Stabilized Approximation to Degenerate Transport Equations via Filtering

V.J. Ervin *   E.W. Jenkins *

Department of Mathematical Sciences, Clemson University, Clemson, SC, 29634-0975, USA.

Abstract

We analyze a stabilization technique for advection dominated flow problems. Of particular interest are coupled parabolic/hyperbolic problems, when the diffusion coefficient is zero in part of the domain. The unstabilized, computed approximations of these problems are highly oscillatory, and several techniques have been proposed and analyzed to mitigate the effects of the subgrid errors that contribute to the oscillatory behavior. In this paper, we modify a time-relaxation algorithm proposed in [1] and further studied in [10]. Our modification introduces the relaxation operator as a post-processing step. The operator is not time-dependent, so the discrete (relaxation) system need only be factored once. We provide convergence analysis for our algorithm along with numerical results for several model problems.

Key words. Advection dominated; stabilization algorithm; time-relaxation

AMS Mathematics subject classifications. 65N30

1   Introduction

Of interest in this paper is the numerical approximation to u satisfying the general transport equation

$$u_t \ + \ \nabla \cdot f(u, \nabla u) \ = \ g(x, t) \, . \tag{1.1}$$

Our interest in this problem is motivated by groundwater flow applications. Single phase, porous media flows can be modeled by the general convection diffusion equation, where the flux function

$$f(u, \nabla u) = -\frac{k(S(u))}{\mu} \nabla u$$

exhibits different behaviors based on values of $k(S(u))$. The unknown u denotes the pressure head, or potential, associated with the fluid. The saturation S is the ratio of the volume of fluid in the medium relative to the volume of the pore space of the medium. Saturation depends on u, as increased values of the pressure push fluid into the previously unsaturated pore spaces. The

---

permeability $k(S(u))$ of the fluid is a measure of the ability of the medium to transport fluid and increases as the medium becomes more saturated. The parameter μ denotes the viscosity of the fluid.

Curves representing saturation and permeability as a function of pressure head [7] are shown in Figure 1.1 for three material types: sand, silt, and clay. The plots were constructed using data generated from the Mualem-van Genuchten pressure-saturation-permeability relationships [20, 23].
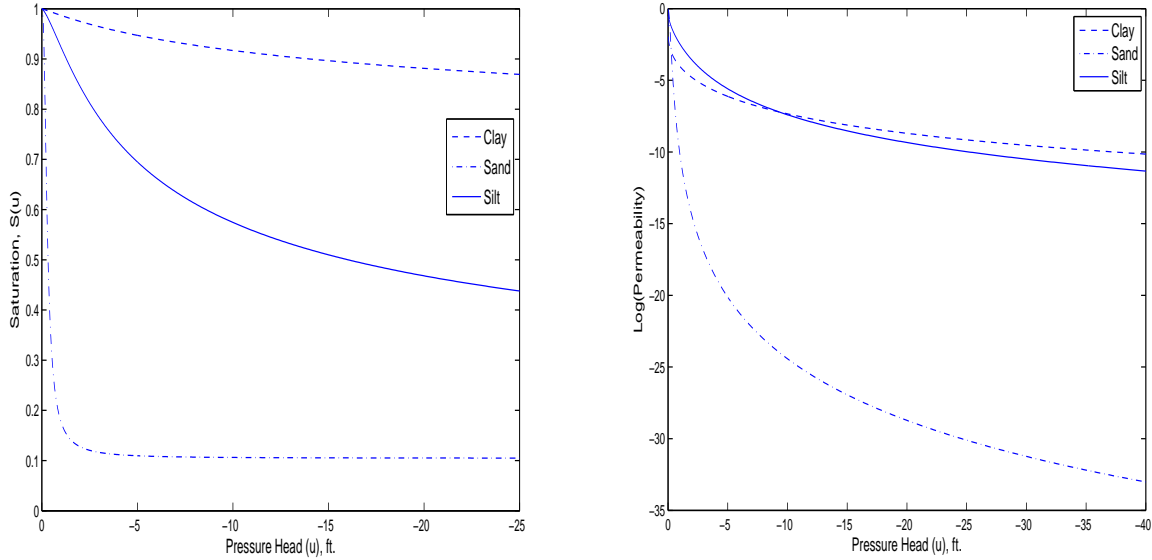

Figure 1.1: Saturation (left) and permeability (right) curves for clay, sand and silt.

These curves differ substantially for different types of media. In the governing equation, the coefficient for the diffusive component depends directly on $k(S(u))$, and the coefficient for the advective component depends, in part, on the derivative of the saturation function with respect to pressure head. As fluid moves through a sand layer, i.e., as saturation values increase, one can see from the steepness of the saturation curve that the governing equation becomes advection dominated. In contrast, the governing equation is more parabolic for flow through clay layers as the values for permeability are several orders of magnitude larger than those for sand, and the saturation curve is much less steep.

There are two cases for which the governing equation changes type when considering porous media flow: (i) the medium is heterogeneous, in which case the governing equation changes type in different fixed regions in space; (ii) infiltration is occuring in the medium, in which case the governing equation is hyperbolic ahead of the front, parabolic along the front, and elliptic behind the front.

Characteristic of these physical problems is: (a) an "uncertain" description of the parameters defining the problem, and (b) sharp transition regions for the flow. Because of the uncertainty in the description of these problems our interest is focused on stable, low order approximation techniques.

The sharp transition regions in the flow field will typically generate non-physical oscillations in the the approximation unless some stabilization is added to the approximation algorithm. There are

many effective stabilization schemes which have been proposed and used. For examples, the method of artifical viscosity (AV), the Streamline Upwind Petrov Galerkin (SUPG) method, characteristic based methods (modified method of characteristics (MMOC-Galerkin) [11], characteristic mixed finite element method [2], Eulerian-Lagrangian localized adjoint methods (ELLAM) [24]), the locally discontinuous Galerkin (LDG) method [6], to mention a few. A more recent approach has been methods which include a "sub-grid" stabilization term, commonly referred to as variational multiscale methods (VMS) [15, 16, 17]. The AV and SUPG methods produce a stable approximation at a cost of reduced asymptotic accuracy of the approximation. Characterstic methods require the tracking, and integration along flow curves with special attention required near the inflow and outflow boundaries of the domain. The LDG method are typically inefficient in diffusion-dominated problems. With the VMS approach stability of the approximation is obtained by adding to the computational modeling equations a term which attempts to model the influence of the unresolved scales on the computational scales. This typically involves *projections* onto locally enriched subspaces.

The use of *filtering* to remove spurious high frequency oscillations in computed approximations to underresolved flow problems has a long history. (See [5, 21] and references therein.) Spectral based filtering methods often have difficulties in handling problems on irregular domains. Filtering methods typically result in the filtered approximation being overly diffused. A fairly recent approach to counter the addition of excess diffusion from simply filtering the approximation is to combine filtering with *deconvolution*, an approximate inverse of the filtering operation. In [1] Adams and Stolz investigated a combined *filter and deconvolution* approach for the numerical approximation of the averaged, time dependent Euler equations. Their approach was very successful in obtaining accurate approximations. This *filter and deconvolution* approach has been studied and extended to Navier-Stokes fluid flow problems by Layton and collaborators [8, 19, 18].

