Numer. Math. (2000) 86: 239–256
Digital Object Identifier (DOI) 10.1007/s002110000151

Numerische
Mathematik

A characteristic-Galerkin approximation to a system of shallow water equations

Clint N. Dawson[1], **Mónica L. Martínez-Canales**[2]

[1] Center for Subsurface Modeling – C0200; Texas Institute for Computational and Applied Mathematics, The University of Texas, Austin, TX 78712, USA
[2] Department of Geological and Environmental Sciences; Braun Hall, Bldg 320, Stanford University, Stanford, CA 94305-2115, USA

Received October 15, 1998 / Revised version received March 13, 1999 /
Published online April 20, 2000 –

Summary. Characteristic methods are known to handle advective flow better than traditional Galerkin methods and allow large time steps to be taken when compared to standard time-stepping methods. In this paper, we investigate a characteristic-Galerkin approximation to the 2-dimensional system of shallow water equations. We derive $\mathcal{L}^\infty\left((0,T);\mathcal{L}^2(\Omega)\right)$ bounds for elevation and velocity, showing these to be optimal for velocity in $\mathcal{L}^2\left((0,T);\mathcal{H}^1(\Omega)\right)$.

Mathematics Subject Classification (1991): 65N30

1. Introduction

Interest in modeling shallow water environments, such as bays, estuaries and other coastal waters, has spawned a generation of shallow water simulators. Because of the complexities of coastal geometries, and the need to allow for domains which incorporate parts of the deep ocean, many of these simulators are based on finite element methodology. We refer, for example, to the ADCIRC (Advanced Circulation) model developed by Luettich et al. (1991), which is a widely used shallow water simulator based on Galerkin finite elements.

In the shallow water equations (SWE), diffusive effects can often be small relative to advective acceleration, for example, in channels and narrow inlets. It is well-known that standard Galerkin schemes do not handle advective flow

Correspondence to: C.N. Dawson

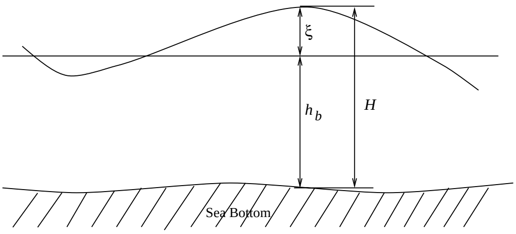

Fig. 1. Vertical cross-section depicting elevation and bathymetry

very well unless small time steps and highly refined grids are used. Most shallow water simulators have some stabilizing mechanisms built-in. For example, the ADCIRC simulator is based on a reformulation of the first-order continuity equation into a second order wave equation, first proposed by Lynch and Gray (1979). This approach allows for the capturing of so-called "$2\Delta x$ waves," but still has problems handling highly advective flow.

In this paper, we propose using characteristics methods as a means to handle advection with the additional benefit of an improved time truncation error when compared to standard finite-difference time-stepping schemes. The method we propose is similar to a Characteristic-Galerkin approximation derived recently by Zienkiewicz and Ortiz (1995a,1995b), with promising numerical results. Their method relies on a Chorin-type projection with fractional time-stepping along the characteristics in the velocity equation. These characteristics are approximated using a Taylor expansion assuming that the foot of the characteristic is very close to the nodal point around which the expansion was taken. Our method differs in that we use finite-difference time-steps along the characteristics in both the continuity and momentum equations, and do not perform a Taylor expansion of the solution at the foot of the characteristic. We will describe and analyze this particular Characteristic-Galerkin finite element method for solving the SWE.

The SWE are obtained by depth averaging the 3-dimensional incompressible Navier-Stokes equations using appropriate free-surface and boundary conditions along with a hydrostatic pressure assumption (Weiyan (1992)). Let $\xi(x_1, x_2, t)$ be the free surface elevation above a reference plane and let $h_{\mathrm{b}}(x_1, x_2)$ be the bathymetric depth under that reference plane (see Fig. 1) so that $H = \xi + h_{\mathrm{b}}$ is the total water column. The SWE are valid in regions where the horizontal length scale L is much greater than the vertical length scale H, and the underwater topography doesn't change too fast. It should be noted that the latter two properties imply that ∇h_{b} is small in the sense that $\left|\frac{\partial h_{\mathrm{b}}}{\partial x_1}\right|, \left|\frac{\partial h_{\mathrm{b}}}{\partial x_2}\right| \approx \frac{d\,|h_{\mathrm{b}}|}{d\,|L|} < \frac{d\,|H|}{d\,|L|} << 1$.

