Locally Conservative Discontinuous Petrov-Galerkin Finite Elements for Fluid Problems

Truman Ellis, Leszek Demkowicz, and Jesse Chan

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

Abstract

1 Intoduction

The discontinuous Petrov-Galerkin (DPG) method with optimal test functions has been under active development for convection-diffusion type systems [4, 5, 6, 1, 7, 2, 8]. In this paper, we develop a theory for the locally conservative formulation of DPG for convection-diffusion type equations and supplement this with extensive numerical results.

1.1 Importance of Local Conservation

Locally conservative methods hold a special place for numerical analysts in the field of fluid dynamics. Perot[9] argues

> Accuracy, stability, and consistency are the mathematical concepts that are typically used to analyze numerical methods for partial differential equations (PDEs). These important tools quantify how well the mathematics of a PDE is represented, but they fail to say anything about how well the physics of the system is represented by a particular numerical method. In practice, physical fidelity of a numerical solution can be just as important (perhaps even more important to a physicist) as these more traditional mathematical concepts. A numerical solution that violates the underlying physics (destroying mass or entropy, for example) is in many respects just as flawed as an unstable solution.

The discontinuous Petrov-Galerkin finite element method has been described as least squares finite elements with a twist. The key difference is that least square methods seek to minimize the residual of the solution in the L^2 norm, while DPG seeks the minimization in a dual norm realized through the inverse Riesz map. Exact mass conservation has been an issue that has plagued least squares finite elements for a long time. Several approaches have been used to try to adress this. Chang and Nelson[3] developed the *restricted LSFEM*[3] by augmenting the least squares equations with a Lagrange multiplier explicitly enforcing mass conservation element-wise. Our conservative formulation of DPG takes a similar approach and both methods share similar negative of transforming a minimization method to a saddle-point problem.

1.2 DPG is a Minimum Residual Method

Roberts *et al.* presents a brief history and derivation of DPG with optimal test functions in [10]. We follow his derivation of the standard DPG method as a minumum residual method. Let U be the trial Hilbert space and V the test Hilbert space for a well-posed variational problem $b(u,v)=l(v)$. In operator form this is $Bu=l$, where $B:U\to V'$. We seek to minimize the residual for the discrete space $U_h\subset U$:

$$u_h=\arg\min_{u_h\in U_h}\frac{1}{2}\left\|Bu_h-l\right\|_{V'}^2\,. \tag{1}$$

Recalling that the Riesz operator $R_V:V\to V'$ is an isometry defined by

$$\langle R_V v,\delta v\rangle=(v,\delta v)_V\,,\quad\forall\delta v\in V,$$

we can use the Riesz inverse to minimize in the V-norm rather than its dual:

$$\frac{1}{2}\left\|Bu_h-l\right\|_{V'}^2=\frac{1}{2}\left\|R_V^{-1}(Bu_h-l)\right\|_V^2=\frac{1}{2}\left(R_V^{-1}(Bu_h-l),R_V^{-1}(Bu_h-l)\right)_V\,. \tag{2}$$

The first order optimality condition for (2) requires the Gâteaux derivative to be zero in all directions $\delta u\in U_h$, i.e.,

$$\left(R_V^{-1}(Bu_h-l),R_V^{-1}B\delta u\right)_V=0,\quad\forall\delta u\in U.$$

By definition of the Riesz operator, this is equivalent to

$$\left\langle Bu_h-l,R_V^{-1}B\delta u_h\right\rangle=0\quad\forall\delta u_h\in U_h\,. \tag{3}$$

Now, we can identify $v_{\delta u_h}:=R_V^{-1}B\delta u_h$ as the optimal test function for trial function δu_h. Define $T:=R_V^{-1}B:U_h\to V$ as the trial-to-test operator. Now we can rewrite (3) as

$$b(u_h,v_{\delta u_h})=l(v_{\delta u_h}). \tag{4}$$

The DPG method then is to solve (4) with optimal test functions $v_{\delta u_h}\in V$ that solve the auxiliary problem

$$(v_{\delta u_h},\delta v)_V=\langle R_V v_{\delta u_h},\delta v\rangle=\langle B\delta u_h,\delta v\rangle=b(\delta u_h,\delta v)\,,\quad\forall\delta v\in V. \tag{5}$$

Using a continuous test basis would result in a global solve for every optimal test function. Therefore DPG uses a discontinuous test basis which makes each solve element-local and much more computationally tractable. Of course, (5) still requires the inversion of the infinite-dimensional Riesz map, but approximating V by a finite dimensional V_h which is of a higher polynomial degree than U_h (hence "enriched space") works well in practice.

No assumptions have been made so far on the definition of the inner product on V. In fact, proper choice of $(\cdot,\cdot)_V$ can make the difference between a solid DPG method and one that suffers from robustness issues.

2 Analysis

We now proceed to develop a locally conservative formulation of DPG for convection-diffusion type problems, but there are a few terms that we need to define first. If Ω is our problem domain, then we can partition it into finite elements K such that

$$\overline{\Omega}=\bigcup_K\bar{K},\qquad K\text{ open},$$

with corresponding *skeleton* Γ_h and *interior skeleton* Γ_h^0,

$$\Gamma_h:=\bigcup_K\partial K\qquad\Gamma_h^0:=\Gamma_h-\Gamma.$$

We define broken Sobolev spaces element-wise:

$$\begin{aligned}H^1(\Omega_h)&:=\textstyle\prod_K H^1(K),\\ \boldsymbol{H}(\mathrm{div},\Omega_h)&:=\textstyle\prod_K\boldsymbol{H}(\mathrm{div},K).\end{aligned}$$

We also need the trace spaces:

$$
\begin{aligned}
H^{\frac{1}{2}}(\Gamma_h) \quad &:= \left\{\hat{v} = \{\hat{v}_K\} \in \textstyle\prod_K H^{1/2}(\partial K) \;:\; \exists v \in H^1(\Omega) : v|_{\partial K} = \hat{v}_K\right\}, \\
H^{-\frac{1}{2}}(\Gamma_h) \quad &:= \left\{\hat{\sigma}_n = \{\hat{\sigma}_{Kn}\} \in \textstyle\prod_K H^{-1/2}(\partial K) \;:\; \exists \boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div}, \Omega) : \hat{\sigma}_{Kn} = (\boldsymbol{\sigma} \cdot \boldsymbol{n})|_{\partial K}\right\},
\end{aligned}
$$

which are developed more precisely in [10].

2.1 Element Conservative Convection-Diffusion

Now that we have briefly outlined the abstract DPG method, let us apply it to the convection-diffusion equation. The strong form of the steady convection-diffusion problem with homogeneous Dirichlet boundary conditions reads

$$
\begin{cases}
\nabla \cdot (\boldsymbol{\beta} u) - \epsilon \Delta u &= f \quad \text{in } \Omega \\
u &= 0 \quad \text{on } \Gamma\,,
\end{cases}
$$

where u is the property of interest, $\boldsymbol{\beta}$ is the convection vector, and f is the source term. Nonhomogeneous Dirichlet and Neumann boundary conditions are straightforward but would add technicality to the following discussion. Let us write this as an equivalent system of first order equations:

$$
\begin{aligned}
\nabla \cdot (\boldsymbol{\beta} u - \boldsymbol{\sigma}) &= f \\
\frac{1}{\epsilon}\boldsymbol{\sigma} - \nabla u &= \mathbf{0}\,.
\end{aligned}
$$

If we then multiply the top equation by some scalar test function v and the bottom equation by some vector-valued test function τ, we can integrate by parts over each element K:

$$
\begin{aligned}
-(\boldsymbol{\beta} u - \boldsymbol{\sigma}, \nabla v)_K + ((\boldsymbol{\beta} u - \boldsymbol{\sigma}) \cdot \mathbf{n}, v)_{\partial K} &= (f, v)_K \\
\frac{1}{\epsilon}(\boldsymbol{\sigma}, \boldsymbol{\tau})_K + (u, \nabla \cdot \boldsymbol{\tau})_K - (u, \tau_n)_{\partial K} &= 0\,.
\end{aligned}
\tag{6}
$$

The discontinuous Petrov-Galerkin method refers to the fact that we are using discontinuous optimal test functions that come from a space differing from the trial space. It does not specify our choice of trial space. Nevertheless, many considerations of DPG in the literature [] associate DPG with the so-called "ultra-weak formulation." We will follow the same derivation for the convection-diffusion equation, but we emphasize that other formulations are available. Thus, we seek field variables $u \in L^2(K)$ and $\boldsymbol{\sigma} \in \mathbf{L^2}(K)$. Mathematically, this leaves their traces on element boundaries undefined, and in a manner similar to the hybridized discontinuous Galerkin method, we define new unknowns for trace $\hat{u}$ and flux $\hat{t}$. Applying these definitions to (6) and adding the two equations together, we arrive at our desired variational problem.

Find $\boldsymbol{u} := (u, \boldsymbol{\sigma}, \hat{u}, \hat{t}) \in \boldsymbol{U} := L^2(\Omega_h) \times \boldsymbol{L}^2(\Omega_h) \times H^{1/2}(\Gamma_h) \times H^{-1/2}(\Gamma_h)$ such that

$$
\begin{aligned}
\underbrace{-(\boldsymbol{\beta} u - \boldsymbol{\sigma}, \nabla v)_K + (\hat{t}, v)_{\partial K} + \frac{1}{\epsilon}(\boldsymbol{\sigma}, \boldsymbol{\tau})_K + (u, \nabla \cdot \boldsymbol{\tau})_K - (\hat{u}, \tau_n)_{\partial K}}_{b(\mathbf{u},\mathbf{v})} = \underbrace{(f, v)_K}_{l(\mathbf{v})} \qquad &\text{in } \Omega \tag{7} \\
u = 0 \qquad &\text{on } \Gamma \tag{8} \\
\tag{9}
\end{aligned}
$$

for all $\boldsymbol{v} := (v, \boldsymbol{\tau}) \in \boldsymbol{V} := H^1(\Omega_h) \times \boldsymbol{H}(\mathrm{div}, \Omega_h)$.

Let $\boldsymbol{U}_h := U_h \times \boldsymbol{S}_h \times \hat{U}_h \times \hat{F}_h \subset L^2(\Omega_h) \times \boldsymbol{L}^2(\Omega_h) \times H^{\frac{1}{2}}(\Gamma_h) \times H^{-\frac{1}{2}}(\Gamma_h)$ be a finite-dimensional subspace, and let $\boldsymbol{u}_h := (u_h.\boldsymbol{\sigma}_h, \hat{u}_h\hat{t}_h) \in \boldsymbol{U}_h$ be a group variable. The element conservative DPG scheme is derived from the Lagrangian:

$$
L(\boldsymbol{u}_h, \lambda_k) = \frac{1}{2}\left\|R_V^{-1}(b(\boldsymbol{u}_h, \cdot) - (f, \cdot))\right\|_{\boldsymbol{V}}^2 - \sum_K \lambda_K (b(\boldsymbol{u}_h, (1_K, \mathbf{0})) - l((1_K, \mathbf{0})))\,,
\tag{10}
$$

where $(1_K, \mathbf{0})$ is the test function in which $v = 1$ on element K and 0 elsewhere and $\boldsymbol{\tau} = \mathbf{0}$ everywhere.

Taking the Gâteaux derivatives as before, we arrive at the following system of equations:

$$\left\{ \begin{aligned} b(\boldsymbol{u}_h, T(\delta\boldsymbol{u}_h)) - \textstyle\sum_K \lambda_K b(\boldsymbol{u}_h, (1_K, \mathbf{0})) &= (f, T(\delta\boldsymbol{u}_h)) \quad \forall \delta\boldsymbol{u}_h \in \boldsymbol{U}_h \\ b(\boldsymbol{u}_h, (1_K, \mathbf{0})) &= (f, (1_K, \mathbf{0})) \quad \forall K\,, \end{aligned} \right. \tag{11}$$

where $T := R_V^{-1} B : \boldsymbol{U}_h \to \boldsymbol{V}$ is the same trial-to-test operator as in the original formulation.