Motivated by the work of Stolz and Adams in [1] we initially investigated a regular finite element approximation to (1.1) with the addition of a stabilizing term, which was a function of the approximate solution and its local (spatial) average. We did not find this approach to be effective in controlling the development of spurious oscillations in the approximations for the problems we investigated. Instead, what we found to be effective in computing a stable approximation was a two step approach where in the first step an intial approximation $W^n(\cdot)$ to $u(n\Delta t, \cdot)$ is computed and then in the second step we filter and deconvolve $W^n(\cdot)$ to remove an spurious oscillations to obtain $U^n(\cdot)$ the approximation to $u(n\Delta t, \cdot)$. The deconvolution in the second step is used to improve the accuracy of the filter quantity.

In our investigation we consider the consider the equation

$$u_t + \mathbf{v} \cdot \nabla u - \nabla \cdot (\alpha \nabla u) = f, \text{ in } \Omega, \tag{1.2}$$

$$\text{subject to } u = 0, \text{ on } \partial\Omega, u(x,0) = u_0(x), \text{ in } \Omega. \tag{1.3}$$

As $\mathbf{v}$ in (1.2) typically denotes a fluid velocity, we assume the incompressibility condition $\nabla \cdot \mathbf{v} = 0$, and $\|\mathbf{v}\|_\infty = C_{\mathbf{v}}$. Additionally, we assume that $\alpha \geq 0$ and $\alpha \in L^\infty(\Omega)$.

The approximation algorithm is:

Approximation Algorithm:
Given $U^0 \in X_h$, for $n = 1, 2, \ldots, N_T$

Step 1. Determine $W^n \in X_h$ satisfying

$$\left(\frac{W^n - U^{n-1}}{\Delta t}, \phi\right) + (\mathbf{v}^n \cdot \nabla W^n, \phi) + (\alpha \nabla W^n, \nabla \phi) = (f^n, \phi), \quad \forall \phi \in X_h. \tag{1.4}$$

Step 2. $U^n \in X_h$ given by

$$U^n = D_h G_h(W^n). \tag{1.5}$$

In (1.5) $G_h(\cdot)$ denotes the discrete filter operator defined in (2.4), and $D_h(\cdot)$ the discrete deconvolution operator defined in (2.5).

For the proposed algorithm the following points are noteworthy.
(i) Ease of implementation, (simple modification of a standard Galerkin approximation).
(ii) Requires no special spatial discretization.
(iii) Gives optimal asymptotic order of convergence for the approximation.

This paper is organized as follows. In Section 2 we present an analysis of the Approximation Algorithm, establishing computability of the method and a priori error estimates. Three examples are presented in Section 3. The first example considered models an infiltration problem where the solution exhibits a moving front propogating through the domain. Ahead of the front the modeling equation is hyperbolic, along the front parabolic, and behind the front the modeling equation is elliptic. Example two studies a moving front problem in a two dimensional domain. The third example, taken from [9] models flow through a heterogeneous medium. In one part of the domain the diffusion coefficient is equal to zero. Within the domain this problem's modeling equation classification changes from parabolic to hyperbolic and hyperbolic back to parabolic.

2 Mathematical Analysis

In this section we show the scheme is numerically stable, and derive asymptotic error estimates. We begin by defining the notation used and presenting some properties of the differential filter studied and the deconvolution operator.

2.1 Mathematical Preliminaries

The $L^2(\Omega)$ norm and inner product will be denoted by $\|\cdot\|$ and $(\cdot, \cdot)$. Likewise, the $L^p(\Omega)$ norms and the Sobolev $W_p^k(\Omega)$ norms are denoted by $\|\cdot\|_{L^p}$ and $\|\cdot\|_{W_p^k}$, respectively. For the semi-norm in $W_p^k(\Omega)$ we use $|\cdot|_{W_p^k}$. H^k is used to represent the Sobolev space W_2^k, and $\|\cdot\|_k$ denotes the norm in H^k. For functions $v(\mathbf{x}, t)$ defined on the entire time interval $(0, T)$, we define

$$\|v\|_{\infty,k} := \sup_{0 < t < T} \|v(t, \cdot)\|_k, \qquad \text{and} \qquad \|v\|_{m,k} := \left(\int_0^T \|v(t, \cdot)\|_k^m \, dt\right)^{1/m}.$$

Let $X := H_0^1(\Omega) = \{f \in H^1(\Omega) : f|_{\partial\Omega} = 0\}$. For our analysis we assume that the solution $u \in X$.

For the discrete approximation, we assume that $\Omega \subset \mathbb{R}^{\acute{d}}$ ($\acute{d} = 2, 3$) is a polygonal domain and T_h is a triangulation of Ω made of triangles (in $\mathbb{R}^2$) or tetrahedrals (in $\mathbb{R}^3$). Additionally, we assume

that there exist constants c_1, c_2 such that

$$c_1 h \leq h_K \leq c_2 \rho_K$$

where h_K is the diameter of triangle (tetrahedral) K, ρ_K is the diameter of the greatest ball (sphere) included in K, and $h = \max_{K \in T_h} h_K$. Let $P_k(A)$ denote the space of polynomials on A of degree no greater than k. Then we define the finite element space X_h as.

$$X_h := \left\{ v \in X \cap C(\bar{\Omega})^2 : v|_K \in P_k(K), \ \forall K \in T_h \right\}.$$

Let Δt be the step size for t so that $t_n = n\Delta t$, $n = 0, 1, 2, \ldots, N_T$, with $T := N_T \Delta t$, $f^n := f(t_n)$, and $d_t f^n := \frac{f(t_n) - f(t_{n-1})}{\Delta t}$. We define the following additional norms:

$$\|\|v\|\|_{\infty,k} := \max_{0 \leq n \leq N_T} \|v^n\|_k, \qquad \|\|v\|\|_{m,k} := \left(\sum_{n=0}^{N_T} \|v^n\|_k^m \Delta t \right)^{1/m}.$$

In addition, for $u(x,t) \in H^{k+1}$ we make use of the following approximation properties:

$$\inf_{S \in X_h} \|u(\cdot, t) - S\|_r \leq C h^{k+1-r} \|u(\cdot, t)\|_{k+1}. \tag{2.1}$$

There are a number of choices possible for the discrete filter operator G_h and the discrete deconvolution operator D_h [3, 12]. We assume that any such operators satisfy the following.

