ICES REPORT 18-15

August 2018

Alternative enriched test spaces in the DPG method for singular perturbation problems

by

Jacob Salazar, Jaime Mora, and Leszek Demkowicz

The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Jacob Salazar, Jaime Mora, and Leszek Demkowicz, "Alternative enriched test spaces in the DPG method for singular perturbation problems," ICES REPORT 18-15, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, August 2018.

Alternative enriched test spaces in the DPG method for singular perturbation problems

Jacob Salazar*, Jaime Mora, and Leszek Demkowicz

Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin, 201 E 24th St, Austin, TX 78712, USA

Abstract

We propose and investigate the application of alternative enriched test spaces in the discontinuous Petrov-Galerkin (DPG) finite element framework for singular perturbation linear problems, with an emphasis on 2D convection-dominated diffusion. Providing robust L^2 error estimates for the field variables is considered a convenient feature for this class of problems, since this norm would not account for the large gradients present in boundary layers. With this requirement in mind, Demkowicz and others [3, 9] have previously formulated special test norms, which through DPG deliver the desired L^2 convergence. However, robustness has only been verified through numerical experiments for *tailored* test norms which are problem-specific, whereas the *quasi-optimal* test norm (not problem specific) has failed such tests due to the difficulty to resolve the optimal test functions sought in the DPG technology. To address this issue (i.e. improve optimal test functions resolution for the quasi-optimal test norm) we propose to discretize the local test spaces with functions that depend on the perturbation parameter ϵ. Explicitly, we work with B-spline spaces defined on a ϵ-dependent Shishkin submesh. Two examples are run using adaptive h-refinement to compare the performance of proposed test spaces with that of standard test spaces. We also include a modified norm and a continuation strategy aiming to improve time performance and briefly experiment with these ideas.

1 Introduction

1.1 DPG fundamentals

Suppose we have an abstract variational problem: find $\mathfrak{u} \in \mathcal{U}$ such that

$$b(\mathfrak{u}, \mathfrak{v}) = \ell(\mathfrak{v}) \quad \forall \mathfrak{v} \in \mathcal{V} \tag{1.1}$$

where both trial space $\mathcal{U}$ and test space $\mathcal{V}$ are Hilbert spaces, $b(\cdot, \cdot) : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ is a bilinear continuous functional that satisfies the inf-sup condition, and $\ell \in \mathcal{V}'$.

In this non-symmetric functional setting (i.e. $\mathcal{U}$ may be different from $\mathcal{V}$), we define the energy norm $\|\cdot\|_E$ on $\mathcal{U}$ through the bilinear form b and the chosen norm on $\mathcal{V}$, $\|\cdot\|_V$, as follows

*Corresponding author. E-mail: jacobess@utexas.edu

$$\|\mathfrak{u}\|_E = \sup_{\mathfrak{v}\in\mathcal{V}} \frac{|b(\mathfrak{u},\mathfrak{v})|}{\|\mathfrak{v}\|_{\mathcal{V}}}. \tag{1.2}$$

If we define an operator $\mathcal{B} : \mathcal{U} \to \mathcal{V}'$ by $\mathcal{B}\mathfrak{u} := b(\mathfrak{u}, \cdot)$, then (1.2) is equivalent to

$$\|\mathfrak{u}\|_E = \sup_{\mathfrak{v}\in\mathcal{V}} \frac{|\langle \mathcal{B}\mathfrak{u}, \mathfrak{v}\rangle|}{\|\mathfrak{v}\|_{\mathcal{V}}} = \|\mathcal{B}\mathfrak{u}\|_{\mathcal{V}'}. \tag{1.3}$$

1.1.1 The ideal DPG

The ideal Discontinuous Petrov-Galerkin (DPG) method is a minimum residual method that delivers the best approximation error in the energy norm, which is the same as minimizing the residual in the dual test space norm; that is, given a discrete trial subspace $\mathcal{U}^h \subset \mathcal{U}$ with dimension N_h, if $\mathfrak{u} \in \mathcal{U}$ is the exact solution, the ideal DPG solution $\mathfrak{u}_h \in \mathcal{U}^h$ satisfies

$$\|\mathfrak{u} - \mathfrak{u}_h\|_E = \inf_{\mathfrak{w}_h\in\mathcal{U}^h} \|\mathfrak{u} - \mathfrak{w}_h\|_E = \inf_{\mathfrak{w}_h\in\mathcal{U}^h} \|\mathcal{B}(\mathfrak{u} - \mathfrak{w}_h)\|_{\mathcal{V}'} = \inf_{\mathfrak{w}_h\in\mathcal{U}^h} \|\ell - \mathcal{B}\mathfrak{w}_h\|_{\mathcal{V}'}. \tag{1.4}$$

Now, let $\mathcal{R}_{\mathcal{V}} : \mathcal{V} \to \mathcal{V}'$ be the Riesz map of $\mathcal{V}$ associated to norm $\|\cdot\|_{\mathcal{V}}$. Then (1.4) turns into

$$\|\mathfrak{u} - \mathfrak{u}_h\|_E = \inf_{\mathfrak{w}_h\in\mathcal{U}^h} \|\mathcal{R}_{\mathcal{V}}^{-1}(\ell - \mathcal{B}\mathfrak{w}_h)\|_{\mathcal{V}}. \tag{1.5}$$

It is shown in [6] how after working out the minimization in (1.5), the ideal DPG methodology corresponds to the following *automatically stable*[1] Petrov-Galerkin formulation: find $\mathfrak{u}_h \in \mathcal{U}^h$ such that

$$b(\mathfrak{u}_h, \mathfrak{v}_h) = \ell(\mathfrak{v}_h) \quad \forall \mathfrak{v}_h \in \mathcal{V}^h := \mathbf{T}\mathcal{U}^h \subset \mathcal{V}, \tag{1.6}$$

with $\mathbf{T} : \mathcal{U} \to \mathcal{V}$ being the *trial-to-test* map, defined by $\mathbf{T} := \mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}$. The functions living in $\mathcal{V}^h$ are referred to as the *optimal test functions*.

By Babuška's theorem, the following stability estimate holds for the ideal DPG [6]:

$$\|u - u_h\|_{\mathcal{U}} \leq \frac{M}{\gamma_h} \inf_{w_h} ||u - w_h||_{\mathcal{U}} \leq \frac{M}{\gamma} \inf_{w_h} ||u - w_h||_{\mathcal{U}}, \tag{1.7}$$

where $M > 0$ is the continuity constant of bilinear functional b, and γ_h is its discrete inf-sup constant. When the trial space is equipped with the energy norm, it can be shown that (by design) $M = 1$ and $\gamma = 1$, where γ is the inf-sup constant in the continuous setting (therefore retrieving (1.4)).

1.1.2 DPG in practice

Unless we are dealing with a very particular test space, for which inverting its Riesz map $\mathcal{R}_{\mathcal{V}}$ is feasible, in general this problem is infinite-dimensional so that it is impossible to invert the Riesz operator without some approximation. That's why we need to work with a finite-dimensional *enriched test space* $\mathcal{V}^r \subset \mathcal{V}$ rather than with the whole space. The dimension of this space is $N_r \geq N_h$. The stability of the ideal DPG is carried over to the practical DPG whenever there exists a Fortin operator $\mathbf{F}_r : \mathcal{V} \to \mathcal{V}^r$, which must satisfy appropriate orthogonality and continuity conditions [6]. The only modification to (1.6) is to replace the trial-to-test

[1] The discrete inf-sup constant is bounded below by the continuous one, $\gamma_h \geq \gamma$

operator by its counterpart into the enriched test space $\mathbf{T}^r := \mathcal{R}_{\mathcal{V}^r}^{-1}\mathcal{B}$; where $\mathcal{R}_{\mathcal{V}^r}^{-1}$ is the (Bubnov-) Galerkin approximation of $\mathcal{R}_{\mathcal{V}}^{-1}$ with $\mathcal{V}^r$. Let C^r be the continuity constant of $\mathbf{F}_r$; then the estimate (1.7) becomes

$$\|u - u_h\|_{\mathcal{U}} \leq \frac{C^r M}{\gamma} \inf_{w_h} ||u - w_h||_{\mathcal{U}}. \tag{1.8}$$

At this point, it is already possible to translate the formulation into an algebraic problem. Let's take a basis of the enriched test space $\{\mathfrak{v}_k^r\}_{k=1}^{N_r}$, hence the near-optimal test functions $\{\mathfrak{v}_{i,opt}\}_{i=1}^{N_h}$ can be computed through

$$\mathfrak{v}_{i,opt} = \sum_{j,k=1}^{N_r} (\mathsf{B}^{\mathsf{T}})_{ij} (\mathsf{G}^{-1})_{jk} \mathfrak{v}_k \quad i = 1, ..., N_h, \tag{1.9}$$

where B is called the enriched stiffness matrix, and G is the Gram matrix of $\mathcal{V}^r$, defined by

$$\mathsf{B}_{ji} := b(\mathfrak{u}_i, \mathfrak{v}_j^r),$$
$$\mathsf{G}_{jk} := (\mathfrak{v}_j^r, \mathfrak{v}_k^r)_{\mathcal{V}},$$

where $\{\mathfrak{u}_i\}_{i=1}^{N_h}$ is a basis for $\mathcal{U}^h$. If the DPG solution is expressed as $\mathfrak{u}^h = \sum \mathfrak{u}_i \mathsf{d}_i$, its coefficient vector d is then the solution to the discrete problem

$$\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{B}\mathsf{d} = \mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{l}, \tag{1.10}$$

with $\mathsf{l}_k := \ell(\mathfrak{v}_k^r)$. It is important to mention that matrices G and $\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{B}$ are Hermitian positive definite and for broken test spaces, introduced later, they become also sparse.

1.2 The optimal test norm

If we wish to minimize the error in a user-prescribed norm, a question arises when looking at (1.6): Which test norm should be defined on $\mathcal{V}$ to make that happen?

In [15] Zitelli et. al. introduced the concept of *the optimal test norm* (a notion of optimality that must not be confused with that of the optimal test functions), which we explain here. Given a norm on our trial space $\|\cdot\|_{\mathcal{U}}$, and under the assumption that conjugate operator $\mathcal{B}' : \mathcal{V}'' \sim \mathcal{V} \to \mathcal{U}'$ is injective, the so called optimal test norm is defined by:

$$\|\mathfrak{v}\|_{\mathcal{V},opt} := \sup_{\mathfrak{u}\in\mathcal{U}} \frac{|b(\mathfrak{u}, \mathfrak{v})|}{\|\mathfrak{u}\|_{\mathcal{U}}}. \tag{1.11}$$

Recalling (1.2), we deduce that the use of this norm on the test space implies (see [15, Prop. 2.1]):

$$\|\mathfrak{u}\|_{\mathcal{U}} = \|\mathfrak{u}\|_E, \tag{1.12}$$

which, in other words, says that if we use the ideal DPG methodology with the optimal test norm, then we can minimize the solution error in any trial norm $\|\cdot\|_{\mathcal{U}}$ we choose.

However, in practical implementations of the method, we encounter some issues that place limitations on that goal. In order to address those limitations, we need to define what we understand by robustness in the present work, and next we focus on the model problem to be later studied numerically.

1.3 Robustness

In the case of singular perturbation problems, we consider a discretization method to be *robust* if the stability constant C relating the actual approximation error and the best approximation error is independent of the perturbation parameter (ϵ):

$$\|\mathfrak{u} - \mathfrak{u}_h\|_{\mathcal{U}^{(1)}} \leq C \inf_{\mathfrak{w}_h \in \mathcal{U}^h} \|\mathfrak{u} - \mathfrak{w}_h\|_{\mathcal{U}^{(2)}}, \tag{1.13}$$

with norms $\|\cdot\|_{\mathcal{U}^{(1)}}$ and $\|\cdot\|_{\mathcal{U}^{(2)}}$ being ideally the same [9]. For this ($\|\cdot\|_{\mathcal{U}^{(1)}} = \|\cdot\|_{\mathcal{U}^{(2)}} = \|\cdot\|_{\mathcal{U}}$) it must hold that the constants in (1.8) (that is, the bounding constants of b and $\mathbf{F}_r$) are all independent of ϵ.

Here lies the challenge of singular perturbation problems. It is desired to be capable of varying the value of ϵ, as in real applications this number ranges several orders of magnitude, usually in much smaller scales than the rest of the data involved. Thus, it is important to show independence of ϵ in the constants, in order to have a method that suits a wide range of values of this perturbation parameter. Coming up with a method that shows this independence is not an easy task, and although DPG is equipped with good tools to approach the problem, there are some complications in the way that must be carefully handled in the process.

2 Choice of the norm

With the purpose of illustrating the challenges present in the solution of singular perturbation problems through DPG, we'll describe a model problem and discuss its *ultraweak variational formulation*, the functional setting, and explain the factors that intervene in the choice of the test norm, concerning both mathematical and computational aspects.

2.1 Model Problem

For the sake of simplicity, we limit ourselves to the two-dimensional space. We will work with an important singular perturbation model problem, the steady-state convection-dominated diffusion, briefly referred to as "confusion" problem. Its strong form is stated as follows: find $u : \Omega \to \mathbb{R}$ such that

$$\begin{cases} -\epsilon \Delta u + \operatorname{div}(\beta u) = f & \text{in } \Omega, \\ u = u_0 & \text{on } \Gamma; \end{cases} \tag{2.1}$$

where $\Omega \subset \mathbb{R}^2$ is a simply connected open set (a domain) with piecewise smooth boundary Γ, $\epsilon > 0$ is the diffusivity or perturbation parameter in this case, β is a divergence-free advection vector field ($\operatorname{div} \beta = 0$) of magnitude $|\beta| = \mathcal{O}(1)$, $f \in L^2(\Omega)$ and $u_0 \in H^{1/2}(\Gamma)$. Rewriting the second order partial differential equation as a system of first order equations by introducing $\sigma := \epsilon \nabla u$, we obtain:

$$\begin{cases} \epsilon^{-1}\sigma - \nabla u = 0 & \text{in } \Omega, \\ -\operatorname{div}(\sigma - \beta u) = f & \text{in } \Omega, \\ u = u_0 & \text{on } \Gamma; \end{cases} \tag{2.2}$$

From (2.2) we can identify the differential operator $\mathcal{A}$ acting on group unknown (σ, u):

$$\mathcal{A} := \begin{pmatrix} \epsilon^{-1} & -\nabla \\ -\operatorname{div} & \beta \cdot \nabla \end{pmatrix}. \tag{2.3}$$

2.2 Ultraweak Variational Formulation

In order to solve (2.2) weakly, we can relax both of the equations in the system so that the required regularity of the solution is minimal. Let's introduce a vector-valued test function τ for the first line, and a scalar-valued test function v for the second one. Thus, the ultraweak variational formulation is the one brought up by defining the following bilinear and linear functionals on $\mathfrak{u} = \left(\begin{smallmatrix} \sigma \\ u \end{smallmatrix}\right)$, and $\mathfrak{v} = \left(\begin{smallmatrix} \tau \\ v \end{smallmatrix}\right)$:

$$b(\mathfrak{u}, \mathfrak{v}) := (\mathfrak{u}, \mathcal{A}^* \mathfrak{v})_\Omega \tag{2.4}$$

$$\ell(\mathfrak{v}) := \langle \tau \cdot n, u_0 \rangle_\Gamma + (f, v)_\Omega \tag{2.5}$$

where

$$\mathcal{A}^* \mathfrak{v} = \begin{pmatrix} \epsilon^{-1} & \nabla \\ \operatorname{div} & -\beta \cdot \nabla \end{pmatrix} \begin{pmatrix} \tau \\ v \end{pmatrix} = \begin{pmatrix} \epsilon^{-1}\tau + \nabla v \\ \operatorname{div} \tau - \beta \cdot \nabla v \end{pmatrix}, \tag{2.6}$$

the pairing $(\cdot, \cdot)_\Omega$ is the $L^2(\Omega)$ inner product (which generates norm $\| \cdot \|_\Omega$), and $\langle \cdot, \cdot \rangle_\Gamma$ denotes the duality pairing of boundary energy spaces $H^{-1/2}(\Gamma) \times H^{-1/2}(\Gamma)$. Having reduced the regularity in the trial variables, for (2.4) to make sense we do require some regularity in the test functions, namely, $\tau \in \boldsymbol{H}(\operatorname{div}, \Omega)$ and $v \in H^1(\Omega)$ with $v = 0$ on Γ (equivalently, $v \in H_0^1(\Omega)$). Notice that because the normal trace of τ, i.e. $(\tau \cdot n)|_\Gamma$, lives in Sobolev space $H^{-1/2}(\Gamma)$, the first term in the linear functional (2.5) is well defined.