Since the trial-to-test operator selectively produces test functions in $H^1(\Omega_h)$ and $\boldsymbol{H}(\text{div}, \Omega_h)$, denote the result $v_{\delta\boldsymbol{u}_h}$ when $T(\delta u_h) \in H^1(\Omega_h)$ and $\boldsymbol{\tau}_{\delta\boldsymbol{u}_h}$ when $T(\delta u_h) \in \boldsymbol{H}(\text{div}, \Omega_h)$. Then, putting (11) into more concrete terms for convection-diffusion, we get:

$$\left\{ \begin{aligned} &-(\boldsymbol{\beta} u - \boldsymbol{\sigma}, \nabla v_{\delta\boldsymbol{u}_h}) + \langle \hat{t}, v_{\delta\boldsymbol{u}_h} \rangle + \tfrac{1}{\epsilon}(\boldsymbol{\sigma}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}) + (u, \nabla \cdot \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}) - \langle \hat{u}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h} \cdot \boldsymbol{n} \rangle \\ &\qquad\qquad\qquad - \textstyle\sum_K \lambda_K (\delta\hat{t}, (1_K, \mathbf{0})) = (f, v_{\delta\boldsymbol{u}_h}) \qquad \forall \delta\boldsymbol{u}_h \in \boldsymbol{U}_h \\ &\qquad\qquad\qquad\qquad\qquad \langle \hat{t}, (1_K, \mathbf{0}) \rangle = (f, (1_K, \mathbf{0})) \quad \forall K\,. \end{aligned} \right. \tag{12}$$

The optimal test functions are determined by solving local problems determined by the choice of test norm. For illustrative purposes we choose the robust test norm developed by Chan *et al.* [2]. With this choice, we get our local problem:

Find $v_{\delta\boldsymbol{u}_h} \in H^{-1}(K)$, $\boldsymbol{\tau}_{\delta\boldsymbol{u}_h} \in \boldsymbol{H}(\text{div}, K)$ such that:

$$\begin{aligned} (\nabla \cdot \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}, \nabla \cdot \delta\boldsymbol{\tau})_K &+ \min\left\{\frac{1}{\epsilon}, \frac{1}{|K|}\right\} (\boldsymbol{\tau}_{\delta\boldsymbol{u}_h}, \delta\boldsymbol{\tau})_K + \epsilon(\nabla v_{\delta\boldsymbol{u}_h}, \nabla \delta v)_K + (\boldsymbol{\beta} \cdot \nabla v_{\delta\boldsymbol{u}_h}, \boldsymbol{\beta} \cdot \nabla \delta v)_K \\ &+ \alpha \min\left\{\sqrt{\frac{\epsilon}{|K|}}, 1\right\} (v_{\delta\boldsymbol{u}_h}, \delta v)_K = b(\delta\boldsymbol{u}_h, (\delta v, \delta\boldsymbol{\tau})) \qquad \forall \delta v \in H^{-1}(K),\ \delta\boldsymbol{\tau} \in \boldsymbol{H}(\text{div}, K)\,, \end{aligned} \tag{13}$$

where, typically, $\alpha = 1$.

We can decompose the discrete flux space $\hat{F}_h$ into element conserving fluxes and some algebraic complement:

$$\hat{F}_h = \hat{F}_h^c \oplus \hat{F}_h^{nc}\,, \quad \hat{F}_h^c := \{\hat{t} : \langle \hat{t}, (1_K, \mathbf{0}) \rangle = 0 \quad \forall K\}\,. \tag{14}$$

This may be accomplished by employing for fluxes in $\hat{F}_h^c$ traces of curls, $\nabla \times \phi$.

In practice, we solve (12) as a fully coupled system of equations, but for didactic purposes, we can examine it one piece at a time. If we restrict $(12)_1$ to $\delta\hat{t} \in \hat{F}_h^{nc}$, the Lagrange multiplier term drops out and we left with a system that we can solve for u_h, $\boldsymbol{\sigma}_h$, $\hat{u}_h$, and $\hat{t}_h^c \in \hat{F}_h^c$.

Restricting ourselves to $\delta\hat{t} \in \hat{F}_h^c$ in $(12)_1$, we obtain a system of equations that can be solved for u_h, $\boldsymbol{\sigma}_h$, $\hat{u}$, and $\hat{t} \in \hat{F}_h^c$. The remaining fluxes, $\hat{t} \in \hat{F}_h^{nc}$ are determined from $(12)_2$. Finally, testing with $\hat{t} \in \hat{F}_h^{nc}$ in $(12)_1$, we obtain a system of equations that can be solved for Lagrange multipliers λ_K.

Additionally, we can pass in local problem (13) with $\alpha \to 0$. The fact that the test functions will be determined then up to a constant does not matter, for $\delta\hat{t} \in \hat{F}_h^c$, equation $(12)_1$ is orthogonal to constants. Mathematically, we are dealing with equivalence classes of functions, but in order to obtain a single function that we can deal with numerically, we replace the alpha term with a zero mean scaling condition to obtain the new test norm,

$$\begin{aligned} (\nabla \cdot \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}, \nabla \cdot \delta\boldsymbol{\tau})_K &+ \min\left\{\frac{1}{\epsilon}, \frac{1}{|K|}\right\} (\boldsymbol{\tau}_{\delta\boldsymbol{u}_h}, \delta\boldsymbol{\tau})_K \\ &+ \epsilon(\nabla v_{\delta\boldsymbol{u}_h}, \nabla \delta v)_K + (\boldsymbol{\beta} \cdot \nabla v_{\delta\boldsymbol{u}_h}, \boldsymbol{\beta} \cdot \nabla \delta v)_K + \frac{1}{|K|^2} \int_K v_{\delta\boldsymbol{u}_h} \delta v\,, \end{aligned} \tag{15}$$

where the $\frac{1}{|K|^2}$ coefficient is needed to make the zero mean term scale correctly with h.

It is convenient to be able to take $\alpha \to 0$ as we will see in some later numerical experiments.

2.2 Proof of Convergence

We cannot directly use Brezzi's theorem, but we can reproduce his argument. First of all, variational formulation (7) is equivalent to:

Find $\boldsymbol{u} \in \boldsymbol{U}$ such that

$$b(\boldsymbol{u}, T(\delta\boldsymbol{u})) = (f, T(\delta\boldsymbol{u})) \quad \forall \boldsymbol{v} \in \boldsymbol{V}\,. \tag{16}$$

This is because the trial-to-test operator $T = R_V^{-1}B$ is an isomorphism.

Restricting ourselves to $\delta\hat{t}_h \in \hat{F}_h^c$, and subtracting (16) from (12), we have the Galerkin orthogonality condition:

$$\begin{aligned}\frac{1}{\epsilon}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}) + (u_h - u, \nabla\cdot\boldsymbol{\tau}_{\delta\boldsymbol{u}_h}) - \langle \hat{u}_h - \hat{u}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}\cdot\boldsymbol{n}\rangle \\ - (\boldsymbol{\beta}(u_h - u) - (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}), \nabla v_{\delta\boldsymbol{u}_h}) + \langle \hat{t}_h - \hat{t}, v_{\delta\boldsymbol{u}_h}\rangle \\ \forall(\delta u_h, \delta\boldsymbol{\sigma}_h, \delta\hat{u}_h, \delta\hat{t}_h) \in U_h \times \boldsymbol{S}_h \times \hat{U}_h \times \hat{F}_h^c\,.\end{aligned} \tag{17}$$

We begin with an equivalent of Brezzi's *discrete inf-sup in kernel condition.*

Lemma 1. *The following condition holds:*

$$\sup_{\substack{(\delta u_h, \delta\boldsymbol{\sigma}_h, \delta\hat{u}, \delta\hat{t})\in \\ U_h\times\boldsymbol{S}_h\times\hat{U}_h\times\hat{F}_h^c}} \frac{\left|b((u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h), (v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}))\right|}{\|(v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h})\|} \geq \alpha\left\|(u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h)\right\|$$

$$\forall(u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h) \in U_h \times \boldsymbol{S}_h \times \hat{U}_h \times \hat{F}_h^c\,. \tag{18}$$

Proof. By the very construction of the trial-to-test operator, the supremum on the left is equal to

$$\left\|T(u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h)\right\|_V = \left\|R_V^{-1}B(u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h)\right\|_V = \left\|B(u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h)\right\|_{V'} \tag{19}$$

The condition follows now from the fact that operator B is bounded below, i.e. bilinear form $b((u_h, \boldsymbol{\sigma}_h, \hat{u}_h, \hat{t}_h), (v, \boldsymbol{\tau}))$ satisfies the inf-sup condition. □

Assume additionally that flux decomposition (14) is orthogonal. We begin by decomposing $\hat{t}_h$,

$$\hat{t}_h = \hat{t}_h^c + \hat{t}_h^{nc}\,. \tag{20}$$

Next, we use the triangle inequality,

$$\begin{aligned}\left\|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \hat{u} - \hat{u}_h, \hat{t} - \hat{t}_h)\right\| &= \left\|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \hat{u} - \hat{u}_h, \hat{t} - (\hat{t}_h^c + \hat{t}_h^{nc}))\right\| \\ &\leq \left\|(u - w_h, \boldsymbol{\sigma} - \boldsymbol{s}_h, \hat{u} - \hat{w}_h, \hat{t} - (\hat{r}_h^c + \hat{t}_h^{nc}))\right\| \\ &\quad + \left\|(w_h - u_h, \boldsymbol{s}_h - \boldsymbol{\sigma}_h, \hat{w}_h - \hat{u}_h, \hat{r}_h^c - \hat{t}_h^c)\right\|\end{aligned} \tag{21}$$

where $w_h \in U_h$, $\boldsymbol{s}_h \in \boldsymbol{S}_h$, $\hat{w}_h \in \hat{U}_h$, and $\hat{r}_h^c \in \hat{F}_h$ are arbitrary. The inf-sup in kernel condition now implies:

$$\left\|(w_h - u_h, \boldsymbol{s}_h - \boldsymbol{\sigma}_h, \hat{w}_h - \hat{u}_h, \hat{r}_h^c - \hat{t}_h^c)\right\| \tag{22}$$

$$\begin{aligned}&\leq \frac{1}{\alpha} \sup_{\substack{(\delta u_h, \delta\boldsymbol{\sigma}_h, \delta\hat{u}, \delta\hat{t})\in \\ U_h\times\boldsymbol{S}_h\times\hat{U}_h\times\hat{F}_h^c}} \frac{\left|b((w_h - u_h, \boldsymbol{s}_h - \boldsymbol{\sigma}_h, \hat{w}_h - \hat{u}_h, \hat{r}_h^c - \hat{t}_h^c), (v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}))\right|}{\|(v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h})\|} \\ &\leq \frac{1}{\alpha} \sup_{\substack{(\delta u_h, \delta\boldsymbol{\sigma}_h, \delta\hat{u}, \delta\hat{t})\in \\ U_h\times\boldsymbol{S}_h\times\hat{U}_h\times\hat{F}_h^c}} \frac{\left|b((w_h - u_h, \boldsymbol{s}_h - \boldsymbol{\sigma}_h, \hat{w}_h - \hat{u}_h, \hat{r}_h^c + \hat{t}_h^{nc} - \hat{t}), (v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h}))\right|}{\|(v_{\delta\boldsymbol{u}_h}, \boldsymbol{\tau}_{\delta\boldsymbol{u}_h})\|} \\ &\leq \frac{1}{\alpha}\left\|(w_h - u_h, \boldsymbol{s}_h - \boldsymbol{\sigma}_h, \hat{w}_h - \hat{u}_h, \hat{t} - (\hat{r}_h^c + \hat{t}_h^{nc}))\right\|\end{aligned} \tag{23}$$

where the second step follows from the Galerkin orthogonality condition (17), and the last step is a consequence of the minimum energy extension norm for fluxes. Cobining with (21), we get the final stability result:

$$\left\|(u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\hat{u}-\hat{u}_h,\hat{t}-\hat{t}_h)\right\| \leq \left(1+\frac{1}{\alpha}\right) \inf_{\substack{(w_h,\boldsymbol{s}_h,\hat{w},\hat{r}^c_h)\in\\ U_h\times\boldsymbol{S}_h\times\hat{U}_h\times\hat{F}^c_h}} \left\|(u-w_h,\boldsymbol{\sigma}-\boldsymbol{s}_h,\hat{u}-\hat{w},\hat{t}-(\hat{r}^c_h+\hat{t}^{nc}_h))\right\| . \tag{24}$$

3 Numerical Experiments

We explore each of the following experiments with two sets of test norms: the standard graph norm, and the aforementioned robust norm.

