ON THE ACCURACY OF THE FINITE ELEMENT METHOD PLUS TIME RELAXATION

J. CONNORS AND W. LAYTON

ABSTRACT. If $\overline{u}$ denotes a local, spatial average of u, then $u' = u - \overline{u}$ is the associated fluctuation. Consider a time relaxation term added to the usual finite element method. The simplest case for the model advection equation $u_t + u_x = f(x,t)$ is:

$$(u_{h,t} + u_{h,x}, v_h) + \chi(u'_h, v'_h) = (f(x,t), v_h).$$

We analyze the error in this and (more importantly) higher order extensions and show that the added time relaxation term not only suppresses excess energy in marginally resolved scales but also increases the accuracy of the resulting finite element approximation.

1. INTRODUCTION

In 1973 Dupont [Du73], in a landmark result, showed that in general the usual, continuous finite element method for first order hyperbolic equations converges suboptimally by one power of the mesh width h, even for infinitely smooth solutions, periodic boundary conditions and uniform meshes (see also Hedstrom [Hed79]). For less smooth solutions, it is also well known that the usual Galerkin FEM can produce highly oscillatory approximate solutions, e.g., [C79]. Even cases (for example linear elements and cubic splines) for which optimal convergence has been proven for periodic boundary conditions on uniform meshes (e.g., Dupont [Du73], Thomée and Wendroff [TW74]), optimal convergence rates are not expected on highly nonuniform meshes. Dupont's result for smooth solutions and the "wiggles" observed in tests for less smooth solutions have motivated the development of many nonstandard Galerkin methods, stabilizations and regularizations for first order hyperbolic problems and associated convection dominated, convection-diffusion equations. Examples include the SUPG method, Hughes [BH80], discontinuous Galerkin methods, Lesaint and Raviart [LR74], sub-grid artificial viscosity methods, e.g., [L02], [L05], Guermond [Guer99], Burman and Hansbo [BH04] and Braack, Burman, John and Lube [BBJL07].

Among these many variations on finite element methods, in complex applications there is a special interest in regularizations that are computational inexpensive, increase accuracy and incorporate numerical realizations of important physical processes omitted in the hyperbolic model equation. This report performs a numerical analysis of one such regularization which is motivated by work of Rosenau

Date: June, 2008.

Key words and phrases. time relaxation, deconvolution, hyperbolic equation, finite element method.

This paper is in final form and no version of it will be submitted for publication elsewhere. The work of both authors was partially supported by NSF grant DMS 0508260.

[R89] and Schochet and Tadmore [ST92] on the regularized Chapman-Enskog expansions of conservation laws. It has been extensively tested by Stolz and Adams and Kleiser in [AS02], [SA99], [SAK01a], [SAK01b], [SAK02] for compressible flows. The regularization is inexpensive and incorporates physical effects by time relaxation, Section 1.1, to damp fluctuations in time induced by marginally resolved scales in conservation laws and convection dominated problems. This regularization is thus physically interesting; it has been proven to truncate scales, [LN07], and is established in the practical computations of Stolz, Adams and Kleiser. In this report we study the complementary accuracy question:

Does this regularization also increase the asymptotic accuracy of the approximation as well as stabilize the approximation of under − resolved solutions?

To reduce the problem to a simple form, consider the advection equation: given smooth, known 1−periodic functions $u_0(x), f(x,t)$, find $u = u(x,t)$ satisfying

$$\begin{aligned} u_t + u_x &= f(x,t), x \in (0,1), 0 < t \leq T < \infty, \\ u(0,t) &= u(1,t), \text{ and } u(x,0) = u_0(x). \end{aligned} \tag{1.1}$$

The simplest example of the discretization.

The simplest example of the family of methods be considered is a small a variation on the usual finite element method. To present it, let

$$X = H^1_{\#}(0,1) := \{v \in L^2(0,1) : \frac{d}{dx}v \in L^2(0,1), v(0) = v(1)\},$$

and let $X_h \subset X$ denote a generic, conforming finite element space based on a mesh with representative mesh-width h and satisfying an approximation property typical of piecewise polynomials of degree k. The semi-discrete approximation begins with a chosen a filter length scale (traditionally denoted δ) and a relaxation parameter χ. Let over-bar denote a discrete local averaging over radius $O(\delta)$ (defined precisely in Section 1.2). Thus, given an approximate solution u_h its discrete average is denoted $\overline{u_h}^h$ and the associated fluctuation is $u_h{}' := u_h - \overline{u_h}^h$. Although our analysis is for a specific filter, it can be studied as well for many other filters. The main properties of averaging used in the analysis are $O(\delta^2)$ accuracy in $L^2(0,1)$ (with $L^2(0,1)$ norm denoted $||\cdot||$) and smoothing in the form

$$||\phi - \overline{\phi}^h|| \leq C\delta^{2-l}||\frac{d^{2-l}\phi}{dx^{2-l}}||, l = 0,1,2, \text{ and}$$

$$\delta^2||\triangle^h\overline{\phi}^h|| + \delta||\frac{d}{dx}\overline{\phi}^h|| + ||\overline{\phi}^h|| \leq C||\phi||.$$

The zeroth order example of the approximations we consider is: find $u_h : [0,T] \to X_h$ satisfying

$$\begin{aligned} &(u_{h,t} + u_{h,x}, v_h) + \chi(u_h{}', v_h') = 0, \forall v_h \in X_h, \\ &u_h(x,0) \text{ approximates } u_0 \text{ well.} \end{aligned} \tag{1.2}$$

This is the usual Galerkin approximation plus a time relaxation/stabilization term intended to damp small fluctuations, see Section 1.1. In other studies, the time relaxation term has often been added in the simpler form $\cdots + \chi(u_h{}', v_h)$. The difference between the term $\cdots + \chi(u_h{}', v_h')$ above and the simpler form $\cdots + \chi(u_h{}', v_h)$ is discussed in Section 5.2.

Adding in the term $\chi(u_h', v_h')$ introduces a consistency error in the discrete equations which, using its $O(\delta^2)$ accuracy, is

$$\textit{consistency error} = \sup_{v_h \in X_h} \frac{\chi(u_h{}', v_h')}{||v_h||} \simeq \sqrt{\chi||u - \overline{u}||^2} = C(u)\sqrt{\chi}\delta^2.$$

For an interesting example, if $\chi = O(h^{-1}), \delta = O(\sqrt{h})$ this consistency error term is $O(\sqrt{h})$. This suggests that the $N = 0$ case, (1.3) above, is not interesting and higher order (generalized) fluctuations are necessary to attain greater accuracy.

The higher order case.

The most important variant of (1.2), analyzed herein and introduced by Stolz, Adams and Kleiser in their computations of turbulent compressible flows [AS02], [SA99], [SAK01a], [SAK01b], [SAK02] (see also [Gue04]), is based on a higher order fluctuation model. Briefly, given a *continuous* averaging operator, denoted $\phi \to \overline{\phi}$ (see Section 3.1), a *continuous* deconvolution operator D_N is a bounded linear operator on $L^2(0,1)$ with the property

$$\phi = D_N\overline{\phi} + O(\delta^{2N+2}) \text{ for smooth } \phi. \tag{1.3}$$

In particular,

$$||\phi - D_N\overline{\phi}|| \leq C(N)\delta^{2N+2}||\frac{d^{2N+2}\phi}{dx^{2N+2}}|| \text{ , for } \phi \in H^{2N+2}_{\#}(0,1).$$

In the *discrete* case, considered herein, let $\overline{\cdot}^h$ and D_N^h denote discrete averaging and deconvolution operators. These act on X_h instead of X, are defined precisely in Section 2 and have properties analogous to the continuous case. The associated higher order[1], (discrete) generalized fluctuation is

$$u_h^* := u_h - D_N^h\overline{u_h}^h.$$

The (higher order) time relaxation discretization is then: find $u_h : [0,T] \to X_h$ satisfying

$$\begin{aligned} &(u_{h,t} + u_{h,x}, v_h) + \chi(u_h{}^*, v_h^*) = 0, \forall v_h \in X_h, \\ &\qquad u_h(x,0) \in X_h \text{ approximates } u_0 \text{ well.} \end{aligned} \tag{1.4}$$

Note that since $\phi = \overline{\phi} + O(\delta^2)$ (1.2) is the $N = 0$ case of (1.4).

The consistency error in the higher order method (1.4) is not limited (n the continuous case) since

$$\textit{consistency error} \simeq \sqrt{\chi||u - D_N\overline{u}||^2} = C(u)\sqrt{\chi}\delta^{2N+2}.$$

For example, let $\chi = O(h^{-1}), \delta = O(\sqrt{h})$ if $N = 0$ the consistency error is $O(\sqrt{h})$ while if $N = 1$ the consistency error is already $O(h^{3/2})$.

If $\chi = 0$, (1.4) reduces to the usual FEM which converges with suboptimal rate $O(h^k)$ with continuous piecewise polynomials of degree k:

$$\begin{aligned} \sup_{0\leq t\leq T} \|u(t) - u_h(t)\| \leq C\,\|u(0) - u_h(0)\| + \\ + C \min_{v_h:[0,T]\to X_h} \sup_{0\leq t\leq T} \{\|u - v_h\| + \|u_t - v_{h,t}\| + \|(u - v_h)_x\|\}. \end{aligned}$$

[1]As N increases to moderate values $\phi \to D_N\overline{\phi}$ becomes quite close to sharp spectral cutoff.

The classic paper of Dupont [Du73] shows that in general this result is unimprovable for the usual Galerkin method: the L^2 convergence rate of $O(h^3)$ is attained for Hermite cubics[2].

1.1. **The genesis and use of time relaxation stabilization.** Many stabilizations are used for convection dominated problems and each has its own advantages and disadvantages. The present time relaxation regularization has minimal effect on the solution's large scales. It thus has promise for longer time calculations. It also does not change the order of the equation. Thus, in more complex problems no extra boundary conditions (either explicit or implicit) are needed and no artificial boundary or interior layers are introduced in the solution or its derivatives. When an evolution equation is solved as a part of a complex application in which an efficient filtering routine is implemented, higher order time relaxation is also efficient in both computer time and programmer effort.

The time relaxation term first arose in theoretical studies of regularizations of Chapman-Enskog expansions of conservation laws in Rosenau [R89], Schochet and Tadmor [ST92]. The higher order time relaxation term was pioneered by Stolz, Adams and Kleiser in their large eddy simulations of compressible turbulence. As a stand alone regularization; it has been successful for the Euler equations for shock-entropy wave interaction and other tests, [AL99], [AS01], [AS02], [SAK01a], [SAK01b], [SAK02], as well as aerodynamic noise prediction and control, Guenaff [Gue04]. It was observed to ensure sufficient numerical entropy dissipation for numerical solution of conservation laws, Adams and Stolz [AS02], p.393. A mathematical foundation for its inclusion in models for turbulent flow has also been derived, [LN07], [ELN07].

2. Preliminaries: averaging and deconvolution.

Averaging and deconvolution present interesting new challenges for the numerical analysis of singularly perturbed differential equations. (They are themselves interesting, discrete, elliptic-elliptic singularly perturbation problems, [RST96], [SW83].) There are surely many ways yet to be discovered to use them to increase the accuracy of approximate solutions to many problems. Let

$$C^\infty_\# := \{\phi \in C^\infty_{loc}(\mathbb{R}) : \frac{d^l\phi}{dx^l}(0) = \frac{d^l\phi}{dx^l}(1), \text{ for all } l \geq 0\}$$

and let $H^k_\# = H^k_\#(0,1)$ denote the of the closure of $C^\infty_\#$ in the H^k norm. We let $X = H^1_\#(0,1)$ and $X_h \subset X$ denote a typical, finite element subspace of X associated with a maximum mesh size h. We shall suppose that the finite element space satisfies the following approximation assumption, typical of piecewise polynomials of degree k: for all $v \in X \cap H^{l+1}(0,1)$

$$\inf_{v_h \in X_h} \{h||\frac{d}{dx}(v - v_h)|| + ||v - v_h||\} \leq Ch^{l+1}|v|_{H^{l+1}}, \text{ for } 1 \leq l \leq k. \tag{2.1}$$

[2]Dupont shows that there exist infinitely smooth solutions for which a lower estimate for the error of $O(h^3)$ holds. His proof also shows that this suboptimal rate of convergence is the generic case.

2.1. **Averaging by continuous and discrete differential filters.** We define next the precise continuous and discrete differential filters used herein. These are related to the Yoshida regularization of semi-groups and to scale space analysis. Differential filters were introduced into flow modeling by Germano [Ger86]. The stabilization used is based on averaging by a *discrete differential filter* Manica and Kaya-Merdan [MM06]. Let $\delta > 0$ (and typically $1 \geq \delta \geq O(h)$) be the selected averaging radius.

Definition 1 (Continuous and discrete differential filter)**.** *Given $\phi \in L^2(0,1)$ its discrete average $G_h\phi = \overline{\phi}^h \in X_h$ is the unique solution of*

$$\delta^2(\overline{\phi}^h_x, v_{h,x}) + (\overline{\phi}^h, v_h) = (\phi, v_h), \forall v_h \in X_h. \tag{2.2}$$

The associated fluctuation is $\phi' := \phi - \overline{\phi}^h$.

The continuous differentially filtered average $G\phi = \overline{\phi} \in X$ is the unique solution of

$$\delta^2(\overline{\phi}_x, v_x) + (\overline{\phi}, v) = (\phi, v), \forall v \in X. \tag{2.3}$$

Associated with (2.2) define the usual discrete Laplacian operator $\triangle^h : L^2(0,1) \to X_h$ and projection $\Pi_h : L^2(0,1) \to X_h$ by

$$(\phi_x, v_{h,x}) = (-\triangle^h\phi, v_h), \forall v_h \in X_h \text{ , and}$$
$$(\phi, v_h) = (\Pi_h\phi, v_h), \forall v_h \in X_h.$$

With these definitions, the discrete filter (2.2) can be written $(-\delta^2\triangle^h + \Pi_h)\overline{\phi} = (\Pi_h\phi)$ or

$$\overline{\phi}^h = G_h\phi = (-\delta^2\triangle^h + \Pi_h)^{-1}(\Pi_h\phi) \ , \tag{2.4}$$

and the continuous filter is $\overline{\phi} = G\phi = (-\delta^2\triangle + I)^{-1}\phi$.

The mathematical stability and accuracy properties of continuous and discrete averaging has been extensively studied in [BIL06], [D04], [DE06], [ELN07], [L07], [LL03], [LL05], [LL06a], [LMNR06], [LMNR08], [LN07], [MM06] for multi-dimensional domains because they are central to large eddy simulation of turbulent flows. Next, we recall and sharpen a few useful results from [LMNR08] specialized to the periodic case.