Notice that in this formulation u and σ come from an L^2 space, which is a very favorable situation in the "confusion" problem, because the large gradients (caused by the steep boundary layers present in the solutions to this equation) won't influence the measure of the error in the chosen L^2 norm [9].

Additionally, notice that for this formulation the optimal test norm is simply

$$\|\mathfrak{v}\|_{\mathcal{V},opt} = \|\mathcal{A}^* \mathfrak{v}\|_{\mathcal{U}} = \|\mathcal{A}^* \mathfrak{v}\|_\Omega. \tag{2.7}$$

Because of this identity, we also call this norm the *adjoint-operator norm*.

An alternative to the adjoint-operator norm is the *adjoint-graph norm* or *quasi-optimal norm*:

$$\|\mathfrak{v}\|^2_{\mathcal{V},qopt} := \|\mathcal{A}^* \mathfrak{v}\|^2_\Omega + \eta \left(\|\tau\|^2_\Omega + \|v\|^2_\Omega \right). \tag{2.8}$$

where η is a scaling coefficient. Assume that operator $\mathcal{A}^*$ is bounded below, i.e., [2]

$$\|\mathcal{A}^* \mathfrak{v}\|^2_\Omega \geq \alpha \|\mathfrak{v}\|^2_\Omega \quad \forall \mathfrak{v} = (\tau, v) \in H(\operatorname{div}, \Omega) \times H_0^1(\Omega).$$

We have trivially,

$$\|\mathfrak{v}\|_{\mathcal{V},opt} \leq \|\mathfrak{v}\|_{\mathcal{V},qopt},$$

and

$$\|\mathfrak{v}\|^2_{\mathcal{V},opt} \leq \|\mathcal{A}^* \mathfrak{v}\|^2_\Omega + \frac{\eta}{\alpha} \|\mathcal{A}^* \mathfrak{v}\|^2 = \left(1 + \frac{\eta}{\alpha}\right) \|\mathcal{A}^* \mathfrak{v}\|^2_\Omega;$$

[2] By the Closed Range Theorem for closed operators [13, Remark 5.19.1] operator $\mathcal{A}$ is also bounded below with the same constant α.

With $\eta/\alpha = \mathcal{O}(1)$, the optimal norm and the adjoint-graph norm are *robustly equivalent* and so are the corresponding energy norms. The ideal DPG method with the adjoint-graph test norm no longer delivers the L^2-projection, but the method is robust in the L^2-norm.

2.3 Broken test spaces

Let Ω_h be a non-overlapping partition of Ω into open elements with Lipschitz boundary, so that $\bar{\Omega} = \cup\{\bar{\mathcal{K}}; \mathcal{K} \in \Omega_h\}$, and $\Gamma_h = \cup\{\partial\mathcal{K}; \mathcal{K} \in \Omega_h\}$ denotes the mesh skeleton. Setting up a variational formulation with *broken* test spaces is a two step process. In the first step, boundary conditions imposed on the test space are removed, giving as a result new boundary unknowns. In the second step we proceed to actually *break* the test functions, replacing the usual energy spaces defined on Ω with spaces defined elementwise as follows (for details on this process see [7]):

$$H^1(\Omega_h) := \left\{v \in L^2(\Omega) : \ v|_{\mathcal{K}} \in H^1(\mathcal{K}), \mathcal{K} \in \Omega_h\right\}, \tag{2.9}$$

$$\boldsymbol{H}(\mathrm{div}, \Omega_h) := \left\{\tau \in \boldsymbol{L}^2(\Omega) : \ \tau|_{\mathcal{K}} \in \boldsymbol{H}(\mathrm{div}, \mathcal{K}), \mathcal{K} \in \Omega_h\right\}. \tag{2.10}$$

The adoption of broken test spaces is what adds the *discontinuous* nature to DPG. It is easy to see that the broken spaces contain the classical energy spaces: $H^1(\Omega) \subset H^1(\Omega_h)$, $\boldsymbol{H}(\mathrm{div}, \Omega) \subset \boldsymbol{H}(\mathrm{div}, \Omega_h)$. From (2.2), now multiplying each line by a vector-valued broken test function τ and a scalar-valued broken test function v and integrating by parts we obtain the *broken ultraweak variational formulation*. In this new formulation, additional unknowns are included to account for reproducing the problem when only testing with functions from the classical test spaces: (1) trace $\hat{u} \in H^{1/2}(\Gamma_h)$ (satisfying the boundary condition $\hat{u} = 0$ on Γ), which represents the skeleton values of "density" u; (2) flux $\hat{\sigma}_n \in H^{-1/2}(\Gamma_h)$ which represents the normal trace of net flux $\beta u - \sigma$ across the mesh skeleton. The bilinear form and linear form of this variational formulation are:

$$b((\mathfrak{u}, \hat{\mathfrak{u}}), \mathfrak{v}) = (\mathfrak{u}, \mathcal{A}^*\mathfrak{v})_{\Omega_h} + \langle \hat{\sigma}_n, v\rangle_{\Gamma_h} - \langle \tau \cdot n, \hat{u}\rangle_{\Gamma_h} \tag{2.11}$$

$$\ell(\mathfrak{v}) = \langle \tau \cdot n, \hat{u}_0\rangle_{\Gamma_h} + (v, f)_{\Omega_h} \tag{2.12}$$

where $\mathfrak{v} \in \mathcal{V} = \boldsymbol{H}(\mathrm{div}, \Omega_h) \times H^1(\Omega_h)$ and $(\mathfrak{u}, \hat{\mathfrak{u}}) \in \mathcal{U} = \boldsymbol{L}^2(\Omega) \times L^2(\Omega) \times H^{-1/2}(\Gamma_h) \times H^{1/2}(\Gamma_h)$ are group variables, $\mathfrak{v} = (\tau, v)$, $\mathfrak{u} = (\sigma, u)$ and $\hat{\mathfrak{u}} = (\hat{\sigma}_n, \hat{u})$. Moreover, $\hat{u}_0 \in H^{1/2}(\Gamma_h)$ is a lift into the interior skeleton of the boundary values u_0. Finally, we are using the pairings on the mesh and the skeleton defined by

$$(\cdot,\cdot)_{\Omega_h} := \sum_{\mathcal{K} \in \Omega_h} (\cdot,\cdot)_{\mathcal{K}} \quad (\text{ this one induces norm } \|\cdot\|_{\Omega_h}),$$

$$\langle \cdot,\cdot\rangle_{\Gamma_h} := \sum_{\mathcal{K} \in \Omega_h} \langle \cdot,\cdot\rangle_{\partial\mathcal{K}}.$$

Now we can see how the optimal test norm for this problem looks like. If our norm on trial space $\mathcal{U}$ is

$$\|(\mathfrak{u}, \hat{\mathfrak{u}})\|^2_{\mathcal{U}} := \|\sigma\|^2_{\Omega} + \|u\|^2_{\Omega} + \|\hat{\sigma}_n\|^2_{H^{-1/2}(\Gamma_h)} + \|\hat{u}\|^2_{H^{1/2}(\Gamma_h)}, \tag{2.13}$$

then, according to [12] the optimal test norm is

$$\|\mathfrak{v}\|^2_{\mathcal{V},opt} = \|\mathcal{A}^*\mathfrak{v}\|^2_{\Omega_h} + \|(\tau \cdot n)|_{\Gamma_h}\|^2_{(H^{1/2}(\Gamma_h))'} + \|v|_{\Gamma_h}\|^2_{(H^{-1/2}(\Gamma_h))'}. \tag{2.14}$$

Finally, in order to discretize this broken variational formulation, we need discrete subspaces not only for the field variables in L^2 but for the trace and flux on the mesh skeleton. Then we need to choose finite-dimensional subspaces $\boldsymbol{Y}^p \subset \boldsymbol{L}^2(\Omega)$, $Y^p \subset L^2(\Omega)$, $\hat{V}^p \subset H^{-1/2}(\Gamma_h)$ and $\hat{W}^p \subset H^{1/2}(\Gamma_h)$ to form the trial discrete subspace $\mathcal{U}^h$.

2.4 Localization of the test norm

For a simpler test norm, we find breaking the test spaces computationally very useful, because matrix G becomes block diagonal, with each block corresponding to an element of the mesh. Then the "inversion" of G can be performed elementwise, and thus we also can compute an element contribution for the DPG stiffness matrix $\mathsf{B}^\mathsf{T}\mathsf{G}^{-1}\mathsf{B}$, which can be assembled globally in the same fashion as in any Galerkin finite element code. This is not the case with norm (2.14), because the skeleton terms involve jumps in the functions across edges, therefore the computation of the norm is no longer fully local.

Instead of attempting to derive an optimal test norm corresponding to the broken variational formulation, we can opt for using the quasi-optimal adjoint-graph norm corresponding to the original formulation (2.8). Due to the presence of the L^2 term, these adjoint-graph norm extends to the broken test space (then it is *localizable*, contrary to the norm (2.14)). The fundamental result in [2] shows that the broken variational formulation remains stable with the stability constant of the same order as α.

2.5 Discrete stability with the quasi-optimal norm

In summary, with the localizable adjoint-graph norm and broken test spaces, we end up with the following estimate for the problem at hand

$$\begin{aligned}\|\sigma-\sigma_h\|_\Omega^2+\|u-u_h\|_\Omega^2+\|\hat{\sigma}_n-\hat{\sigma}_{n,h}\|_{flux}^2+\|\hat{u}-\hat{u}_h\|_{trace}^2 \leq C\Big\{&\inf_{\varsigma_h\in\boldsymbol{Y}^p}\|\sigma-\varsigma_h\|_\Omega^2+\inf_{u_h\in Y^p}\|u-w_h\|_\Omega^2\\ &+\inf_{\hat{\varsigma}_{n,h}\in\hat{V}^p}\|\hat{\sigma}_n-\hat{\varsigma}_{n,h}\|_{flux}^2\\ &+\inf_{\hat{w}_h\in\hat{W}^p}\|\hat{u}-\hat{w}_h\|_{trace}^2\Big\},\end{aligned} \tag{2.15}$$

where $\|\cdot\|_{flux}$ and $\|\cdot\|_{trace}$ are special norms for fluxes and traces implied by the quasi-optimal test norm, stability constant C is of order 1 and independent of ϵ, and the term inside the brackets represents the best approximation error of $\mathfrak{u}$ in space $\mathcal{U}^h$. The good news is that we control fields u, σ in the desired L^2-norm in a robust way. The bad news is that, in the process of breaking test functions, we have introduced new unknowns which must be measured in a norm implied by our quasi-optimal test norm, which we have to accept. For a longer discussion on the subject, see [2].

2.6 Quasi-optimal test norm vs. robust test norm

The optimal test functions can be seen as the solutions to the dual problem $\mathcal{A}^*\mathfrak{v}=\mathfrak{u}$ with homogeneous boundary conditions [6], then it is no surprise that the optimal test functions develop boundary layers and their resolution is difficult. There have been multiple developments concerning the design of norms that

produce test functions which are easier to approximate. Demkowicz and Heuer [9] present the analysis of robustly designed norms for the "confusion" problem, which involve mesh dependent coefficients and weighted L^2 norms for some terms (such weights are data dependent). Later, Chan et al. [3], showed that by setting appropriate boundary conditions the use of weights is unnecessary, resulting in the following robust test norm (specific to the "confusion" problem):

$$\|\mathfrak{v}\|^2_{\mathcal{V},D} := \|\operatorname{div}\tau\|^2_{\Omega_h} + \epsilon\|\nabla v\|^2_{\Omega_h} + \|\beta\cdot\nabla v\|^2_{\Omega_h} + \|C_\tau\tau\|^2_{\Omega_h} + \|C_v v\|^2_{\Omega_h}; \tag{2.16}$$

$$C_v|_{\mathcal{K}} := \min\left\{\sqrt{\epsilon/|\mathcal{K}|}, 1\right\}, \tag{2.17}$$

$$C_\tau|_{\mathcal{K}} := \min\left\{\sqrt{1/\epsilon}, \sqrt{1/|\mathcal{K}|}\right\} \quad \forall\mathcal{K}\in\Omega_h; \tag{2.18}$$

where $|\mathcal{K}|$ is the area of element $\mathcal{K}$. In Section 4, norm (2.16) will be commonly referred to as *the robust* test norm. The robust test norm separates v and τ so the local Gram matrix becomes block-diagonal. The mesh dependent factors eliminate dominance of diffusion by reaction in (2.16) which results in optimal test functions with "weak" boundary layers that can be easily resolved.

Although these norms have robust features, having to design a norm for each specific boundary value problem is less desirable than using the optimal (or quasi-optimal) norm; which is a general concept that can be applied on any problem given in its ultraweak variational formulation. However, we can try a simpler compromise approach.

Due to the theoretical difficulty and the lack of generality that these designed norms represent, and despite their virtues, we wish to keep trying the more general adjoint-graph norm but looking for another strategy to ease the resolution of the optimal test functions. The strategy of our choice is to implement boundary-layer-adapted test spaces, as it is explained in detail in section 3. We will later compare the numerical performance of the distinct norms introduced above and draw some conclusions from the results.

2.7 Testing new norms and strategies

Due to the difficulty to solve problems with $\epsilon < 10^{-4}$ we explored additional strategies.

The first strategy is to define a modified test norm. With a big simplification of the stability analysis provided in [3, 9], we adopt the mesh-dependent coefficient in the quasi-optimal norm, and replace the small perturbation parameter ϵ with a value depending on the element size h. For each element, define $\epsilon^* := \max\{ch, \epsilon\}$, with c being a positive constant (in our examples set to $c = 10^{-2}$). Thus, our *modified adjoint-graph norm* (referred to as the *EPS STAR* norm in the plots for the results) is defined as

$$\|(\tau, v)\|^2_{\mathcal{V},eps-star} := \|\frac{1}{\epsilon^*}\tau + \nabla v\|^2_{\Omega_h} + \|\operatorname{div}\tau - \beta\cdot\nabla v\|^2_{\Omega_h} + \|v\|^2_{\Omega_h}; \tag{2.19}$$

where we have expanded (2.8), with $\eta = 1$ (although removing the extra $\|\tau\|$ term), and incorporated the new coefficient ϵ^*.

Another proposed strategy which we briefly explore numerically, could be interpreted as a continuation technique with respect to parameter ϵ. This strategy is understood in the context of adaptive refinements process. The motivation is to allow for an approximate layer-capturing mesh to quickly develop. In order to do this we link ϵ to the previous iteration residual, so that it adopts a more lenient value when the residual is high and progressively approaches its original value (ϵ_{orig}) as the residual decreases. At each refinement

we replace ϵ with ϵ_i defined by:

$$\epsilon_i = \max\left\{\frac{1}{2}\left(c\rho_{i-1} + \epsilon_{i-1}\right), \epsilon_{orig}\right\}, \tag{2.20}$$

where ϵ_i is the perturbation value used for refinement i (with $i = 1, 2, ...$) , $\rho_{i-1} = \|\mathcal{B}\mathfrak{u}_{i-1}^h - \ell\|_{\mathcal{V}'}$ is the residual obtained after solving the previous refinement step, c is a positive constant and we set $\rho_0 = 0$ and $\epsilon_0 = \epsilon_{orig}$.