Assumption DG1: The discrete filter operator G_h and the discrete deconvolution operator D_h satisfy:

$$\|D_h G_h\|_{L^2 \to L^2} \leq 1, \quad \text{and} \quad \|I - D_h G_h\|_{L^2 \to L^2} \leq 1. \tag{2.2}$$

We investigate the simplest differential filter defined by: $\bar{u} := G(u) \in X$, where for a given parameter δ (referred to as the filter radius)

$$\delta^2 (\nabla \bar{u}, \nabla v) + (\bar{u}, v) = (u, v), \quad \forall v \in X. \tag{2.3}$$

Computationally we implement (2.3) on a finite dimensional subspace of X which gives rise to the discrete differential filter $\bar{u}_h := G_h(u) \in X_h$ defined as

$$\delta^2 (\nabla \bar{u}_h, \nabla v) + (\bar{u}_h, v) = (u, v), \quad \forall v \in X_h. \tag{2.4}$$

Having filtered the approximation, to dampen non-physical oscillations in the approximation, we then apply deconvolution to increase the accuracy of the filtered quantity. Herein we use the van Cittert family of approximate deconvolution operators.

We denote the Nth order van Cittert continuous and discrete deconvolution operators as D and D_h, repectively, where

$$D\phi := \sum_{n=0}^{N} (I - G)^n \phi, \quad \text{and} \quad D_h \phi := \sum_{n=0}^{N} (I - G_h)^n \phi. \tag{2.5}$$

Remark: 1. For the deconvolution parameter $N = 0$, $D_h = I$, and thus $D_h G_h(u) = G_h(u) = \overline{u}_h$ given in (2.4). For $N = 1$, from (2.5), $D_h G_h(u) = 2\overline{u}_h - \overline{\overline{u}}_{hh}$, where $\overline{\overline{u}}_{hh} := G_h(\overline{u}_h)$.
2. G_h and D_h defined in (2.4) and (2.5), respectively, satisfy (2.2) [22].

The difference between a function and its filtered/deconvolved approximation is given in the next lemma.

Lemma 1 *[18] For smooth ϕ the discrete approximate deconvolution operator satisfies*

$$\|\phi - D_h G_h \phi\| \leq C_1 \, \delta^{2N+2} \|\bar{\phi}\|_{H^{2N+2}} + C_2 \left(\delta h^k + h^{k+1}\right) \left(\sum_{n=1}^{N+1} |G^n(\phi)|_{k+1}\right), \qquad (2.6)$$

where $\bar{\phi}_h := G_h(\phi)$.

The dependence of the $|G^n(\phi)|_{k+1}$ terms in (2.6) upon the *filter radius* δ, for a general smooth function ϕ, is not fully understood. In the case of ϕ periodic the $|G^n(\phi)|_{k+1}$ are independent of δ. Also, for ϕ satisfying homogeneous boundary conditions, with the additional property that $\Delta^j \phi = 0$ on $\partial\Omega$ for $0 \leq j \leq \left[\frac{k+1}{2}\right] - 1$, the $|G^n(\phi)|_{k+1}$ are independent of δ. (See [18].)

Note that for $N \leq 2$ and ϕ satisfying homogeneous boundary conditions, the $|G^n(\phi)|_{k+1}$ terms are independent of δ.

Below in the analysis we make the following assumption.
Assumption DG2: The $|G^n(\phi)|_{k+1}$ terms in (2.6) are independent of δ, and

$$\|\phi - D_h G_h \phi\| \leq C_1 \, \delta^{2N+2} \|\phi\|_{H^{2N+2}} + C_2 \left(\delta h^k + h^{k+1}\right) \|\phi\|_{k+1}, \qquad (2.7)$$

The discrete Gronwall's lemma plays an important role in the following analysis.

Lemma 2 (Discrete Gronwall's Lemma) *[14] Let Δt, H, and a_n, b_n, c_n, γ_n (for integers $n \geq 0$) be nonnegative numbers such that*

$$a_l + \Delta t \sum_{n=0}^{l} b_n \leq \Delta t \sum_{n=0}^{l} \gamma_n a_n + \Delta t \sum_{n=0}^{l} c_n + H \quad \text{for } l \geq 0.$$

Suppose that $\Delta t \, \gamma_n < 1$, for all n, and set $\sigma_n = (1 - \Delta t \, \gamma_n)^{-1}$. Then,

$$a_l + \Delta t \sum_{n=0}^{l} b_n \leq \exp\left(\Delta t \sum_{n=0}^{l} \sigma_n \gamma_n\right) \left\{\Delta t \sum_{n=0}^{l} c_n + H\right\} \quad \text{for } l \geq 0. \qquad (2.8)$$

We begin our analysis by establishing computability of the approximation scheme (1.4),(1.5), and an a priori bound for the approximation.

Lemma 3 *For the approximation scheme (1.4),(1.5) we have that W^l, $D_h G_h(W^l)$, U^l, $l = 1, \ldots N_T$ exist at each iteration. In addition, $\|U^l\|^2 \leq \|W^l\|^2$, and (for $\Delta t < 1$):*

$$\|W^l\|^2 + \sum_{n=1}^{l} \|W^n - U^{n-1}\|^2 + 2\sum_{n=1}^{l} \Delta t \|\alpha^{1/2} \nabla W^n\|^2 \leq \exp(T) \left(\|\|f\|\|_{2,0}^2 + \|U^0\|^2\right). \qquad (2.9)$$

6

Additionally,

$$\|\|W\|\|_{\infty,0}^2 \leq \exp(T)\left(\|\|f\|\|_{2,0}^2 + \|U^0\|^2\right) \quad and \quad \|\|W\|\|_{0,0}^2 \leq T\exp(T)\left(\|\|f\|\|_{2,0}^2 + \|U^0\|^2\right). \quad (2.10)$$

Proof: Equation (1.4) can be equivalently rewritten as

$$a(W^n, \phi) = b(\phi), \quad \forall \phi \in X_h, \quad (2.11)$$

$$\text{where } a(\psi, \phi) := (\psi, \phi) + \Delta t(\mathbf{v}^n \cdot \nabla \psi, \phi) + \Delta t(\alpha \nabla \psi, \nabla \phi)$$
$$\text{and } b(\phi) := (U^{n-1}, \phi) + \Delta t(f^n, \phi).$$

Noting that for $\psi \neq 0$,

$$a(\psi, \psi) = \|\psi\|^2 + \Delta t \int_\Omega \mathbf{v}^n \cdot \nabla(\frac{1}{2}\psi^2)\, dA + \Delta t\,\|\alpha^{1/2}\nabla\psi\|^2$$
$$= \|\psi\|^2 - \frac{1}{2}\Delta t \int_\Omega \psi^2\, \nabla \cdot \mathbf{v}^n\, dA + \Delta t\,\|\alpha^{1/2}\nabla\psi\|^2$$
$$= \|\psi\|^2 + \Delta t\,\|\alpha^{1/2}\nabla\psi\|^2 \quad (\text{ using } \nabla \cdot \mathbf{v}^n = 0\,)$$
$$> 0,$$

and that (1.4) represents a (square) linear system of equations, positivity of $a(\psi, \psi)$ implies existence and uniqueness of W^n satisfying (1.4).

The existence and uniqueness of $D_h G_h(W^n)$ and U^n follows directly from the Riesz representation theorem and the definition of D_h.