Lemma 1 (Stability, smoothing and accuracy of averaging)**.** *For $\phi \in X$ we have*

$$\delta^2||\triangle^h\overline{\phi}^h|| + \delta||\frac{d}{dx}\overline{\phi}^h|| + ||\overline{\phi}^h|| \leq C||\phi||, \text{ and } ||\frac{d}{dx}\overline{\phi}^h|| \leq C||\frac{d}{dx}\phi||.$$

If $\phi \in X$ and $\triangle\phi \in L^2(0,1)$

$$\delta^2||\frac{d}{dx}(\phi - \overline{\phi}^h)||^2 + ||\phi - \overline{\phi}^h||^2 \leq C \inf_{v_h \in X_h}\{\delta^2||\frac{d}{dx}(\phi - v_h)||^2 + ||\phi - v_h||^2\} + C\delta^4||\triangle\phi||^2.$$

For all $\phi \in X$

$$\delta^2||\frac{d}{dx}(\overline{\phi} - \overline{\phi}^h)||^2 + ||\overline{\phi} - \overline{\phi}^h||^2 = \min_{v_h \in X_h}\{\delta^2||\frac{d}{dx}(\overline{\phi} - v_h)||^2 + ||\overline{\phi} - v_h||^2\}. \tag{2.5}$$

Under the approximation assumption (2.1) and for $\overline{\phi} \in X \cap H^{k+1}(0,1)$

$$\delta^2||\frac{d}{dx}(\overline{\phi} - \overline{\phi}^h)||^2 + ||\overline{\phi} - \overline{\phi}^h||^2 \leq C(\delta^2 h^{2k} + h^{2k+2})|\overline{\phi}|^2_{H^{k+1}}. \tag{2.6}$$

Further, for $\overline{\phi} \in X \cap H^{k+1}(0,1)$

$$
\begin{aligned}
(2.7) \qquad ||\overline{\phi} - \overline{\phi}^h|| &\leq C\frac{h}{\delta} \min_{v_h \in X_h} \{\delta^2 ||\frac{d}{dx}(\overline{\phi} - v_h)||^2 + ||\overline{\phi} - v_h||^2\}^{\frac{1}{2}} \\
&\leq C(h^{k+1} + \delta^{-1}h^{k+2})|\overline{\phi}|_{H^{k+1}}.
\end{aligned}
$$

$$
\begin{aligned}
(2.8) \quad ||\overline{\phi} - \overline{\phi}^h||_{H^{-1}(0,1)} &\leq Ch \min_{v_h \in X_h} \{\delta^2 ||\frac{d}{dx}(\overline{\phi} - v_h)||^2 + ||\overline{\phi} - v_h||^2\}^{\frac{1}{2}} \\
&\leq C(\delta h^{k+1} + h^{k+2})|\overline{\phi}|_{H^{k+1}} \ , \textit{for } k \geq 1 \ , \\
(2.9) \quad ||\overline{\phi} - \overline{\phi}^h||_{H^{-1}(0,1)} &\leq Ch(\frac{h}{\delta}) \min_{v_h \in X_h} \{\delta^2 ||\frac{d}{dx}(\overline{\phi} - v_h)||^2 + ||\overline{\phi} - v_h||^2\}^{\frac{1}{2}} \\
&\leq C(h^{k+2}(1 + \frac{h}{\delta}))|\overline{\phi}|_{H^{k+1}} \ , \textit{for } k \geq 2 \ .
\end{aligned}
$$

Proof. The first two claims were proven in Lemma 2.11 and 2.12 in [LMNR08]. The third and fourth claim will follow from normal finite element error analysis, e.g., [SW83], accounting for the dependence upon δ. We give a brief proof next. Given ϕ, the equations for $\overline{\phi}$ and $\overline{\phi}^h$ are, respectively,

$$
\begin{aligned}
(2.10) \qquad \delta^2(\overline{\phi}^h_x, v_{h,x}) + (\overline{\phi}^h, v_h) &= (\phi, v_h), \forall v_h \in X_h, \\
\delta^2(\overline{\phi}_x, v_x) + (\overline{\phi}, v) &= (\phi, v), \forall v \in X.
\end{aligned}
$$

In other words, $\overline{\phi}^h$ is the usual Galerkin approximation of a symmetric and coercive problem so the third claim (2.5) holds (see, e.g., [SW83] for more details about estimates for elliptic-elliptic singular perturbation problems). The fourth (2.6) follows from the third plus the approximation assumption. The fifth follows from a classical duality argument for (2.10) as follows. Let Ψ be the unique $1-$periodic solution of

$$
(2.11) \qquad -\delta^2 \triangle \Psi + \Psi = \overline{\phi} - \overline{\phi}^h, \text{ on } (0,1) \text{ and } \Psi(0) = \Psi(1).
$$

The solution Ψ has the following regularity, for example [L07],

$$
\delta^2 ||\triangle \Psi|| + \delta ||\frac{d}{dx}\Psi|| + ||\Psi|| \leq C||\overline{\phi} - \overline{\phi}^h||.
$$

Subtracting the continuous and discrete equations in (2.10) above gives a standard Galerkin orthogonality condition

$$
\delta^2(\frac{d}{dx}(\overline{\phi} - \overline{\phi}^h), v_{h,x}) + (\overline{\phi} - \overline{\phi}^h, v_h) = 0, \forall v_h \in X_h
$$

The variational formulation of the dual problem (2.11) is

$$
\delta^2(\Psi_x, v_x) + (\Psi, v) = (\overline{\phi} - \overline{\phi}^h, v), \forall v \in X.
$$

Setting $v = \overline{\phi} - \overline{\phi}^h$ and using the Galerkin orthogonality gives, $\forall v_h \in X_h$,

$$
\begin{aligned}
||\overline{\phi} - \overline{\phi}^h||^2 &= \delta^2((\Psi - v_h)_x, (\overline{\phi} - \overline{\phi}^h)_x) + (\Psi - v_h, \overline{\phi} - \overline{\phi}^h) \leq \\
&\leq [\delta^2 ||(\Psi - v_h)_x||^2 + ||\Psi - v_h||^2]^{\frac{1}{2}} [\delta^2 ||(\overline{\phi} - \overline{\phi}^h)_x||^2 + ||\overline{\phi} - \overline{\phi}^h||^2]^{\frac{1}{2}}.
\end{aligned}
$$

Choosing v_h appropriately, using the approximation assumption (2.1) and the regularity of Ψ yields

$$
||\overline{\phi} - \overline{\phi}^h||^2 \leq C[(\frac{h}{\delta})^2 + (\frac{h}{\delta})^4]^{\frac{1}{2}} ||\overline{\phi} - \overline{\phi}^h|| [\delta^2 ||(\overline{\phi} - \overline{\phi}^h)_x||^2 + ||\overline{\phi} - \overline{\phi}^h||^2]^{\frac{1}{2}},
$$

so that, using the approximation assumption gives

$$\begin{aligned} ||\overline{\phi}-\overline{\phi}^h|| &\leq C\frac{h}{\delta}[\delta^2||(\overline{\phi}-\overline{\phi}^h)_x||^2+||\overline{\phi}-\overline{\phi}^h||^2]^{\frac{1}{2}} \leq \\ &\leq Ch^{k+1}[1+\frac{h}{\delta}]|\overline{\phi}|_{H^{k+1}}, \end{aligned}$$

which is the claimed L^2 error bound.

The H^{-1} estimate will follow similarly by duality. Indeed, let $\beta \in H^1_{\#}(0,1)$ be fixed but arbitrary and let Ψ be the $1-$periodic solution of

$$-\delta^2 \triangle \Psi + \Psi = \beta, \text{ on } (0,1) \text{ and } \Psi(0)=\Psi(1). \tag{2.12}$$

It is known, e.g., [L07], (and easily proven by Fourier series) that the solution Ψ satisfies

$$\delta^2||\frac{d^3}{dx^3}\Psi||+\delta||\frac{d^2}{dx^2}\Psi||+||\frac{d}{dx}\Psi|| \leq C||\frac{d}{dx}\beta||.$$

The variational formulation of the dual problem (2.12) is that

$$\delta^2(\Psi_x,v_x)+(\Psi,v)=(\overline{\phi}-\overline{\phi}^h,v), \forall v \in X.$$

Setting $v=\overline{\phi}-\overline{\phi}^h$ and using the Galerkin orthogonality (2.10) gives, $\forall v_h \in X_h$,

$$\begin{aligned} (\beta,\overline{\phi}-\overline{\phi}^h) &= \delta^2((\Psi-v_h)_x,(\overline{\phi}-\overline{\phi}^h)_x)+(\Psi-v_h,\overline{\phi}-\overline{\phi}^h) \leq \\ &\leq [\delta^2||(\Psi-v_h)_x||^2+||\Psi-v_h||^2]^{\frac{1}{2}}[\delta^2||(\overline{\phi}-\overline{\phi}^h)_x||^2+||\overline{\phi}-\overline{\phi}^h||^2]^{\frac{1}{2}}. \end{aligned}$$

Using the approximation assumption and the regularity of Ψ yields

$$\begin{aligned} (\beta,\overline{\phi}-\overline{\phi}^h) &\leq Ch||\frac{d}{dx}\beta||[\delta^2||(\overline{\phi}-\overline{\phi}^h)_x||^2+||\overline{\phi}-\overline{\phi}^h||^2]^{\frac{1}{2}}, \text{ if } k \geq 1 \text{ and,} \\ (\beta,\overline{\phi}-\overline{\phi}^h) &\leq Ch(\frac{h}{\delta})||\frac{d}{dx}\beta||[\delta^2||(\overline{\phi}-\overline{\phi}^h)_x||^2+||\overline{\phi}-\overline{\phi}^h||^2]^{\frac{1}{2}}, \text{ if } k \geq 2. \end{aligned}$$

The last two results follows by dividing by $||\frac{d}{dx}\beta||$ and taking the supremum of $\beta \in H^1_{\#}(0,1)$. □

2.2. **Deconvolution.** The deconvolution problem, central in image processing, e.g., [BB98], is

given $\overline{\phi}$ *(+noise)*, find ϕ *(approximately).*

One of the most basic deconvolution methods is the van Cittert algorithm, [vC31], [BB98]. For continuous deconvolution, it is equivalent to N steps of first order Richardson iteration for the equivalent problem

$$\text{given } \overline{u} \text{ solve } u=u+\{\overline{u}-Gu\} \text{ for } u.$$

Algorithm 1 (van Cittert Approximate Deconvolution)**.** *Set* $v_0=\overline{u}$ *and fix* N
for $n=1,2,\cdots,N-1$, *perform*
$v_{n+1}=v_n+\{\overline{u}-Gv_n\}$
Define $D_N\overline{u}:=v_N$.

The discrete van Cittert deconvolution operator is defined by substituting G_h for G and $\overline{u}^h$ for $\overline{u}$ in the above algorithm. The discrete van Cittert deconvolution operator will be denoted by D_N^h. By eliminating the intermediate steps, the N^{th} van Cittert deconvolution operators D_N and D_N^h are given explicitly by

$$D_N\phi := \sum_{n=0}^{N}(I-G)^n\phi \text{ and } D_N^h\phi := \sum_{n=0}^{N}(I-G_h)^n\phi. \tag{2.13}$$

The van Cittert operator acts like an extrapolation in scale space from resolved to unresolved scales. For example, the approximate de-convolution operator corresponding to $N = 0, 1, 2$ are:

$$\begin{aligned} D_0\overline{u} &= \overline{u}, \\ D_1\overline{u} &= 2\overline{u} - \overline{\overline{u}}, \\ D_2\overline{u} &= 3\overline{u} - 3\overline{\overline{u}} + \overline{\overline{\overline{u}}}. \end{aligned}$$

Definition 2 (Deconvolution error)**.** *Given ϕ, the deconvolution error is, respectively, in the continuous or discrete cases given by*

$$\textit{deconvolution error } = \phi - D_N\overline{\phi} \;\; \textit{or} \;\; = \phi - D_N^h\overline{\phi}^h \;\; .$$

The deconvolution error plays a fundamental role in the consistency error inherent in algorithms based on deconvolution methods. Although there remain many open questions about even these simplest deconvolution operators, a theory is beginning to develop. We summarize next a few points, sharpening the results, as necessary, for the hyperbolic problem.

Proposition 1 (Stability and accuracy of deconvolution)**.** *Let D_N and D_N^h be given by the van Cittert algorithm. Then D_N and $D_N^h : L^2(\Omega) \to L^2(\Omega)$ are bounded, self-adjoint positive operators as are $I - D_N$ and $I - D_N^h$. Further,*

$$\begin{aligned} ||D_N\overline{v}|| &\leq C(N)||v||, \forall v \in X, \\ ||D_N^h\overline{v}^h|| &\leq C(N)||v||, \forall v \in X, \\ ||\frac{d}{dx}D_N^h\overline{v}^h|| &\leq C(N)||\frac{d}{dx}v||, \forall v \in X. \end{aligned}$$

In the case of differential filters $A := -\delta^2\triangle + 1, A_h := (-\delta^2\triangle^h + \Pi_h)\Pi_h, G = A^{-1}$, and $G_h = (A_h)^{-1}\Pi_h$. Then

$$\phi - D_N\overline{\phi} = \delta^{2N+2}(-\triangle)^{N+1}(A)^{-(N+1)}\phi, \forall\phi \in L^2(\Omega), \tag{2.14}$$

$$\phi_h - D_N^h\overline{\phi}^h = \delta^{2N+2}(-\triangle^h)^{N+1}(A_h)^{-(N+1)}\phi, \forall\phi_h \in X_h. \tag{2.15}$$

Proof. The stability claims are in Lemma 2.11 of [LMNR08]. The first accuracy results (2.14) was proven by Stoltz and Adams [SA99] and independently by Dunca [D04] and Dunca and Epshteyn [DE06], see also [BIL06]. The second (2.15) is in Lemma 2.10 in [LMNR08]. □

Lemma 2 (smoothing property)**.** *For any $\phi_h \in X_h$*

$$\delta^2||\triangle^h(D_N^h\overline{\phi_h}^h)|| + \delta||\frac{d}{dx}(D_N^h\overline{\phi_h}^h)|| + ||D_N^h\overline{\phi_h}^h|| \leq C(N)||\phi_h||.$$

Proof. Consider the van Cittert algorithm. We have at its initiation $\phi_0 = \overline{\phi_h}^h$ which satisfies

$$\delta^2(\overline{\phi}_{h,x}^h, v_{h,x}) + (\overline{\phi_h}^h, v_h) = (\phi_h, v_h), \forall v_h \in X_h. \tag{2.16}$$

Setting $\phi_h = v_h$ and using various inequalities gives for $\phi_0 = \overline{\phi_h}^h$

$$|||\phi_0||| :=^{def} \{\delta^4||\triangle^h(D_N^h\phi_0)||^2+\delta^2||\frac{d}{dx}(D_N^h\phi_0)||^2+||D_N^h\phi_0||^2\}^{\frac{1}{2}} \leq C||\phi_h||. \tag{2.17}$$

For the first step of van Cittert

$$\phi_1 = \phi_0 + \{\overline{\phi_0}^h - \overline{\phi_h}^h\}.$$

Thus, by the triangle inequality and the last estimate

$$\begin{aligned} |||\phi_1||| &\leq |||\phi_0||| + |||\overline{\phi_0}^h||| + |||\overline{\phi_h}^h||| \\ &\leq C|||\phi_h||| + |||\overline{\phi_0}^h||| \end{aligned} \tag{2.18}$$

If the above estimate (2.17) is applied with ϕ_h replaced by ϕ_0 on the RHS we have

$$|||\overline{\phi_0}^h||| \leq C||\phi_0|| = C||\overline{\phi_h}^h|| \leq C||\phi_h||,$$

again by (2.17). Thus,

$$|||\phi_1||| \leq C||\phi_h||.$$

For the general case we proceed by induction. □

One question regards boundedness of the right-hand side of (2.14) and (2.15). Some partial answers, summarized below, are in [L07] and [LMNR08].