3 Test space discretization

We want to choose appropriate finite-dimensional subspaces $\mathcal{U}^h \subset \mathcal{U}$ and $\mathcal{V}^r \subset \mathcal{V}$ for implementing the DPG methodology. Though we could generalize this exercise to elements of "all" shapes , for the purpose of the current study we limit ourselves to working with quadrilateral elements only. As explained in the previous section, the most challenging part of applying DPG to the "confusion" problem is to accurately resolve the optimal test functions. Then, we decide to discretize the trial space with classical polynomial spaces, but for the test space the idea is to use specially tailored functions, that may account for the eventual boundary or internal layers generated by the convection-dominated physics of the problem.

We will employ the first exact sequence of polynomial spaces for the Nédélec quadrilateral element, and for that purpose the following definitions are required. Let $\mathcal{I} \subset \mathbb{R}$ be the unit open interval $(0, 1)$. Thus $\widetilde{\mathcal{K}} := \mathcal{I}^2$ represents the unit square, which is the master element for every element in our partition. We assume standard shape regularity assumptions avoiding shape degenerate elements. Consequently, for each element $\mathcal{K} \in \Omega_h$ there is a diffeomorphic element map $\mathbf{x}_\mathcal{K} : \widetilde{\mathcal{K}} \longrightarrow \bar{\mathcal{K}}$ with nonsingular jacobian $\mathcal{J}_\mathcal{K}$. For all energy spaces of interest defined on $\mathcal{K}$, there are well-known Piola maps that relate them to spaces defined over $\widetilde{\mathcal{K}}$ [14, 10]:

$$H^1(\widetilde{\mathcal{K}}) \ni \widetilde{v} \longmapsto T_\mathcal{K}^{\text{grad}}\widetilde{v} := \widetilde{v} \circ \mathbf{x}_\mathcal{K}^{-1} = v \in H^1(\mathcal{K}) \tag{3.1}$$

$$H(\text{div}, \widetilde{\mathcal{K}}) \ni \widetilde{\tau} \longmapsto T_\mathcal{K}^{\text{div}}\widetilde{\tau} := (\det \mathcal{J}_\mathcal{K}^{-1}\mathcal{J}_\mathcal{K}\widetilde{\tau}) \circ \mathbf{x}_\mathcal{K}^{-1} = \tau \in H(\text{div}, \mathcal{K}) \tag{3.2}$$

$$L^2(\widetilde{\mathcal{K}}) \ni \widetilde{u} \longmapsto T_\mathcal{K}\widetilde{u} := (\det \mathcal{J}_\mathcal{K}^{-1}\widetilde{u}) \circ \mathbf{x}_\mathcal{K}^{-1} = u \in L^2(\mathcal{K}) \tag{3.3}$$

$$\boldsymbol{L}^2(\widetilde{\mathcal{K}}) \ni \widetilde{\sigma} \longmapsto \boldsymbol{T}_\mathcal{K}\widetilde{\sigma} := (\det \mathcal{J}_\mathcal{K}^{-1}\widetilde{\sigma}) \circ \mathbf{x}_\mathcal{K}^{-1} = \sigma \in \boldsymbol{L}^2(\mathcal{K}) \tag{3.4}$$

The conformity wanted in the subspaces makes possible to derive similar maps for the trace spaces, where we can exploit the fact that $H^{1/2}(\Gamma_h)$ is formed by traces of $H^1(\Omega)$ and $H^{-1/2}(\Gamma_h)$ contains the normal traces of $H(\text{div}, \Omega)$. Clearly, Γ_h is the union of all the edges that exist in the mesh. One can show that pullbacks of traces on an edge depend only on the restriction of element maps to the edge and therefore, they are well defined, i.e. independent of which element map is being used. The inverse image of e through $\mathbf{x}_{\mathcal{K}_e}$ is denoted $\widetilde{e}$, with unit tangent $\mathbf{t}_{\widetilde{e}}$. Hence, the Piola maps for the trace spaces correspond edgewise to:

$$H^{1/2}(\Gamma_h)\big|_{\widetilde{e}} \ni \widetilde{u} \longmapsto T_e^{\text{tr,grad}}\widetilde{u} := \widetilde{u} \circ \mathbf{x}_{\mathcal{K}_e}^{-1} = v \in H^{1/2}(\Gamma_h)\big|_e \tag{3.5}$$

$$H^{-1/2}(\Gamma_h)\big|_{\widetilde{e}} \ni \widetilde{\sigma}_n \longmapsto T_e^{\text{tr,div}}\widetilde{\sigma}_n := (|\mathcal{J}_{\mathcal{K}_e}\mathbf{t}_{\widetilde{e}}|^{-1}\widetilde{\sigma}_n) \circ \mathbf{x}_{\mathcal{K}_e}^{-1} = \sigma_n \in H^{-1/2}(\Gamma_h)\big|_e \tag{3.6}$$

These Piola maps are applied in the definition of the polynomial subspaces that will be used in the DPG implementation. For any integers $q_1, q_2 \geq 0$, and $\mathcal{P}^{q_1}(\mathcal{I})$ denoting the space of univariate polynomials defined on $\mathcal{I}$ of degree up to q_1, we define the tensor product polynomial space $\mathcal{Q}^{q_1,q_2}(\widetilde{\mathcal{K}}) := \mathcal{P}^{q_1}(\mathcal{I}) \otimes \mathcal{P}^{q_2}(\mathcal{I})$. With

these definitions, for some integer $p \geq 1$ our discrete trial space for the broken ultraweak formulation is chosen as follows:

$$\begin{aligned}
\mathcal{U}^h :=& \boldsymbol{Y}^p \times Y^p \times \hat{W}^p \times \hat{V}^p \quad \subset \quad \boldsymbol{L}^2(\Omega) \times L^2(\Omega) \times H^{1/2}(\Gamma_h) \times H^{-1/2}(\Gamma_h) = \mathcal{U}, \text{ where} \\
\boldsymbol{Y}^p :=& \left\{ \sigma \in \boldsymbol{L}^2(\Omega) : \sigma|_{\mathcal{K}} \in \boldsymbol{T}_{\mathcal{K}} \left(\mathcal{Q}^{p-1,p-1}(\widetilde{\mathcal{K}}) \right)^2 \ \forall \mathcal{K} \in \Omega_h \right\}, \\
Y^p :=& \left\{ u \in L^2(\Omega) : u|_{\mathcal{K}} \in T_{\mathcal{K}} \mathcal{Q}^{p-1,p-1}(\widetilde{\mathcal{K}}) \ \forall \mathcal{K} \in \Omega_h \right\}, \\
\hat{W}^p :=& \left\{ \hat{u} \in C(\partial\Omega^h) : \hat{u}|_e \in T_e^{\text{tr,grad}} \mathcal{P}^p(\widetilde{e}) \forall \text{ edge } e \in \partial\Omega^h \right\}, \\
\hat{V}^p :=& \left\{ \hat{\sigma}_n \in L^2(\partial\Omega^h) : \hat{\sigma}_n|_e \in T_e^{\text{tr,div}} \mathcal{P}^{p-1}(\widetilde{e}) \forall \text{ edge } e \in \partial\Omega^h \right\}.
\end{aligned}$$

The shape functions used for these spaces are Legendre and integrated-Legendre (Lobatto) polynomials (for the edges), and tensor products of those (for the quadrilaterals), as exposed in [11].

In most DPG implementations the enriched test subspace has been chosen following the same guidelines, after increasing the polynomial degree to $p_r = p + \Delta p$, with $\Delta p \geq 0$. Should us want to adopt such a strategy, our discrete test space would be:

$$\begin{aligned}
\mathcal{V}_L^r :=& V_L^{p_r} \times W_L^{p_r} \quad \subset \quad H(\text{div}, \Omega^h) \times H^1(\Omega^h) = \mathcal{V}, \text{ where} \\
V_L^{p_r} :=& \left\{ \tau \in H(\text{div}, \Omega^h) : \tau|_{\mathcal{K}} \in T_{\mathcal{K}}^{\text{div}} \left(\mathcal{Q}^{p_r,p_r-1}(\widetilde{\mathcal{K}}) \times \mathcal{Q}^{p_r-1,p_r}(\widetilde{\mathcal{K}}) \right) \ \forall \mathcal{K} \in \Omega^h \right\}, \\
W_L^{p_r} :=& \left\{ v \in H^1(\Omega^h) : v|_{\mathcal{K}} \in T_{\mathcal{K}}^{\text{grad}} \mathcal{Q}^{p_r,p_r}(\widetilde{\mathcal{K}}) \ \forall \mathcal{K} \in \Omega^h \right\}.
\end{aligned}$$

In the just defined spaces, the subscript L indicates that they are constructed using Legendre and Lobatto polynomials. However, these test spaces are not so well suited for the "confusion" problem, as they do not incorporate the perturbation parameter by any means, hence they hardly perceive the degenerate boundary layers generated unless the element size is similar in magnitude to the parameter. Next, we are proposing alternative spaces that in some way account for the perturbation parameter, aiming to get a higher solution quality when resolving the optimal test functions for the "confusion" problem.

3.1 Design of enriched test functions

Two notions are central when designing these enriched test functions: Shishkin meshes and B-spline spaces.

3.1.1 Shishkin mesh

The application of Shishkin meshes for the implementation of singularly perturbed problems with DPG was first made in [12]. In that work, each quadrilateral element was partitioned into a 3×3 grid, giving a special width to the outermost subelements, as shown in Figure 3.1. The resulting test functions are especially favorable for capturing boundary layers in *hp*-finite element methods, therefore it was a reasonable heuristic choice to adopt the idea for the local problem of solving DPG's optimal test functions.

The reference mentioned above [12] used a locally conforming space of Bernstein polynomials, enforcing C^0 continuity across subelements of the Shishkin mesh. We are trying a slight variation of this approach, which will help reducing local cost and will allow us to assess whether higher-continuity local spaces can be of help for better resolving the optimal DPG test functions.

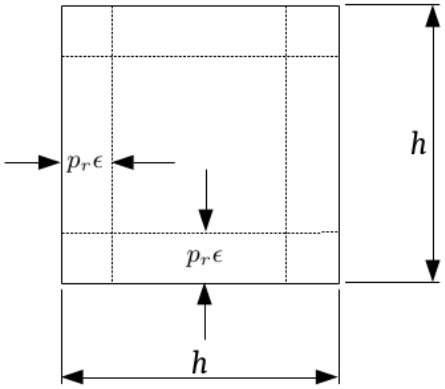

Figure 3.1: Shishkin subgrid in each element (showing physical dimensions).

3.1.2 B-splines

B-splines are piecewise polynomial functions defined over an interval $[a, b]$, that possess several properties that have made them a successful tool both in numerical analysis and in computer graphics [4]. In order to define a set of B-splines that may be used as shape functions, we are required to set an integer order $m \geq 0$, another integer $n \geq m + 1$ that corresponds to the number of B-splines, and a set of parameters referred to as the knot vector:

$$\Xi = \{\xi_1, ..., \xi_{n+m+1}\}, \tag{3.7}$$

where $a = \xi_1 = ... = \xi_{m+1} < \xi_{m+2} \leq ... \leq \xi_n < \xi_{n+1} = ... = \xi_{n+m+1} = b$ are real numbers. Each different value in Ξ is called a knot (we order them increasingly and denote them ζ_l, for $l = 0, ..., N$), and the number of times that the same value is present in the vector is called multiplicity (denoted M_l, for $l = 0, ..., N$). Using this knot vector, we define the $i - th$ B-spline of order m, $B_{m,i}$, as follows:

$$B_{0,i}(\xi) = \begin{cases} 1 & \text{if } \xi_i \leq xi < \xi_{i+1}, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } m = 0; \tag{3.8}$$

$$B_{m,i}(\xi) = \frac{\xi - \xi_i}{\xi_{i+m} - \xi_i} B_{m-1,i}(\xi) + \frac{\xi_{i+m+1} - \xi}{\xi_{i+m+1} - \xi_{i+1}} B_{m-1,i+1} \qquad \text{for } m \geq 1; \tag{3.9}$$

for $i = 1, 2, ..., n$. The formula above is known as the Cox-de Boor recursion formula [4]. Additionally, it is known that continuity of B-splines across knots (that is, at $\xi = \zeta_l$) is of order $\alpha_l = m - M_l$, $l = 0, ..., N$. Let $\boldsymbol{\alpha} = \{\alpha_0, ..., \alpha_N\}$ be the regularity vector and define the B-spline space:

$$\boldsymbol{B}^m_{\boldsymbol{\alpha}}([a, b]) := \text{span}\, \{B_{m,i}\}_{i=1}^n$$

Let $[a, b] = [0, 1]$, so we can simplify the notation $\boldsymbol{B}^m_{\boldsymbol{\alpha}} = \boldsymbol{B}^m_{\boldsymbol{\alpha}}([0, 1])$.

Buffa et al. [1] have introduced exact sequence B-spline spaces for tensor product elements in two dimensions. The logic of such spaces follows the same ideas as for tensor product polynomial spaces. In this case, the two-dimensional tensor product B-spline space on the master element is defined as:

$$\mathcal{S}^{m_1,m_2}_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} := \boldsymbol{B}^{m_1}_{\boldsymbol{\alpha}_1} \otimes \boldsymbol{B}^{m_2}_{\boldsymbol{\alpha}_2}.$$

3.1.3 Specifications of the enriched test space

In order to define a tensor product B-spline space based on the Shishkin mesh presented in Figure 3.1, we have the following data:

$$
\begin{aligned}
N = N_1 = N_2 &= 3;\\
\zeta_0 &= 0;\\
\zeta_1 &= \min\{0.25, \epsilon p_r h^{-1}\};\\
\zeta_2 &= 1 - \min\{0.25, \epsilon p_r h^{-1}\};\\
\zeta_3 &= 1;\\
\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 &= \{-1, p_r - 1, p_r - 1, -1\};\\
\boldsymbol{\alpha} - \mathbf{1} &= \{-1, p_r - 2, p_r - 2, -1\}.
\end{aligned}
$$

With this data and the definitions above, we can reconstruct the corresponding knot vectors and hence the alternative enriched spaces:

$$
\begin{aligned}
\mathcal{V}_B^r :=& V_B^{p_r} \times W_B^{p_r} \quad \subset \quad H(\mathrm{div}, \Omega^h) \times H^1(\Omega^h) = \mathcal{V}, \text{ where}\\
V_B^{p_r} :=& \left\{ \tau \in H(\mathrm{div}, \Omega^h) : \tau|_{\mathcal{K}} \in T_{\mathcal{K}}^{\mathrm{div}} \left(\mathcal{S}^{p_r, p_r - 1}_{\boldsymbol{\alpha}, \boldsymbol{\alpha} - \mathbf{1}} \times \mathcal{S}^{p_r - 1, p_r}_{\boldsymbol{\alpha} - \mathbf{1}, \boldsymbol{\alpha}} \right) \ \forall \mathcal{K} \in \Omega^h \right\},\\
W_B^{p_r} :=& \left\{ v \in H^1(\Omega^h) : v|_{\mathcal{K}} \in T_{\mathcal{K}}^{\mathrm{grad}} \mathcal{S}^{p_r, p_r}_{\boldsymbol{\alpha}, \boldsymbol{\alpha}} \ \forall \mathcal{K} \in \Omega^h \right\}.
\end{aligned}
$$

In the subspaces above, the subscript B is for B-splines. Unlike the choice found in [12], we have a higher continuity of the shape functions across knots and a smaller space, which reduces the computational complexity at the element level. If we used C^0 regularity at the knots, the dimension of $\mathcal{V}^r$ restricted to one element would be $2(3p_r+1)(3p_r-2)+(3p_r+1)^2 = 3(3p_r+1)^2-6(3p_r+1)$. In the current work we propose an enriched space which is piecewise of the same polynomial degree as the ones in the cited reference, but with a local dimension of $\dim \mathcal{V}_B^r\big|_{\mathcal{K}} = 2(p_r+3)(p_r+2)+(p_r+3)^2 = 3(p_r+3)^2 - 2(p_r+3)$, which is smaller than the one above whenever $p_r > 1$. On the other hand, if we see the dimension of the polynomial counterpart we have $\dim \mathcal{V}_L^r\big|_{\mathcal{K}} = 3(p_r+1)^2 - 2(p_r+1)$. Even though the main focus is not to analyze the computational cost of this implementation, the fact that the spaces herein proposed are of a dimension much lower than with other Shishkin-mesh-based enriched space is certainly an appealing feature. To illustrate this aspect, we will later show results on computation time.