3.1 Erickson-Johnson Model Problem

The Erickson-Johnson problem is one of the few convection-diffusion problems with a known analytical solution. Take a domain $\Omega=[0,1]^2$, with $\boldsymbol{\beta}=(1,0)^T$, $f=0$ and boundary conditions $\hat{t}=u_0$ when $\beta_n\leq 0$, and u_0 is the trace of the exact solution. For $n=1,2,\cdots$, let $\lambda_n=n^2*\pi^2*\epsilon$, $r_n=\frac{1+\sqrt{1+4\epsilon\lambda_n}}{2\epsilon}$, and $s_n=\frac{1-\sqrt{1+4\epsilon\lambda_n}}{2\epsilon}$, the exact solution is

$$u(x,y)=C_0+\sum_{n=1}^{\infty}C_n\frac{\exp(s_n(x-1))-\exp(r_n(x-1))}{r_n\exp(-s_n)-s_n\exp(-s_n)}\cos(n\pi y)\,. \tag{25}$$

Taking $\epsilon=0.01$, $C_1=1$, and $C_{n\neq 1}=0$, the exact solution looks like Figure 1. We solve this problem with both conservative and nonconservative formulations of DPG with both graph and robust test norms.

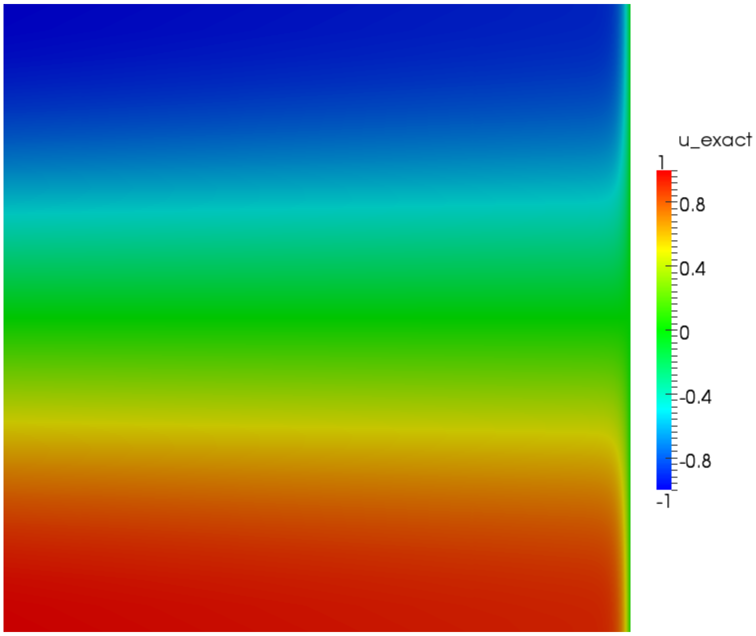
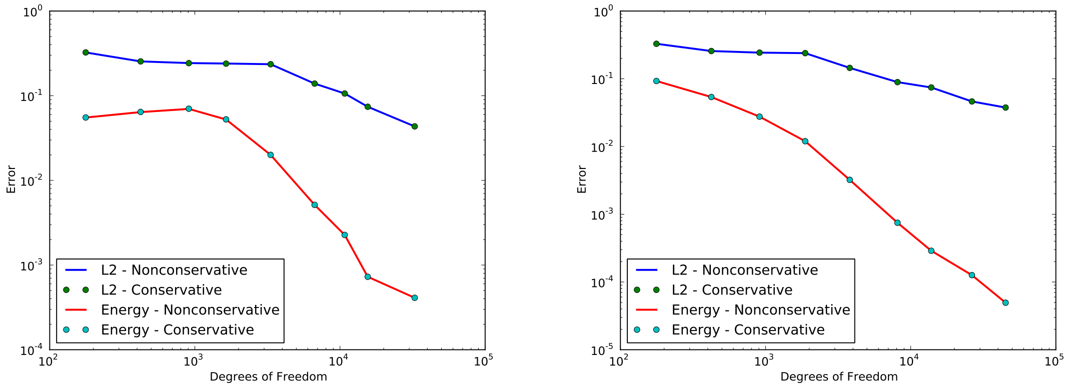


Figure 1: Erickson-Johnson exact solution

We plot the L^2 and energy error for the various methods in Figures 2a and 2b. Encouragingly, the conservative and nonconservative plots appear to lay right on top of each other. This indicates that the conservative formulation at least doensn't hurt our convergence significantly.

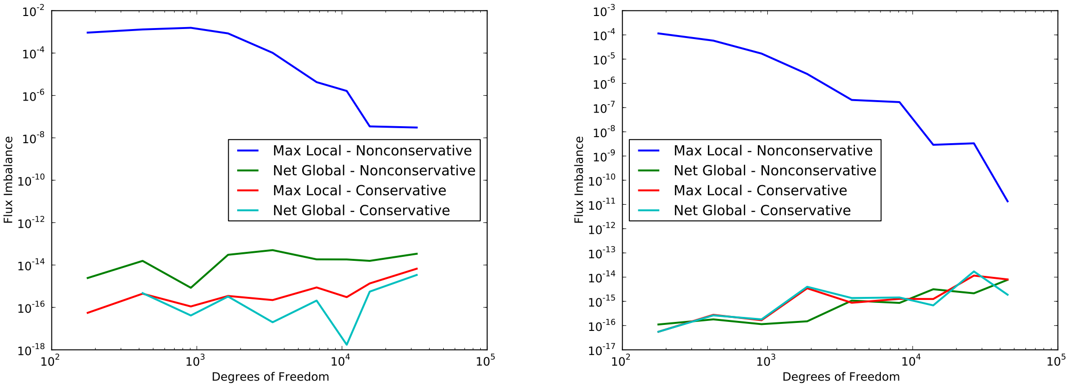
We plot the flux imbalance errors in Figures 3a and 3b. As expected, the element conservative formulation brings these errors to the order of machine precision, but it is worth pointing out that the flux imbalances for standard DPG are reasonable, especially under refinement.



(a) With graph norm

(b) With robust norm

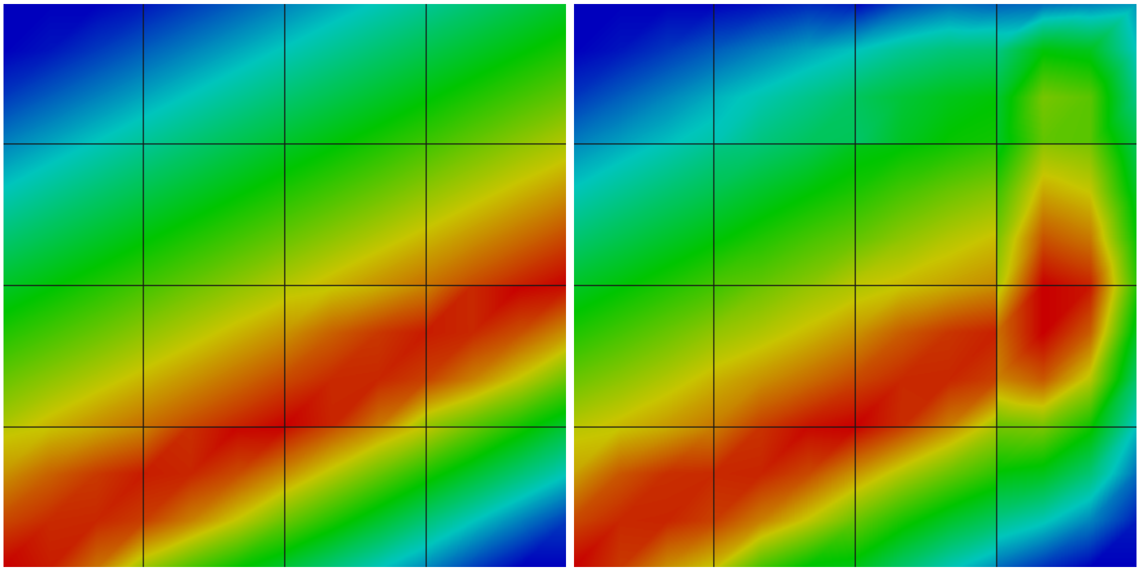
Figure 2: Error in Erickson-Johnson solutions



(a) With graph norm

(b) With robust norm

Figure 3: Flux imbalance in Erickson-Johnson solutions



(a) Conservative with graph norm

(b) Conservative with robust norm

Figure 4: Skewed convection-diffusion problem with initial mesh

3.2 Skewed Convection-Diffusion Problem

This is a standard convection-diffusion problem with a skewed convection vector relative to the mesh. The problem setup is identical to the Erickson-Johnson problems except that $\boldsymbol{\beta} = (2, 1)$. The following results are for $\epsilon = 10^{-4}$.

On our initial mesh, we see no noticeable difference between conservative and nonconservative results, though the choice of test norm affects things dramatically. The graph norm produces a solution which is comparable to an L2 project of the exact solution. In this case, the field variable don't show the very small boundary layer along the right and top surfaces, though a visualization of the traces would show that the wall boundary condition is in fact being enforced. As we refine, the field variables will begin to reflect the boundary layer. On the other hand, the robust norm does show a boundary layer being formed in the field variables.

As we refine, the three results, nonconservative with graph or robust norm, and conservative with graph norm converge to a an essentially identical solution (to the eyeball norm), as represented by Figure 5a.

Finally, we plot the flux imbalances for the several methods in Figure 6a and Figure 6b. Interestingly, it appears the effectiveness of our local conservation formulation drops as we refine with the graph norm.