With the choice $\phi = W^n$ in (1.4) and using $(U^{n-1}, W^n) = 1/2\|W^n\|^2 + 1/2\|U^{n-1}\|^2 - 1/2\|W^n - U^{n-1}\|^2$, we have that

$$\frac{1}{\Delta t}\|W^n\|^2 - \frac{1}{\Delta t}\|U^{n-1}\|^2 + \frac{1}{\Delta t}\|W^n - U^{n-1}\|^2 + 2\|\alpha^{1/2}\nabla W^n\|^2 \leq \|f^n\|^2 + \|W^n\|^2,$$

i.e.

$$\|W^n\|^2 - \|U^{n-1}\|^2 + \|W^n - U^{n-1}\|^2 + 2\Delta t\|\alpha^{1/2}\nabla W^n\|^2 \leq \Delta t\|W^n\|^2 + \Delta t\|f^n\|^2. \quad (2.12)$$

From (1.5) we have

$$\|U^n\| = \|D_h G_h(W^n)\| \leq \|W^n\|, \quad (2.13)$$

as $\|D_h G_h\| \leq 1$. Using (2.13) with $n \to n-1$, and summing from $n = 1$ to l we obtain

$$\|W^l\|^2 + \sum_{n=1}^{l}\|W^n - U^{n-1}\|^2 + 2\sum_{n=1}^{l}\Delta t\|\alpha^{1/2}\nabla W^n\|^2 \leq \sum_{n=1}^{l}\Delta t\|W^n\|^2 + \sum_{n=1}^{l}\Delta t\|f^n\|^2 + \|U^0\|^2$$
$$(2.14)$$

from which (2.9) follows via the discrete Gronwall lemma (2.8).

The estimates in (2.10) follow immediately from (2.9) and the definition of the norms. ∎

Remark: Lemma 3 establishes the uniform boundedness of $\|W^l\|$, $\|U^l\|$ and $\sum_{n=1}^{l}\Delta t\|\alpha^{1/2}\nabla W^n\|^2$. From a physical point of view the filtering step in **Step 2** of the algorithm serves to dampen any high frequency oscillations generated in W^n.

Theorem 1 *For* $u \in L^\infty(0, T; H^{k+1}(\Omega)) \cap L^2(0, T; H^{2N+2}(\Omega))$, $u_t \in L^2(0, T; H^{k+1}(\Omega))$, $u_{tt} \in L^2(0, T; L^2(\Omega))$, *satisfying (1.2),(1.3), and* U^n, W^n *given by (1.4)-(1.5) we have that for* $\Delta t < 1$

$$\||u - U|\|_{\infty,0} + \||u - W|\|_{\infty,0} + \left(\Delta t \sum_{n=1}^{N_T} \|\alpha^{1/2}\nabla(u^n - W^n)\|^2 \right)^{1/2}$$

$$\leq C \left(h^{k+1}\||u|\|_{\infty,k+1} + h^k \||u|\|_{2,k+1} + h^{k+1}\|u_t\|_{2,k+1} + \Delta t \|u_{tt}\|_{2,0} + (\Delta t)^{-1} h^{k+1} \||u|\|_{2,k+1} \right.$$

$$\left. + (\Delta t)^{-1} \left(\delta^{2N+2}\||u|\|_{2,2N+2} + (\delta h^k + h^{k+1})\||u|\|_{2,k+1} \right) \right). \tag{2.15}$$

Proof: We have that the true solution $u^n := u(n\Delta t, x)$ satisfies

$$\left(\frac{u^n - u^{n-1}}{\Delta t}, \phi \right) + (\mathbf{v}^n \cdot \nabla u^n, \phi) + (\alpha \nabla u^n, \nabla\phi) = (f^n, \phi) - \left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, \phi \right), \quad \forall \phi \in X_h. \tag{2.16}$$

With $\varepsilon^n := u^n - W^n$, $e^n := u^n - U^n$, subtracting (1.4) from (2.16) we have

$$\left(\frac{\varepsilon^n - e^{n-1}}{\Delta t}, \phi \right) + (\mathbf{v}^n \cdot \nabla\varepsilon^n, \phi) + (\alpha \nabla\varepsilon^n, \nabla\phi) = - \left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, \phi \right), \quad \forall \phi \in X_h. \tag{2.17}$$

Let $S^n \in X_h$. Additionally, define $\Lambda^n := u^n - S^n$, $F^n := S^n - W^n$ and $E^n := S^n - U^n$. Noting that $\varepsilon^n = \Lambda^n + F^n$, $e^n = \Lambda^n + E^n$, with the choice $\phi = F^n$ (2.17) becomes

$$\left(\frac{F^n - E^{n-1}}{\Delta t}, F^n \right) + (\mathbf{v}^n \cdot \nabla F^n, F^n) + (\alpha \nabla F^n, \nabla F^n) =$$

$$- \left(\frac{\Lambda^n - \Lambda^{n-1}}{\Delta t}, F^n \right) - (\mathbf{v}^n \cdot \nabla\Lambda^n, F^n) - (\alpha \nabla\Lambda^n, \nabla F^n)$$

$$- \left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, F^n \right). \tag{2.18}$$

We need a second equation for F^n and E^n. As W^n and U^n are connected through the *filter and deconvolve* equation (1.5), we use that equation. The true solution $u(\cdot, t^n) = u^n$ satisfies

$$u^n = D_h G_h u^n + (I - D_h G_h)u^n. \tag{2.19}$$

Subtracting (1.5) from (2.19) yields

$$e^n = D_h G_h \varepsilon^n + (I - D_h G_h)u^n \tag{2.20}$$

$$\text{i.e. } E^n = D_h G_h F^n - (I - D_h G_h)\Lambda^n + (I - D_h G_h)u^n. \tag{2.21}$$

Substituting (2.21) (with n replaced by $n - 1$) into (2.18) and rearranging yields

$$\left(\frac{F^n - D_h G_h F^{n-1}}{\Delta t}, F^n \right) + (\mathbf{v}^n \cdot \nabla F^n, F^n) + (\alpha \nabla F^n, \nabla F^n)$$

$$= - \left(\frac{\Lambda^n - \Lambda^{n-1}}{\Delta t}, F^n \right) - (\mathbf{v}^n \cdot \nabla\Lambda^n, F^n) - (\alpha \nabla\Lambda^n, \nabla F^n)$$

$$- \left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, F^n \right) - \left(\frac{(I - D_h G_h)\Lambda^{n-1}}{\Delta t}, F^n \right) + \left(\frac{(I - D_h G_h)u^{n-1}}{\Delta t}, F^n \right), \tag{2.22}$$

8

i.e., using $\nabla \cdot \mathbf{v}^n = 0$,

$$\|F^n\|^2 - \|D_h G_h F^{n-1}\|^2 + 2\Delta t \|\alpha^{1/2}\nabla F^n\|^2 \leq -2\left((\Lambda^n - \Lambda^{n-1}), F^n\right) - 2\Delta t\left(\mathbf{v}^n \cdot \nabla \Lambda^n, F^n\right)$$
$$- 2\Delta t\left(\alpha \nabla \Lambda^n, \nabla F^n\right) - 2\Delta t\left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, F^n\right) - 2\left((I - D_h G_h)\Lambda^{n-1}, F^n\right)$$
$$+ 2\left((I - D_h G_h)u^{n-1}, F^n\right). \tag{2.23}$$