4 Numerical Results

We present numerical results to two problems, with which we compare the performance of discrete test spaces based on Legendre and Lobatto polynomials (hereinafter referred to as 'Lobatto' only) with B-splines.

For the implementation of the present method we used the Cox-de Boor formula for evaluation of B-splines, and for numerical integration we used Gaussian quadrature over each sub-element of the Shishkin mesh. Our test functions were mounted on top of the existing *2Dhp* code infrastructure originally developed by Demkowicz [5]. This high order hp-finite element code was previously expanded to include DPG capabilities.

Making use of the error estimator built into the DPG methodology at use [8], we use the error convergence of adaptive h-refinements as comparison criterion. We show numerical convergence results in the L^2 norm for u and σ.

For both problems, the domain is the unit square. The first one (Eriksson-Johnson) has constant advection vector and the second one (herein referred to as curved streamlines) variable advection vector. These problems have boundary conditions more complicated than the one of the simplified presentation above, and this fact may have important theoretical implications, but that does not interfere with the implementation of the test norms and spaces herein developed. For a deep analysis on the effects of boundary condition see [3, 6].

For each problem we initially compare three diffusivity values ($\epsilon = 10^{-2}, 10^{-3}, 10^{-4}$) showing the L^2-error evolution with respect to the total number of degrees of freedom (DOF). Additionally we include evolution with respect to time of computation.

Two polynomial degrees were used ($p = 2, 3$); enrichment values for Lobatto test spaces are $\Delta p = 0, 1, 2$, while for B-splines only $\Delta p = 0$ is used. The initial mesh is a single *quad* element with homogeneous refinements.

To provide a uniform picture of the refinement processes under different parameters (polynomial degree, enrichment, ϵ) we limit results to 10^6 DOF, 10^4 seconds and 200 refinement steps.

Additionally, in order to explore the behaviour of proposed EPS STAR norm, we analyze both problems with a more demanding diffusivity $\epsilon = 10^{-5}$, extending our computing limits to larger values.

4.1 Example 1: Eriksson-Johnson problem

Consider the following convection-diffusion equation:

$$\begin{aligned} -\epsilon\Delta u + \nabla\cdot(\beta u) = f \quad \text{in} \quad \Omega = (0,1)^2, \\ u = u_0 \quad \text{on} \quad \Gamma, \end{aligned} \tag{4.1}$$

with $\beta = (1,0)^\intercal$. The following manufactured solution corresponds to a homogeneous case ($f = 0$):

$$u(x,y) = \frac{e^{r_1(x-1)} - e^{r_2(x-2)}}{e^{-r_1} - e^{-r_2}} \cos\pi y; \qquad r_{1,2} = \frac{-1\sqrt{1+4\epsilon^2\pi^2}}{-2\epsilon}; \tag{4.2}$$

which develops a boundary layer next to the boundary at $x = 1$ (see Figure 4.1).

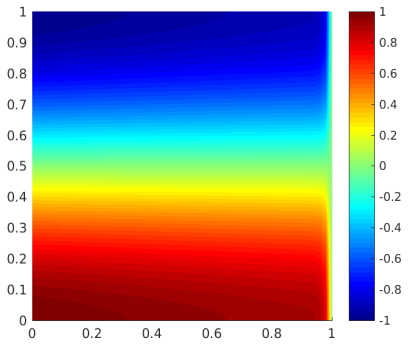

Figure 4.1: Exact solution, Eriksson-Johnson problem, $\epsilon = 10^{-2}$

It is important to mention that the numbers shown in the legend labels, at the end of each data series, is the refinement step corresponding to the last data point included in the graph.

For $\epsilon = 10^{-2}$, a somewhat 'benign' problem (Figures 4.2 and 4.3), we see no big differences. With B-splines (BSP in the labels) holding the same error at about the same number of DOF. However, when comparing computation time, for $p = 2$ Lobatto with $\Delta p = 2$ requires less time than B-splines; and for $p = 3$ Lobatto with $\Delta p = 0, 1$ requires less time than B-splines.

Now for $\epsilon = 10^{-3}$ (Figures 4.4 and 4.5), $p = 2$ B-splines clearly provide better results both ways, error vs DOF and error vs time. For $p = 3$, B-splines provide similar results to Lobatto with $\Delta p = 2$.

With $\epsilon = 10^{-4}$ (Figures 4.6 and 4.7), we start to observe interesting behaviours. For Lobatto not all values of Δp converge; they ($\Delta p = 1, 2$ for $p = 2$ and $\Delta p = 0, 2$ for $p = 3$) fail to decrease the L^2-error with subsequent refinement steps.

Therefore, for this problem, we can argue B-splines provide better results than any Lobatto (p, Δp combination) set up.

In Figures 4.9 and 4.9 the L^2-error to energy error ratio is shown. It is clear that B-splines tend to the same value (close to the million DOF) across the tested ϵ values; whereas Lobatto does not clearly converge to a single number and its values are at least an order of magnitude greater.

4.1.1 EPS STAR norm

Figures 4.10 and 4.11 show that for this problem, the EPS STAR norm behaves very similarly to the robust norm, except after the million DOF where robust norm with B-splines present a sudden error reduction. Regarding this problem, quasi-optimal test norm shows a slightly better behaviour (in particular B-splines)

At the same time, this verifies the improved behaviour of B-splines under robust norm.

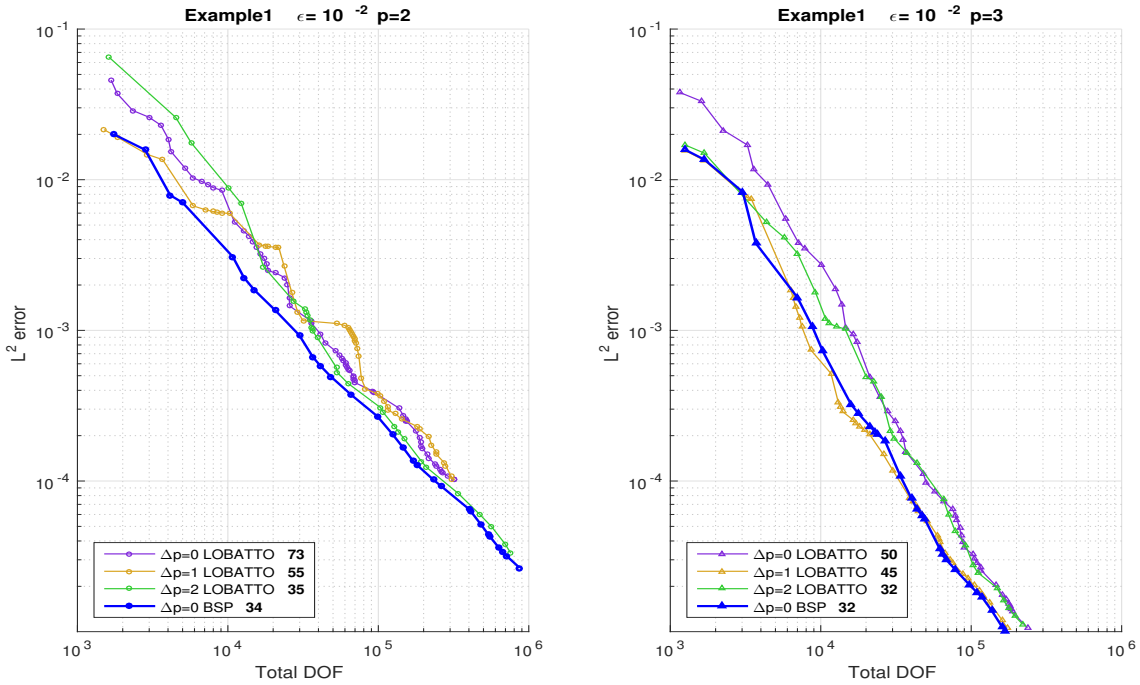

Figure 4.2: L^2-error vs DOF. $\epsilon = 10^{-2}$. $p = 2,\ p = 3$.

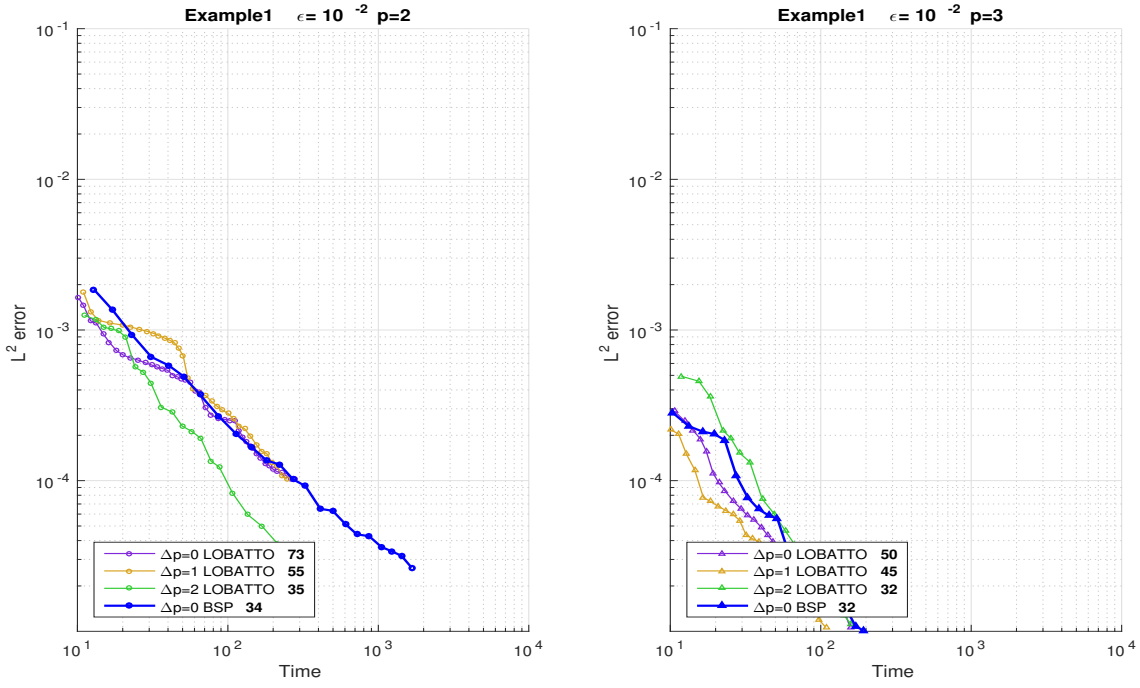

Figure 4.3: L^2-error vs time. $\epsilon = 10^{-2}$. $p = 2,\ p = 3$.

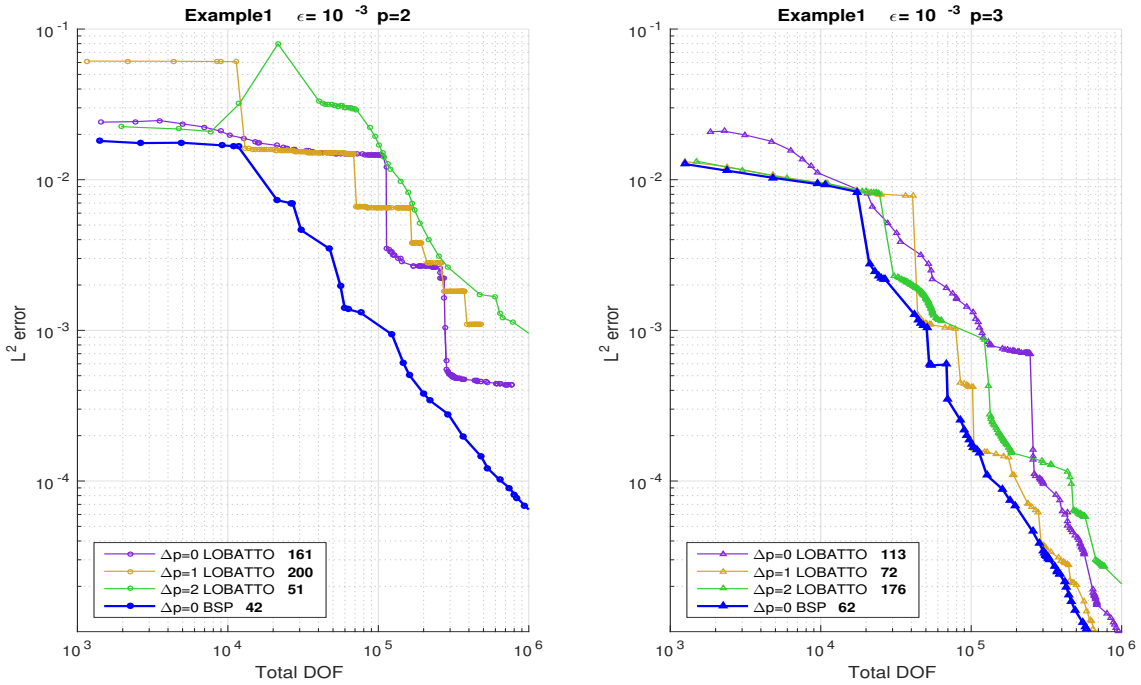

Figure 4.4: L^2-error vs DOF. $\epsilon = 10^{-3}$. $p = 2,\ p = 3$.

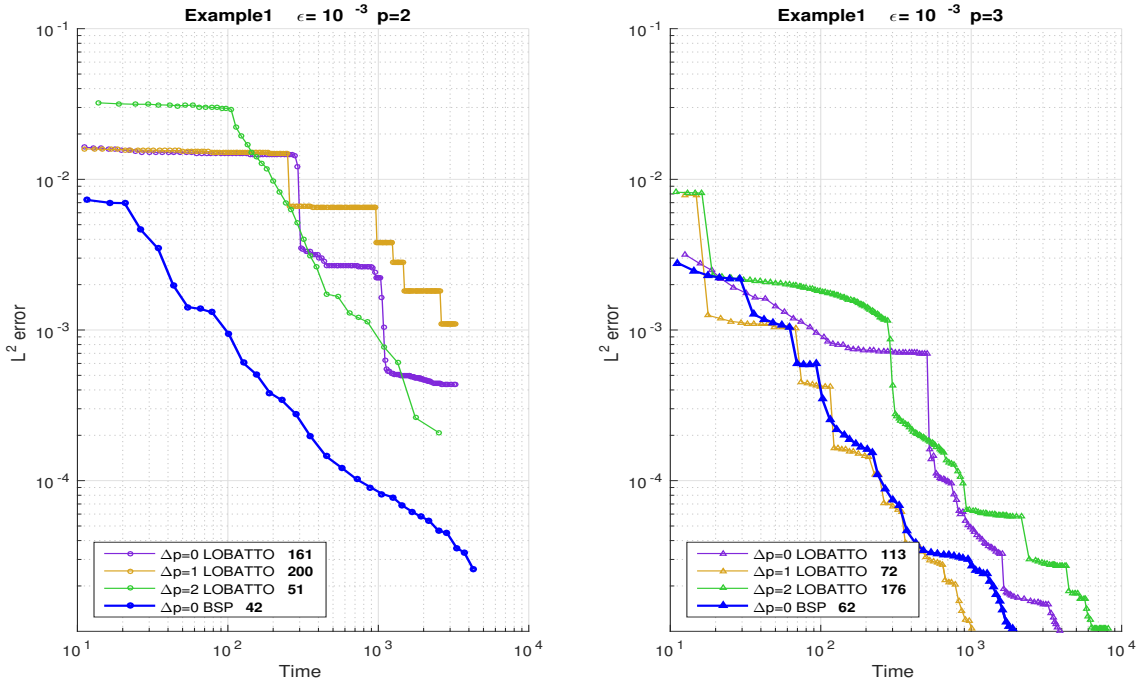

Figure 4.5: L^2-error vs time. $\epsilon = 10^{-3}$. $p = 2,\ p = 3$.