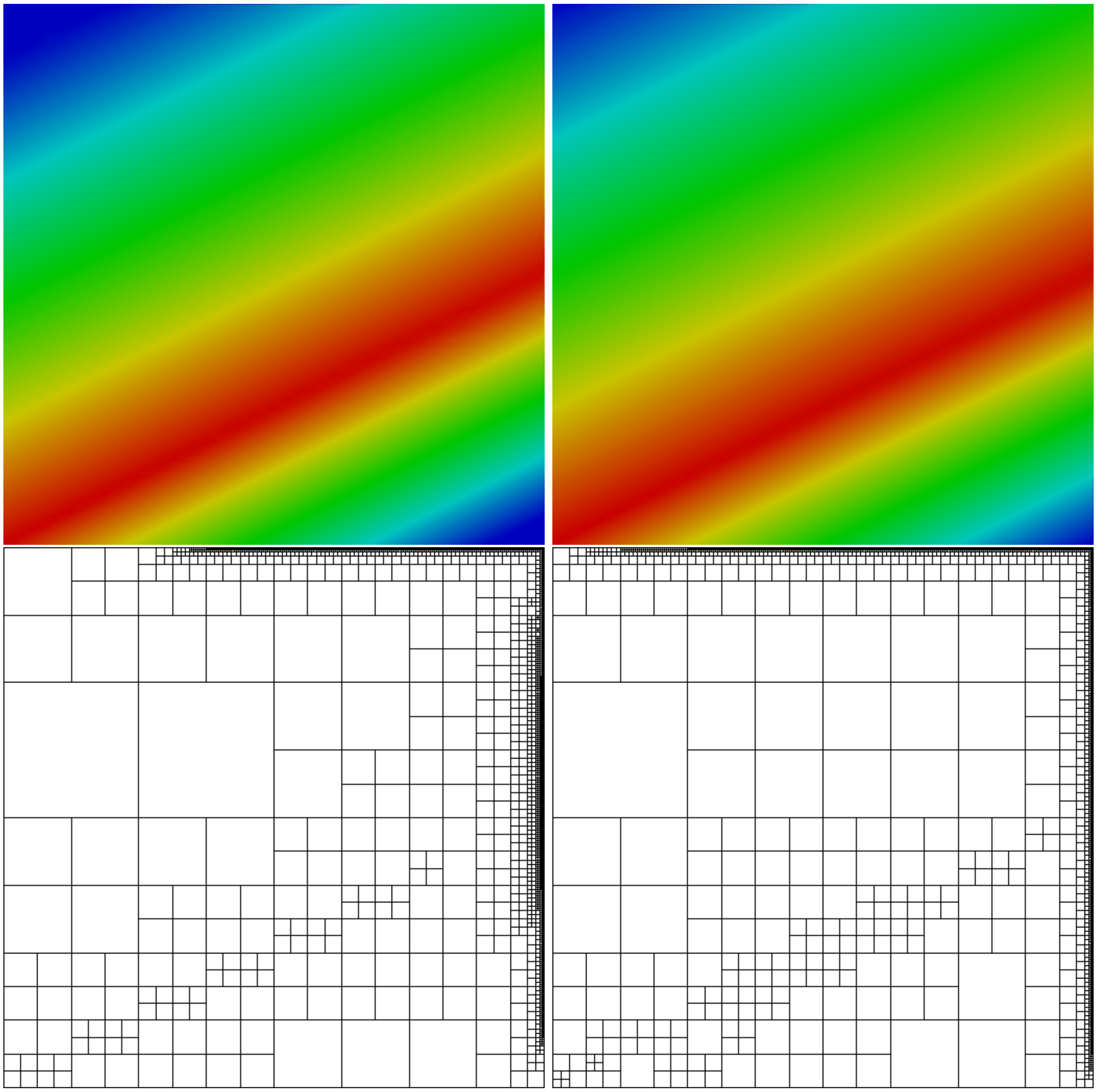
3.3 Double Glazing Problem

This nominal problem definition takes the same unit square domain with

$$\boldsymbol{\beta} = \begin{pmatrix} 2(2y-1)(1-(2x-1)^2) \\ -2(2x-1)(1-(2y-1)^2) \end{pmatrix}; \quad f = 0; \quad \hat{u} = \begin{cases} 1 & \text{on } \Gamma_{right} \\ 0 & \text{else} \end{cases},$$

except that this definition for the boundary is not legal for $\hat{u} \in H^{\frac{1}{2}}(\Gamma_h)$ which is continuous. Therefore we use a ramp of width $\sqrt{\epsilon}$ on the right edge to go from 0 to 1. The following results are for $\epsilon = 10^{-2}$. Note that $\boldsymbol{\beta} = (0, 0)^T$ at the center of the domain.

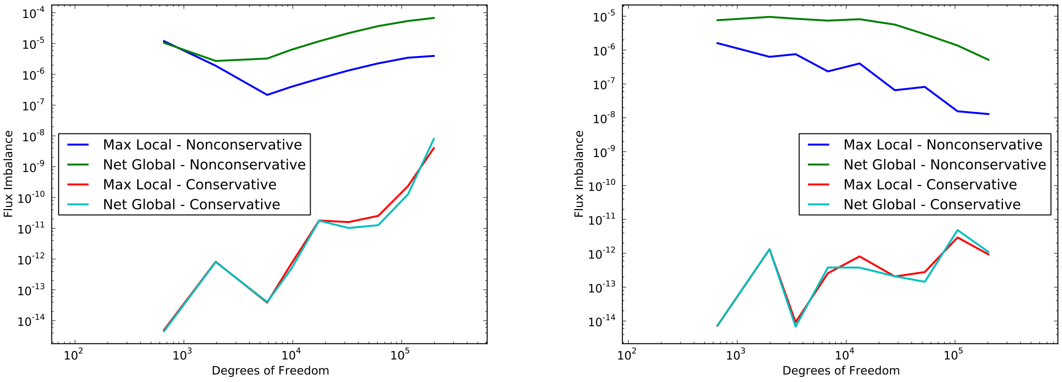
The choice of test norm appears to have a small effect on the choice mesh refinement choices as shown in Figure 7. Other than that, the conservative and nonconservative solutions behave almost exactly to the eye. The flux imbalances in Figure 8 are what we would expect. The nonconservative formulation appears to become more conservative with refinement.



(a) Conservative with graph norm

(b) Conservative with robust norm

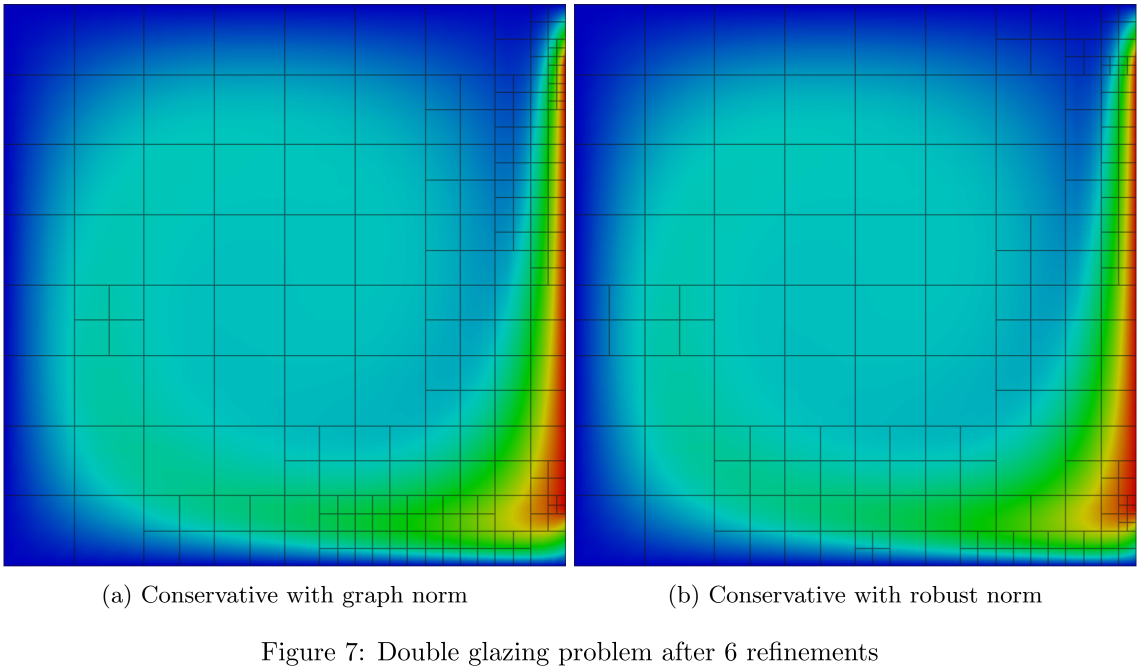
Figure 5: Skewed convection-diffusion problem after 8 refinements



(a) With graph norm

(b) With robust norm

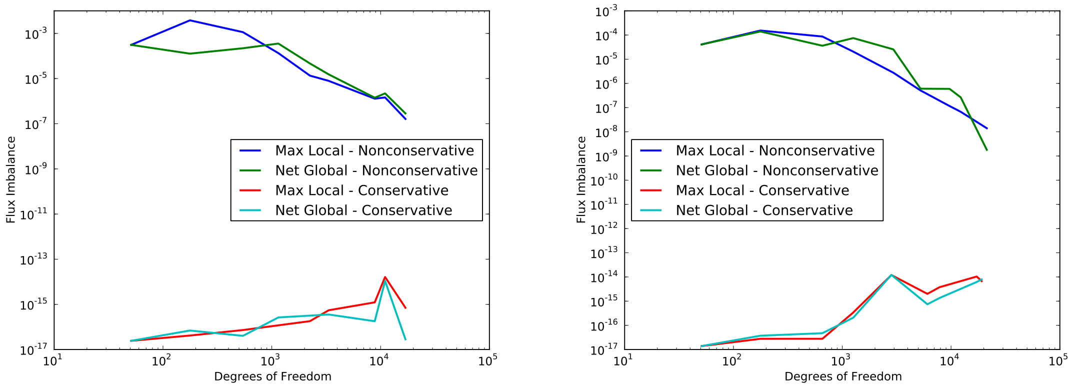
Figure 6: Flux imbalance in skewed confusion solutions



(a) Conservative with graph norm

(b) Conservative with robust norm

Figure 7: Double glazing problem after 6 refinements



(a) With graph norm

(b) With robust norm

Figure 8: Flux imbalance in double glazing solutions

3.4 Vortex Problem

This problem models a mildly diffusive vortex convecting fluid in a circle. We deal with domain $\Omega = [-1, 1]^2$ with $\epsilon = 10^{-4}$, $\boldsymbol{\beta} = (-y, x)^T$, and $f = 0$. Note that $\boldsymbol{\beta} = 0$ at the domain center. We have an inflow boundary condition when $\boldsymbol{\beta} \cdot \mathbf{n} < 0$, in which case we set $\hat{t} = \boldsymbol{\beta} \cdot \mathbf{n} \cdot u_0$ where $u_0 = \frac{\sqrt{x^2+y^2}-1}{\sqrt{2}-1}$ which will vary from 0 at the center of boundary edges to 1 at corners. We don't enforce an outflow boundary.

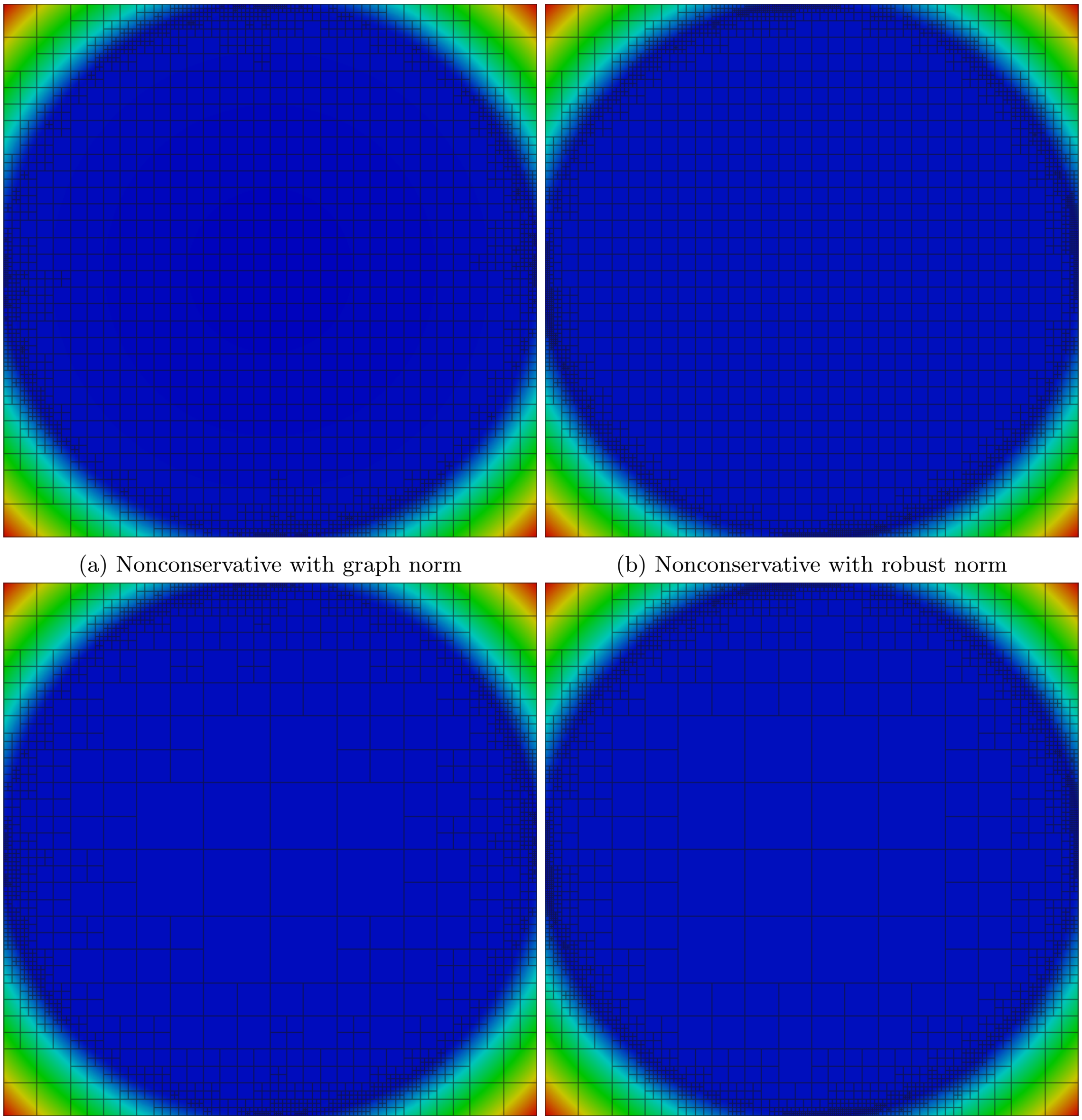
Each of the four methods seems to have it's own specific refinement pattern, but in general we can see that the nonconservative formulations seem to place much more refinements in the center of the vortex where the solution is constant. The conservative formulations seem to recognize that more error is concentrated on the edges of the vortex where we have a sharp change from constant values to linear growth in terms of radius. Again we see that standard DPG appears to approach conservation with refinement, and unfortunately the conservative formulation appears to be losing enforcement under the robust norm.