Next we investigate the terms on the right hand side of (2.23).

$$\left((\Lambda^n - \Lambda^{n-1}), F^n\right) = \frac{1}{5}\Delta t\,\|F^n\|^2 + \frac{5}{4}\frac{1}{\Delta t}\|\Lambda^n - \Lambda^{n-1}\|^2$$
$$\leq \frac{1}{5}\Delta t\,\|F^n\|^2 + \frac{5}{4}\int_\Omega \left(\int_{(n-1)\Delta t}^{n\Delta t} |\Lambda_t|^2\,dt\right)d\Omega$$
$$= \frac{1}{5}\Delta t\,\|F^n\|^2 + \frac{5}{4}\int_{(n-1)\Delta t}^{n\Delta t} \|\Lambda_t\|^2\,dt. \tag{2.24}$$

Using the boundness of $\mathbf{v}$,

$$2\Delta t\left(\mathbf{v}^n \cdot \nabla \Lambda^n, F^n\right) \leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5\Delta t\,\|\mathbf{v}^n \cdot \nabla \Lambda^n\|^2$$
$$\leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5\,C_{\mathbf{v}}\,\Delta t\,\|\nabla \Lambda^n\|^2. \tag{2.25}$$

$$2\Delta t\left(\alpha \nabla \Lambda^n, \nabla F^n\right) \leq \Delta t\|\alpha^{1/2}\nabla F^n\|^2 + \Delta t\|\nabla \Lambda^n\|^2. \tag{2.26}$$

$$2\Delta t\left(u_t^n - \frac{u^n - u^{n-1}}{\Delta t}, F^n\right) \leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5\Delta t\,\|u_t^n - \frac{u^n - u^{n-1}}{\Delta t}\|^2$$
$$\leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5\Delta t\,\Delta t\int_{(n-1)\Delta t}^{n\Delta t}\int_\Omega |u_{tt}|^2\,d\Omega\,dt$$
$$= \frac{1}{5}\Delta t\,\|F^n\|^2 + 5(\Delta t)^2\int_{(n-1)\Delta t}^{n\Delta t} \|u_{tt}\|^2\,dt. \tag{2.27}$$

Using $\|(I - D_h G_h)\| \leq 1$,

$$-2\left((I - D_h G_h)\Lambda^{n-1}, F^n\right) \leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5(\Delta t)^{-1}\,\|(I - D_h G_h)\Lambda^{n-1}\|^2$$
$$\leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5(\Delta t)^{-1}\,\|\Lambda^{n-1}\|^2. \tag{2.28}$$

Similarly,

$$2\left((I - D_h G_h)u^{n-1}, F^n\right) \leq \frac{1}{5}\Delta t\,\|F^n\|^2 + 5(\Delta t)^{-1}\,\|(I - D_h G_h)u^{n-1}\|^2. \tag{2.29}$$

9

Combining estimates (2.24)-(2.29) with equation (2.23), summing from $n = 1$ to $n = l$, and using $\|D_h G_h\| \leq 1$, $\|F^0\| = 0$, we obtain

$$\|F^l\|^2 + \Delta t \sum_{n=1}^{l} \|\alpha^{1/2} \nabla F^n\|^2 \leq \Delta t \sum_{n=1}^{l} \|F^n\|^2 + 5C_{\mathbf{v}} \Delta t \sum_{n=1}^{l} \|\nabla \Lambda^n\|^2$$

$$+ \frac{5}{4} \sum_{n=1}^{l} \int_{(n-1)\Delta t}^{n\Delta t} \|\Lambda_t\|^2 \, dt + 5(\Delta t)^2 \sum_{n=1}^{l} \int_{(n-1)\Delta t}^{n\Delta t} \|u_{tt}\|^2 \, dt + 5(\Delta t)^{-1} \sum_{n=1}^{l} \|\Lambda^{n-1}\|^2$$

$$+ 5(\Delta t)^{-1} \sum_{n=1}^{l} \|(I - D_h G_h)u^n\|^2 . \tag{2.30}$$

The terms on the RHS can be further simplified using (2.1), (2.7).

$$5C_{\mathbf{v}} \Delta t \sum_{n=1}^{l} \|\nabla \Lambda^n\|^2 \leq C h^{2k} \Delta t \sum_{n=1}^{l} \|u^n\|_{k+1}^2 = C h^{2k} \|\|u\|\|_{2,k+1}^2 . \tag{2.31}$$

$$\frac{5}{4} \sum_{n=1}^{l} \int_{(n-1)\Delta t}^{n\Delta t} \|\Lambda_t\|^2 \, dt \leq C h^{2k+2} \|u_t\|_{2,k+1}^2 . \tag{2.32}$$

$$5(\Delta t)^2 \sum_{n=1}^{l} \int_{(n-1)\Delta t}^{n\Delta t} \|u_{tt}\|^2 \, dt \leq 5(\Delta t)^2 \|u_{tt}\|_{2,0}^2 . \tag{2.33}$$

$$5(\Delta t)^{-1} \sum_{n=1}^{l} \|\Lambda^{n-1}\|^2 \leq C (\Delta t)^{-2} h^{2k+2} \Delta t \sum_{n=1}^{l} \|u^{n-1}\|_{k+1}^2 \leq C (\Delta t)^{-2} h^{2k+2} \|\|u\|\|_{2,k+1}^2 . \tag{2.34}$$

$$5(\Delta t)^{-1} \sum_{n=1}^{l} \|(I - D_h G_h)u^n\|^2 \leq C (\Delta t)^{-2} \Delta t \sum_{n=1}^{l} \left(\delta^{4N+4} \|u^n\|_{2N+2}^2 + (\delta^2 h^{2k} + h^{2k+2}) \|u^n\|_{k+1}^2 \right)$$

$$= C (\Delta t)^{-2} \left(\delta^{4N+4} \|\|u\|\|_{2,2N+2}^2 + (\delta^2 h^{2k} + h^{2k+2}) \|\|u\|\|_{2,k+1}^2 \right) . \tag{2.35}$$

Using the bounds (2.31)–(2.35), together with Gronwall's Lemma, for $\Delta t < 1$ from (2.30) we obtain

$$\|F^l\|^2 + \Delta t \sum_{n=1}^{l} \|\alpha^{1/2} \nabla F^n\|^2$$

$$\leq C \exp(T) \left(h^{2k} \|\|u\|\|_{2,k+1}^2 + h^{2k+2} \|u_t\|_{2,k+1}^2 + (\Delta t)^2 \|u_{tt}\|_{2,0}^2 + (\Delta t)^{-2} h^{2k+2} \|\|u\|\|_{2,k+1}^2 \right.$$

$$\left. + (\Delta t)^{-2} \left(\delta^{4N+4} \|\|u\|\|_{2,2N+2}^2 + (\delta^2 h^{2k} + h^{2k+2}) \|\|u\|\|_{2,k+1}^2 \right) \right) . \tag{2.36}$$

With the triangle inequality and (2.36) we obtain the estimate for $(u - W)$ given in (2.15).