Proposition 2 (Deconvolution error estimates)**.** *For all* $\phi \in L^2(0,1)$

$$||\phi - D_N^h\overline{\phi}^h|| \leq C(N)||\phi - \overline{\phi}^h||,$$

while if $\phi \in X$ *and* $\triangle\phi \in L^2(0,1)$

$$||\phi - D_N^h\overline{\phi}^h|| \leq C(N) \inf_{v_h \in X_h} \{\delta^2||\frac{d}{dx}(\phi - \overline{\phi}^h)||^2 + ||\phi - \overline{\phi}^h||^2\}^{\frac{1}{2}} + C(N)\delta^2||\triangle\phi||.$$

Further, under the approximation assumption (2.1) and for $\phi \in X \cap H^{2N+2}(0,1) \cap H^{k+1}(0,1)$

$$||\phi - D_N^h\overline{\phi}^h|| \leq C(N)(\delta h^k + h^{k+1})||\phi||_{k+1} + C\delta^{2N+2}|\phi|_{2N+2}.$$

Under the same conditions,

$$||\phi - D_N^h\overline{\phi}^h|| \leq C(N)(h^{k+1} + \delta^{-1}h^{k+2})||\phi||_{k+1} + C\delta^{2N+2}|\phi|_{2N+2}.$$

Proof. The first two inequalities are proven in Lemmas 2.12 and 2.13 in [LMNR08]. The third combines Lemma 2.14 and Remark 2.14 in [LMNR08] with [L07]. The remainder is a sharpening of Lemma 2.14 in [LMNR08]. Since the proof of the shortened result has same structure as that of Lemma 2.14 in [LMNR08], we shall outline the proof where it is identical to that of Lemma 2.14 in [LMNR08] and give the details where it deviates.

To begin, we rewrite $\phi - D_N^h\overline{\phi}^h = (I - D_N^hG_h)\phi$ as

$$(I - D_N^hG^h)\phi = (I - D_NG)\phi + (D_N - D_N^h)G\phi + D_N^h(G - G^h)\phi. \tag{2.19}$$

This is a small but critical reordering of the decomposition of the corresponding one in the proof Lemma 2.14 in [LMNR08]. We know $||D_N^h|| \leq C(N)$ so

$$||D_N^h(G - G^h)\phi|| \leq C(N)||\overline{\phi} - \overline{\phi}^h|| \leq Ch^{k+1}(1 + \frac{h}{\delta})|\overline{\phi}|_{k+1}.$$

The bound of the RHS has been sharpened in Lemma 1 above from the corresponding one in [LMNR08]. The first term in (2.19) is bounded in Proposition 2 giving

$$||(I - D_N G)\phi|| \leq C\delta^{2N+2}||\triangle^{N+1}\phi|| \leq \delta^{2N+2}|\phi|_{2N+2}.$$

Consider the term $(D_N - D_N^h)G\phi$. Adding and subtracting terms in this order yields $G\phi$ (rather than $G_h\phi$) which is at least as smooth as ϕ. Thus, we estimate $(D_N - D_N^h)G\phi$ using the argument in [LMNR08] (Lemma 2.14). Indeed, with $\psi = G\phi$

$$(D_N - D_N^h)G\phi = \sum_{n=0}^{N}[(I-G)^n - (I-G_h)^n]\psi.$$

The $n = 0$ term vanishes while the $n = 1$ term is $\overline{\psi} - \overline{\psi}^h$, bounded in Lemma 1 above. Consider

$$\sum_{n=0}^{N}[(I-G)^n - (I-G_h)^n]\psi.$$

For $n = 2$ we have, adding and subtracting $(I - G_h)(I - G)$ (instead of the reverse $(I - G)(I - G_h)$ used in [LMNR08]), gives

$$(I-G)^2\psi - (I-G_h)^2\psi = (G_h - G)(I-G)\psi + (I-G_h)(G_h-G)\psi.$$

It is not difficult to show, using the spectral mapping theorem that both G_h and $I - G_h$ are SPD and $||I - G_h||_{L(L^2\to L^2)} \leq 1$. Thus,

$$\begin{aligned} ||(I-G)^2\psi - (I-G_h)^2\psi|| &\leq ||(G_h-G)(I-G)\psi|| + ||(G_h-G)\psi|| \\ &\leq C(h^{k+1} + \delta^{-1}h^{k+2})\{|(I-G)\overline{\phi}|_{k+1} + |\overline{\phi}|_{k+1}\}. \end{aligned}$$

Since $|\overline{\phi}|_{k+1} \leq C||\phi||_{k+1}$, $|\overline{\overline{\phi}}|_{k+1} \leq C||\phi||_{k+1}$ we have

$$||(I-G)^2\psi - (I-G_h)^2\psi|| \leq C(h^{k+1} + \delta^{-1}h^{k+2})||\phi||_{k+1},$$

and the result holds for $n = 2$. To complete the proof we continue by induction for $3 \leq n \leq N$ as in [LMNR08] only adding and subtracting terms in the same order as the above $n = 2$ case. □

The error in discrete deconvolution is bounded by the error in the best approximation in X_h. If filtering is itself inexpensive to perform then the van Cittert algorithm is economical in both computer time and programmer effort because it only requires repeated filtering. Various other deconvolution methods are also being developed for similar purposes such as Tikhonov [S07], [MS07] and optimized van Cittert [LS07], and could also be considered. It is also possible to define means and fluctuations by projections into hierarchical finite element spaces. This idea leads to more methods that are similar in motivation but whose analysis would be different in detail.

3. A quasi-static projection.

The reason for the improvement in accuracy for the time relaxation discretization is captured already in the analysis of an equilibrium projection in this section. The projection

$$Q : X \to X_h$$

is defined after some necessary notation as follows. When $N = 0$, $D_0^h = \Pi_h$ and $(I - D_0^h G_h)u = u - \overline{u}^h$ is the fluctuation about the mean (normally denoted u').

Analogously, for $N > 0$, $u - D_N^h G_h u = u - D_N^h \overline{u}^h$ represents a higher order, generalized fluctuation (for example [LMNR07], [LN07]) that we will denote by u^*.

Definition 3 (Higher order fluctuations)**.** *The generalized (higher order) fluctuation, denoted u^*, is*

$$u^* := u - D_N^h G_h u.$$

Given $\chi > 0$ and $\delta > 0$ and for $w \in X$, the projection $Q : X \to X_h$ by $w_h := Qw \in X_h$ is (unique) solution of the finite dimensional linear problem:

$$((w - w_h)_x + (w - w_h), v_h) + \chi((w - w_h)^*, v_h^*) = 0, \forall v_h \in X_h. \tag{3.1}$$

Lemma 3. *Let $\chi > 0$, $\delta \geq 0$. (3.1) has a unique solution. $Q : X \to X_h$ is a well-defined projection operator.*

Proof. The equations (3.1) defining Q reduce to a linear system for the projection so existence is implied by triviality of $ker(Q)$. To verify triviality, let $w = 0$, and set $v_h = w_h$. This gives

$$(w_{h,x}, w_h) + ||w_h||^2 + \chi||w_h^*||^2 = 0.$$

Under periodic boundary conditions $(w_{h,x}, w_h) = 0$ and thus $w_h = 0$. That $Q^2 = Q$ follows similarly. □

If $\chi = 0$, (3.1) reduces to the usual finite element method for a two point boundary value problem. Finite element error analysis shows that $w - Qw$ satisfies

$$||w - Qw|| \leq C \inf_{v_h \in X_h} \{||\frac{d}{dx}(w - v_h)|| + ||w - v_h||\}.$$

This estimate also leads to an asymptotic rate of convergence sub-optimal by one power of h, sharp for some elements, [L83], (as for hyperbolic problems, Dupont [Du73]). For other, special elements on uniform meshes and for smoother solutions this extra power of h can be recovered by a cancellation argument, e.g., Axelsson and Gustafsson [AG79]. On the other hand, if, as is commonly manifested as oscillations at the smallest resolved scale, this loss of accuracy comes in the behavior of the error at the smallest resolved scales, it is plausible that the extra control of small scales in the time relaxation discretization for $\chi > 0$ will lead to an increase in accuracy as well as stability.

Theorem 1 (projection error)**.** *Let D_N^h be the discrete van Cittert deconvolution operator. Let $w \in X$ and let Q be the projection (3.1). Then, the error in the projection satisfies*

$$||w - Qw||^2 + \chi||(w - Qw)^*||^2 \leq C(N) \inf_{v_h \in X_h} \{(1 + \delta^{-2})||w - v_h||^2 + \\ + \chi^{-1}||\frac{d}{dx}(w - v_h)||^2 + \chi||(w - v_h)^*||^2\}.$$

Proof. Let $\widetilde{w} \in X_h$ be arbitrary (for the moment) and write

$$w - Qw = (w - \widetilde{w}) - (Qw - \widetilde{w}) = \eta - \phi_h$$

where $\eta := w - \widetilde{w}$ and $\phi_h := Qw - \widetilde{w} \in X_h$. The projection equations can be rewritten as

$$(\phi_{h,x} + \phi_h, v_h) + \chi({\phi_h}^*, v_h^*) = (\eta_x + \eta, v_h) + \chi(\eta^*, v_h^*) = 0, \forall v_h \in X_h.$$

Setting $v_h = \phi_h$, using $(\phi_{h,x}, \phi_h) = 0$ and the usual inequalities gives

$$||\phi_h||^2 + \chi||\phi_h^*||^2 \leq 2(\eta_x, \phi_h) + ||\eta||^2 + \chi||\eta^*||^2. \tag{3.2}$$

The key term is (η_x, ϕ_h). The idea is to separate scales in this term and treat the large and small scales differently. The small scales are controlled by the stabilization in the time relaxation term and the large scales are treated as a consistency error term. (This idea has been used for other stabilizations in, for example, [L02], [L05] and Guermond [Gue04].) We treat the $N = 0$ case and the general case separately to make the ideas clear.

The N = 0 case.

In the $N = 0$ case this term is split into means and fluctuations and integrated by parts:

$$N = 0 : (\eta_x, \phi_h) = (\eta_x, \overline{\phi_h}^h + \phi_h') = -(\eta, (\overline{\phi_h}^h)_x) + (\eta_x, \phi_h').$$

Inserting this in the right hand side of (3.2) and using standard inequalities gives

$$||\phi_h||^2 + \chi||\phi_h'||^2 \leq ||\eta||^2 + \chi||\eta'||^2 + \frac{\chi}{2}||\phi_h'||^2 + \frac{C}{2\chi}||\eta_x||^2 + 2||\eta||||(\overline{\phi_h}^h)_x||.$$

We use the á priori bound $||(\overline{\phi_h}^h)_x|| \leq \frac{1}{2\delta}||\phi_h||$ from Lemma 1 (tracking the constant through its proof) and $2||\eta||||(\overline{\phi_h}^h)_x|| \leq ||\eta||\frac{1}{\delta}||\phi_h|| \leq \frac{1}{2}||\phi_h||^2 + \frac{\delta^{-2}}{2}||\eta||^2$. This yields

$$||\phi_h||^2 + \chi||\phi_h'||^2 \leq (2 + \delta^{-2})||\eta||^2 + 2\chi||\eta'||^2 + C\chi^{-1}||\eta_x||^2 + 2||\eta||.$$

Thus, by the triangle inequality

$$||w-Qw||^2+\chi||(w-Qw)'||^2 \leq C(N) \inf_{v_h \in X_h} \{(1+\delta^{-2})||w-v_h||^2+\chi^{-1}||\frac{d}{dx}(w-v_h)||^2+\chi||(w-v_h)'||^2\}.$$

The case of higher N≥ 1.

The proof for $N \geq 1$ follows the above $N = 0$ case only the splitting uses generalized fluctuations. Indeed, split

$$(\eta_x, \phi_h) = (\eta_x, D_N^h(\overline{\phi_h}^h) + \phi_h^*) = -(\eta, (D_N^h(\overline{\phi_h}^h))_x) + (\eta_x, \phi_h^*)$$

Both terms on the RHS are handled analogously to the $N = 0$ case with the second term treated identically. The first term on the RHS $(\eta, (D_N^h(\overline{\phi_h}^h))_x)$ is treated like the term $(\eta, (\overline{\phi_h}^h)_x)$ of the $N = 0$ case using the á priori estimate from Lemma 2 that $||(D_N^h(\overline{\phi_h}^h))_x|| \leq C\delta^{-1}||\phi_h||$. The remainder of the proof follows exactly the $N = 0$ case. □

Corollary 1. *Let X_h satisfy the approximation assumption (2.1). Let $w \in X$ be smooth and let Q be the projection (3.1). Then, the error in the projection satisfies*

$$||w - Qw||^2 + \chi||(w - Qw)^*||^2 \leq C(N, w)\{(1 + \delta^{-2})h^{2k+2} + \chi^{-1}h^{2k} + \chi h^{2k+2}\}.$$

Thus if $\delta = O(h^{\frac{1}{2}}), \chi = O(h^{-1})$

$$\begin{aligned} ||w - Qw|| &\leq C(N, w)h^{k+\frac{1}{2}}, \\ ||(w - Qw)^*|| &\leq C(N, w)h^{k+\frac{3}{2}} \end{aligned}$$

Proof. Use $||(w-v_h)^*|| \leq C||w-v_h||$ and the approximation assumption (2.1). □

4. Numerical analysis of the time dependent problem.

This section proves the following error estimate for the method (1.4) which (roughly speaking) states that the error in the method consists of the error in the equilibrium projection plus a consistency error term. With proper choice of χ, δ and N we obtain the improved rate of convergence of $O(h^{k+\frac{1}{2}})$ which seems to be typical of stabilized methods.