3.5 Wedge Problem

We devised this problem to examine some of the issues we were having blow up of the solution under the robust test norm. The domain is a rectangle $[-0.5, 0.5] \times [-1, 0.5]$ where the triangle connecting the bottom edges with the origin at $(0, 0)$ is removed. The convection vector $\boldsymbol{\beta} = (1, 0)^T$, $\epsilon = 10^{-1}$, and we apply boundary conditions $\hat{t} = 0$ on the left inflow edge, $\hat{u} = 0$ on the top edge, $\boldsymbol{\beta} \cdot \mathbf{n} \cdot \hat{u} - \hat{t} = \sigma_n = 0$ on the right outflow edge, $\hat{u} = 1$ on the leading edge of the wedge, and $\hat{t} = \boldsymbol{\beta} \cdot \mathbf{n} \cdot 1$ on the trailing edge of the wedge.

When the colorscale is scaled from 0 to 1, the macro view of all of the solutions looks nearly identical as seen in Figure 11. However, we start to see some differences in the refinement patterns in Figure 12 indicating that the various methods distribute error in different ways. We see more evidence of this when we zoom in on the tip of the wedge where we get a change in boundary conditions. The graph norm solutions don't exhibit any significant problems here. The nonconservative and conservative solutions vary smoothly in the range $(-0.00127, 1.03)$, plus or minus some less significant digits. The nonconservative robust solution jumps from -19.6 to 20.3 at the tip, though this appears to be a very localized phenominon. Adding local conservation appears to exacerbate this situation as the solution varies from -20.0 to 23.2 and appears to polute a larger area of the problem. We implemented a crude hp-adaptive refinement strategy in order to try to resolve any singularities present in the solution. If the cell measure dips below 5×10^{-4}, we switch from h-adaptivity to p-adaptivity. Figure 14 shows the resultant polynomial orders after refinement. It really appears that the apparent singularities in the robust solutions are not physical phenomena that just need resolution, but a weakness in the method with the robust test norm.

We plot the flux imbalances in Figure 15. Despite the extreme overshoots and undershoots in the conservative method with the robust test norm, we still appear to maintain acceptable levels of conservation.



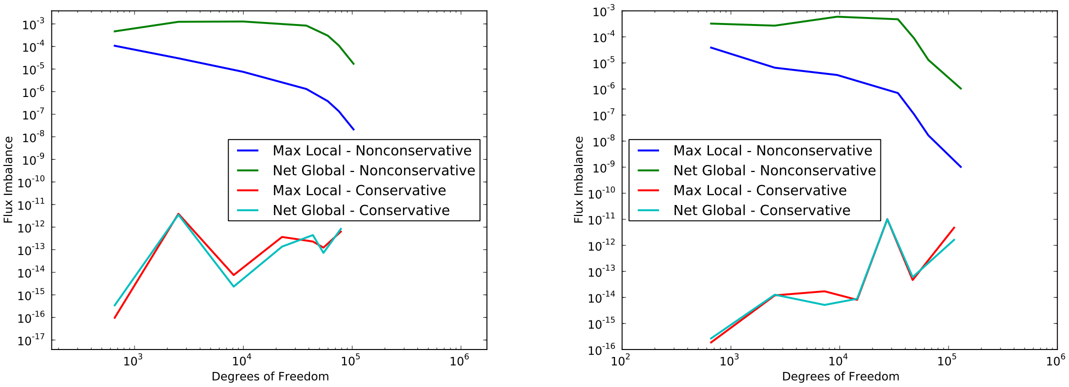
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 9: Vortex problem after 6 refinements



(a) With graph norm

(b) With robust norm

Figure 10: Flux imbalance in vortex solutions

3.6 Inner Layer Problem

3.7 Discontinuous Source Problem

3.8 Hemker Problem

The Hemker problem is defined on a domain $\Omega = \{[-3,9]\times[-3,3]\}\{(x,y) : x^2+y^2 < 1\}$, essentially a simplified model of flow past an infinite cylinder. We start with the initial mesh shown in Figure 21 in which we approximate the circle by 32 linear segments. Our code does not fully support curvilinear elements at this time, so as we refine, we just subdivide the initial linear segments. The boundary conditions are $\hat{t} = \boldsymbol{\beta}\cdot\mathbf{n}\cdot 1$ on the left inflow boundary, $\boldsymbol{\beta}\cdot\mathbf{n}\cdot\hat{u} - \hat{t} = \sigma_n = 0$ on the right outflow boundary, $\hat{t} = 0$ on the top and bottom edges, and $\hat{u} = 1$ on the cylinder.

After some early refinement steps in which the graph norm solutions exhibit some strange behavior (see Figure 22) all four methods appear to converge toward a decent solution as shown in Figure 23. The different methods seem to prioritize different parts of the mesh for refinement, but with sufficient refinement, the differences in the actual solution are fairly insignificant. TODO: Explain why $\hat{u}$ boundary is not nice on inflow.

References

[1] J. Chan, L. Demkowicz, R. Moser, and N. Roberts. A class of Discontinuous Petrov–Galerkin methods. Part V: Solution of 1D Burgers and Navier–Stokes equations. Technical Report 25, ICES, 2010. submitted to J. Comp. Phys.

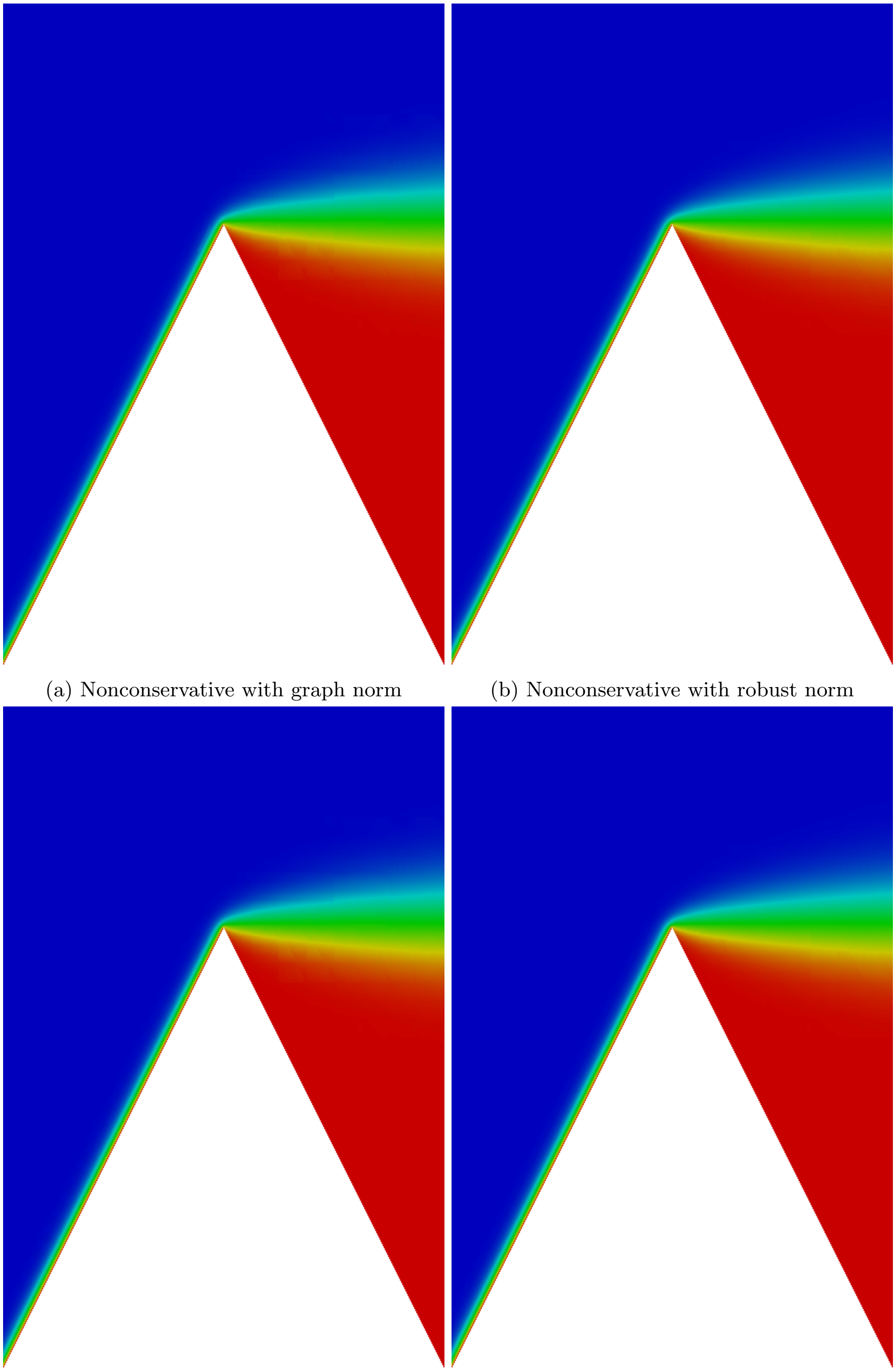
[2] J. Chan, N. Heuer, T Bui-Thanh, and L. Demkowicz. Robust DPG method for convection-dominated diffusion problems ii: a natural inflow condition. Technical Report 21, ICES, 2012.

[3] C. L. Chang and John J. Nelson. Least-squares finite element method for the stokes problem with zero residual of mass conservation. *SIAM J. Num. Anal.*, 34:480–489, 1997.

[4] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation. *Comput. Methods Appl. Mech. Engrg.*, 2009. accepted, see also ICES Report 2009-12.

[5] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part II: Optimal test functions. *Numer. Meth. Part. D. E.*, 2010. in print.

[6] L. Demkowicz, J. Gopalakrishnan, and A. Niemi. A class of discontinuous Petrov-Galerkin methods. Part III: Adaptivity. Technical Report 1, ICES, 2010. submitted to App. Num. Math.