Note that from (2.21), and $\|D_h G_h\| \leq 1$,

$$\|E^n\| \leq \|F^n\| + \|\Lambda^n\| + \|(I - D_h G_h)u^n\| . \tag{2.37}$$

This estimate, combined with the triangle inequality and (2.36), then gives the bound for $(u - U)$ in (2.15).

∎

In the case of a continuous, piecewise linear finite element approximation we obtain the following.

Corollary 1 *For u and satisfying the hypothesis of Theorem 1, with $k = 1$, $\delta = h$, $N = 1$, we have that for $\Delta t < 1$*

$$\||u - U|\|_{\infty,0} + \||u - W|\|_{\infty,0} + \left(\Delta t \sum_{n=1}^{N_T} \|\alpha^{1/2}\nabla(u^n - W^n)\|^2 \right)^{1/2} \leq C \left((\Delta t) + h^1 + (\Delta t)^{-1}h^2 \right) . \tag{2.38}$$

∎

3 Numerical Results

We numerically investigate the Approximation Algorithm using three examples. Two of the examples exhibit a moving front propagating through a domain; the first (Example 1) is in one spatial dimension, the second (Example 2) is in two spatial dimensions. The third example uses a flux function $u - \alpha u_x$, where $\alpha = 1$ in the parabolic part of the domain and 0 in the hyperbolic part. This problem was studied in [9], where the authors used local discontinuous Galerkin methods to resolve the interface.

The Approximation Algorithm does not require that the stabilized approximation in **Step 2** be computed on the same grid as the approximation in **Step 1**.

There are two modifications of the Approximation Algorithm that we have found to be computationally useful. The first is to filter the approximation W^n to obtain U^n on a refined grid, or equivalently, in an enriched approximation space $X_{h/2}$. This has the effect of reducing the spatial width of large transition in the approximation. The second modification is to let U^n be a (convex) linear combination of the extrapolated approximation (using U^{n-1} and U^{n-2}) plus the filtered approximation. The extrapolated approximation is used instead of W^n to avoid the introduction of spurious oscillations into the approximation. These modification may be summarized as:
Step 2mod. $U^n \in X_h$ satisfies

$$\tilde{U}^n = R_{h/2 \to h} D_{h/2} G_{h/2}(W^n) , \tag{3.1}$$
$$U^n = (1 - \chi)(2U^{n-1} - U^{n-2}) + \chi\tilde{U}^n . \tag{3.2}$$

where $R_{h/2 \to h}$ denotes a restriction operator from $X_{h/2}$ to X_h.

For the three examples considered below we construct a continuous piecewise linear approximation to the solution, i.e. $k = 1$, use $\delta = h$ for the filter radius, $N = 1$ for the deconvolution operator, and $\chi = 0.9$ in (3.2).

11

3.1   Example 1

This example models the propogation of a moving front through a domain. The modeling equation is

$$u_t \; + \; (v\,u \; - \; b\,u_x)_x \;\; = \;\; 0\,, \; -1 < x < 3\,, \; t > 0\,. \tag{3.3}$$

The true solution is $u(x,t) \;\; = \;\; a - c\,tanh(c(x - at)/(2b))$, where in (3.3) $v = u/2$ and a, b, c are constants. The initial condition and boundary conditions are chosen to match the true solutions.

For the simulations presented in Figures 3.1, 3.2 the values used for a, b, and c are 1.0, 0.01 and 1.0, respectively.

The true solution is shown on the left in Figure 3.1, and the unstabilized solution is shown on the right. The unstabilized approximation was computed on a grid with mesh spacing $h = 1/8$ using a time step size of $\Delta t = 1/16$. The profiles are shown at times $T = 0, 1, 1.5$, and 2.

The unstabilized solution exhibits large oscillations along the front. Notice that the smoothing operators resolve the oscillations, but when simply using **Step 2**, this resolution occurs at the expense of having the transition region spread over a larger portion of the domain. The overshoot and undershoot are well resolved when using **Step 2mod**, where the filtering and deconvolution is performed on a finer mesh, $h = 1/16$.
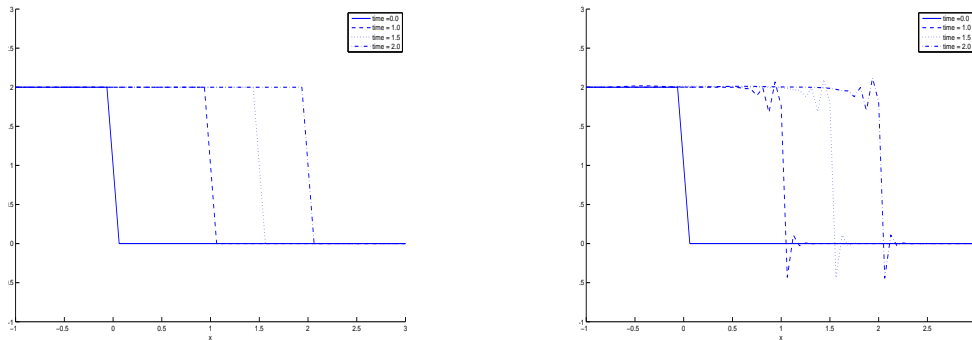

Figure 3.1: True solution (left), unstabilized solution (right).

3.2   Example 2

Next we consider the propogation of a moving front in two spatial dimensions. The modeling equation studied is

$$u_t \; + \; \mathbf{v} \cdot \nabla u \; - \; \mu \Delta u \;\; = \;\; 0\,, (x, y) \in (0, 1) \times (0, 1)\,, \;\; t > 0\,. \tag{3.4}$$
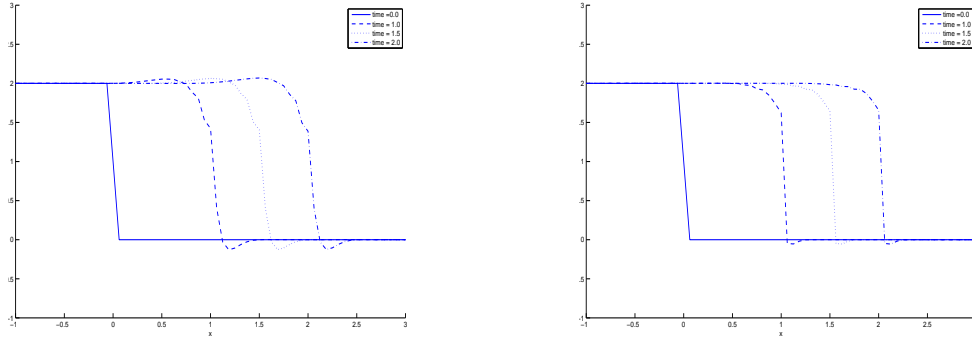
Figure 3.2: Stabilized approximation using **Step 2** (left) and **Step 2mod** (right).