Theorem 2 (Convergence of the method with time relaxation)**.** *Let Q be the equilibrium projection (3.1). Let X_h satisfy 2.1. For $0 \le t \le T < \infty$ the error in the method (1.4) satisfies*

$$\sup_{[0,T]} ||u-u_h||^2 + \int_0^T \chi||(u-u_h)^*||^2 dt \le C(T,N)\{||u(0)-u_h(0)||^2 + \sup_{[0,T]} ||u-Qu||^2 + \int_0^T [\chi||(u-Qu)^*||^2 + ||u_t - Qu_t||^2 + \chi||u^*(t)||^2]dt\}$$

Remark 1. *The consistency error $\int_0^T \chi||u^*(t)||_*^2 dt$ is directly related to the deconvolution error and will be bounded using Proposition 2. Since $(Qu)_t = Q(u_t)$, the remaining terms will be estimated using Theorem 1.*

Proof. Subtraction shows that the error $e(t) := u(t) - u_h(t)$ satisfies

$$(e_t + e_x, v_h) + \chi(e^*, v_h^*) = \chi(u^*, v_h^*), \forall v_h \in X_h,$$

which is driven by the methods consistency error on the right-hand side. As usual, split the error as

$$e = \eta - \phi_h, \eta = u - Qu, \phi_h = u_h - Qu.$$

Rearranging the error equation, we then have, for any $v_h \in X_h$,

$$(\phi_{h,t}+\phi_{h,x}, v_h)+\chi(\phi_h^*, v_h^*) = (\eta_t-\eta, v_h)+\chi(u^*, v_h^*)+[(\eta_x+\eta, v_h)+\chi(\eta^*, v_h^*)]. \tag{4.1}$$

By the definition of the projection operator Q the bracketed term on the right-hand side vanishes. Setting $v_h = \phi_h$ gives

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}||\phi_h||^2 + \chi||\phi_h^*||^2 &= (\eta_t - \eta, \phi_h) + \chi(u^*, \phi_h^*) \\ &\le \frac{1}{2}[||\eta_t||^2 + ||\eta||^2] + \frac{\chi}{2}||u^*||^2 + \frac{\chi}{2}||\phi_h^*||^2 + ||\phi_h||^2. \end{aligned}$$

Equivalently,

$$\frac{d}{dt}||\phi_h||^2 + \chi||\phi_h^*||_*^2 \le [||\eta_t||^2 + ||\eta||^2] + \frac{\chi}{2}||u^*||_*^2 + 2||\phi_h||^2.$$

Gronwall's' inequality implies that for $0 \le t \le T < \infty$

$$||\phi_h(t)||^2 + \int_0^t \chi||\phi_h^*||^2 dt' \le ||\phi_h(0)||^2 + C(T)\int_0^t ||\eta_t(t')||^2 + ||\eta(t')||^2 + \frac{\chi}{2}||u^*(t')||^2 dt'.$$

The trial inequality then yields the claimed result. □

Using the error estimate for the projection in Theorem 1 gives the following.

Corollary 2. *For $0 \le t \le T < \infty$ the error in the method (1.4) satisfies*

$$\sup_{[0,T]} ||u-u_h||^2 + \int_0^T \chi||(u-u_h)^*||^2 dt \le C(T,N)\{||u(0)-u_h(0)||^2 +$$

$$\inf_{v_h:[0,T]\to X_h} \{\sup_{[0,T]}[(1+\delta^{-2}) \sup_{[0,T]} ||u-v_h||^2 + \chi^{-1}||\frac{d}{dx}(u-v_h)||^2 + \chi||(u-v_h)^*||^2] +$$

$$+ \int_0^T (1+\delta^{-2})||(u-v_h)_t||^2 + \chi^{-1}||\frac{d}{dx}(u-v_h)_t||^2 + \chi||(u-v_h)_t^*||^2 dt\} +$$

$$+ \int_0^T \chi||u^*(t)||^2 dt\}.$$

Proof. This follows from Theorems 1 and 2. □

By Proposition 2 we can estimate the consistency error which is the square root of the last term in the above error estimate. Indeed, under the approximation assumption (2.1)

$$\begin{aligned} \int_0^T \chi||u^*(t)||^2 dt &= \int_0^T \chi||(I - D_N^h G_h)u(t)||^2 dt \le \\ &\le \int_0^T C\chi h^{2k+2}(1+(\frac{h}{\delta})^2)||u(t)||_{k+1}^2 + C\chi\delta^{4N+4}|u(t)|_{2N+2}^2 dt \qquad (4.2) \\ &\le C(u)\chi\{h^{2k+2}(1+(\frac{h}{\delta})^2) + \delta^{4N+4}\}. \end{aligned}$$

Rates of convergence then follow.

Corollary 3. *Suppose the assumptions of Theorem 2 hold, u is smooth and the approximation assumption (2.1) hold. Then,*

$$\sup_{[0,T]} ||u-u_h||^2 + \int_0^T \chi||(u-u_h)^*||^2 dt \le C(T,N,u)\{||u(0)-u_h(0)||^2 +$$

$$+ (1+\delta^{-2})h^{2k+2} + \chi^{-1}h^{2k} + \chi h^{2k+2} + \chi h^{2k+2}(\frac{h}{\delta})^2 + \chi\delta^{4N+4}\}.$$

Proof. This is immediate. □

Taking square roots and collecting the leading terms (when $1 \ge \delta \ge h, \chi \ge 1$) in the above error estimate gives

$$error(t) \lesssim error(0) + \delta^{-1}h^{k+1} + \chi^{-\frac{1}{2}}h^k + \chi^{\frac{1}{2}}h^{k+1} + \chi^{\frac{1}{2}}h^{k+1}(\frac{h}{\delta}) + \chi^{\frac{1}{2}}\delta^{2N+2}\}$$

The error is optimized by

$$\delta \simeq \sqrt{h}\ ,\ \chi \simeq h^{-1} \text{ and } N \ge k.$$

These choices attains the accuracy for smooth solutions

$$error(t) \lesssim error(0) + Ch^{k+\frac{1}{2}}$$

which is suboptimal by one-half power of h for the L^2 error but super-optimal for the error in the generalized fluctuation.

5. Possible extension to another time relaxation term

The theory of the discretization has been developed for a simplified problem: one space dimension, periodic boundary conditions, no time discretization and a convenient form of the time relaxation term. We shall consider non trivial boundary conditions through a computational illustration in Section 6. Time discretization of the relaxation term is less understood. In the case of implicit methods, if (1.4) is discretized in time by the $(1,1)$-Padé / trapezoid method it is straightforward (although longer than the continuous time case) to prove stability, convergence and even superconvergence. Thus, there remains efficiency, which is critical in 3 dimensional problems for which (1.1) is a common simplified model. In full trapezoidal discretizations of the method (1.4), at each time step a linear system must be solved which, if assembled, is full due to the (non-local) filtering terms in the deconvolution operator D_N. When an iterative method is used to solve this linear system, this filtering occurs in a residual calculation and can be implemented *without* assembly. The action of D_N in this residual requires N solves with the coefficient matrix $(-\delta^2 \triangle^h + 1)$, which has condition number $O(1 + (\frac{\delta}{h})^2)$. On the other hand, if the term $\chi({u_h}^*,\ v_h^*) = \chi({u_h}^{**},\ v_h)$ is treated explicitly (as tested by Guenaff [Gue04] for $N = 0$), no special care is needed since this term involves filtering a known function $2N$ times. This suggests that for complex $3d$ problems, some combination of implicit (for stiff terms) and explicit (for the time relaxation term) methods would be the most efficient.

This section considers alternate forms of the time relaxation term. If the generalized fluctuation operator $v \to v^*$ is positive semi-definite (as with the van Cittert deconvolution operator $I - D_N^h G_h$), the added term in the method is often simplified to $\chi({u_h}^*, v_h)$. In this case some improvement in the error over the usual Galerkin FEM can be shown. This is sketched next.

When $I - D_N^h G_h$ is symmetric and positive

$$u, v \to ((I - D_N^h G_h)u, v)$$

defines a semi-inner product and semi-norm.

Definition 4. $(u,v)_* := ((I - D_N^h G_h)u, v)$ *, and* $||u||_* := (u,u)_*^{\frac{1}{2}}$*. Given* $\chi > 0$ *and* $\delta > 0$*, define the modified projection* $\widetilde{Q} : X \to X_h$ *by* $w_h := \widetilde{Q}w \in X_h$ *is unique solution of the finite dimensional linear problem:*

$$((w - w_h)_x + (w - w_h), v_h) + \chi((w - w_h)^*, v_h) = 0, \forall v_h \in X_h. \tag{5.1}$$

Following the analysis in Section 3, we have the following projection error estimate.

Theorem 3 (The modified projection's error)**.** *Consider the modified projection (5.1). Then, the error in the projection satisfies*

$$||w - \widetilde{Q}w||^2 + \chi||w - \widetilde{Q}w||_*^2 \le C(N) \inf_{v_h \in X_h} \{(1 + \delta^{-2})||w - v_h||^2 + \\ + \chi^{-1}||\frac{d}{dx}(w - v_h)||^2 + \chi||w - v_h||_*^2\}.$$

Proof. Let $\widetilde{w} \in X_h$ be arbitrary (for the moment) and write

$$w - \widetilde{Q}w = (w - \widetilde{w}) - (\widetilde{Q}w - \widetilde{w}) = \eta - \phi_h$$

where $\eta := w - \widetilde{w}$ and $\phi_h := Qw - \widetilde{w} \in X_h$. The projection equations can be rewritten as

$$(\phi_{h,x} + \phi_h, v_h) + \chi(\phi_h{}^*, v_h) = (\eta_x + \eta, v_h) + \chi(\eta^*, v_h) = 0, \forall v_h \in X_h.$$

Setting $v_h = \phi_h$, using $(\phi_{h,x}, \phi_h) = 0$, $(\eta^*, \phi_h) = (\eta, \phi_h)_* \leq ||\eta||_* ||\phi_h||_*$ and the usual inequalities gives

$$||\phi_h||^2 + \chi||\phi_h||_*^2 \leq 2(\eta_x, \phi_h) + ||\eta||^2 + \chi||\eta||_*^2. \tag{5.2}$$

The key term is (η_x, ϕ_h). This term is split into means and fluctuations and integrated by parts. Indeed, split

$$(\eta_x, \phi_h) = (\eta_x, D_N^h(\overline{\phi_h}^h) + \phi_h^*) = -(\eta, (D_N^h(\overline{\phi_h}^h))_x) + (\eta_x, \phi_h^*)$$

Inserting this in the right hand side of (5.2) and using standard inequalities gives

$$||\phi_h||^2 + \chi||\phi_h||_*^2 \leq ||\eta||^2 + \chi||\eta||_*^2 + \frac{\chi}{2}||\phi_h||_*^2 + \frac{C}{2\chi}||\eta_x||^2 + 2||\eta||||(D_N^h\overline{\phi_h}^h)_x||.$$

Using the á priori bound $||(D_N^h(\overline{\phi_h}^h))_x|| \leq C\delta^{-1}||\phi_h||$ yields

$$||\phi_h||^2 + \chi||\phi_h||_*^2 \leq (2 + \delta^{-2})||\eta||^2 + 2\chi||\eta||_*^2 + C\chi^{-1}||\eta_x||^2 + 2||\eta||.$$

Finally, the triangle inequality completes the proof. □

Consider the modified method: given $\chi > 0$, find $u_h : [0, T] \to X_h$ satisfying

$$\begin{aligned}(u_{h,t} + u_{h,x}, v_h) + \chi(u_h{}^*, v_h) &= 0, \forall v_h \in X_h, \\ &\quad u_h(x, 0) \text{ approximates } u_0(x) \text{ well}\end{aligned} \tag{5.3}$$

Adapting the error analysis in Theorem 2 yields the following result.

Theorem 4. *Let $\widetilde{Q}$ be the modified equilibrium projection. For $0 \leq t \leq T < \infty$ the error in the method (5.3) satisfies*

$$\begin{aligned}\sup_{[0,T]} ||u - u_h||^2 + \int_0^T \chi||u - u_h||_*^2 dt \leq C(T, N)\{||u(0) - u_h(0)||^2 + \sup_{[0,T]} ||u - \widetilde{Q}u||^2 \\ + \int_0^T [\chi||u - \widetilde{Q}u||_*^2 + ||u_t - \widetilde{Q}u_t||^2 + \chi^2||u^*(t)||^2]dt\}\end{aligned}$$

Proof. The error $e(t) := u(t) - u_h(t)$ satisfies

$$(e_t + e_x, v_h) + \chi(e^*, v_h) = \chi(u^*, v_h), \forall v_h \in X_h.$$

Split the error as $e = \eta - \phi_h, \eta = u - \widetilde{Q}u, \phi_h = u_h - \widetilde{Q}u$. Rearranging the error equation, gives, for any $v_h \in X_h$,

$$(\phi_{h,t} + \phi_{h,x}, v_h) + \chi(\phi_h^*, v_h) = (\eta_t - \eta, v_h) + \chi(u^*, v_h) + [(\eta_x + \eta, v_h) + \chi(\eta^*, v_h)].$$

Due to $\widetilde{Q}$ the bracketed term vanishes. Setting $v_h = \phi_h$ gives

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}||\phi_h||^2 + \chi||\phi_h||_*^2 &= (\eta_t - \eta, \phi_h) + \chi(u^*, \phi_h) \\ &\leq \frac{1}{2}[||\eta_t||^2 + ||\eta||^2] + \frac{1}{2}\chi^2||u^*||^2 + C||\phi_h||^2.\end{aligned}$$

In the critical step we use the inequality $\chi(u^*, \phi_h) \leq \frac{1}{2}\chi^2||u^*||^2 + \frac{1}{2}||\phi_h||^2$. The remainder of the proof follows the same as that of Theorem 2. □

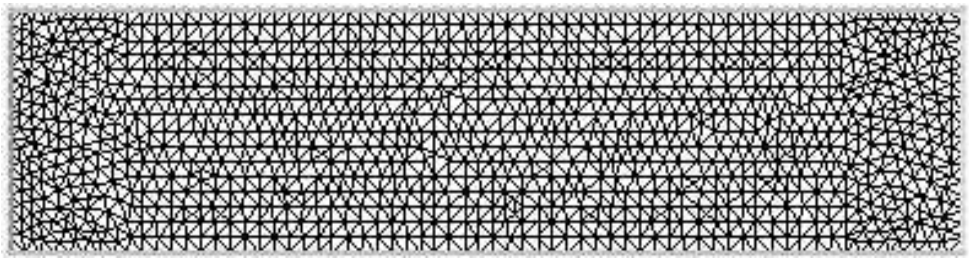

FIGURE 1. One Delaunay mesh used

The consistency error is bounded similarly with χ replaced by χ^2 as

$$\int_0^T \chi^2 ||u^*(t)||^2 dt \leq C(u)\chi^2\{h^{2k+2}(1+(\frac{h}{\delta})^2)+\delta^{2N+2}\}.$$

This gives, with the optimal choices $\delta \simeq \sqrt{h}$, $\chi \simeq h^{-\frac{2}{3}}$ and $N \geq k+1$, that the accuracy for smooth solutions is predicted to be somewhat less than the previous case:

$$error(t) \lesssim error(0) + Ch^{k+\frac{1}{3}}.$$

Remark 2. *We believe this estimate might be improvable. For example, one can try instead* $\chi(u^*,\phi_h) = \chi(u,\phi_h)_* \leq \frac{1}{2}\chi||u||_*^2 + \frac{1}{2}\chi||\phi_h||_*^2$. *Another possibility is that the consistency error estimate might be improvable through estimates in negative Sobolev norms because, for example,*

$$\begin{aligned}\int_0^T \chi||u||_*^2 dt &= \int_0^T \chi(u,u)_* dt = \int_0^T \chi(u^*,u)dt \leq \\ &\leq \sqrt{\int_0^T \chi||u||_{H^1(0,1)}^2 dt}\sqrt{\int_0^T \chi||u^*||_{H^{-1}(0,1)}^2 dt}.\end{aligned}$$

This is an open problem.