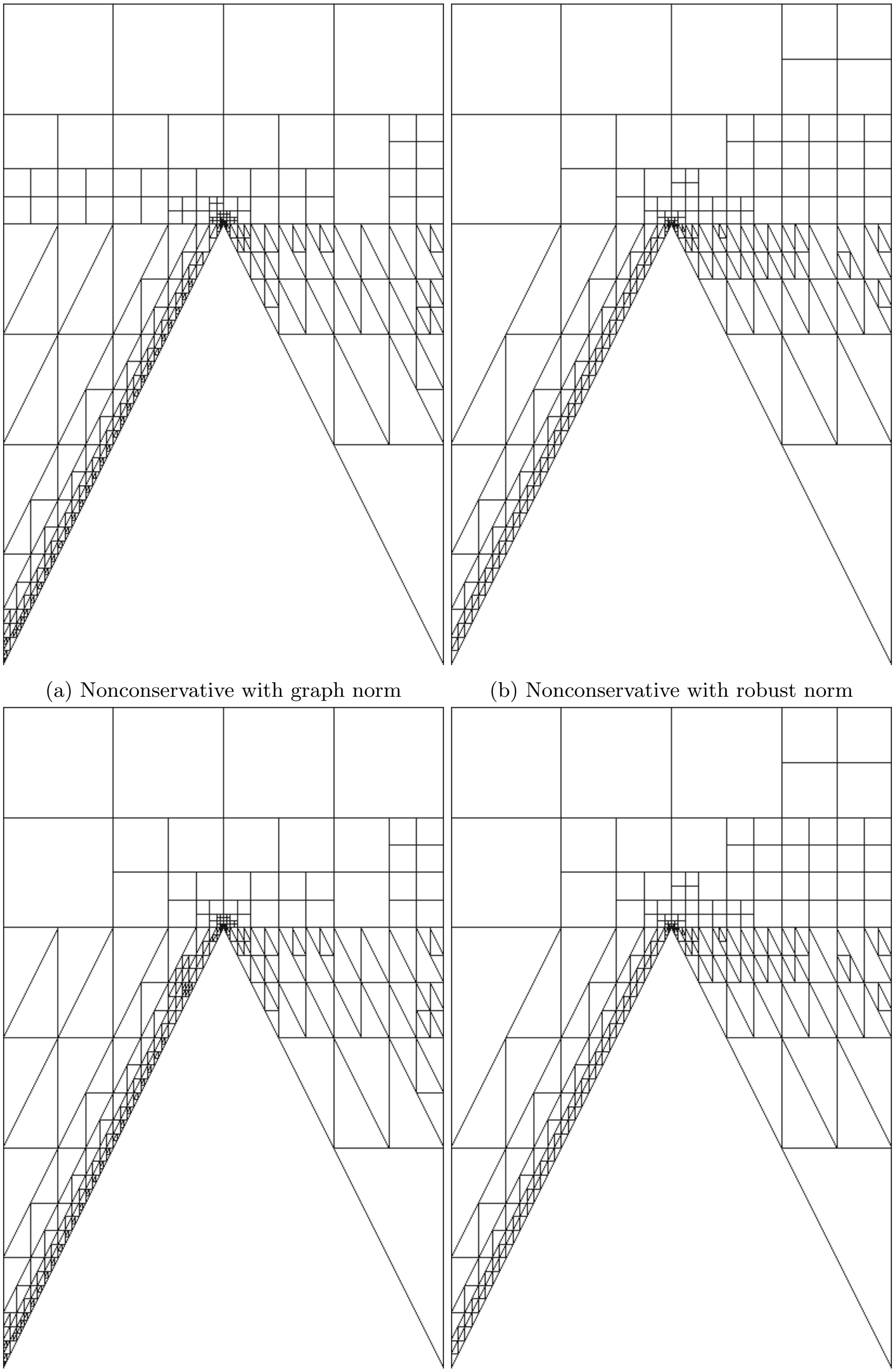
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 11: Wedge problem after 16 refinements



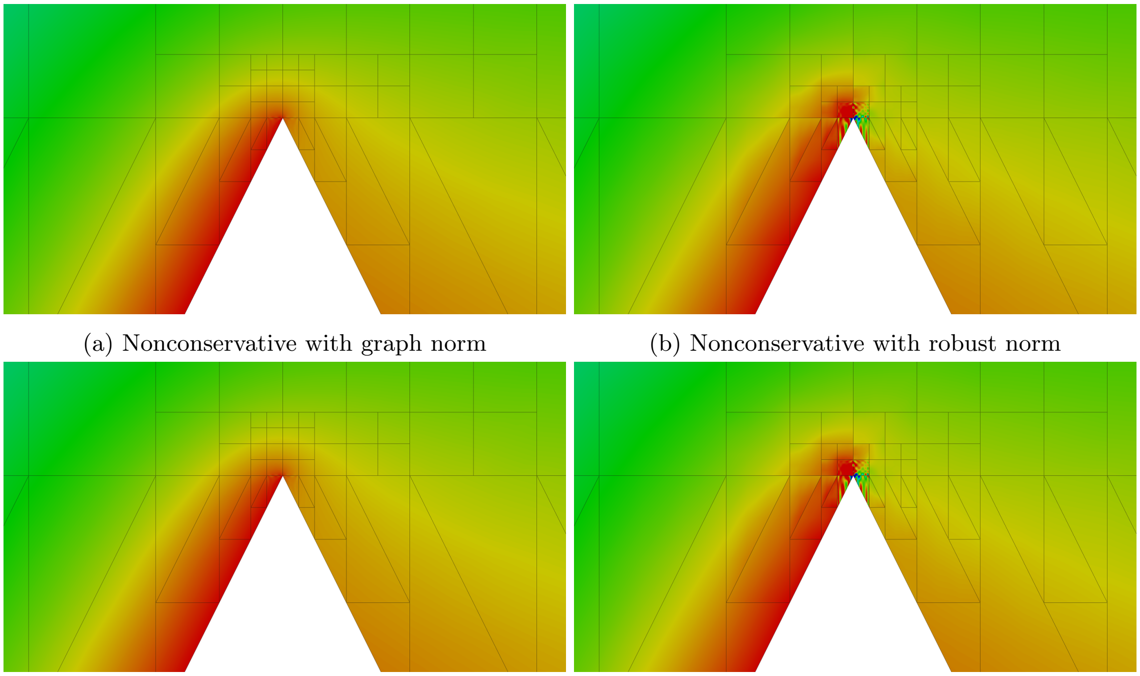
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 12: Wedge problem mesh after 16 refinements



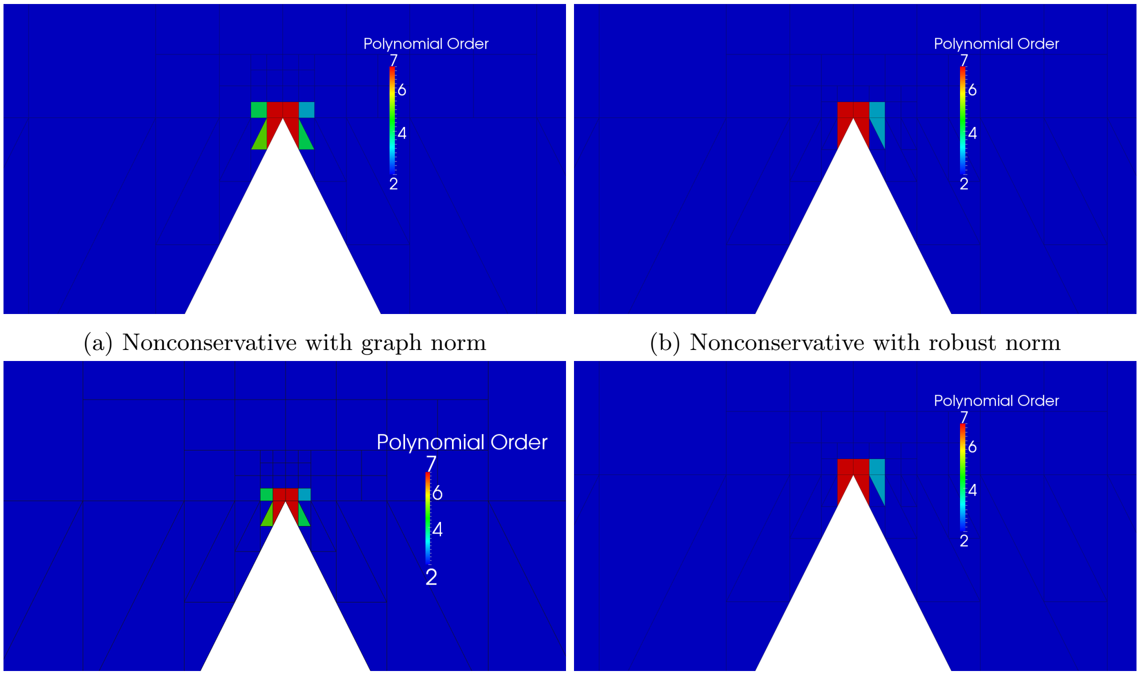
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 13: Wedge problem zoomed in after 16 refinements



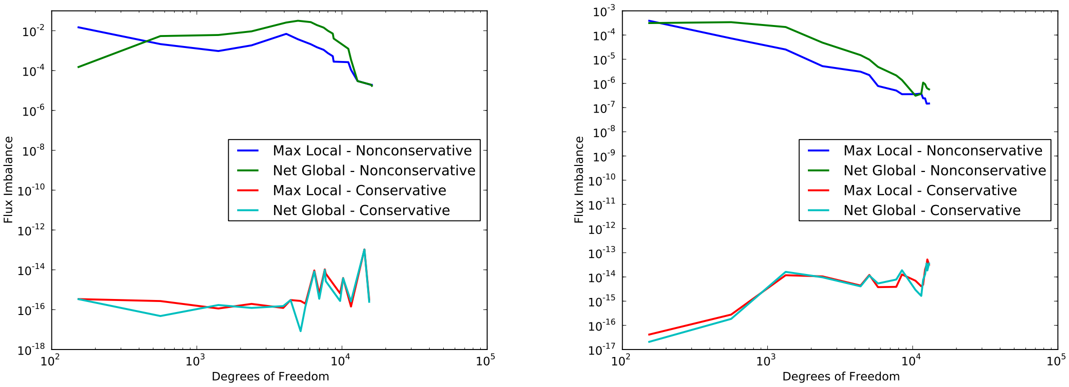
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 14: Wedge problem poly after 16 refinements



(a) With graph norm

(b) With robust norm

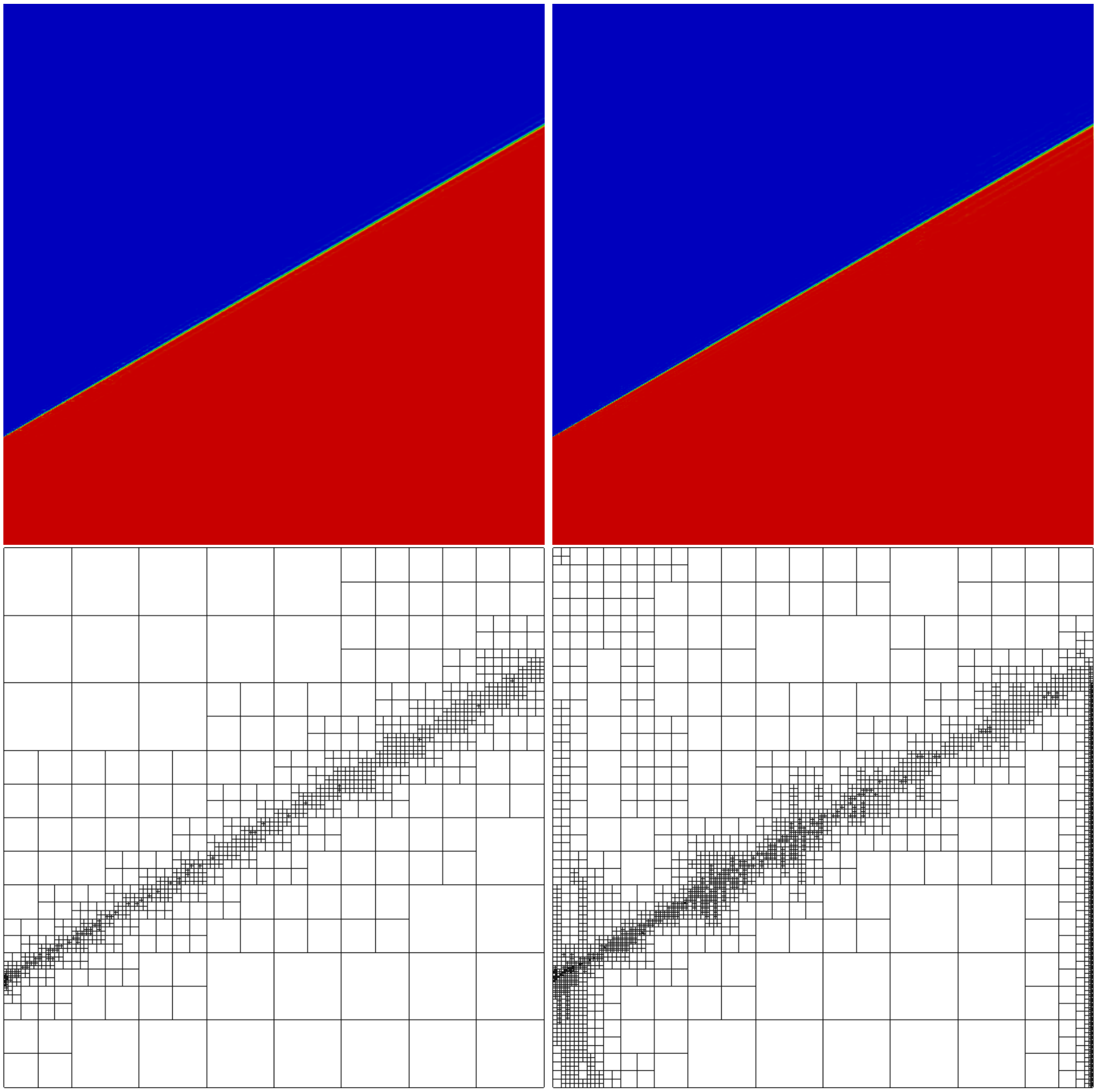
Figure 15: Flux imbalance in wedge solutions

[7] L. Demkowicz and N. Heuer. Robust DPG method for convection-dominated diffusion problems. Technical Report 33, ICES, 2011.