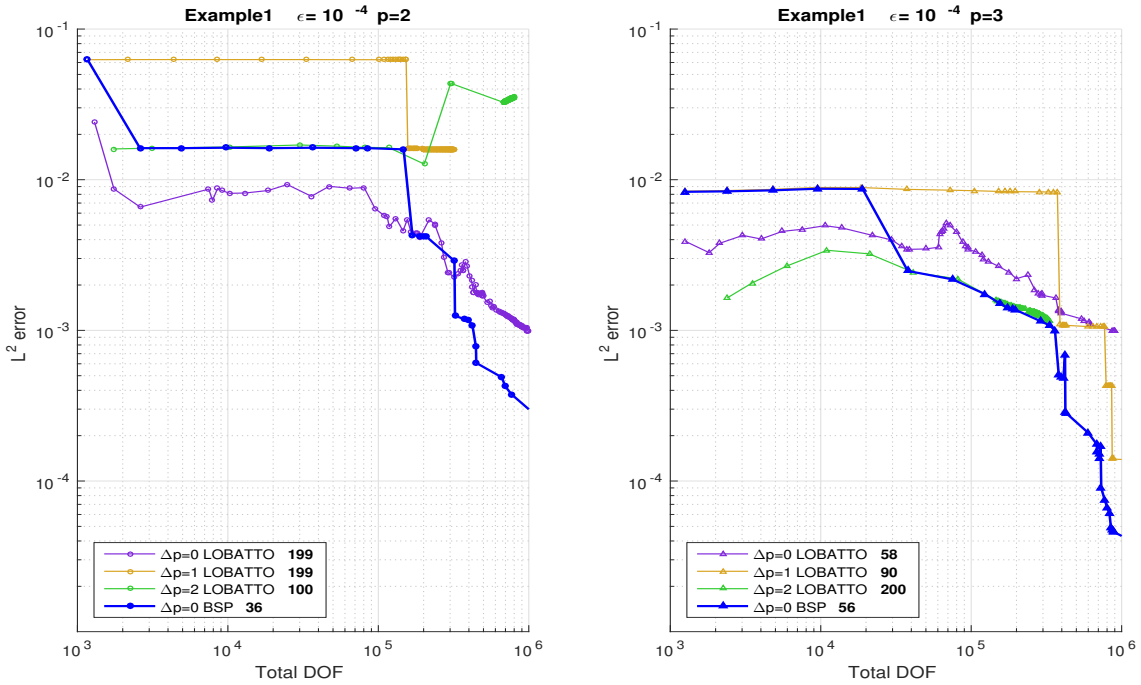

Figure 4.6: L^2-error vs DOF. $\epsilon = 10^{-4}$. $p = 2,\ p = 3$.

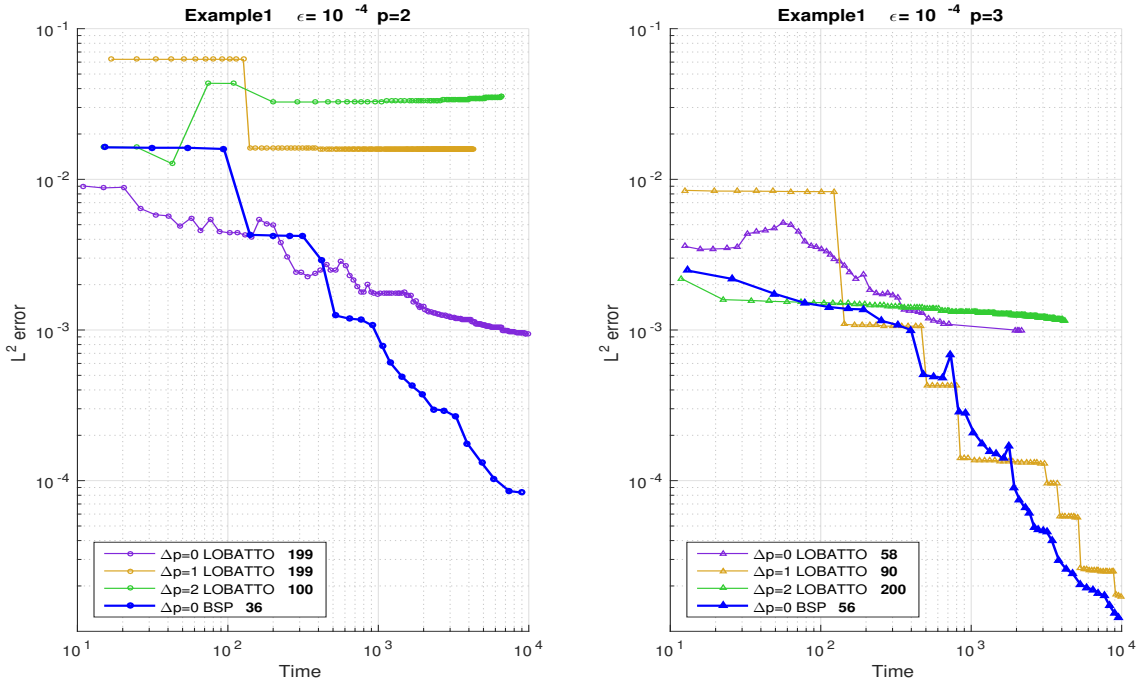

Figure 4.7: L^2-error vs time. $\epsilon = 10^{-4}$. $p = 2,\ p = 3$.

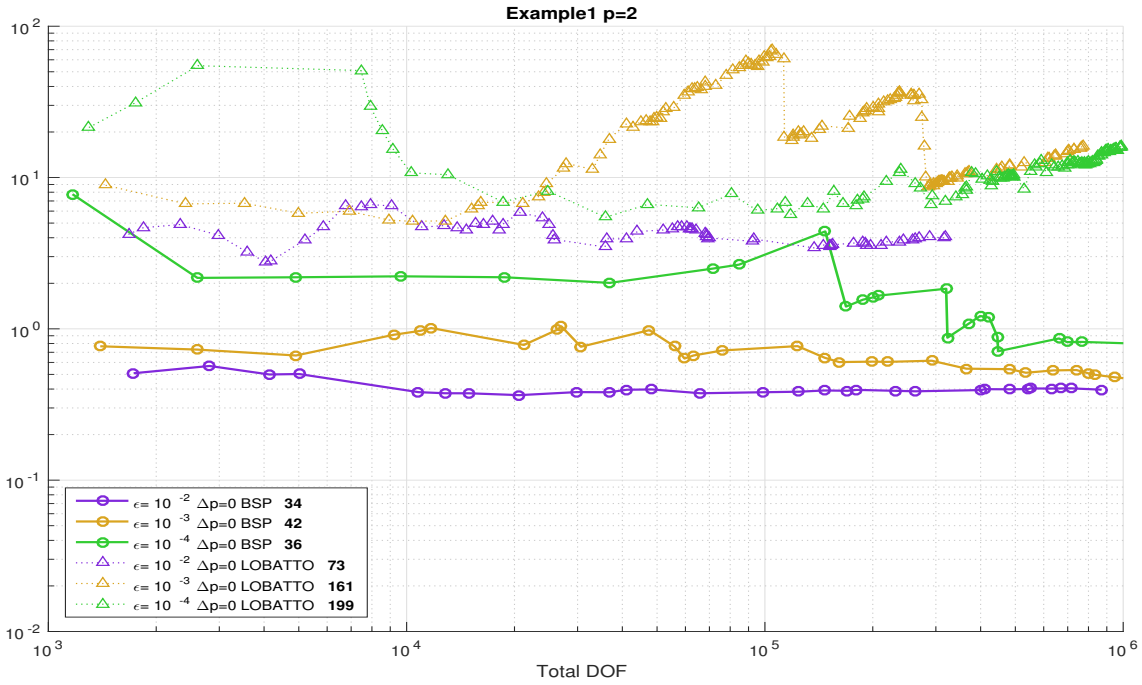

Figure 4.8: Ratio of L^2 and energy norms. $p = 2$, $\Delta p = 0$.

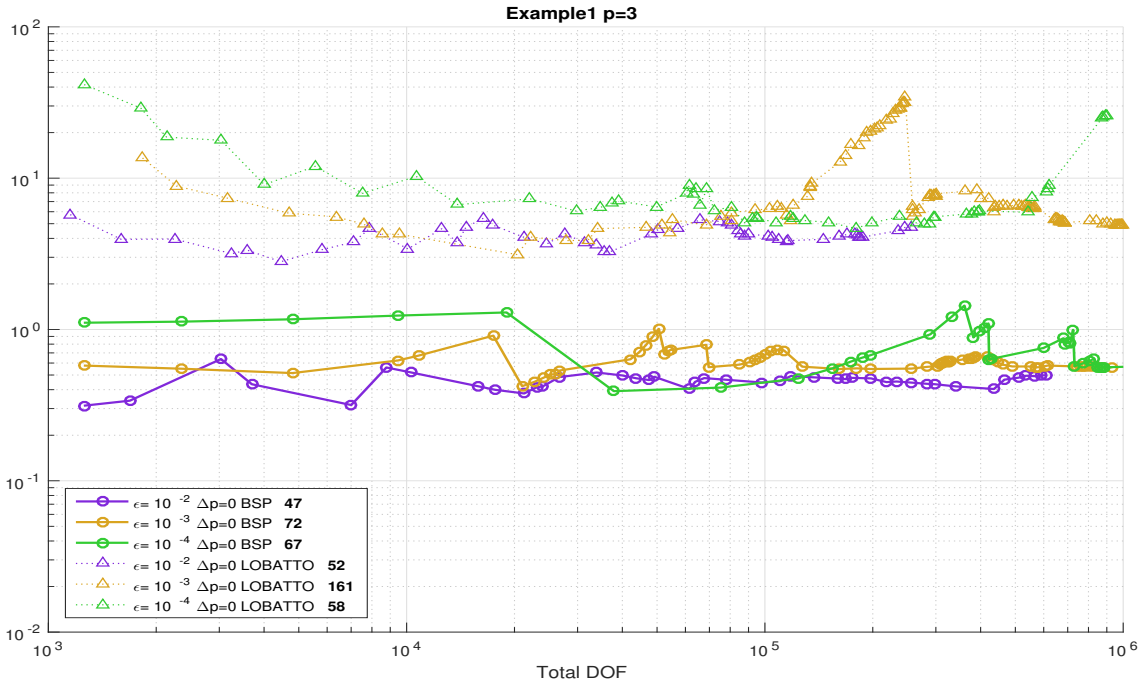

Figure 4.9: Ratio of L^2 and energy norms. $p = 3$, $\Delta p = 0$.

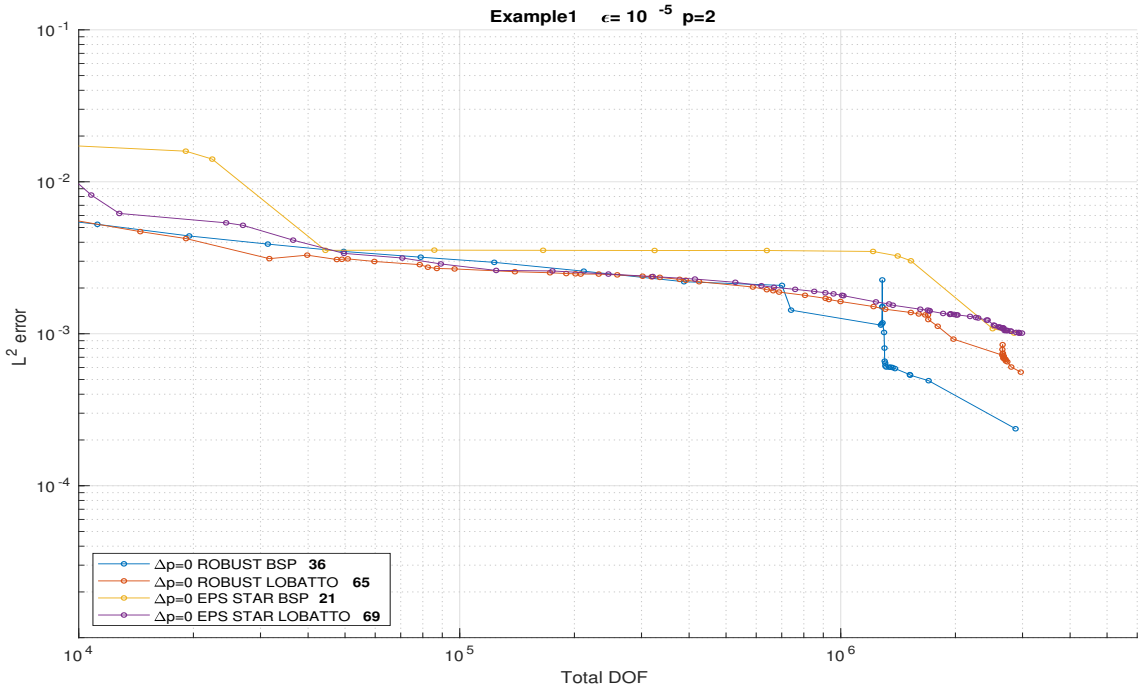

Figure 4.10: Comparison of EPS STAR norm with robust norm: L^2-error vs DOF. $\epsilon = 10^{-5}$. $p = 2$.

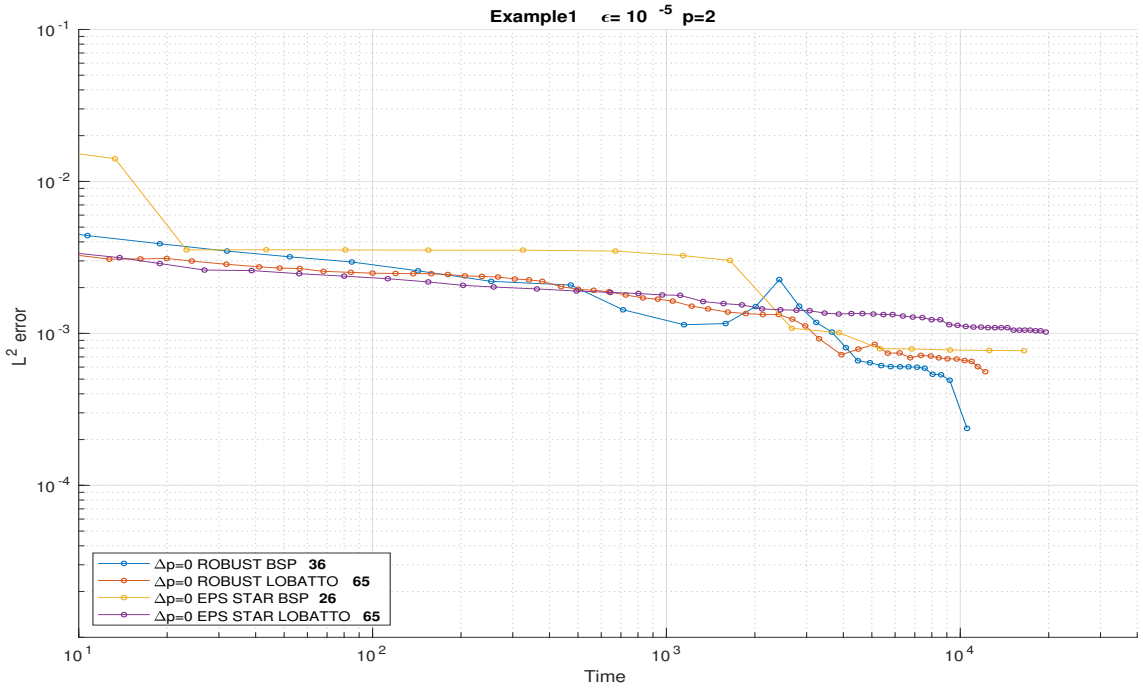

Figure 4.11: Comparison of EPS STAR norm with robust norm: L^2-error vs time. $\epsilon = 10^{-5}$. $p = 2$.