6. Two Computational Illustrations

There have been several simplifying assumptions in the formulation of the model hyperbolic problem including that the problem is linear, is in one space dimension and has periodic boundary conditions. Nonlinear conservation laws require spacial techniques that are beyond this report so the tests will focus upon the behavior of the method when the other two simplifications are relaxed. We let the spacial domain be a rectangle in $2d$. It is well known that many methods can have special (usually favorable) properties on uniform meshes and on meshes that are aligned exactly with the convection direction. To remove this effect, we always use meshes generated by a Delaunay algorithm. A typical example is plotted next in Figure 1.

All calculations were done using FreeFEM++, [HePi]. With unstructured meshes, we can select the convecting velocity to be $(1,0)$. Thus we have the following test problem with right hand side chosen so that the true solution is

$$u_{true}(x,y,t) = sin(4\pi y)sin(\pi x)sin(t),$$

given by find $u = u(x, y, t)$ satisfying

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} &= f(x, y, t), \text{ on } \Omega = (0, 1) \times (0, \frac{1}{4}), \text{ and } t > 0, \\ u(x, y, 0) &= 0, \text{ on } \Omega, \; u(0, y, t) = 0 \text{ for } t > 0, 0 < y < \frac{1}{4}. \end{aligned}$$

In the first table we give the error in the usual finite element method using quadratics on triangles with maximum triangle diameter given as h for the Delaunay mesh generated. The rates were calculated in the table in the standard way using the error at two successive h's and supposing $error(h) = Ch^a$ and then solving for the exponent a.

(6.1)

$h =$	L^2error	rate
4.62798e-2	2.41113e-4	—-
2.54801e-2	6.66330e-5	2.155
1.30216e-2	1.64784e-5	2.081
6.52674e-3	3.83958e-6	2.110
3.44127e-3	9.62077e-7	2.162

Usual FEM errors

A line of best fit through a log-log plot of errors gives convergence rate 2.116 for the usual FEM. This is consistent with an $O(h^2)$ error estimate for quadratics for the usual FEM. Next we add a time relaxation term to this test problem to test if, with the scaling predicted by the theory (in the simplified context) the accuracy does increase.

The fact that the theory is in a simplified context is potentially important. The averaging operator is the solution of a second order problem and the PDE is only first order and thus the PDE does not have enough boundary conditions for the averaging operator. It can be argued that many convection dominated problems contain small amounts of diffusion and that including this and the accompanying boundary conditions would completely resolve this issue. We, however, wanted to see how the method could perform for the pure hyperbolic limit and test the limitations of the error analysis. We also note that the question of boundary conditions for finite element methods for even 2×2 systems contains extra subtleties, [L83b] and Gunzburger [G77].

As a first (reasonable) guess for the extra boundary conditions needed we exploited the fact that in our chosen time stepping method we were always averaging a known function. Thus, given a function ϕ vanishing at the inflow boundary $x = 0$, we calculated its average $\phi \to \overline{\phi}^h$ by the usual FEM approximation of

$$\begin{aligned} -\delta^2 \triangle \overline{\phi} + \overline{\phi} &= \phi, \text{ in } \Omega, \\ \overline{\phi}(0, y) &= \phi(= 0), \text{ on the inflow boundary} \\ \overline{\phi} &= \phi, \text{ on the rest of the boundary.} \end{aligned}$$

Other definitions are possible and can be explored. The parameter values tested were $\delta = 0.1\sqrt{h}, \chi = 1/h$, and $N = 2$ deconvolution steps. These choices agree with the theoretical predictions of values that increase accuracy with quadratic elements. The following errors were observed. A line of best fit to the log-log plot of errors gives convergence rate 2.668 , consistent with the theoretical prediction of 2.5. A similar test (omitted for compactness) also showed no convergence rate

improvement over the usual FEM when using $N = 0$ and $N = 1$ deconvolution orders, also as predicted.

(6.2)

$h =$	L^2error	rate
4.62798e-2	1.50848e-4	—
2.54801e-2	2.08445e-5	3.316
1.30216e-2	3.10657e-6	2.836
6.52674e-3	5.53691e-7	2.497
3.44127e-3	9.59315e-8	2.739

Time Relax FEM errors

We note that although the exact solution is chosen to be exactly zero on the boundary, the approximate solution is zero only on the inflow boundary and only approximately zero on the rest of the boundary. One can ask if the extra boundary conditions in the averaging operator played any role in the errors. This can also be easily tested using the knowledge that the true solution is exactly zero on $\partial\Omega$. To do this we repeated the test but redefined the averaging operator $\phi \to \overline{\phi}^h$ by the usual FEM approximation of

$$-\delta^2 \triangle \overline{\phi} + \overline{\phi} = \phi, \text{ in } \Omega, \ \overline{\phi} = 0, \text{ on } \partial\Omega.$$

With the same parameters we observed the following errors.

(6.3)

$h =$	L^2error	rate
4.62798e-2	1.13634e-4	—-
2.54801e-2	1.87568e-5	3.018
1.30216e-2	2.55018e-6	2.972
6.52674e-3	3.20682e-7	3.002
3.44127e-3	4.28390e-8	3.145

Time Relax FEM errors

using extra BCs

A line of best fit to log-log plot of errors gives convergence rate 3.034. This convergence rate is beyond the theory.

The second test is a problem with a non-smooth solution. For the same domain, test problem, meshes, algorithmic parameters (δ, χ, N), pick the body forces to be identically zero, $f(x,t) \equiv 0$. We choose a discontinuous boundary condition, for $j = 1, 2$,

$$\begin{aligned} u(0,y,t) &= 0, 0 \le x \le \frac{1}{8}, \\ u(0,y,t) &= 1, \frac{1}{8} < x \le \frac{1}{4}. \end{aligned}$$

The initial condition was

$$\begin{aligned} u(x,y,0) &= 0, 0 \le x \le \frac{1}{8}, \\ u(x,y,0) &= e^{-x}, \frac{1}{8} < x \le \frac{1}{4}. \end{aligned}$$

The exact solution is a discontinuity that moves across the domain so that after $t = 1$ it reaches a steady discontinuous step. The usual FEM on a general unstructured mesh without any upwinding or limiters is particularly unsuitable for problems like this one. Thus, we do not expect excellent solution quality from any variation of the

methods studied. we test this problem to see if the time relaxation term gives any improvement at all for non smooth solutions. We give next approximate solution plots at $t = 1$ beginning with the (expected bad) usual FEM in Figure 2.

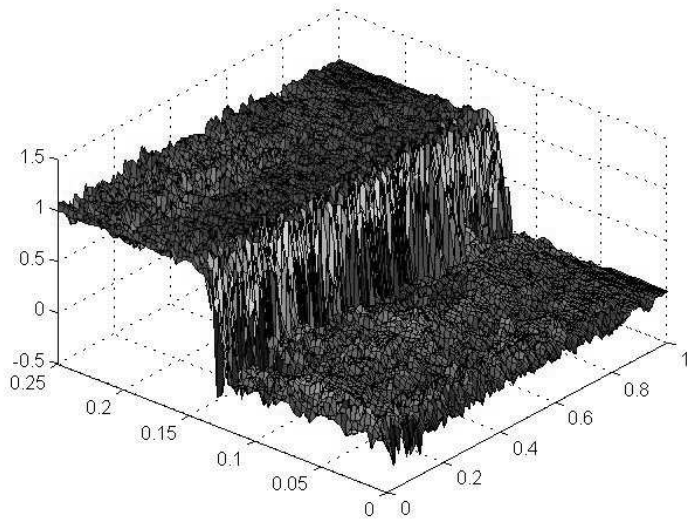

FIGURE 2. Usual FEM at t=1

The behavior is actually worse than the above seems to indicate. This is revealed when one zooms in to the solution in Figure 3.

Next we give the approximate solution obtained using the FEM+time relaxation for the same data and algorithmic parameters in Figure 4. Oscillations still occur. (The aim of time relaxation is not to prevent all oscillations but rather to damp those that do occur.) However, compared with Figure 2, the oscillations are much smaller than without time relaxation and the larger ones are confined to a neighborhood of the discontinuity. Zooming in to the worst subregion in Figure 5, compared to Figure 3, is consistent with the above description.

By computing the solution for various values of δ, we found the best value of the averaging radius for this problem to be to $\delta = 0.03\sqrt{h}$, Figure 6, below.

7. Conclusions

The theory of the regularized Chapman-Enskog expansions of conservation laws in Rosenau [R89] and Schochet and Tadmor [ST92] suggest scaling $\chi \simeq \delta^{-1}$. Numerical analysis of the error in the method (for smooth solutions) suggests that there is a difference in the discrete case and instead $\chi \simeq \delta^{-2}$. If one views the mesh-width h as the induced filter width instead, we recover the scaling $\chi \simeq (filter\ width)^{-1}$. On the other hand, in simulations of turbulent compressible flow it is

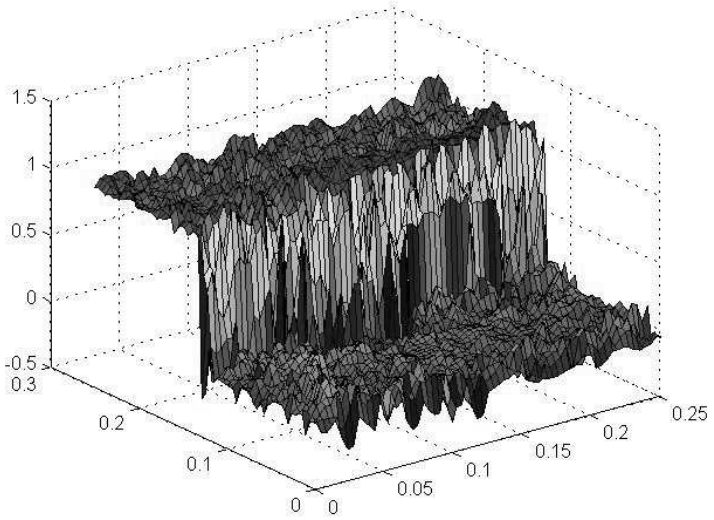

FIGURE 3. A zoom of the usual FEM solution

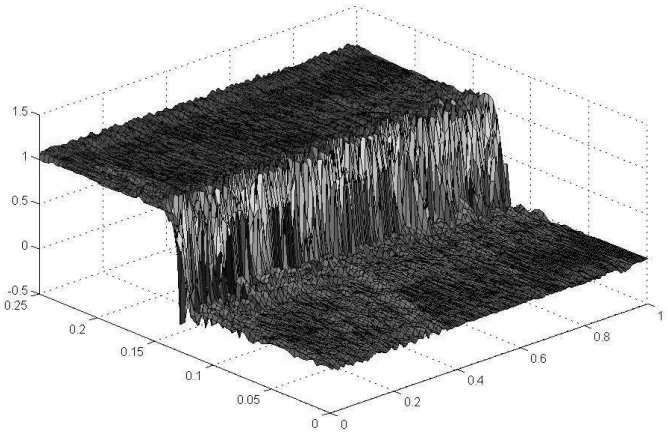

FIGURE 4. FEM+time relaxation solution

common practice to take $\delta = O(h)$ (for example, $\delta = 3h$) to try to squeeze maximum information from a given resolution. The numerical analysis herein suggests this might be over reaching and predicts $\delta \simeq \sqrt{h}$ as more accurate on the large scales. This is confirmed by the initial and very limited tests herein.

The last table also suggests that greater accuracy than proven herein might be hiding in the method (possibly for special elements) with a better averaging or

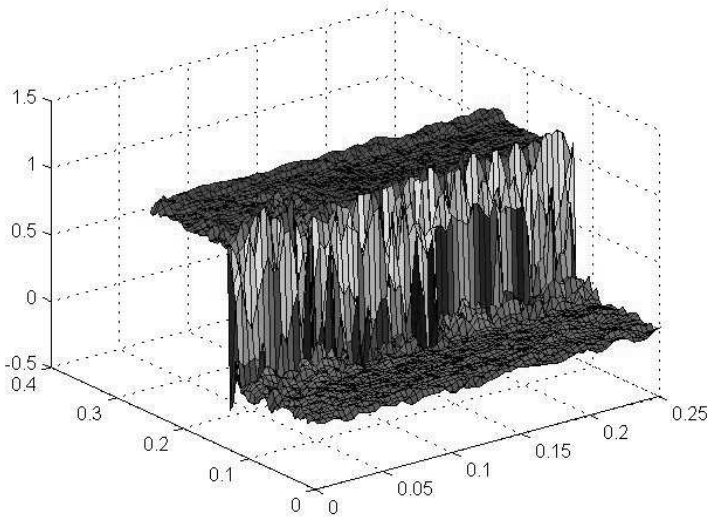

FIGURE 5. Zoom of the FEM+time relaxation solution

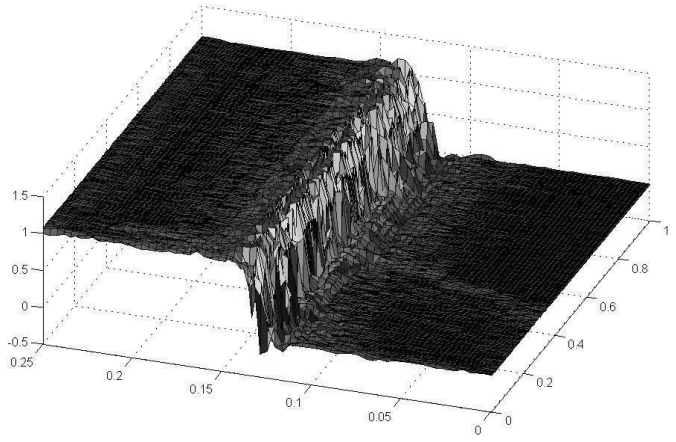

FIGURE 6. Time relaxation with $\delta = 0.03\sqrt{h}$

specification of extra boundary conditions. There are many (and obvious) possibilities testable. There are also cases where the $O(h^{k+\frac{1}{2}})$ error estimate can be provably improved to $O(h^{k+1})$. When the finite element space consist of even order splines on a uniform mesh, super convergence of the time relaxation discretization has been proven, [L07b], and also implies optimality in L^2. More generally, on a uniform or near the uniform mesh, the cancellation argument of Dupont [Du73] can be adapted provided a finite element basis exists which is symmetric about

the node associated with each basis function. This includes continuous, piecewise linears and cubic splines, Dupont [Du73], continuous quadratics and cubics, Axelsson and Gustafsson [AG79], but not C^1 Hermite cubics, Dupont [Du73], Hedstrom [Hed79].

