|_q&=\frac{|T_j|^{1+\frac{1}{q}}}{(q+1)^{\frac{1}{q}}}.
\end{aligned}$$

Thus, on each element (and therefore globally), we have

$$\|\Pi(v)'\|_q\leq\|\Pi_1(v)'\|_q+|\alpha_j|\|Q_j'\|_q\leq\|v'\|_q+C_q\|v'-\Pi_1(v)'\|_q\leq(1+2C_q)\|v'\|_q,$$

where the constant $C_q=6/(q+1)^{\frac{1}{q}}$ is mesh-independent. □

Acknowledgment: Ignacio Muga and Kristoffer van der Zee thank Leszek Demkowicz, Jay Gopalakrishnan, Paul Houston, Weifeng Qui and Sarah Roggendorf for helpful discussions. They also thank the unknown reviewers for their insightful suggestions and, in particular, for their encouragement to prove the *surjectivity* Lemma 2.6.

Funding: The work by Ignacio Muga was done in the framework of Chilean FONDECYT research project No. 1160774. Ignacio Muga was also partially supported by the European Union's Horizon 2020, research and innovation program under the Marie Sklodowska-Curie grant agreement No. 777778. Matthew Tyler and Kristoffer van der Zee are grateful for the support provided by the London Mathematical Society (LMS) Undergraduate Research Bursary Grant *"Advanced discontinuous discretisation techniques for multiscale partial differential equations"* 17-18 103, and thank Donald Brown for his contributions. Kristoffer van der Zee also thanks the support provided by the Royal Society International Exchanges Scheme/Kan Tong Po Visiting Fellowship Programme, and the above FONDECYT project.

References

[1] P. Azérad, *Analyse des équations de Navier–Stokes en bassin peu profond et de l'équation de transport*, PhD thesis, Université de Neuchâtel, Neuchâtel, 1996.
[2] P. Azérad and J. Pousin, Inégalité de Poincaré courbe pour le traitement variationnel de l'équation de transport, *C. R. Acad. Sci. Paris Sér. I Math.* **322** (1996), no. 8, 721–727.
[3] C. Bardos, A. Y. le Roux and J.-C. Nédélec, First order quasilinear equations with boundary conditions, *Comm. Partial Differential Equations* **4** (1979), no. 9, 1017–1034.
[4] H. Beirão da Veiga, Existence results in Sobolev spaces for a stationary transport equation, *Ric. Mat.* **36** (1987), suppl., 173–184.
[5] H. Beirão da Veiga, Boundary-value problems for a class of first order partial differential equations in Sobolev spaces and applications to the Euler flow, *Rend. Semin. Mat. Univ. Padova* **79** (1988), 247–273.
[6] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts Appl. Math. 15, Springer, New York, 2008.
[7] D. Broersen, W. Dahmen and R. P. Stevenson, On the stability of DPG formulations of transport equations, *Math. Comp.* **87** (2018), no. 311, 1051–1082.
[8] T. Bui-Thanh, L. Demkowicz and O. Ghattas, Constructively well-posed approximation methods with unity inf-sup and continuity constants for partial differential equations, *Math. Comp.* **82** (2013), no. 284, 1923–1952.
[9] P. Cantin, Well-posedness of the scalar and the vector advection-reaction problems in Banach graph spaces, *C. R. Math. Acad. Sci. Paris* **355** (2017), no. 8, 892–902.
[10] P. Cantin and N. Heuer, A DPG framework for strongly monotone operators, *SIAM J. Numer. Anal.* **56** (2018), no. 5, 2731–2750.
[11] C. Carstensen, P. Bringmann, F. Hellwig and P. Wriggers, Nonlinear discontinuous Petrov–Galerkin methods, *Numer. Math.* **139** (2018), no. 3, 529–561.
[12] J. Chan, L. Demkowicz and R. Moser, A DPG method for steady viscous compressible flow, *Comput. Fluids* **98** (2014), 69–90.
[13] J. Chan, J. A. Evans and W. Qiu, A dual Petrov–Galerkin finite element method for the convection-diffusion equation, *Comput. Math. Appl.* **68** (2014), no. 11, 1513–1529.
[14] A. Cohen, W. Dahmen and G. Welper, Adaptivity and variational stabilization for convection-diffusion equations, *ESAIM Math. Model. Numer. Anal.* **46** (2012), no. 5, 1247–1273.
[15] W. Dahmen, C. Huang, C. Schwab and G. Welper, Adaptive Petrov–Galerkin methods for first order transport equations, *SIAM J. Numer. Anal.* **50** (2012), no. 5, 2420–2445.
[16] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Vol. 5: Evolution Problems. I*, Springer, Berlin, 1992.
[17] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Vol. 6: Evolution Problems. II*, Springer, Berlin, 1993.
[18] L. Demkowicz and J. Gopalakrishnan, A class of discontinuous Petrov–Galerkin methods. Part I: The transport equation, *Comput. Methods Appl. Mech. Engrg.* **199** (2010), no. 23–24, 1558–1572.
[19] L. Demkowicz and J. Gopalakrishnan, A class of discontinuous Petrov–Galerkin methods. Part II: Optimal test functions, *Numer. Methods Partial Differential Equations* **27** (2011), no. 1, 70–105.