Let $\boldsymbol{u} = (u(x_1, x_2, t)\ , v(x_1, x_2, t))^{\mathrm{T}}$ be the depth-averaged horizontal velocities. Then, the SWE are given by the continuity equation (CE)

$$\frac{\partial \xi}{\partial t} + \nabla \cdot (\boldsymbol{u} H) = 0 \tag{1}$$

and the momentum equations (NCME), written here in non-conservative form,

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} + g \nabla \xi - \mu \Delta \boldsymbol{u} + \tau_{\mathrm{b}} \boldsymbol{u} + \boldsymbol{F} = 0. \tag{2}$$

Here, $0 < \mu$ is viscosity, g is acceleration due to gravity, $\tau_{\mathrm{b}}(\xi, \boldsymbol{u}) = c_{\mathrm{f}} \frac{\sqrt{u^2+v^2}}{H}$ is a bottom friction function, and $\boldsymbol{F}$ is a forcing function consisting of surface and body forces such as Coriolis effects, surface wind stress, surface atmospheric pressure and tide potentials; for instance, $\boldsymbol{F} = (f_{\mathrm{c}} \boldsymbol{k} \times \boldsymbol{u} - \frac{1}{H} \boldsymbol{\tau}_{\mathrm{ws}} + \nabla p_{\mathrm{a}} - g \nabla \mathcal{N})$. This form of $\boldsymbol{F}$ is due to Luettich et al. (1991), and we have shown in Chippada et al. (1999) how such terms can be analyzed in a finite element setting. It should be noted that the final form of the viscosity term is a point of contention in the literature - other forms of it are $\frac{\mu}{H} \Delta(\boldsymbol{u} H)$ (Bernardi and Pironneau (1991)) and $\frac{\mu}{H} \nabla \cdot H \nabla \boldsymbol{u}$ (Gent (1993)).

The rest of this paper is outlined as follows. In Sect. 2, we introduce notation and definitions. In Sect. 3, we review the characteristic formulation of the SWE, introduce the discrete weak formulation, and describe the assumptions we will need in our analysis. In Sect. 4, we introduce the finite element model used to approximate the SWE as well as additional assumptions we will need. In Sect. 5, we review the characteristic equation and properties therein. In Sect. 6, we derive an a priori error estimate based on a discrete $\mathcal{L}^2$ projection. The proof of the error estimate relies on an induction argument to obtain $\mathcal{L}^\infty$ boundedness of the Galerkin approximations.

2. Preliminaries

Let Ω be a bounded polygonal domain in $\mathbb{R}^2$ and $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$. Moreover, let $\bar{\Omega} = \Omega \cup \partial\Omega$ where $\partial\Omega$ is the boundary of $\Omega \subset \mathbb{R}^2$.

The $\mathcal{L}^2$ inner product is denoted by

$$(\varphi, \omega) = \int_\Omega \varphi \diamond \omega \, dx, \qquad \varphi, \omega \in [\mathcal{L}^2(\Omega)]^n,$$

where "$\diamond$" refers to either multiplication, dot product, or double dot product as appropriate. We denote the $\mathcal{L}^2$ norm by $||\varphi|| = ||\varphi||_{\mathcal{L}^2(\Omega)} = (\varphi, \varphi)^{1/2}$. In $\mathbb{R}^n$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ is an n-tuple with nonnegative integer components,

$$D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n} = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$$

and $|\alpha| = \sum_{i=1}^{n} \alpha_i$.