4.2 Example 2: Curved Streamlines

$$-\epsilon\Delta u + \nabla\cdot(\beta u) = f \quad \text{in} \quad \Omega = (0,1)^2, \qquad u = u_0 \quad \text{on}\,\Gamma \tag{4.3}$$

with $\beta = \nabla(e^x \sin y) = e^x(\sin y, \cos y)$ and f, u_0 such that:

$$u(x,y) = \arctan\left(\frac{1-|(x,y)|}{\epsilon}\right) \tag{4.4}$$

with inflow boundary at $x = 0$ and $y = 0$

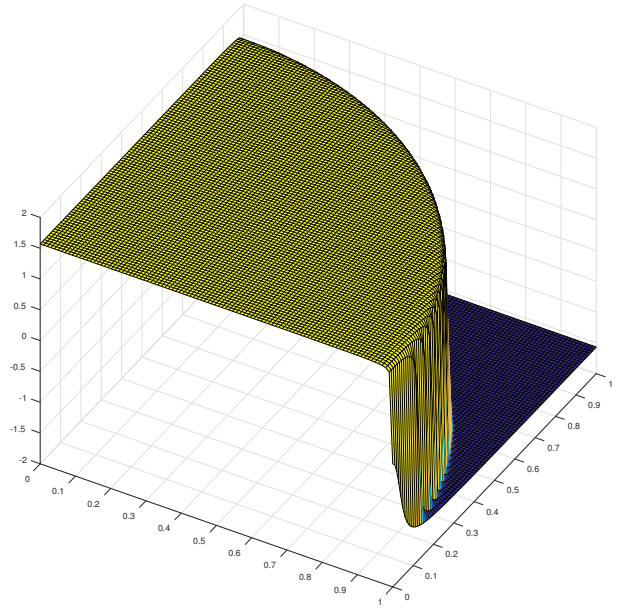

Figure 4.12: Exact solution, Curved Streamlines

For this problem, B-splines and Lobatto (for all Δp values) behave similarly. A few noticeable differences appear for p=2, where Lobatto with $\Delta p = 1$ shows slower convergence for the $\epsilon = 10^{-3}$ problem (Figure 4.16). And for the $\epsilon = 10^{-4}$ one (Figure 4.16) convergence stalls at about 3×10^5 DOF. Also in the $\epsilon = 10^{-4}$ case (Figure 4.18), we observe a slightly better L^2-error with B-splines at 10^6 DOF.

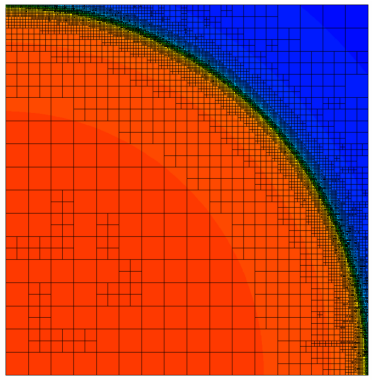

Figure 4.13: L^2-error vs DOF. $\epsilon = 10^{-2}$. P=3

Time registers (Figures 4.15, 4.17, 4.19) are only available for $\Delta p = 0$, and we observe no much difference, with B-splines only slightly better in some cases but slightly worse in others.

In Figures 4.20 and 4.21 the ratio of L^2-error for the field variables and the energy norm error is shown. In both cases, B-splines approach the same value when close to 10^6 DOF, being this value a whole order of magnitude lower than Lobatto polynomials. The ratio stays uniformly bounded from below, bu not from above, however it does converge.

4.2.1 EPS STAR norm

Results for $p = 2$ are shown in Figures 4.22 and 4.23. EPS STAR norm behaves considerably better than the robust norm. At 2 millioin DOF B-splines reach an error close to 0.1 with the EPS STAR norm and 0.3 for the robust norm. Corresponding values for Lobatto polynomials are 0.3 and 0.6. Also convergence over time is improved with the EPS STAR nomr, with B-splines trailing behind Lobatto.

With $p = 3$ (Figures 4.24 and 4.25) EPS STAR norm with B-splines observes improved convergence with respect to time and DOF. By 2×10^6 DOF, B-splines wiht EPS STAR norm reaches less than half the L^2-error than B-splines with robust norm and even better than Lobatto polynomials. It must also be noted that B-splines with EPS STAR norm shows the best behaviour when accounting for the computational time.

4.2.2 Continuation strategy

This strategy was briefly explored with Example 2, $p = 2$ and $\epsilon = 10^{-5}$. Constant c was set to 10^-3 and results are shown in Figures 4.26 and 4.27 along with the corresponding results using the robust norm for comparison. The continuation strategy presents slightly smaller error at 4 million degrees of freedom, however, a significant improvement is observed when comparing computation time. The time it takes the Continuation strategy to reach an L^2-error under 10^{-2} is about 1/5 the time it takes the ROBUST norm (for B-splines).

It is important to highlight that continuation results are actually comparable only in the region where ϵ_i has reached the ϵ_{orig} value, indicated in the figures to the right of the thick vertical line near 8×10^{-3} L^2-error (refer to subsection 2.7 for details on this strategy). Although these results are by no means conclusive, we consider this approach presents an interesting behavior, worth being communicated and more deeply studied in the future.

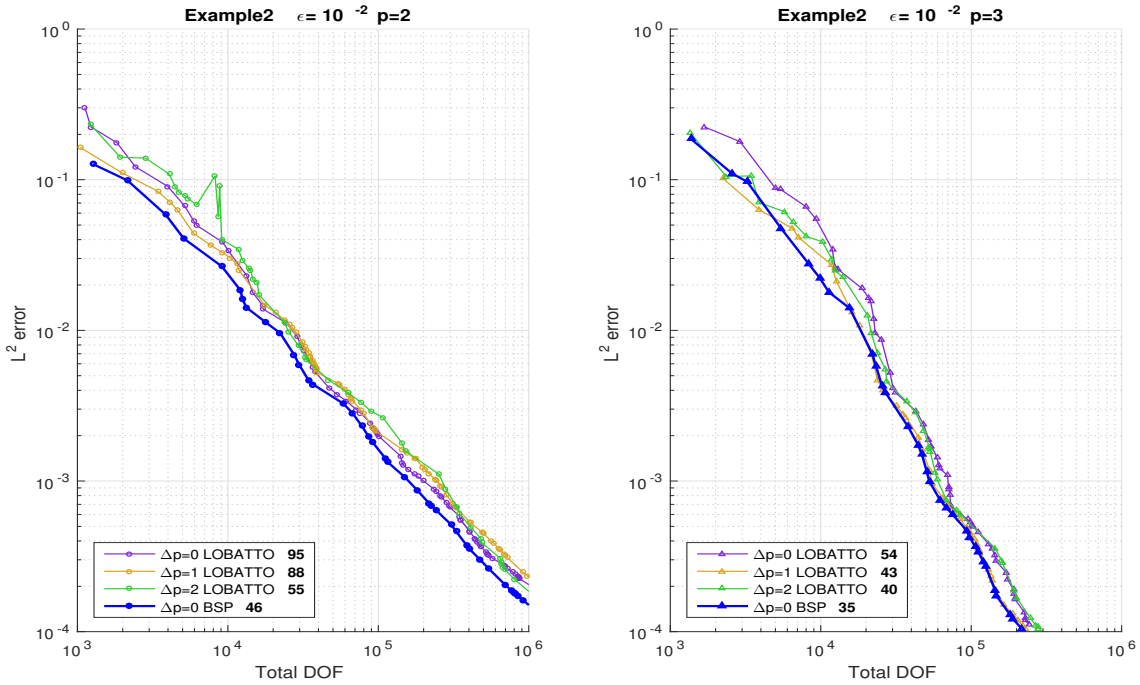

Figure 4.14: L^2-error vs DOF. $\epsilon = 10^{-2}$. $p = 2,\ p = 3$.

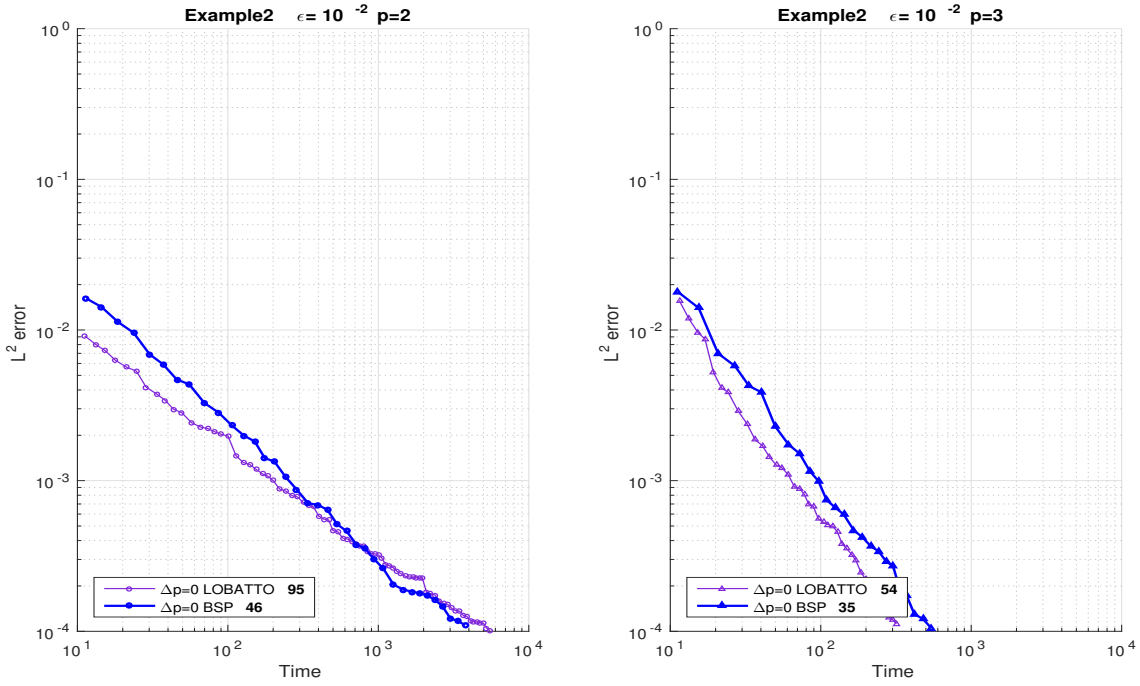

Figure 4.15: L^2-error vs time. $\epsilon = 10^{-2}$. $p = 2,\ p = 3$.

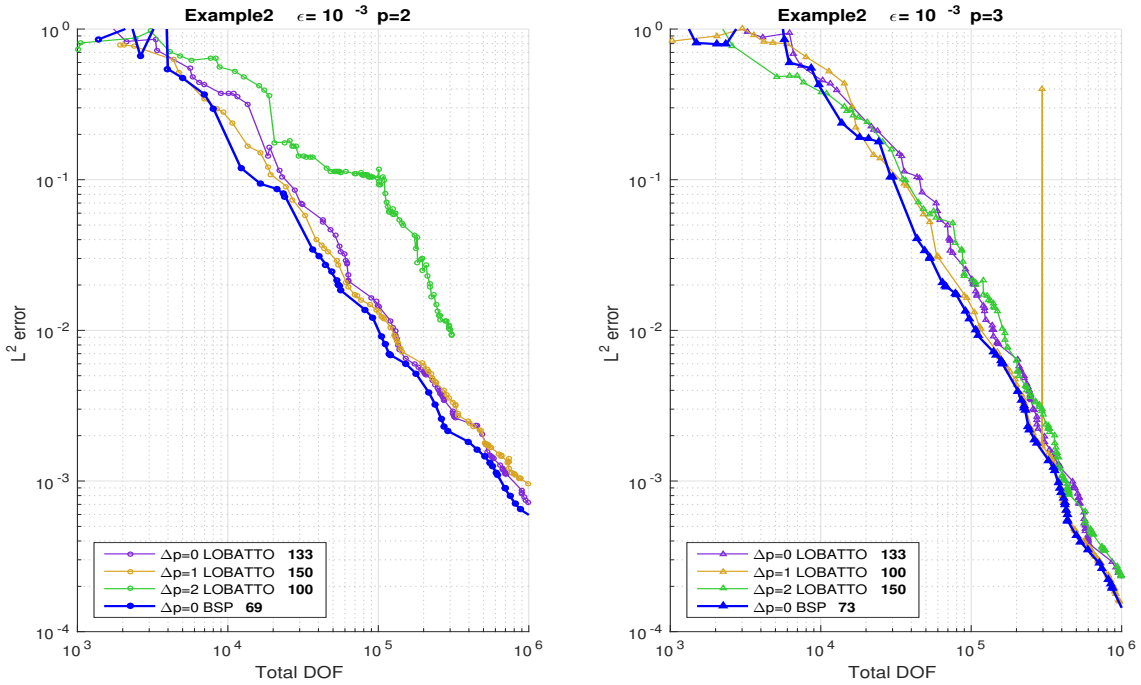

Figure 4.16: L^2-error vs DOF. $\epsilon = 10^{-3}$. $p = 2,\ p = 3$.

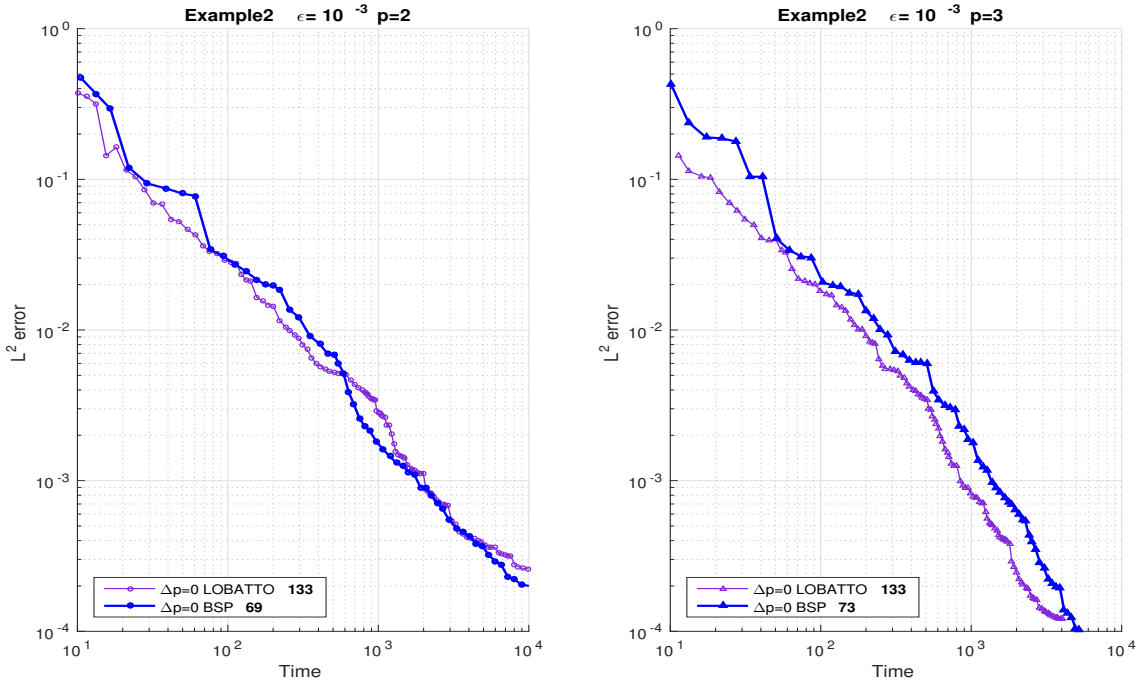

Figure 4.17: L^2-error vs time. $\epsilon = 10^{-3}$. $p = 2,\ p = 3$.