[20] L. Demkowicz and J. Gopalakrishnan, An overview of the discontinuous Petrov Galerkin method, in: *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations: 2012 John H Barrett Memorial Lectures*, IMA Vol. Math. Appl. 157, Springer, Cham (2014), 149–180.
[21] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl. (Berlin) 69, Springer, Heidelberg, 2012.
[22] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Appl. Math. Sci. 159, Springer, New York, 2004.
[23] A. Ern and J.-L. Guermond, Discontinuous Galerkin methods for Friedrichs' systems. I. General theory, *SIAM J. Numer. Anal.* **44** (2006), no. 2, 753–778.
[24] A. Ern and J.-L. Guermond, A converse to Fortin's lemma in Banach spaces, *C. R. Math. Acad. Sci. Paris* **354** (2016), no. 11, 1092–1095.
[25] A. Ern and J.-L. Guermond, Finite element quasi-interpolation and best approximation, *ESAIM Math. Model. Numer. Anal.* **51** (2017), no. 4, 1367–1385.
[26] V. Girault and L. Tartar, L^p and $W^{1,p}$ regularity of the solution of a steady transport equation, *C. R. Math. Acad. Sci. Paris* **348** (2010), no. 15–16, 885–890.
[27] J. Gopalakrishnan, P. Monk and P. Sepúlveda, A tent pitching scheme motivated by Friedrichs theory, *Comput. Math. Appl.* **70** (2015), no. 5, 1114–1135.
[28] J. Gopalakrishnan and W. Qiu, An analysis of the practical DPG method, *Math. Comp.* **83** (2014), no. 286, 537–552.
[29] J.-L. Guermond, Stabilization of Galerkin approximations of transport equations by subgrid modeling, *M2AN Math. Model. Numer. Anal.* **33** (1999), no. 6, 1293–1316.
[30] J. L. Guermond, A finite element technique for solving first-order PDEs in L^P, *SIAM J. Numer. Anal.* **42** (2004), no. 2, 714–737.
[31] H. Holden and N. H. Risebro, *Front Tracking for Hyperbolic Conservation Laws*, 2nd ed., Appl. Math. Sci. 152, Springer, Heidelberg, 2015.
[32] J. E. Lavery, Solution of steady-state one-dimensional conservation laws by mathematical programming, *SIAM J. Numer. Anal.* **26** (1989), no. 5, 1081–1089.
[33] I. Muga and K. G. van der Zee, Discretization of linear problems in Banach spaces: Residual minimization, nonlinear Petrov–Galerkin, and monotone mixed methods, preprint (2018), http://arxiv.org/abs/1511.04400.
[34] T. Piasecki, Steady transport equation in Sobolev-Slobodetskii spaces, *Colloq. Math.* **154** (2018), no. 1, 65–76.
[35] E. B. Saff and S. Tashev, Gibbs phenomenon for best L_p approximation by polygonal lines, *East J. Approx.* **5** (1999), no. 2, 235–251.
[36] A. Stern, Banach space projections and Petrov–Galerkin estimates, *Numer. Math.* **130** (2015), no. 1, 125–133.
[37] R. Verfürth, *A Posteriori Error Estimation Techniques for Finite Element Methods*, Numer. Math. Sci. Comput., Oxford University Press, Oxford, 2013.