We have as a true solution to (3.4) $u(x, y, t) = w(x, t)w(y, t)$ (from [4]), where $\mathbf{v} = [w(x, t) \;\; w(y, t)]^T$, and

$$w(x, t) = \frac{0.1A(x, t) + 0.5B(x, t) + C(x, t)}{A(x, t) + B(x, t) + C(x, t)}, \quad \text{with} \quad A(x, t) = e^{-0.05(x-0.05+4.95t)/\mu},$$

$$B(x, t) = e^{-0.25(x-0.05+0.75t)/\mu}, \quad C(x, t) = e^{-0.5(x-0.375)/\mu}.$$

For the simulations presented below we use $\mu = 10^{-2.25}$. We impose Dirichlet boundary conditions, and the initial condition, chosen to match the true solution.

The true solution and the unstabilized approximation are shown in Figure 3.3 for time $T = 0.6$. As in the one-dimensional example, the computed solution exhibits nonphysical oscillations along the front. The solution is computed using the FreeFem [13] environment on an unstructured mesh with 15 nodal points along each of the boundary edges, resulting in a total of 298 nodal points, and using $\Delta t = 0.01$.
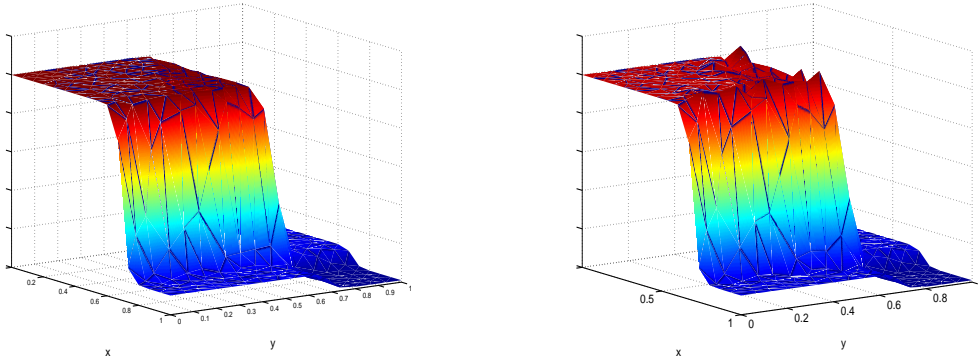

Figure 3.3: True solution (left), unstabilized solution (right)

The stabilized approximations are shown in Figure 3.4. Again the nonphysical oscillations are controlled by the filtering and deconvolution step, with **Step 2mod** resulting in a sharper transition region.
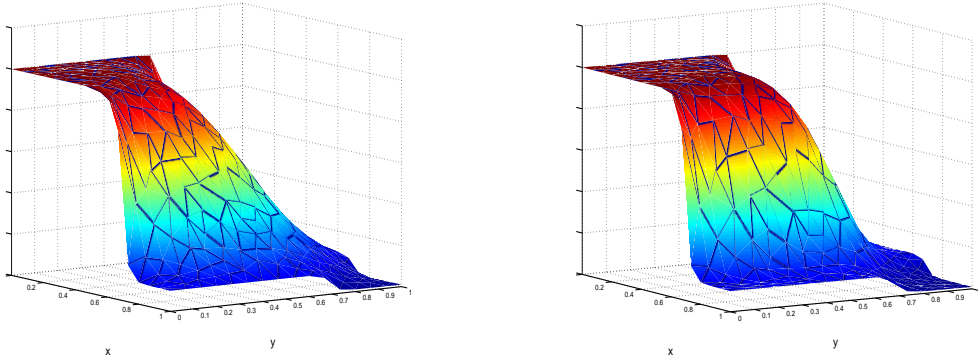
Figure 3.4: Stabilized approximation using **Step 2** (left) and **Step 2mod** (right).

Presented in Table 3.1 are the errors associated with the approximation using **Step 2**. The experimental convergence rate is in good agreement with the predicted convergence rate from Corollary 2.38 of 1.

m	N_T	$\|\|u - U\|\|_{2,2}$	Cvg. rate	$\|\|\nabla(u - W)\|\|_{2,2}$	Cvg. rate
20	80	9.0925E-002	0.85	1.5696E-001	0.38
25	100	7.5161E-002	0.88	1.4419E-001	0.34
30	120	6.4029E-002	1.00	1.3554E-001	0.47
35	140	5.4914E-002	1.08	1.2604E-001	0.54
40	160	4.7563E-002	1.16	1.1725E-001	0.62
45	180	4.1510E-002	1.28	1.0902E-001	0.74

Table 3.1: Experimental convergence rates for the approximations of Example 2 using **Step 2**.

3.3 Example 3

In this example, taken from [9], we investigate the approximation of $u(x,t)$ satisfying

$$u_t + (u - \alpha u_x)_x = f(x), \quad 0 < x < 2, \ t > 0, \tag{3.5}$$

where $\alpha = 1$ on $(0,1) \cup (1.5,2)$ and $\alpha = 0$ on $(1,1.5)$.

The equation is parabolic on $[0,1)$, hyperbolic on $[1,1.5)$, and parabolic again on $[1.5,2]$. We use as the solution to (3.5)

$$u(x,t) = \begin{cases} -\sin(x-1) + \exp(x-1) + t(x-1)^2, & 0 \le x < 1 \\ (x-1)^2(t+1) + 1, & 1 \le x < 1.5 \\ 1 + (x-1.5)\left(x^2 + (0.25t - 3.5)x + (2.75 - 0.625t)\right), & 1.5 \le x \le 2 \end{cases},$$

which is discontinuous at $x = 1.5$, the hyperbolic-parabolic interface. The flux, $(u - \alpha u_x)$, is continuous through the domain. For the boundary condition at $x = 0$ we specify the flux. At the outflow boundary condition $x = 2$ we use $u_x = 0$, which model a *far field* boundary.

14

The unstabilized approximation exhibits highly oscillatory behavior, Figure 3.5, whereas both of the stabilized approximations (corresponding to using **Step 2** and **Step 2mod**) remove these spurous oscillations, Figure 3.6.
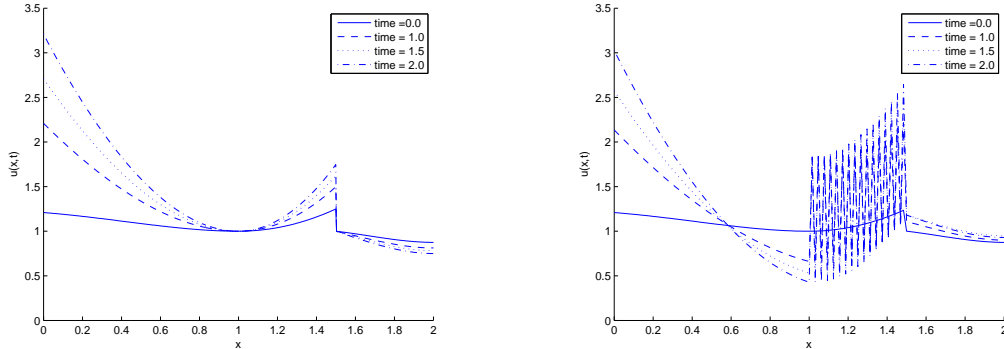

Figure 3.5: True solution (left), unstabilized approximation (right).