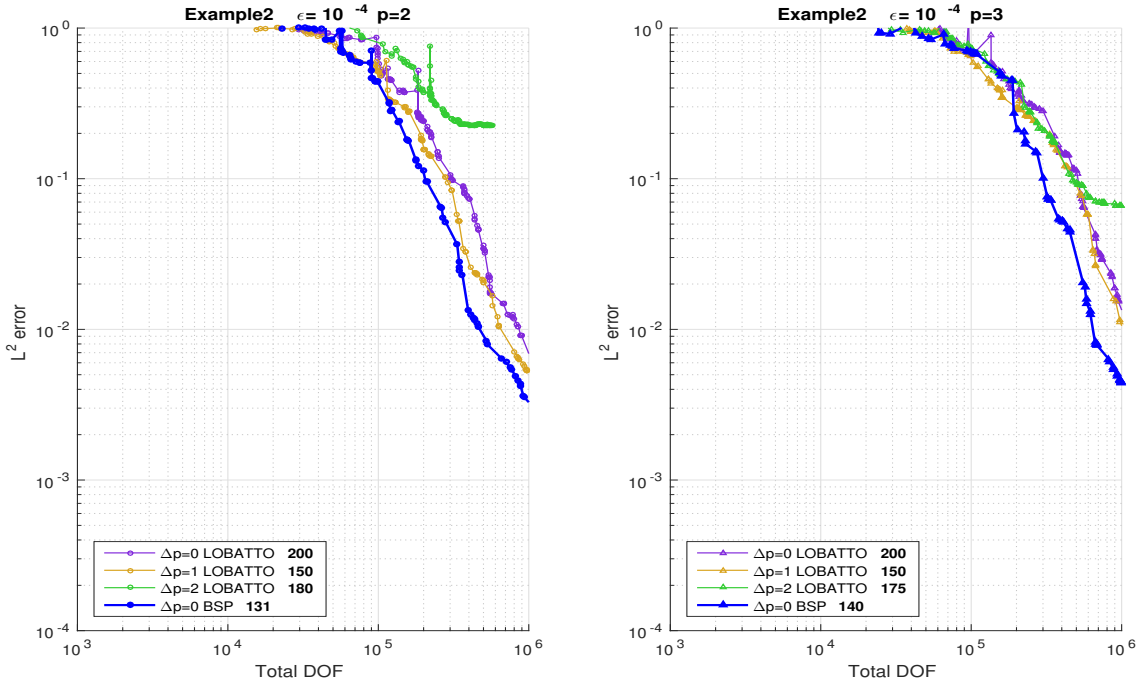

Figure 4.18: L^2-error vs DOF. $\epsilon = 10^{-4}$. $p = 2,\ p = 3$.

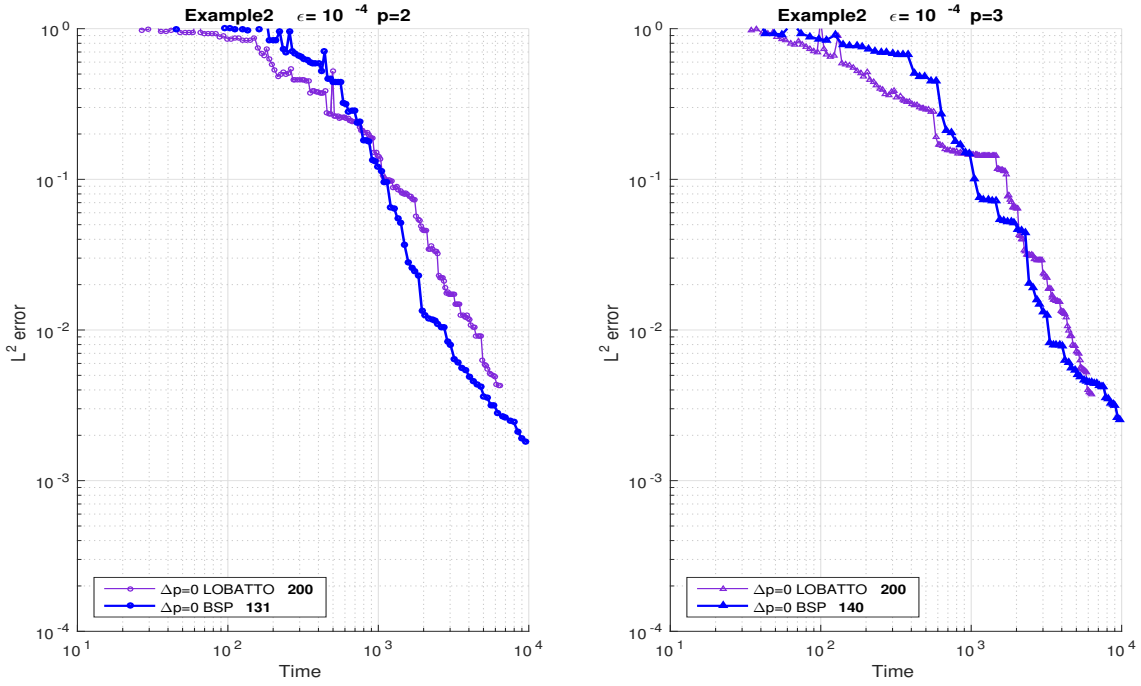

Figure 4.19: L^2-error vs time. $\epsilon = 10^{-4}$. $p = 2,\ p = 3$.

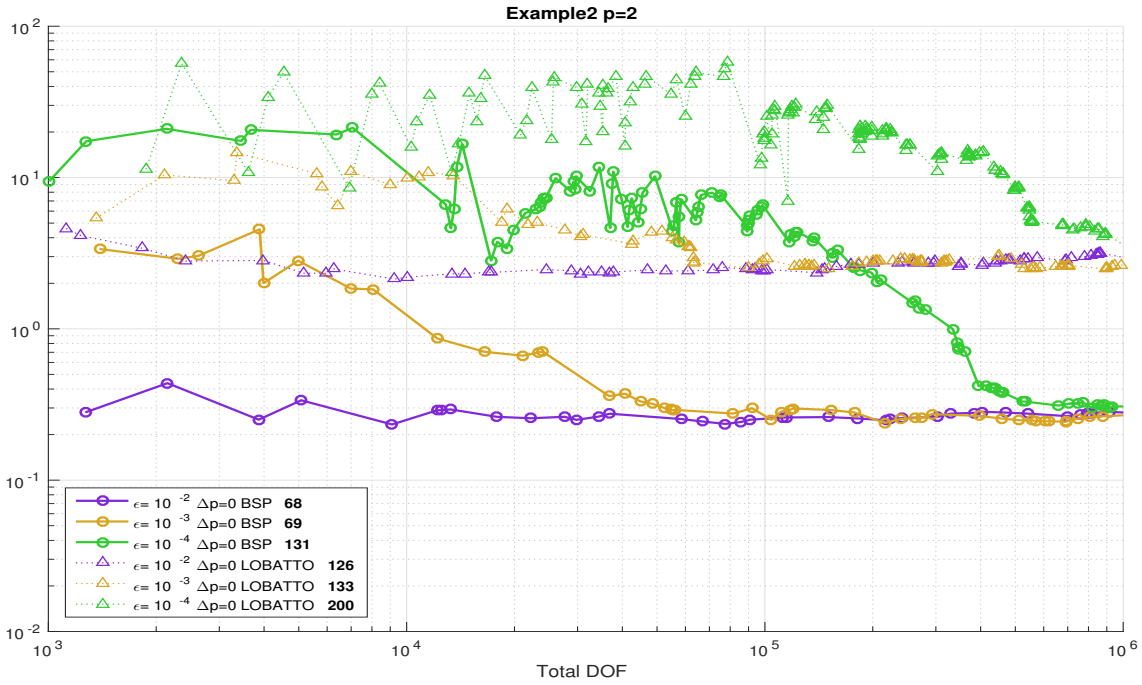

Figure 4.20: Ratio of L^2 and energy norms. $p = 2$, $\Delta p = 0$.

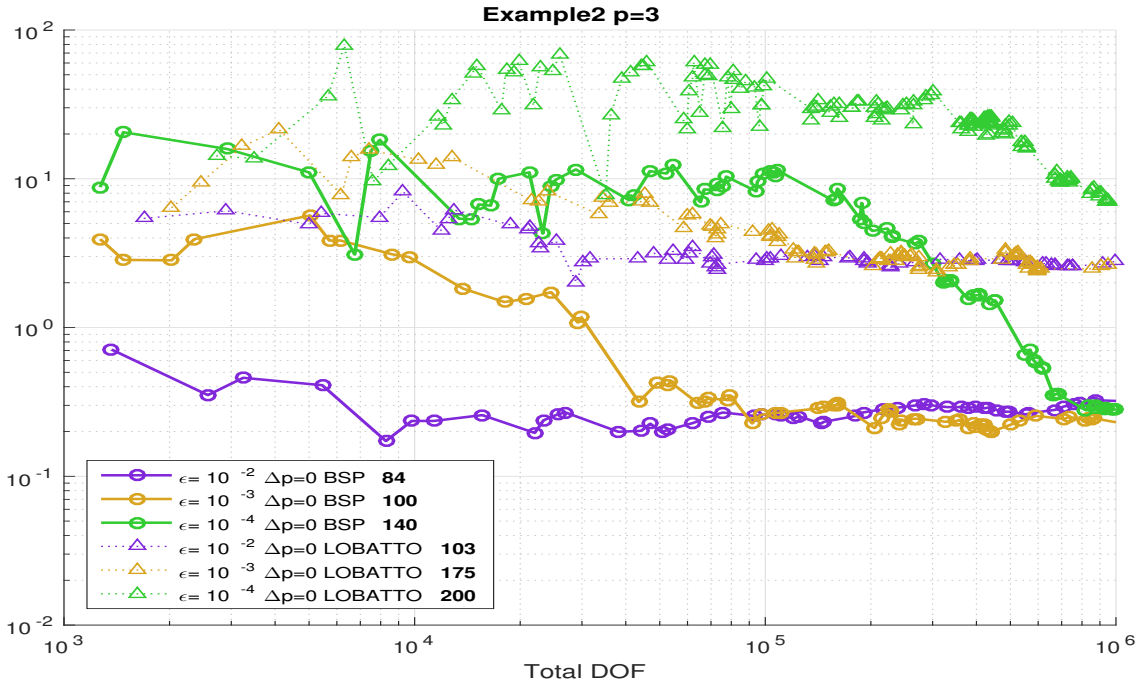

Figure 4.21: Ratio of L^2 and energy norms. $p = 3$, $\Delta p = 0$.

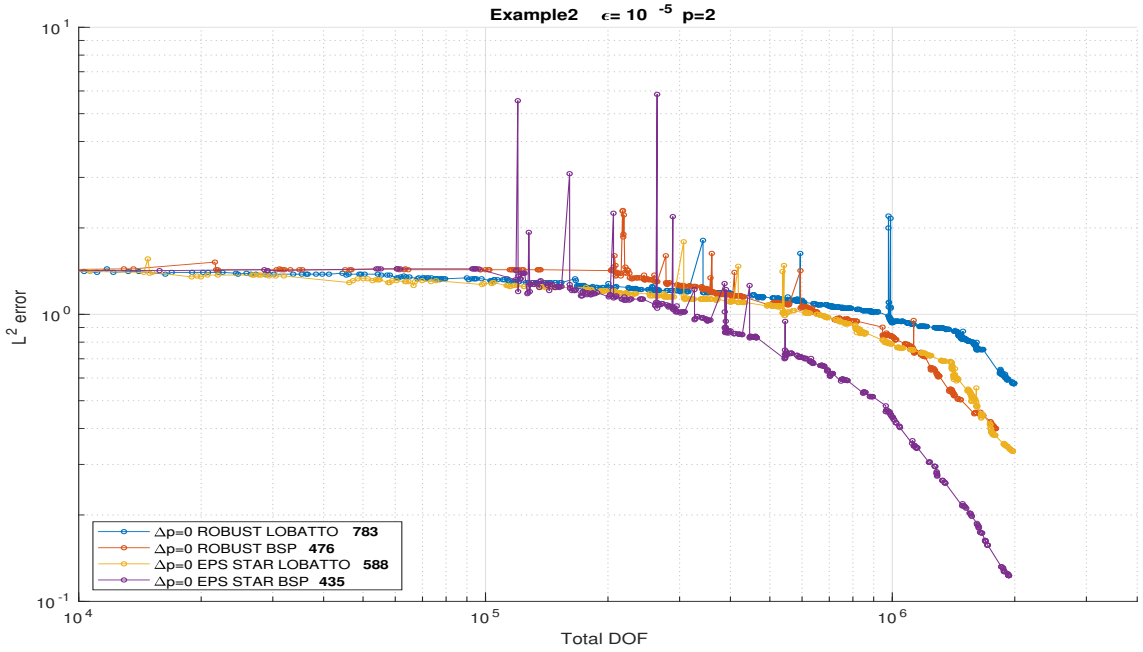

Figure 4.22: Comparison of EPS STAR norm with robust norm: L^2-error vs DOF. $\epsilon = 10^{-5}$. $p = 2$.

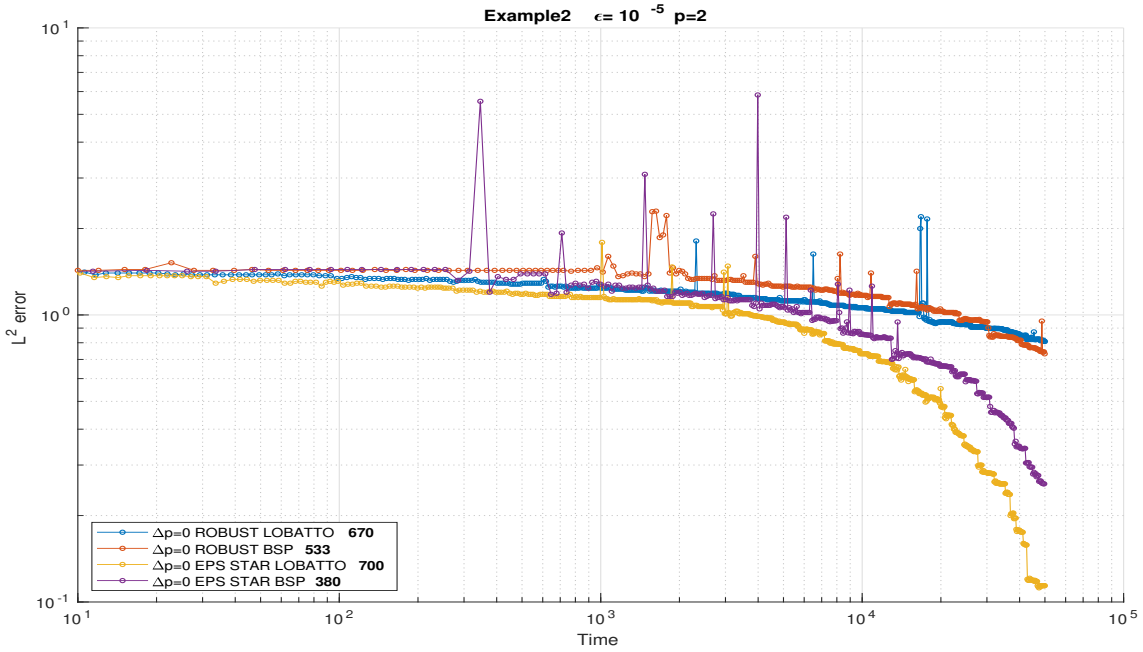

Figure 4.23: Comparison of EPS STAR norm with robust norm: L^2-error vs time. $\epsilon = 10^{-5}$. $p = 2$.