[8] D. Moro, N.C. Nguyen, and J. Peraire. A hybridized discontinuous Petrov-Galerkin scheme for scalar conservation laws. *Int.J. Num. Meth. Eng.*, 2011. in print.

[9] J. B. Perot. Discrete conservation properties of unstructured mesh schemes. *Annual Review of Fluid Mechanics*, 43:299–318, 2011.

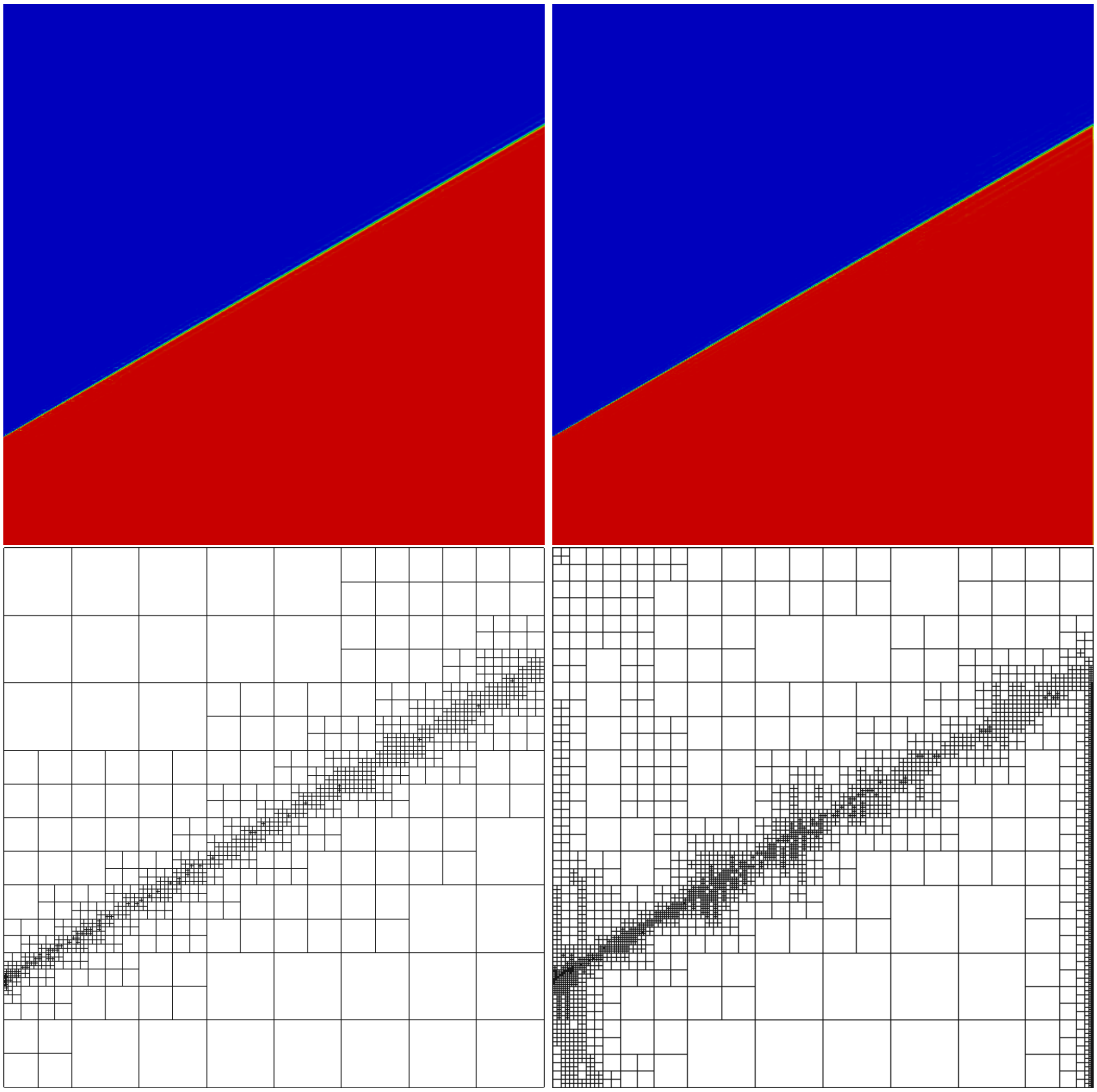
[10] Nathan V. Roberts, Tan Bui-Thanh, and Leszek F. Demkowicz. The DPG method for the Stokes problem. Technical Report 12-22, ICES, 2012.



(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

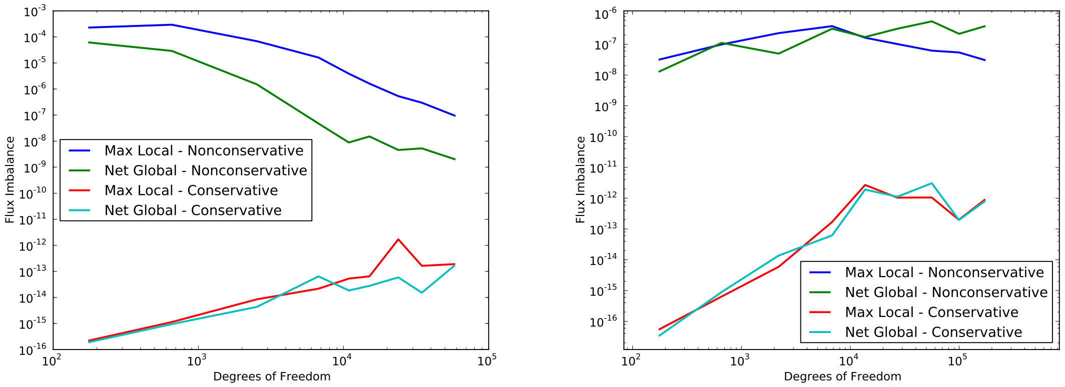
Figure 16: Inner layer problem after 8 refinements



(a) Conservative with graph norm

(b) Conservative with robust norm

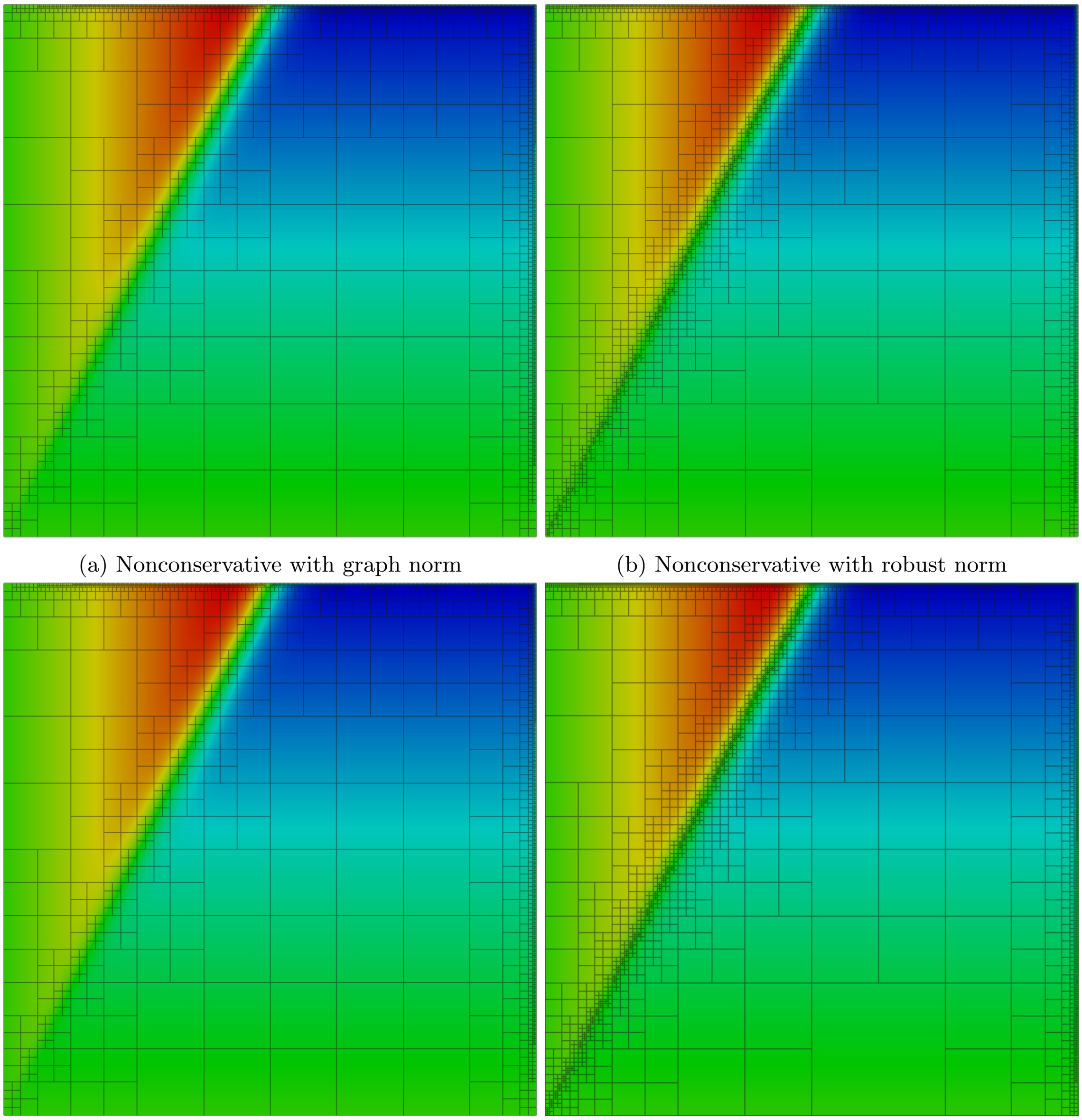
Figure 17: Inner layer problem after 8 refinements