The approximation computed using **Step 1** and **Step 2** is computed on a grid with $h = \frac{1}{32}$ and $\Delta t = \frac{1}{128}$. The approximation computed using **Step 2mod**, shown on the right in Figure 3.6, uses in the second step a grid with $h = \frac{1}{64}$.
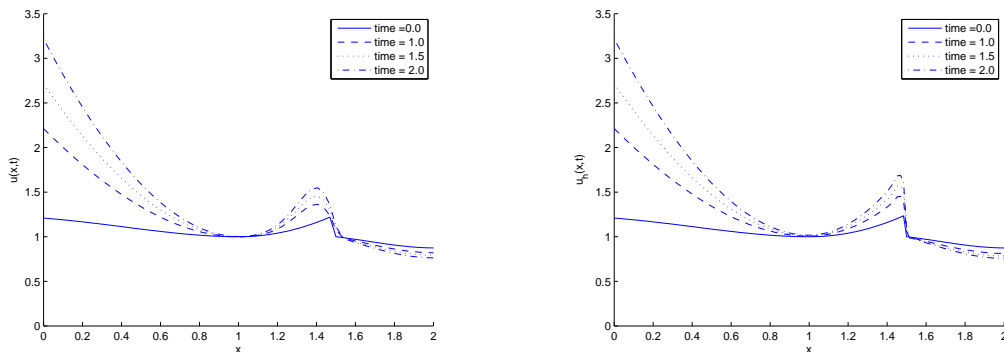

Figure 3.6: Stabilized approximation using **Step 2** (left) and **Step 2mod** (right)

4   Conclusions

In this paper we have investigated a filter-deconvolution stabilization method for advection dominated flow problems. The stabilization is a post-processing step applied to the approximation after each time step. As such, it can be easily incorporated into existing approximation procedures. Stability of the algorithm and optimal convergence rates have been shown. Numerical experiments are given which demonstrate the effectiveness of the algorithm.

Future work will investigate the use of other filter and deconvolution operators. In particular, for

fluid flow applications the Stokes filter [18] is very attractive as it preserves incompressibility of the velocity field.

References

[1] N. ADAMS AND S. STOLZ, *A subgrid-scale deconvolution approach for shock capturing*, J. Comput. Phys., 178 (2002), pp. 391–426.

[2] T. ARBOGAST AND M. WHEELER, *A characteristics-mixed finite element method for advection-dominated transport problems*, SIAM J. Numer. Anal., 32 (1995), pp. 404–424.

[3] L. BERSELLI, T. ILIESCU, AND W. LAYTON, *Large Eddy Simulation*, Springer, Berlin, 2004.

[4] M. BERZINS, *Temporal error control for convection-dominated equations in two space dimensions*, SIAM J. Sci. Comput., 16 (1995), pp. 558–580.

[5] J. BOYD, *Two comments on filtering (artificial viscosity) for Chebyshev and Legendre spectral and spectral element methods: Preserving boundary conditions and interpretation of the filter as a diffusion*, J. Comput. Phys., 143 (1998), pp. 283–288.

[6] B. COCKBURN AND C.-W. SHU, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM J. Numer. Anal., 35 (1998), pp. 2440–2463.

[7] G. DE MARSILY, *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*, Academic Press, Orlando Florida, 1986.

[8] A. DUNCA AND Y. EPSHTEYN, *On the Stolz-Adams deconvolution model for the large-eddy simulation of turbulent flows*, SIAM J. Math. Anal., 37 (2006), pp. 1890–1902.

[9] A. ERN AND J. PROFT, *Multi-algorithmic methods for coupled hyperbolic-parabolic problems*, Int. J. Numer. Anal. Model., 3 (2006), pp. 94–114.

[10] V. ERVIN, W. LAYTON, AND M. NEDA, *Numerical analysis of a higher order time relaxation model of fluids*, Int. J. Numer. Anal. Model., 4 (2007), pp. 648–670.

[11] R. EWING, T. RUSSELL, AND M. WHEELER, *Convergence analysis of an approximation of miscible displacement in porous media by mixed finite elements and a modified method of characteristics*, Comput. Methods Appl. Mech. Engrg., 47 (1984), pp. 73–92.

[12] M. GERMANO, *Differential filters of elliptic type*, Phys. Fluids, 29 (1986), pp. 1757–1758.

[13] F. HECHT, O. PIRONNEAU, A. L. HYARIC, AND K. OHTSUKA, *FreeFem++*. http://www.freefem.org/ff++, 2005.

[14] J. HEYWOOD AND R. RANNACHER, *Finite element approximation of the nonstationary Navier–Stokes problem. Part IV: Error analysis for second-order time discretization*, SIAM J. Numer. Anal., 2 (1990), pp. 353–384.

[15] T. HUGHES, L. MAZZEI, AND K. JANSEN, *Large eddy simulation and the variational multiscale method*, Comput. Vis. Sci., 3 (2000), pp. 47–59.

[16] V. John, S. Kaya, and W. Layton, *A two-level variational multiscale method for convection-dominated convection-diffusion equations*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 4594–4603.

[17] P. Knobloch and G. Lube, *Local projection stabilization for advection–diffusion–reaction problems: One-level vs. two-level approach*, Appl. Numer. Math., 59 (2009), pp. 2891–2907.

[18] W. Layton, C. Manica, M. Neda, and L. Rebholz, *Numerical analysis and computational testing of a high-order Leray-deconvolution turbulence model*, Num. Meth. Part. Diff. Eq., 24 (2008), pp. 555–582.

[19] W. Layton and M. Neda, *Truncation of scales by time relaxation*, J. Math. Anal. Appl., 325 (2007), pp. 788–807.

[20] Y. Mualem, *A new model for predicting the hydraulic conductivity of unsaturated porous media*, Water Resour. Res., 12 (1976), pp. 513–522.

[21] J. Mullen and P. Fischer, *Filtering techniques for complex geometry fluid flows*, Comm. Numer. Methods Engrg., 15 (1999), pp. 9–18.

[22] I. Stanculescu, *Existence theory of abstract approximate deconvolution models of turbulence*, Annali dell'Universitá di Ferrara, 54 (2008), pp. 145–168.

[23] M. van Genuchten, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil Sci. Soc. Am., 44 (1980), pp. 892–898.

[24] H. Wang, X. Shi, and R. Ewing, *An ELLAM scheme for multidimensional advection-reaction equations and its optimal-order error estimate*, SIAM J. Numer. Anal., 38 (2001), pp. 1846–1885.