The convergence result herein extends immediately to multi-dimension Friedrichs systems with non-periodic boundary conditions. Some extensions to convection dominated, convection diffusion problems are possible but possibly more delicate due to boundary and interior layers. Extension to nonlinear conservation laws is more delicate further and depends on structure of the nonlinearity in the specific conservation law and should be done in connection with limiters. Showing that the L^2 accuracy for slightly viscous Navier Stokes equations is greater than that proven [ELN07] is also an interesting and important open problem.

7.1. **Averaging and deconvolution operators.** To introduce the time relaxation discretization, the treatment of boundary conditions by the differential filter must be specified. With second order differential filters extra boundary conditions, beyond those of the first order continuous problem, must be supplied for the differential filter. This difficulty does not occur when solving convection diffusion equations or other (linear or nonlinear) second order problems. The simplest idea of specifying $\overline{\phi} = \phi$ on $\partial\Omega$ when filtering a known function ϕ was used herein and seemed to work.

The differential filter used herein is natural for finite element methods, second order problems and the well developed tools of finite element error analysis. However, it is also only *approximately* local. For hyperbolic problems, it is quite possible that a purely local averaging is preferable. Small but global averaging effects couple the approximate solution away from the step to the behavior at the discontinuity. Thus, it might contribute to the small background oscillations seen away from the discontinuity in Figures 4 and 3. This needs to be tested.

Since deconvolution is a well known and important ill posed problem there are very many deconvolution operators available for testing. Bertero and Boccacci [BB98] give many examples and we note the very interesting construction of Geurts [Geu97].

References

[AL99] N.A. Adams and A. Leonard, *Deconvolution of subgrid scales for the simulation of shock-turbulence interaction*, p. 201 in: *Direct and Large Eddy Simulation III*, (eds.: P. Voke, N.D. Sandham and L. Kleiser), Kluwer, Dordrecht, 1999.

[AS01] N. A. Adams and S. Stolz, *Deconvolution methods for subgrid-scale approximation in large eddy simulation*, *Modern Simulation Strategies for Turbulent Flow*, R.T. Edwards, 2001.

[AS02] N. A. Adams and S. Stolz, *A subgrid-scale deconvolution approach for shock capturing*, J.C.P., 178 (2002),391-426.

[AG79] A.O.H. Axelsson and J. Gustafsson, *Quasioptimal finite element approximation of first order hyperbolic and convection-diffusion equations*, pp. 273-281 in: *Analytical and Numerical Approaches to Asymptotic Problems in Analysis*, (eds: A.O.H. Axelsson, L.S. Frank, and A. van der Sluis) North Holland, Amsterdam, 1980.

[BIL06] L. C. Berselli, T. Iliescu and W. Layton, *Mathematics of Large Eddy Simulation of Turbulent Flows*, Springer, Berlin, 2006

[BB98] M. Bertero and B. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing Ltd.,1998.

[BBJL07] M. Braack, E. Burman, V. John and G. Lube, *Stabilized FEMs for the generalized Oseen problem*, CMAME 196(2007) 853-866.

[BH80] A. Brooks and T.J.R. Hughes, *Streamline upwind Petrov-Galerkin methods for advection-dominated flows*, in: 3rd Int. Conf. on FEM in Fluid Flow, 1980.

[BH04] E. Burman and P. Hansbo, *Edge stabilizations for Galerkin approximations of convection-diffusion-reaction problems*, CMAME 193(2004) 1437-1453.

[C79] G.F. Carey, *An analysis of stability and oscillations in convection-diffusion computations*, 63-71 in: FEM for Convection Dominated Flows, T.J.R. Hughes, ed., ASME, vol. 34 , 1979.

[D04] A. Dunca, *Space averaged Navier-Stokes equations in the presence of walls*, PhD Thesis, University of Pittsburgh, 2004.

[DE06] A. Dunca and Y. Epshteyn, *On the Stolz-Adams de-convolution model for the large eddy simulation of turbulent flows*, SIAM J. Math. Anal. 37(2006), 1890-1902.

[Du73] T. Dupont, *Galerkin methods for first order hyperbolics: an example*, SINUM 10(1973)890-899.

[ELN07] V. Ervin, W. Layton and M. Neda, *Numerical analysis of a higher order time relaxation model of fluids*, Int. J. Numer. Anal. and Modeling, 4(2007) 648-670.

[Gue04] R. Guenanff, *Non-stationary coupling of Navier-Stokes/Euler for the generation and radiation of aerodynamic noises*, PhD thesis: Dept. of Mathematics, Universite Rennes 1, Rennes, France, 2004.

[Guer99] J.-L. Guermond, *Subgrid stabilization of Galerkin approximations of monotone operators*, C. R. Acad. Sci. Paris, Série I, 328 7 (1999) 617-622.

[Ger86] M. Germano, *Differential filters of elliptic type*, Phys. Fluids, 29(1986), 1757-1758.

[Geu97] B. J. Geurts, *Inverse modeling for large eddy simulation*, Phys. Fluids, 9(1997), 3585.

[G77] M. Gunzburger, *On the stability of Galerkin methods for Initial-Boundary Value Problems for hyperbolic systems*, Math. Comp. 31 (1977) 661-675.

[HePi] F. Hecht and O. Pironneau, *FreeFem++* , webpage: http://www.freefem.org.

[Hed79] G. W. Hedstrom, *The Galerkin Method Based on Hermite Cubics*, SINUM,16(1979), 385-393.

[H99] T.J. Hughes, *Multiscale phenomena: the Dirichlet to Neumann formulation, subgrid models, bubbles and the origin of stabilized methods*, CMAME 127(1999), 387-401..

[L83] W. Layton, *Galerkin Methods for Two-Point Boundary Value Problems for First Order Systems*, SINUM, 20 (1983),161-171.

[L83b] W. Layton, *Stable Galerkin Methods for Hyperbolic Systems,* SINUM, 20, (1983) 221–233.

[L07] W Layton, *A remark on regularity of an elliptic-elliptic singular perturbation problem,* technical report, http://www.mathematics.pitt.edu/research/technical-reports.php, 2007.

[L07b] W. Layton , *Superconvergence of finite element discretization of time relaxation models of advection*, BIT, 47(2007) 565-576.

[L05] W. Layton, *Model reduction by constraints and an induced pressure stabilization*, J. Numer. Linear Algb. and Applications, 12[2005]547-562.

[L02] W Layton, *A connection between subgrid-scale eddy viscosity and mixed methods*, Applied Math and Computing, 133[2002],147-157

[LL03] W. Layton and R. Lewandowski, *A simple and stable scale similarity model for large eddy simulation: energy balance and existence of weak solutions*, Applied Math. letters 16(2003) 1205-1209.

[LL05] W. Layton and R. Lewandowski, *Residual stress of approximate deconvolution large eddy simulation models of turbulence. Journal of Turbulence*, 46(2): 1-21, 2006.

[LL06a] W. Layton and R Lewandowski, *On a well posed turbulence model*, Discrete and Continuous Dynamical Systems - Series B, 6(2006) 111-128.

[LMNR08] W. Layton, C. Manica, M. Neda and L. Rebholz, *Numerical analysis of a high accuracy Leray-deconvolution model of turbulence,* Numerical Methods for PDEs, 24(2008) 555-582.

[LMNR06] W. Layton, C. Manica, M. Neda and L. Rebholz, *The joint Helicity-Energy cascade for homogeneous, isotropic turbulence generated by approximate deconvolution models ,* to appear: Adv. and Appls. in Fluid Mechanics, 2007.

[LS07] W. Layton and I. Stanculescu, *K41 optimized deconvolution models*, to appear in: Int. J. Computing and Mathematics, 2007.

[LN07] W. Layton and M. Neda, *Truncation of scales by time relaxation*, JMAA 325(2007), 788-807.

[LR74] P. Lesaint and P.A. Raviart, *On a finite element method for solving the neutron transport equation*, 89-112 in: *Mathematical aspects of finite elements in partial differential equations*, C. de Boor, ed., Academic Press, N.Y., 1974.

[MM06] C. C. Manica and S. Kaya-Merdan, *Convergence Analysis of the Finite Element Method for a Fundamental Model in Turbulence*, tech. report, http://www.math.pitt.edu/techreports.html, 2006.

[MS07] C.C. Manica and I. Stanculescu, *Numerical Analysis of Tikhonov Deconvolution Model,* University of Pittsburgh Technical Report, http://www.math.pitt.edu/techreports.html, 2007.

[RST96] H.-G. Roos, M. Stynes and L. Tobiska, *Numerical methods for singularly perturbed differential equations*, Springer, Berlin, 1996.

[R89] Ph. Rosenau, *Extending hydrodynamics via the regularization of the Chapman-Enskog expansion*, Phys. Rev.A 40 (1989), 7193.

[S01] P. Sagaut, *Large eddy simulation for Incompressible flows,* Springer, Berlin, 2001.

[SW83] A.H. Schatz and L. Wahlbin, *On the finite element method for a singularly perturbed reaction-diffusion equation in two and one dimension*, Math. Comp.., 40(1983) 47-79.

[ST92] S. Schochet and E. Tadmor, *The regularized Chapman-Enskog expansion for scalar conservation laws*, Arch. Rat. Mech. Anal. 119 (1992), 95.

[S07] I. Stanculescu, *Existence Theory of Abstract Approximate Deconvolution Models of Turbulence,* to appear: Annali dell'Universitá di Ferrara, 2007.

[SA99] S. Stolz and N. A. Adams, *On the approximate deconvolution procedure for LES*, Phys. Fluids, II(1999),1699-1701.

[SAK01a] S. Stolz, N. A. Adams and L. Kleiser, *The approximate deconvolution model for LES of compressible flows and its application to shock-turbulent-boundary-layer interaction*, Phys. Fluids 13 (2001),2985.

[SAK01b] S. Stolz, N. A. Adams and L. Kleiser, *An approximate deconvolution model for large eddy simulation with application to wall-bounded flows*, Phys. Fluids 13 (2001),997.

[SAK02] S. Stolz, N. A. Adams and L. Kleiser, *The approximate deconvolution model for compressible flows: isotropic turbulence and shock-boundary-layer interaction*, in: Advances in LES of complex flows (editors: R. Friedrich and W. Rodi) Kluwer, Dordrecht, 2002.

[TW74] V. Thomée and B. Wendroff, *Convergence estimates for Galerkin methods for variable coefficient initial value problems*, SINUM 11(1973), 1059-1068.

[vC31] P. van Cittert, *Zum Einfluss der Spaltbreite auf die Intensitats verteilung in Spektrallinien II*, Zeit. fur Physik 69 (1931), 298-308.

8. Appendix 1: The FreeFEM++ code

We give in this additional section the FreeFEM++ program used to generate the numerical results herein. It follows next.

```
ï»¿border S1(t=0,1){ x=t; y=0; label=1;}
border S2(t=0,0.25){ x=1; y=t; label=2;}
border S3(t=1,0){ x=t; y=0.25; label=3;}
border S4(t=0.25,0){ x=0; y=t; label=4;}
int i,n=16,k,q,p; // n number of intervals on S2,S4
real T=1;
int N=2; //Deconvolution order
mesh Th= buildmesh(S1(4*n+1)+S2(n+1)+S3(4*n+1)+S4(n+1));
fespace Xh(Th,P2); // quadratic elements
/* //output mesh data
{ ofstream ff("mesh"+n+".txt");
for (int r=0;r<Th.nt;r++)
{ for (int z=0;z<3;z++)
```

```
ff<<Th[r][z].x<<" "<<Th[r][z].y<<endl;
}
}
*/
Xh e1,h=hTriangle;
Xh u1,v1,u1barh,u1barholder,phi1old;
Xh u1barhold,zeta1,u1old,f1,f1old;
Xh phi1,phiextr1,w1,u1FiltOld;
int numTri = Th.nt,MaxIters=12;
real delta,dt=0.00125,s,Chi;
int NumPts=T/dt;
real ItErr,TOL=1e-15,area=int2d(Th)(1.);
delta=0.03*sqrt(h[].max);
Chi=1/(h[].max);
s=0;
cout<<"Mesh size ="<<h[].max<<endl;
problem TRFEM (u1,v1) =
int2d(Th)(
( u1*v1 )*(1/dt)
+ ( dx(u1)*v1 )/2)
-int2d(Th)( (u1old/dt)*v1
-( dx(u1old)*v1 )/2
- Chi*(phiextr1*v1)/2
-Chi*(u1FiltOld*v1)/2)
+on(4,u1=(y>0.125));
problem dfiltz(u1barh,v1)
=int2d(Th)(u1barh*v1
+ (delta^2)*(dx(u1barh)*dx(v1)+dy(u1barh)*dy(v1)))
-int2d(Th)(zeta1*v1)
+on(1,2,3,u1barh=zeta1)
+on(4,u1barh=(y>0.125));
u1=(y>0.125)*exp(-x); //FIXED POINT SUB-ITERATIONS
for (p=0;p<2;p++) //USED FOR ACCURATE FIRST,SECOND TIMESTEP
{
u1old=u1;
w1=u1;
e1=TOL+1;
s=s+dt;
q=1;
while (e1[].max > TOL)
{
if ((q==1)&&(p==0))
{
zeta1=u1;
dfiltz;
zeta1=u1barh;
phi1=u1barh;
for (k=0;k<N;k++)
```

```
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=u1;
phiextr1=phiextr1 - zeta1;
zeta1=phiextr1;
dfiltz;
zeta1=u1barh;
phi1=u1barh;
for (k=0;k<N;k++)
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=phiextr1 - zeta1;
u1FiltOld=phiextr1;
}
TRFEM;
zeta1=u1;
dfiltz;
zeta1=u1barh;
phi1=u1barh;
for (k=0;k<N;k++)
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=u1;
phiextr1=phiextr1 - zeta1;
zeta1=phiextr1;
dfiltz;
zeta1=u1barh;
phi1=u1barh;
for (k=0;k<N;k++)
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=phiextr1 - zeta1;
e1 = abs(u1 - w1);
cout<< "e1[].max " << e1[].max << endl ;
cout<< "e1[].min " << e1[].min << endl ;
cout<< " " << endl ;
cout<<"Fixed Point Iter. Number "<<" "<<q<<endl;
w1=u1;
q=q+1;
if (q>MaxIters)
```

```
{
break;
}
} //END WHILE LOOP AND FIRST TIMESTEP
cout<<"Current time = "<<s<<endl;
if (p==0)
{
u1barhold=u1FiltOld;
u1FiltOld=phiextr1;
phiextr1=2*phiextr1-u1barhold;
}
//NOTE "barh" has meaning "star-star"
if (p>0)
{
u1barholder=u1barhold;
u1barhold=u1FiltOld;
u1FiltOld=phiextr1;
phiextr1=3*phiextr1-3*u1barhold+u1barholder;
}
} //END FOR LOOP AND SECOND TIMESTEP
for (i=2;i<T/dt;i++) //Proceed with 3-rd order extrapolation
{
u1old=u1;
s=s+dt;
TRFEM;
zeta1=u1;
dfiltz;
phi1=u1barh;
zeta1=u1barh;
for (k=0;k<N;k++)
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=u1-zeta1;
zeta1=phiextr1;
dfiltz;
zeta1=u1barh;
phi1=u1barh;
for (k=0;k<N;k++)
{
dfiltz;
zeta1=zeta1 + phi1 - u1barh;
}
phiextr1=phiextr1 - zeta1;
u1barholder=u1barhold;
u1barhold=u1FiltOld;
u1FiltOld=phiextr1;
```

```
phiextr1=3*phiextr1-3*u1barhold+u1barholder;
cout<<"Current time = "<<s<<endl;
};
cout<<"Mesh size ="<<h[].max<<endl;
{ ofstream ff("vel"+n+".txt");
for (int r=0;r<Th.nt;r++)
{ for (int z=0;z<6;z++)
ff<<u1[][Xh(r,z)] <<endl;
}
}
```