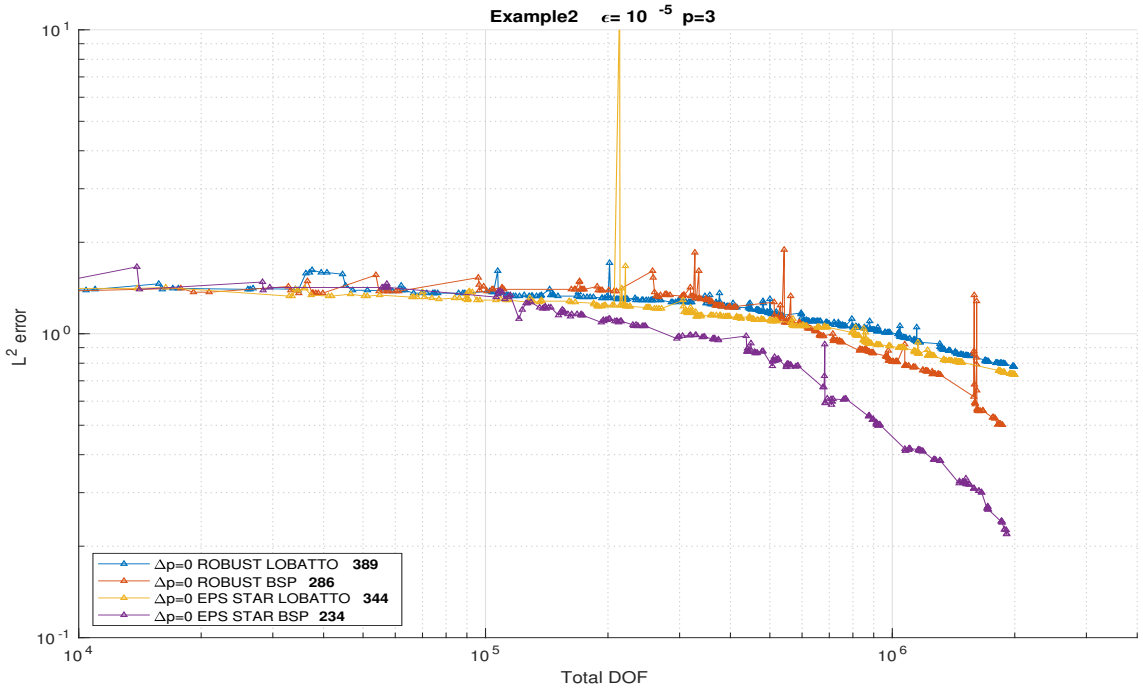

Figure 4.24: Comparison of EPS STAR norm with robust norm: L^2-error vs DOF. $\epsilon = 10^{-5}$. $p = 3$.

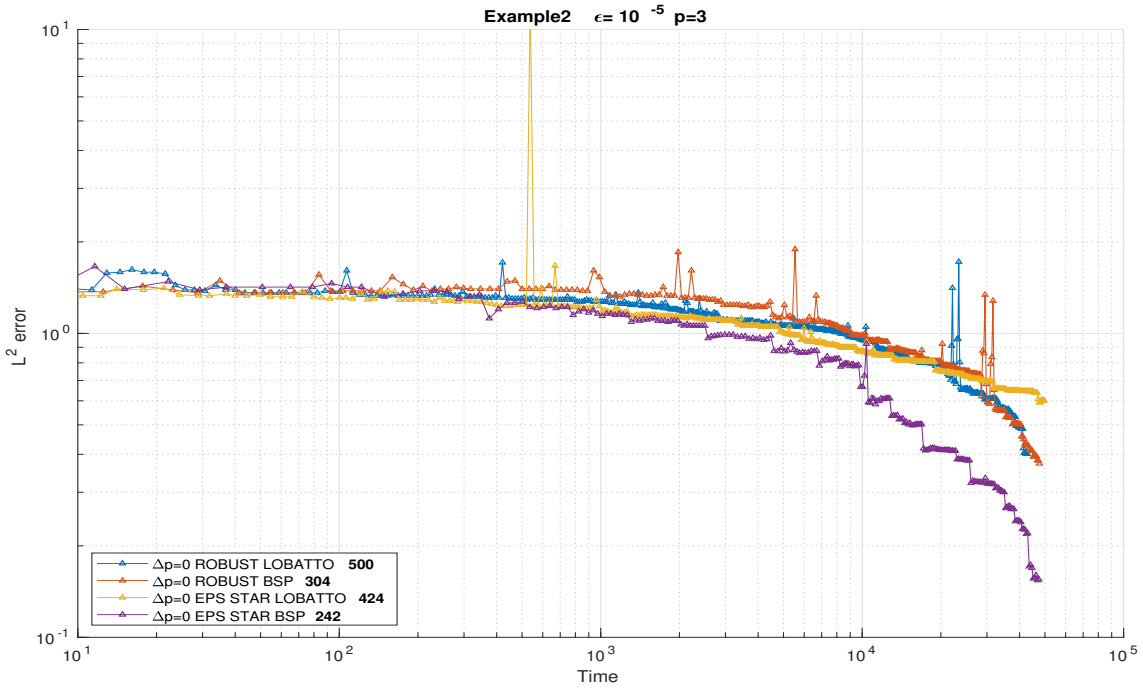

Figure 4.25: Comparison of EPS STAR norm with robust norm: L^2-error vs time. $\epsilon = 10^{-5}$. $p = 3$.

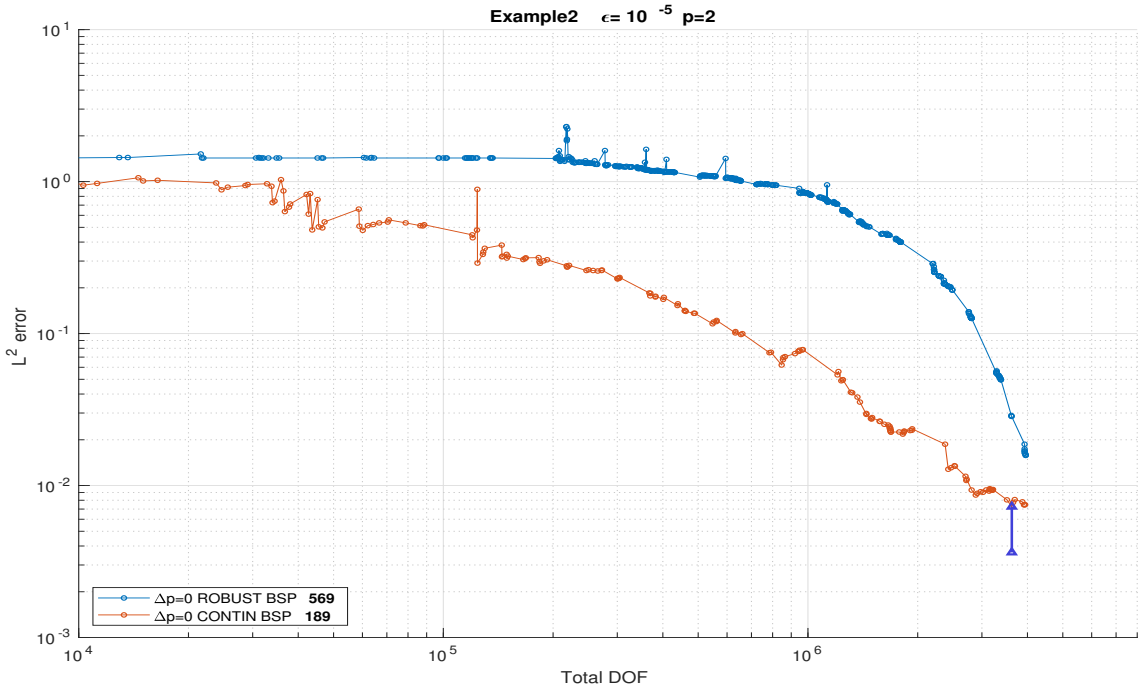

Figure 4.26: Continuation strategy with robust norm results for comparison. L^2-error vs DOF, $p = 2$. Remark: $\epsilon = 10^{-5}$ to the right of the thick vertical line for continuation graph.

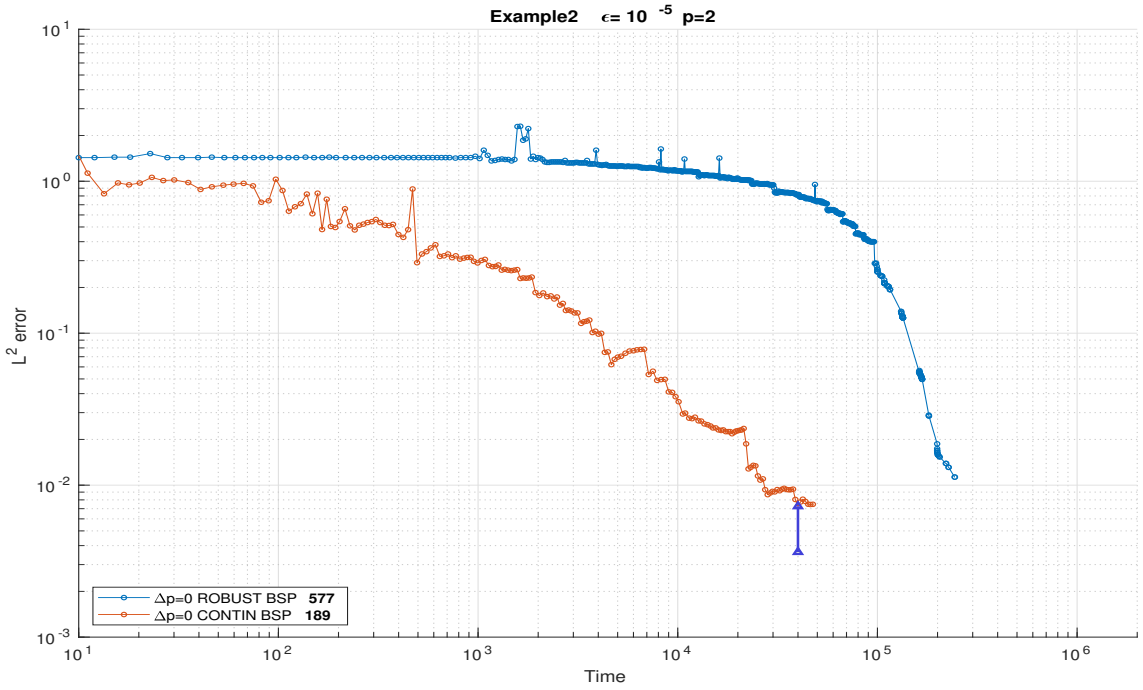

Figure 4.27: Continuation strategy with robust norm results for comparison. L^2-error vs time, $p = 2$. Remark: $\epsilon = 10^{-5}$ to the right of the thick vertical line for continuation graph.

5 Conclusions

We have presented alternative test spaces using B-splines, aiming for robustness of the DPG methodology with optimal test functions for singular perturbation problems. This has previously been numerically verified for specific problems by using specifically tailored norms; while the general (non-problem specific) *quasi-optimal* test norm has failed this purpose.

When using the *quasi-optimal* test norm, robustness is automatically given by the DPG methodology (in the ultraweak variational formulation), hence its importance. However, the difficulty shifts to the resolution of the optimal test functions since they develop boundary layers [9].

In order to tackle this hurdle, we propose using a Shishkin sub-mesh and B-splines to discretize the test space. Numerical results for two problems show B-splines to be a very suitable alternative to the commonly used Legendre and Lobatto polynomials (simply referred to as 'Lobatto' only). B-splines results are at worst similar to Lobatto, but much better in some cases (under adaptive refinements). To summarize the results we recall the following findings:

- Regarding **DOF**, B-splines show better convergence under any of the tested scenarios (although several were very similar to Lobatto).
- Regarding **time**, only for one test were Lobatto results (slightly) better than B-splines ($\Delta p = 2$ in Figure 4.3 left), whereas in a couple (Figures 4.5 and 4.7) B-splines were considerably better.
- Regarding the **L^2-error to energy error ratio**, in both examples we saw B-splines converging (for three different ϵ) to a constant value one order of magnitude smaller than that of Lobatto spaces, implying a better control of the L^2-error.
- Another remarkable outcome is that while for Lobatto test spaces we needed to explore several values of Δp according to each scenario, when using our alternative test spaces $\Delta p = 0$ was a good choice in all cases, therefore removing one variability factor in the method.

Recalling that the B-splines integration strategy herein used was rather non-optimal, leads us to believe that under more efficient integration schemes the time saving could be substantial, and therefore achieving a better overall performance than the conventional test space discretization.

We additionally proposed the EPS STAR norm (modified Adjoint Graph Norm) and a continuation strategy, as means to improve the solution of the most demanding conditions here presented (Example 2 with $\epsilon = 10^{-4}, 10^{-5}$). They appear to have good behaviour but certainly require further analysis and experimentation.

References

[1] Buffa, A., Rivas, J., Sangalli, G., and Vázquez, R. (2011). Isogeometric discrete differential forms in three dimensions. *SIAM Journal on Numerical Analysis*, 49(2):818–844.

[2] Carstensen, C., Demkowicz, L., and Gopalakrishnan, J. (2016). Breaking spaces and forms for the DPG method and applications including Maxwell equations. *Comput. Math. Appl.*, 72(3):494–522.

[3] Chan, J., Heuer, N., Bui-Thanh, T., and Demkowicz, L. (2014). A robust DPG method for convection-dominated diffusion problems II: Adjoint boundary conditions and mesh-dependent test norms. *Comput. Math. Appl.*, 67(4):771–795.

[4] Cottrell, J. A., Hughes, T. J., and Bazilevs, Y. (2009). *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons.

[5] Demkowicz, L. (2006). *Computing with hp Finite Elements. I. One and Two Dimensional Elliptic and Maxwell Problems*. Chapman & Hall/CRC Press, New York.

[6] Demkowicz, L. and Gopalakrishnan, J. (2014). An overview of the DPG method. In Feng, X., Karakashian, O., and Xing, Y., editors, *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations*, volume 157 of *The IMA Volumes in Mathematics and its Applications*, pages 149–180. Springer.

[7] Demkowicz, L. and Gopalakrishnan, J. (2015). Discontinuous Petrov-Galerkin (DPG) method. ICES Report 15-20, The University of Texas at Austin.

[8] Demkowicz, L., Gopalakrishnan, J., and Niemi, A. H. (2012). A class of discontinuous Petrov–Galerkin methods. Part III: Adaptivity. *Appl. Numer. Math.*, 62(4):396–427.

[9] Demkowicz, L. and Heuer, N. (2013). Robust DPG method for convection-dominated diffusion problems. *SIAM J. Numer. Anal.*, 51(5):2514–2537.

[10] Demkowicz, L., Kurtz, J., Pardo, D., Paszyński, M., Rachowicz, W., and Zdunek, A. (2007). *Computing with hp Finite Elements. II. Frontiers: Three Dimensional Elliptic and Maxwell Problems with Applications*. Chapman & Hall/CRC, New York.

[11] Fuentes, F., Keith, B., Demkowicz, L., and Nagaraj, S. (2015). Orientation embedded high order shape functions for the exact sequence elements of all shapes. *Comput. Math. Appl.*, 70(4):353–458.

[12] Niemi, A. H., Collier, N. O., and Calo, V. M. (2013). Automatically stable discontinuous Petrov–Galerkin methods for stationary transport problems: Quasi-optimal test space norm. *Computers & Mathematics with Applications*, 66(10):2096–2113.

[13] Oden, J. T. and Demkowicz, L. (2017). *Applied Functional Analysis*. Chapman and Hall/CRC.

[14] Rognes, M. E., Kirby, R. C., and Logg, A. (2009). Efficient assembly of H(div) and H(curl) conforming finite elements. *SIAM Journal on Scientific Computing*, 31(6):4130–4151.

[15] Zitelli, J., Muga, I., Demkowicz, L., Gopalakrishnan, J., Pardo, D., and Calo, V. M. (2011). A class of discontinuous Petrov–Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1d. *J. Comput. Phys.*, 230(7):2406–2432.