(a) With graph norm

(b) With robust norm

Figure 18: Flux imbalance in inner layer solutions



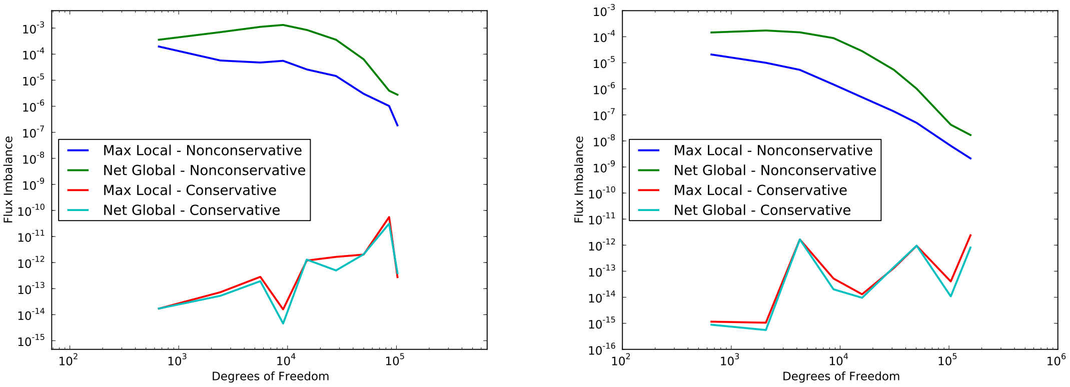
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 19: Discontinuous source problem after 8 refinements



(a) With graph norm

(b) With robust norm

Figure 20: Flux imbalance in discontinuous source solutions

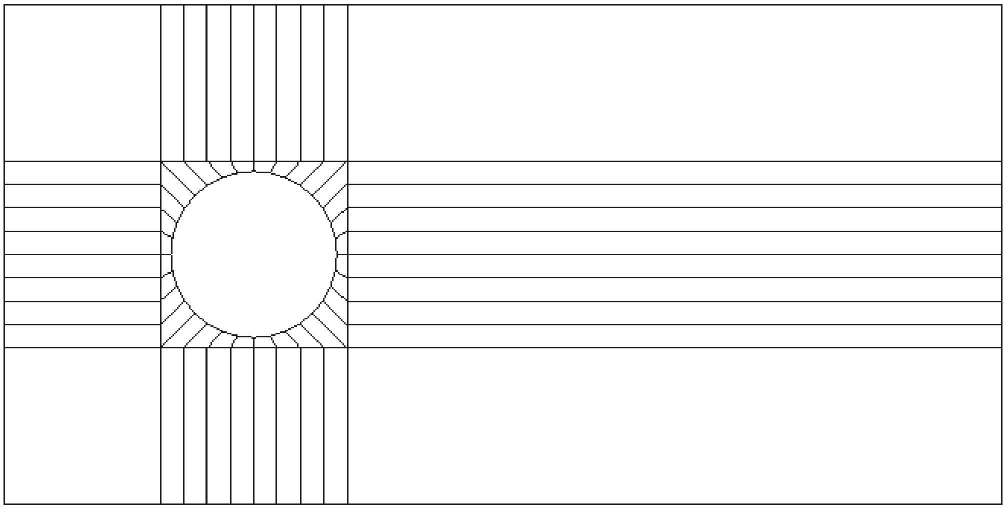
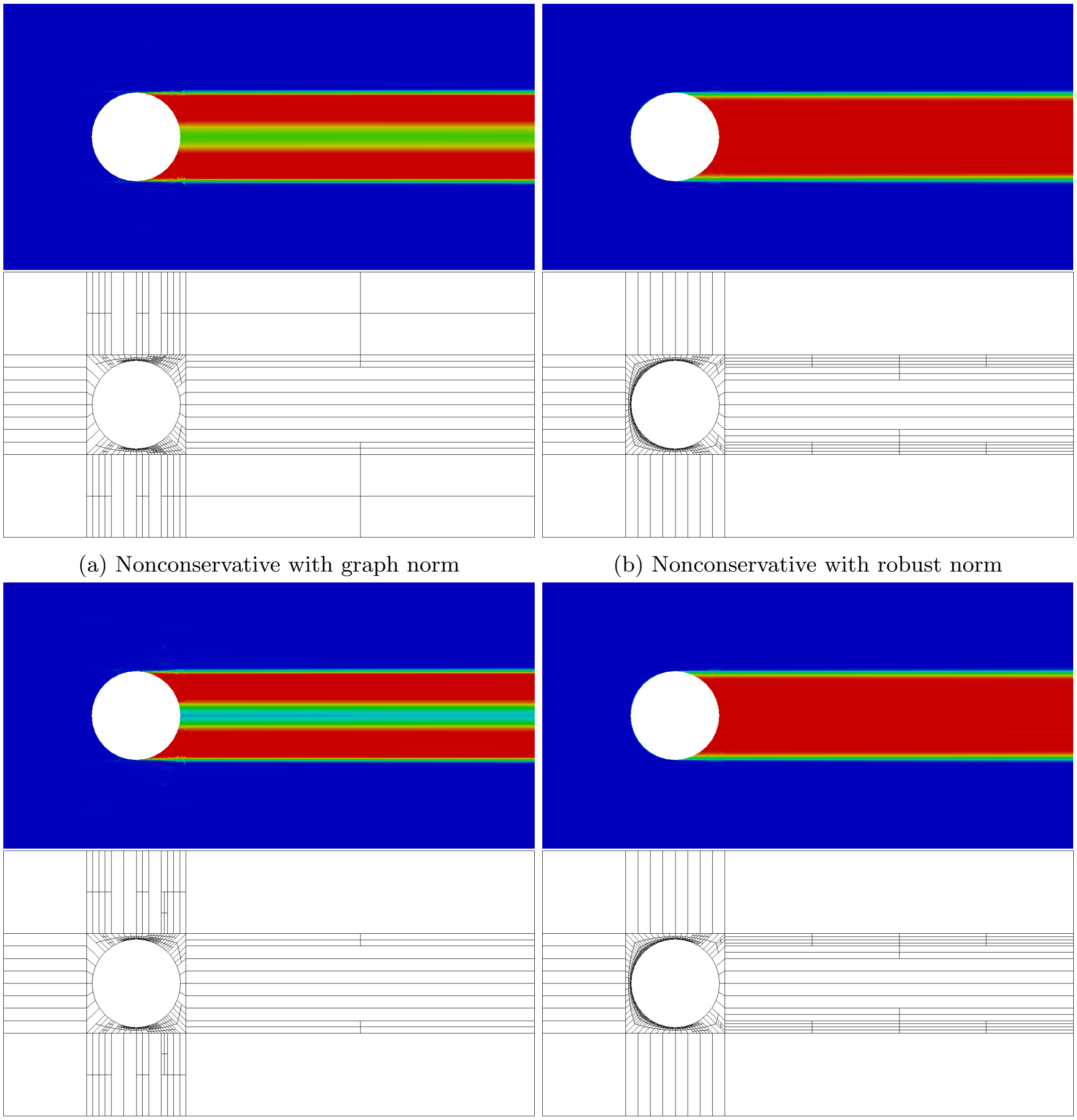


Figure 21: Initial mesh for the Hemker problem



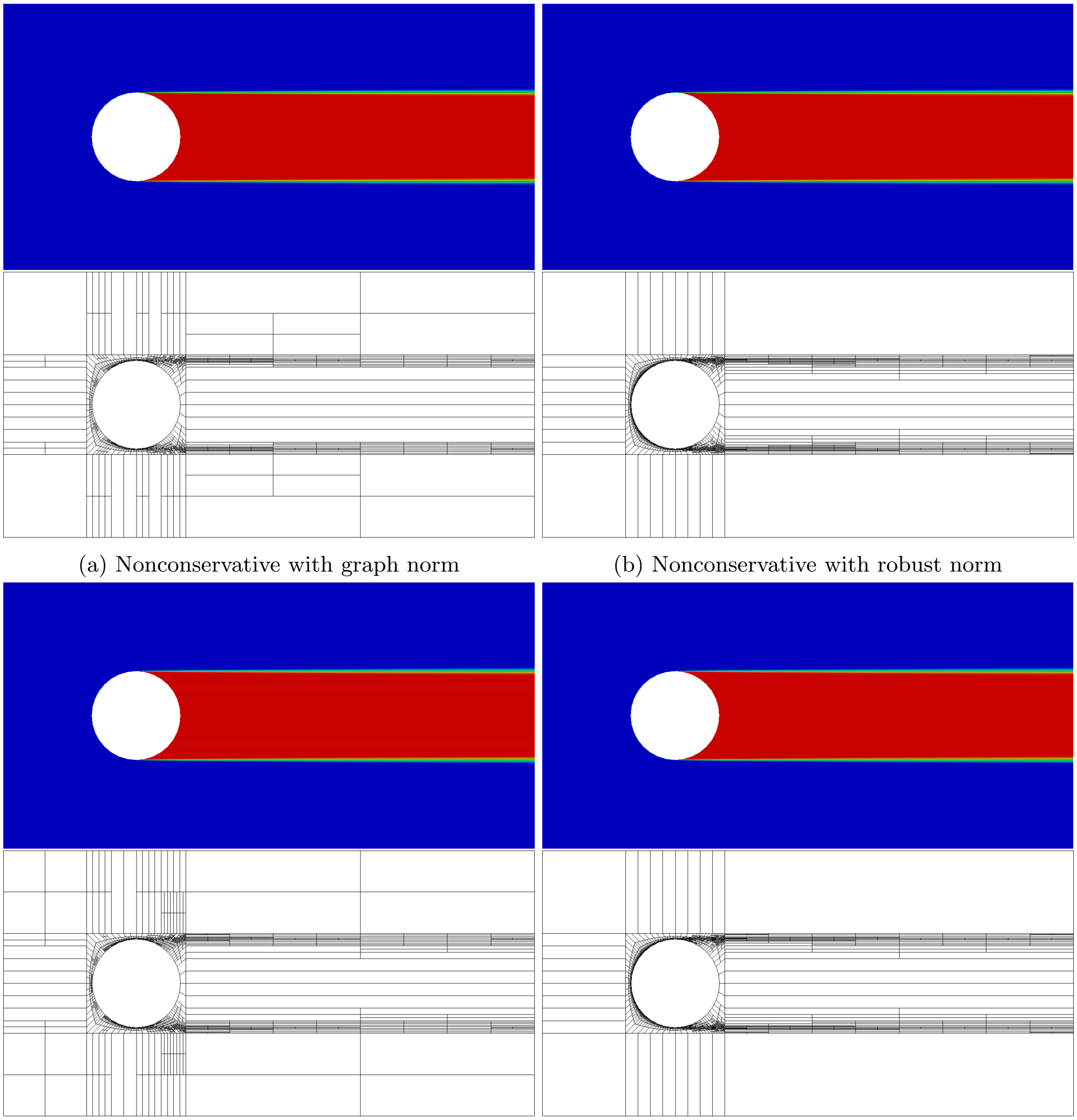
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 22: Hemker problem after 4 refinements



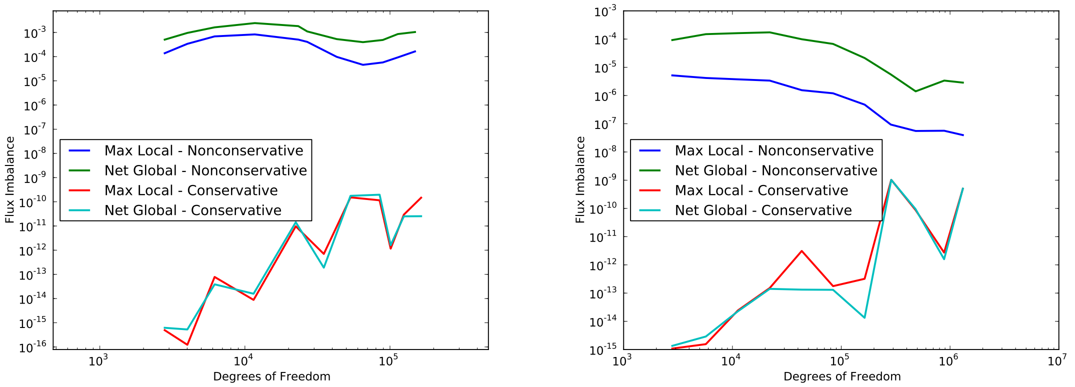
(a) Nonconservative with graph norm

(b) Nonconservative with robust norm

(c) Conservative with graph norm

(d) Conservative with robust norm

Figure 23: Hemker problem after 8 refinements



(a) With graph norm

(b) With robust norm

Figure 24: Flux imbalance in Hemker solutions