9. Appendix 2: Some comments about boundary conditions in a hyperbolic system

Consider the second order wave equation written as the following diagonal 2×2 hyperbolic system coupled through the boundary conditions. (This test was motivated by the example of Gunzburger [G77].) For $0 \le x \le 1, 0 \le t \le T < \infty$ find $u_1(x,t)$, $u_2(x,t)$ satisfying

$$\frac{\partial u_1}{\partial t} + \frac{\partial u_1}{\partial x} = f_1(x,t), \text{ and } \frac{\partial u_2}{\partial t} - \frac{\partial u_2}{\partial x} = f_2(x,t), \tag{9.1}$$

on $0 < x < 1, 0 < t \le T < \infty$. The initial condition for $u_j(x,0)(j = 1,2)$ is specified. Boundary conditions must be specified for (9.1). The inflow boundary for u_1 is $x = 0$ and for u_2 is $x = 1$. The general, well posed boundary condition is that the inflow variables are linear combinations of the outflow variables (e.g., Kreiss and Oliger [**?**]). For (9.1) this becomes, for $t > 0$ and some α, β,

$$u_1(0,t) = \alpha u_2(0,t) \text{ and } u_2(1,t) = \beta u_1(1,t). \tag{9.2}$$

Interestingly, the usual Galerkin method for this simple problem can be catastrophically unstable, Gunzburger [G77], so subtleties occur even for this simple problem. For our example, we shall impose the extra conditions that

$$|\alpha| < 1, \text{ and } |\beta| < 1. \tag{9.3}$$

To explain the meaning of these two conditions, we check semi-boundedness of the operator generating the semi-group associated with (9.1):

$$\int_0^1 \frac{\partial u_1}{\partial x} u_1 - \frac{\partial u_2}{\partial x} u_2 dx = \frac{1}{2}[(1-\beta^2)u_1^2(1) + (1-\alpha^2)u_2^2(0)].$$

This is nonnegative under the data condition (9.3) above. (In particular, this proving stability of the usual Galerkin method, see [L83b].) Further, the strict inequality in the data condition (9.3) implies that the above is: *strictly positive definite in the outflow variables on the boundary.* Under this condition it was proven by Lesaint in his Ph.D. thesis that on a uniform mesh the error in the usual Galerkin approximation to (9.1), (9.2) using linear elements is $O(h^2)$[3]. On the other hand, the best estimate that seems to be provable with continuous quadratics is the suboptimal $O(h^2)$ rate of convergence.

[3]To our knowledge, this result was never published. It can be shown that the error with continuous cubics or cubic splines is also optimal, $O(h^4)$, by combining Lesaint's argument with Axelsson and Gustafsson's.

We thus consider quadratic elements and weakly imposed boundary conditions in our tests of FEM discretization of (9.1), (9.2) with and without time relaxation. Let $X := H^1(0,1)$ (with no boundary conditions imposed) and $X^h \subset X$ denote a (scalar) finite element space associated with a mesh width h. We chose continuous quadratics. The *base* FEM discretization we consider is to find $u_j^h : [0,T] \to X^h (j = 1,2)$ satisfying, $\forall v_j^h \in X^h (j = 1,2)$,

$$(u_{1,t}^h + u_{1,x}^h - f_1, v_1^h) + (u_1^h(0,t) - \alpha u_2^h(0,t))v_1^h(0) = 0, \tag{9.4}$$

$$(u_{2,t}^h - u_{2,x}^h - f_2, v_2^h) + (u_2^h(1,t) - \beta u_1^h(0,t))v_2^h(1) = 0. \tag{9.5}$$

Note that in (9.4), (**??**) the coupling through the boundary conditions is imposed weakly. It is also not difficult to prove stability of this method by setting $v_j^h = u_j^h$ and integration by parts.

Lemma 4. *Consider the method (9.4), (9.5). If $|\alpha| < 1$ and $|\beta| < 1$ the method has a unique solution and its solution satisfies the energy inequality*

$$\frac{1}{2}\frac{d}{dt}\{||u_1^h||^2 + ||u_2^h||^2\} + \{(1-\beta^2)u_2^h(1,t)^2 + (1-\alpha^2)u_2^h(0,t)^2 + \\ + (1-\beta^2)u_1^h(1,t)^2 + (1-\alpha^2)u_1^h(0,t)^2\} \le (f_1, u_1^h) + (f_2, u_2^h).$$

Proof. Set $v_j^h = u_j^h$ and integrate by parts. The energy inequality and the Cauchy Schwarz inequality prove stability since the boundary terms are non-negative. Since the problem is a linear system of ODE's stability also proves existence and uniqueness. □

To introduce the time relaxation discretization, the treatment of boundary conditions by the differential filter must be specified. In some sense second order differential filters are not natural for first order hyperbolics. Extra boundary conditions, beyond those of the continuous problem, must be supplied for the differential filter because of the difference in equation order between the filter and the initial value problem. This difficulty does not occur when solving convection diffusion equations or other (linear or nonlinear) second order problems.

9.1. **A differential filter with incomplete boundary conditions.** There are very many ways to specify extra boundary conditions in the definition of the differential filter. Three examples come to mind immediately. Given a function ϕ to be filtered, take as boundary conditions for $\overline{\phi}$:

- $\overline{\phi} = \phi$ on $\partial\Omega$ or,
- $\nabla\overline{\phi} \cdot n = \nabla\phi \cdot n$ on $\partial\Omega$ (n = outward unit normal) or
- a third kind boundary condition defined variationally below, or
- some combination of the above.

The third method seems a natural attempt: we define the differential filter variationally and then deduce the boundary conditions imposed as natural boundary conditions. To illustrate, given $\phi = (\phi_1, \phi_2)$ satisfying the above boundary condition (9.2)

$$\phi_1(0) = \alpha\phi_2(0), \phi_2(1) = \beta\phi_1(1),$$

define $\overline{\phi}^h = (\overline{\phi}_1^h, \overline{\phi}_2^h)$ as the solution in $X \times X$ of

$$\delta^2(\overline{\phi}_{1,x}^h, v_{1,x}^h) + (\overline{\phi}_1^h, v_1^h) + (\overline{\phi}_1^h(0,t) - \alpha\overline{\phi}_2^h(0,t))v_1^h(0) = (\phi_1, v_1^h), \tag{9.6}$$

$$\delta^2(\overline{\phi}_{2,x}^h, v_{2,x}^h) + (\overline{\phi}_2^h, v_2^h) + (\overline{\phi}_2^h(1,t) - \beta\overline{\phi}_1^h(1,t))v_2^h(1) = (\phi_2, v_2^h), \tag{9.7}$$

for all $v_j^h \in X^h, j = 1, 2$. Note that the BCs have been imposed variationally as the last term on the RHS of each equation. Integration by parts shows that this formulation imposes approximately the following boundary conditions on the averages:

$$\begin{aligned} -\delta^2 \overline{\phi}^h_{1,x}(0,t) + (\overline{\phi}^h_1(0,t) - \alpha \overline{\phi}^h_2(0,t)) &= 0 \text{ , and } -\delta^2 \nabla \overline{\phi}^h_2(0,t) \cdot n = 0, \\ +\delta^2 \overline{\phi}^h_{2,x}(1,t) + (\overline{\phi}^h_2(1,t) - \beta \overline{\phi}^h_1(1,t)) &= 0 \text{ , and } +\delta^2 \nabla \overline{\phi}^h_1(1,t) \cdot n = 0. \end{aligned}$$

The first boundary conditions are $O(\delta^2)$ approximations of (9.2). The second in each pair of BCs seems improvable.

9.2. **Constructing a Test Problem.** Traditionally, the first test is an academic study of convergence rates for smooth solutions. One can, for example, pick

$$\begin{aligned} u_1(x,t) &= 0.8 \cdot a(t) \cdot \cos(\pi x), \\ u_2(x,t) &= a(t) \cdot \cos(3\pi x) \cdot e^{-x}. \end{aligned}$$

These satisfy the boundary conditions (9.2) with

$$\alpha = \frac{u_1(0,t)}{u_2(0,t)} \equiv 0.8, \beta = \frac{u_2(1,t)}{u_1(1,t)} \equiv \frac{0.8}{e} (< 1).$$

Inserting this into the partial differential equations, we find

$$\begin{aligned} f_1(x,t) &= 0.8 \cdot \frac{da(t)}{dt} \cos(\pi x) - 0.8 \cdot \pi \cdot a(t) \sin(\pi x), \\ f_2(x,t) &= \frac{da(t)}{dt} \cos(3\pi x) e^{-x} - a(t)[3\pi \sin(3\pi x) + \cos(3\pi x)] e^{-x}. \end{aligned}$$

The time relaxation discretization is we consider is to find $u_j^h : [0,T] \to X^h (j = 1, 2)$ satisfying , $\forall v_j^h \in X^h (j = 1, 2)$,

$$(9.8)\ (u^h_{1,t} + u^h_{1,x} - f_1, v^h_1) + \chi(u^{h*}_1, v^{h*}_1) + (u^h_1(0,t) - \alpha u^h_2(0,t)) v^h_1(0) = 0,$$

$$(9.9)\ (u^h_{2,t} - u^h_{2,x} - f_2, v^h_2) + \chi(u^{h*}_2, v^{h*}_2) + (u^h_2(1,t) - \beta u^h_1(0,t)) v^h_2(1) = 0.$$

The initial condition is then given by

$$\begin{aligned} u_1(x,0) &= 0.8 \cdot a(0) \cdot \cos(\pi x), \\ u_2(x,0) &= a(0) \cdot \cos(3\pi x) \cdot e^{-x}. \end{aligned}$$

We do not persue this problem further here.

10. More about Averaging or Filtering

Averaging is a very common procedure in turbulence modeling. The classic example is time averaging or Reynolds averaging (after Osbourne Reynolds) over a finite time window :

$$\langle u \rangle_{[0,T]}(x,T) := \frac{1}{T} \int_0^T u(x,t) dt, \text{ or } \langle u \rangle_{[t-T,t]}(x,t) := \frac{1}{T} \int_{t-T}^t u(x,t') dt'.$$

Motivated by practical computations, the key idea is that in an under resolved simulation a computed velocity should properly represent an average over one (or a few) mesh cells. This leads to a space filtered $\overline{u}(x,t)$, such as

$$\overline{u}(x,t) := \frac{1}{\delta^3} \int_{[-\frac{\delta}{2}, +\frac{\delta}{2}]^3} u(x - x', t) dx'. \tag{10.1}$$

Alternately, equivalently and mathematically clearer, weighed local averages are defined by convolution or filtering with the chosen filter kernel, such as a top hat kernel for the above average or a Gaussian:

$$\begin{aligned} \overline{u}(x,t) &: = g_\delta \star u(x,t), \text{ where } g_\delta \star u(x,t) := \int_{\mathbb{R}^3} g_\delta(x')u(x-x',t)dx', \\ g_\delta(x) &: = \delta^{-3} g(x/\delta), \text{ and } g(x) := \text{Gaussian}. \end{aligned}$$

Viewing averaging from the point of view of scale space, one thinks of a fluid velocity as naturally composed of a scale of velocities (scaled by the filter length scale δ)

$$u = u(x,t;\delta)(= (g_\delta \star u)(x,t)).$$

In practical computing, the averaging radius δ is related to the (possibly locally varying) mesh width and filters are often chosen based on computational convenience.

10.1. **What is the right filter?** Ignoring computational convenience for the moment, it is interesting also to consider filtering from the ideal continuum point of view. If the averages are viewed as containing information of physical meaning, there already results on acceptable filters beginning with the famous paper of Koenderink[4]. Scale space analysis[5] begins with some basic postulates that any physically reasonable filter should satisfy. There are various ways to develop the essential result. One begins by assuming the filter is

Condition 1. *linearity:*

$$\overline{\alpha u + \beta v} = \alpha \overline{u} + \beta \overline{v}$$

Condition 2. *spacial invariance (no preferred location):*

$$\overline{u(x-a)} = \overline{u}(x-a)$$

Condition 3. *isotropy (no preferred orientation): for all rotations* R *:*

$$\overline{u}(x) = R^* \overline{u(Rx)}$$

Condition 4. *scale invariance, or the semigroup property (no preferred size): if* $\overline{u}(x,t) := g_\delta \star u(x,t)$ *(so* $\overline{\overline{u}} := g_\delta \star g_\delta \star u$*) then*

$$\overline{\overline{u}} = g_{\sqrt{2}\delta} \star u.$$

For example, in 1984 Koenderink proved the following.

Theorem 5 (Koenderink 1984)**.** *The only filter satisfying C1, C2, C3 and C4 is the Gaussian filter.*

Thus, there is really only one mathematical correct filter: the Gaussian. The same conclusion can be arrived at by other plausible conditions on the filter such as filtering not creating new structures such as local extrema and causality in the scale variable δ. Convolution with the Gaussian is quite expensive computational so that in spite of this strong uniqueness result usually other filters are used. Koenderink's result indicate that other filters for large eddy simulation must be assessed as approximations to the Gaussian. One filtering method which is very convenient

[4]J.J. Koenderink, *The structure of images, Biol. Cybernetics*, 50 (1984) 363-370.

[5]See, e.g., T. Lindeberg, *Scale-space theory in computer vision, Kluwer, Dordrecht, 1994.*

mathematical and not too unreasonable computational is a special differential filter. Define the velocity averages by:

$$\overline{u}(x,t) := (-\delta^2 \triangle + 1)^{-1} u(x,t). \tag{10.2}$$

Differential filters were proposed for large eddy simulation by Germano[6] and the above differential filter is a well known regularization of evolution equations. The close connection of the above differential filter to the Gaussian filter can be seen two ways. In Galdi and Layton[7], it was derived as an approximation of the Gaussian filter by Padé approximations of exponentials as follows. Fourier transformation of

$$\overline{u}(x,t) := g_\delta \star u(x,t)$$

gives

$$\widehat{\overline{u}}(k,t) := \widehat{g_\delta}\widehat{u}(k,t) \text{ , where } \widehat{g_\delta} = e^{-|\delta k|^2}.$$

Since $e^{-|\delta k|^2} \to 0$ as $|k| \to \infty$, the filter suppresses fluctuations; in other words it is smoothing. This is a fundamental property that must be preserved under approximation. The simplest rational approximation preserving this is the $(0,1)$ Padé approximation given by

$$\begin{aligned} e^{-\theta} &= \frac{1}{1+\theta} + O(\theta^2) \text{ , as } \theta \to 0, \text{ so} \\ e^{-|\delta k|^2} &= \frac{1}{1+|\delta k|^2} + O(|\delta k|^4). \end{aligned}$$

Using this approximation in the above for $\widehat{g_\delta}$ and inverting the Fourier transform recovers the differential filter (#).

Alternately, the Gaussian is the heat kernel. Thus, one way to compute the average velocity $\overline{u}(x,t) = g_\delta \star u(x,t)$ is to solve the following evolution equation:

$$\begin{aligned} v_s(x,s) &= \triangle v(x,s) \text{ for } s > 0, \\ v(x,0) &= u(x), \end{aligned}$$

then set

$$\overline{u}(x,t) := v(x,s)|_{s=\delta^2}.$$

This gives the exact Gaussian filtered velocity $\overline{u}(x,t) := g_\delta \star u(x,t)$. Since the averaging radius δ is small, δ^2 is smaller still and we can reasonably approximate $v(x,\delta^2)$ by one step of backward Euler, leading back to the differential filter

$$\overline{u}(x,t) := (-\delta^2 \triangle + 1)^{-1} u(x,t).$$

In comparison with the Gaussian, the differential filter is only approximately scale invariant. Indeed, if we compute $\overline{\overline{u}}$ we find that it fails the semigroup property by $O(\delta^4)$

$$\begin{aligned} \overline{\overline{u}} &= (-\delta^2 \triangle + 1)^{-1}(-\delta^2 \triangle + 1)^{-1} u \\ &\neq (-(\sqrt{2}\delta)^2 \triangle + 1)^{-1} u, \\ &\quad \text{but rather for smooth } u, \\ \overline{\overline{u}} &= (-(\sqrt{2}\delta)^2 \triangle + 1)^{-1} u + O(\delta^4). \end{aligned}$$

Many important theoretical and practical questions remain at this very first step.

---

[6] M. Germano, *Differential filters of elliptic type*, Phys. Fluids, 29(1986), 1757-1758.

[7] G. P. Galdi and W. J. Layton, *Approximation of the large eddies in fluid motion II: A model for space-filtered flow*, Math. Models and Methods in the Appl. Sciences, 10(2000), 343-350.

11. More on Deconvolution

If we can invert the filter exactly (so called, *exact deconvolution*), then the closure problem is solved in principle. For example, the differential filter has an exact filter inverse [8]

$$A := -\delta^2 \triangle + 1$$

which exists as an *unbounded* operator on $L^2(\Omega)$ with dense domain and closed range. To obtain a useful regularization, however, information must be lost by *approximate* deconvolution. This means that the extra terms should be smoothing in some sense. Our intuition is that this is accomplished when the approximate deconvolution operator is a *bounded operator*.

Alternately, an unbounded deconvolution operator will suffer from small divisor problems: *the (inevitable) noise from data and discretization will be magnified by the model rather than damped.* Unfortunately, it is well-known to be impossible for an interesting filter to have a bounded (exact) inverse[9]. Consider the deconvolution problem

$$\text{given } \overline{u} \ (+ \ noise) \text{ solve } \overline{u} = Gu \text{ for } u.$$

Theorem 6. *Let X be a Hilbert space and $G : X \to X$ be a compact linear operator. Then, if G^{-1} is bounded then $dim(X)$ is finite.*

The following is an immediate consequence.

Corollary 4. *Let $G\phi = \overline{\phi}$ denote the filtering operator. If G is smoothing so $G : L^2(\Omega) \to H^s(\Omega)$ is bounded for some $s > 0$ then G cannot have an exact inverse that is a bounded linear operator : $L^2(\Omega) \to L^2(\Omega)$.*

The problem of inverting the filter G is is ill-posed and thus the closure problem itself must be ill posed. This does not mean that accurate approximate closure is impossible. There are after all many good methods for approximate solution of ill posed problems!

11.1. **Approximate deconvolution as an ill-posed a problem.** The filtering or convolution problem is: given ϕ compute $\phi \to G\phi = \overline{\phi} := g \star \phi$. The de-filtering or deconvolution problem is: given $\overline{\phi}$ (possibly $+noise$) solve the following equation approximately for ϕ

$$\text{given } \overline{\phi} \text{ solve } \overline{\phi} = G\phi \text{ for } \phi.$$

Definition 5. *An approximate deconvolution operator D is a bounded linear operator $D : L^2(\Omega) \to L^2(\Omega)$ satisfying*

$$\textit{for smooth functions } \phi :$$

$$\phi = D\ \overline{\phi} + O(\delta^\alpha)\ \delta \to 0 \textit{ for some } \alpha \geq 2.$$

The deconvolution error is

$$e_{DCV}(\phi) = \phi - D_N \overline{\phi}.$$

[8] The same argument can be made for any filter with $\widehat{g}(k) \neq 0$.

[9] L. C. Berselli, T. Iliescu, and W. Layton, *Mathematics of Large Eddy Simulation of Turbulent Flows*. Springer, Berlin, 2006.

AND

M. Bertero and B. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing Ltd., 1998.

The deconvolution problem is an important problem in image processing so there are many methods specifically directed at the deconvolution problem. We shall consider a few examples.

Example 1. *Tichonov-Lavrentiev regularization.*

Let $X = L^2(\Omega)$ and $(\cdot,\cdot), ||\cdot||$ denote the inner product and induced norm on X. If G is a symmetric, positive definite operator, solving the deconvolution problem

$$\text{given } \overline{\phi} \text{ solve } \overline{\phi} = G\phi \text{ for } \phi.$$

in the Hilbert space X is formally equivalent to minimizing the quadratic functional

$$\begin{aligned} \phi &= \arg\min_{v\in X} J(v) \\ J(v) &: = \frac{1}{2}(Gv,v) - (\overline{\phi},v). \end{aligned}$$

The exact solution is obviously $v = \phi$. However, when noise $\varepsilon \in X$ is present it is easy to construct simple examples in which the minimization problem

$$\phi = \arg\min_{v\in X} \frac{1}{2}(Gv,v) - (\overline{\phi}+\varepsilon,v)$$

has no solution.

Tichonov-Lavrentiev regularization picks a regularization parameter $\mu > 0$ and computes the approximation of the deconvolution problem by the approximate minimization problem:

$$\phi_\mu = \arg\min_{v\in X} \frac{1}{2}(Gv,v) - (\overline{\phi},v) + \frac{\mu}{2}||v||^2.$$

The classical Tichonov-Lavrentiev method has error even on the largest scales. C. Manica and I. Stanculescu recently gave an important refinement which reduces the associated error on the large scales significantly. The modification is

$$\begin{aligned} \phi_\mu &= \arg\min_{v\in X} J_\mu(v) \\ J_\mu(v) &= {}^{def}(1-\mu)\{\frac{1}{2}(Gv,v) - (\overline{\phi},v)\} + \frac{\mu}{2}||v||^2. \end{aligned}$$

The Euler-Lagrange equations of the above is

$$\phi_\mu = ((1-\mu)G + \mu I)^{-1}\overline{\phi}$$

so the approximate deconvolution operator induced by Tichonov regularization is

$$D_\mu = ((1-\mu)G + \mu I)^{-1}.$$

When the filter operator is not a SPD, the deconvolution problem is converted into the SPD deconvolution problem

$$\text{given } \overline{\phi} \text{ solve } G^*\overline{\phi} = G^*G\phi \text{ for } \phi.$$

by least squares and the associated deconvolution operator of this is then

$$D_\mu = (G^*G + \mu I)^{-1}G^*.$$

This is the (full) Tikhonov regularization. For example, with the differential filter

$$\overline{u}(x,t) := (-\delta^2\triangle + 1)^{-1}u(x,t),$$

we have formally

$$D_\mu = ((-\delta^2\triangle + 1)^{-1} + \mu I)^{-1}.$$

Thus given a filtered variable $\overline{\phi}$, its deconvolved variable is calculated by solving:

$$\begin{aligned} \{\mu(-\delta^2\triangle+1)+1\}\phi_\mu &= (-\delta^2\triangle+1)\overline{\phi}\ \text{, or, equivalently} \\ \phi_\mu &= \frac{1}{\mu}\overline{\phi}-\frac{1}{\mu}\{\mu(-\delta^2\triangle+1)+1\}^{-1}\overline{\phi}. \end{aligned}$$

Example 2 (The van Cittert algorithm)**.** *In 1931 (!) van Cittert studied a very simple approximate deconvolution algorithm. The van Cittert algorithm is equivalent to first order Richardson iteration for solving the ill posed operator equation $G\phi=\overline{\phi}$ or simple iteration in*

$$\textit{given } \overline{u} \textit{ solve } u=u+\{\overline{u}-A^{-1}u\} \textit{ for } u.$$

Algorithm 2 (van Cittert Approximate Deconvolution)**.** *Set $v_0=\overline{u}$,*
for $n=0,1,2,...,N-1$, perform
$v_{n+1}=v_n+\{\overline{u}-A^{-1}v_n\}$
Define $D_N\overline{u}:=v_N$.

By eliminating the intermediate steps, the N^{th} de-convolution operator D_N is given explicitly by

$$D_N\phi := \sum_{n=0}^{N}(I-A^{-1})^n\phi. \tag{11.1}$$

For example, the approximate de-convolution operator corresponding to $N=0,1,2$ are:

$$\begin{aligned} D_0\overline{u} &= \overline{u}, \\ D_1\overline{u} &= 2\overline{u}-\overline{\overline{u}}, \\ D_2\overline{u} &= 3\overline{u}-3\overline{\overline{u}}+\overline{\overline{\overline{u}}}. \end{aligned}$$

It is known that D_N is bounded, SPD and an asymptotic filter inverse of accuracy $O(\delta^{2N+2})$:

$$\phi = D_N\overline{\phi}+O(\delta^{2N+2})\ \text{, for smooth } \phi.$$

Proposition 3. *Let $G=A^{-1}$ be the differential filter 10.2. Then, both G and $I-G$ are SPD; further $0\le\lambda(G)\le 1$ and $1\le\lambda(D_N)\le N+1$. The operator D_N is bounded*

$$||D_N||_{L(L^2(\Omega)\to L^2(\Omega))}\le N+1.$$

Further,

$$\phi = D_N\overline{\phi}+O(\delta^{2N+2})\ \textit{, for smooth } \phi.$$

One immediate consequence of the above asymptotic result is convergence as $\delta\to 0$ for fixed N. This is typical for filtering, (see the recent work of Stanculescu).

Corollary 5. *For $\phi\in L^2(\Omega), D_N\overline{\phi}\to\phi$ as $\delta\to 0$ for fixed N.*

Proof. Let $\varepsilon>0$ be given and let ψ be a smooth function with $||\phi-\psi||<[1+(N+1)||G||]^{-1}\frac{\varepsilon}{3}$. Write

$$||\phi-D_N\overline{\phi}||\le||\phi-\psi||+||\psi-D_N\overline{\psi}||+||D_N\overline{(\phi-\psi)}||,$$

so that

$$\begin{aligned} ||\phi - D_N\overline{\phi}|| &\leq ||\phi - \psi|| + O(\delta^{2N+2}) + ||D_N||||G||||\phi - \psi|| \leq \\ &\leq \frac{\varepsilon}{3} + O(\delta^{2N+2}) < \varepsilon, \text{ for } \delta \text{ small enough.} \end{aligned}$$

□

Example 3 (Optimized van Cittert deconvolution)**.** *Since the van Cittert algorithm is equivalent to first order Richardson for the operator equation $G\phi = \overline{\phi}$, relaxation parameters can be introduced and no extra cost. with proper choice of the optimization parameters significant improvement of accuracy is possible.*

Algorithm 3 (van Cittert deconvolution with relaxation)**.** *Set $v_0 = \overline{u}$,*
For $n = 0, 1, 2, ..., N-1$,
select relaxation parameter ω_n and compute
$v_{n+1} = v_n + \omega_n\{\overline{u} - A^{-1}v_n\}$
Define $D_N^{\omega}\overline{u} := v_N$.

Optimization of the parameters ω_n depends on the objective and the exact choice of filters two cases have been studied by Stanculescu, insert references.

Case 1 (K41 optimized deconvolution)**.** *In the work of Stanculescu the optimal parameters were derived to minimize the norm of the deconvolution error $||\phi - D_N\overline{\phi}||$ over the resolved scales for velocity fields coming from turbulent flows with the inertial range energy spectrum typical of homogeneous isotropic turbulence:*

$$\begin{gathered} \text{Find } (\omega_0, \omega_1, \cdots, \omega_N) \text{ minimizing} \\ ||u - D_N\overline{u}|| \\ \text{subject to } \widehat{E}(k) = \alpha\varepsilon^{\frac{2}{3}}k^{-\frac{5}{3}}. \end{gathered}$$

Case 2 (Optimization for general velocity fields)**.** *Another possibility is to optimize the norm of the deconvolution error over general, square integrable velocity fields. This leads to the minimax problem*

$$\min_{\omega_j} \max_{\phi \in L^2} ||\phi - D_N\overline{\phi}||$$

which was solved by Stanculescu as well.

Example 4 (Geurts' approximate filter inverse operators)**.** *Exploiting special features of the top hat filter, Geurts*[10] *constructed efficient and ingenious approximate filter inverses of varying degrees of accuracy of the top hat filter.*

Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA,
The work of both authors was partially supported by NSF grant DMS 0508260.
E-mail address: jmc116@pitt.edu, wjl@pitt.edu
URL: http://www.math.pitt.edu/~wjl

[10]B. J. Geurts, *Inverse modeling for large eddy simulation*, Phys. Fluids, 9(1997), 3585.