For ℓ any nonnegative integer, let

$$\mathcal{W}_m^\ell \equiv \{\varphi \in \mathcal{L}^m(\Omega) \mid D^\alpha \varphi \in \mathcal{L}^m(\Omega) \text{ for } |\alpha| \leq \ell\}$$

be the Sobolev space with norm

$$||\varphi||_{\mathcal{W}_m^\ell(\Omega)} = \left(\sum_{|\alpha| \leq \ell} ||D^\alpha \varphi||_{\mathcal{L}^m(\Omega)}^m \right)^{1/m} .$$

Moreover, let

$$\mathcal{W}_\infty^\ell \equiv \{\varphi \in \mathcal{L}^\infty(\Omega) \mid D^\alpha \varphi \in \mathcal{L}^\infty(\Omega) \text{ for } |\alpha| \leq \ell\}$$

be the Sobolev space with norm

$$||\varphi||_{\mathcal{W}_\infty^\ell(\Omega)} = \max_{|\alpha| \leq \ell} ||D^\alpha \varphi||_{\mathcal{L}^\infty(\Omega)}.$$

We will also use the special spaces $\mathcal{H}^\ell = \mathcal{W}_2^\ell$. For relevant properties of these spaces, please refer to Adams (1978).

Furthermore, observe that $\mathcal{H}^\ell$ are spaces of $\mathbb{R}$-valued functions. Spaces of $\mathbb{R}^n$-valued functions will be denoted in boldface type, but their norms will not be distinguished. Thus, $\boldsymbol{\mathcal{L}^2}(\Omega) = [\mathcal{L}^2(\Omega)]^n$ has norm $||\boldsymbol{\varphi}||^2 = \sum_{i=1}^{n} ||\varphi_i||^2$; $\boldsymbol{\mathcal{H}^1}(\Omega) = [\mathcal{H}^1(\Omega)]^n$ has norm $||\boldsymbol{\varphi}||_{\mathcal{H}^1(\Omega)}^2 = \sum_{i=1}^{n} \sum_{|\alpha| \leq 1} ||D^\alpha \varphi_i||^2$; etc.

For X, a normed space with norm $|| \cdot ||_X$ and a map $f\colon [0,T] \to X$, define

$$||f||_{\mathcal{L}^2((0,T);X)}^2 = \int_0^{\mathrm{T}} ||f(\cdot,t)||_X^2 \, dt,$$
$$||f||_{\mathcal{L}^\infty((0,T);X)} = \sup_{0 \leq t \leq T} ||f(\cdot,t)||_X.$$

We define a temporal subdomain of $[0,T]$ by $J_{\Delta t} = \{t^k \mid t^k \in [0,T],\ t^k = k\Delta t,\ k = 0,\ldots,N,\ N\Delta t = T,\ \Delta t \geq 0\}$.

Let $\mathcal{T}$ be a quasi-uniform triangulation of Ω into elements $\omega_i,\ i = \{1,\ldots,n_{\mathcal{T}}\}$, with $\mathrm{diam}(\omega_i) = h_i$ and $h = \max_i h_i$. Let $\mathcal{S}_h$ denote a finite dimensional subspace of $\mathcal{H}^1(\Omega)$ defined on this triangulation consisting of piecewise polynomials of degree less than or equal to $s_1 - 1$, and satisfying the standard approximation property

$$\inf_{\varsigma \in \mathcal{S}_h} ||\phi - \varsigma||_{\mathcal{H}^{s_0}(\Omega)} \leq K_0 h^{\ell - s_0} ||\phi||_{\mathcal{H}^\ell(\Omega)}, \quad \phi \in \mathcal{H}^1(\Omega) \cap \mathcal{H}^\ell(\Omega),$$

for integers s_0, ℓ and $0 \leq s_0 \leq \ell \leq s_1$ and where K_0 is a constant independent of h and ϕ.

Moreover, we will use the following standard results.

Lemma 1 (Inverse estimate). *(See Brenner and Scott (1994).) Let* $h \in (0,1]$ *and* $\mathcal{S}_h \subset \mathcal{W}_p^r(\Omega) \cap \mathcal{W}_q^m(\Omega)$*, where* Ω *is a polyhedral domain in* $\mathbb{R}^n$*,* $1 \le p \le \infty$*,* $1 \le q \le \infty$*, and* $0 \le m \le r$*, then there exists a* $K_0 = K_0(r,p,q)$ *such that* $\forall v \in \mathcal{S}_h$*, we have*

$$||v||_{\mathcal{W}_p^r(\Omega)} \le K_0 h^{m-r+\min(0,\frac{n}{p}-\frac{n}{q})} ||v||_{\mathcal{W}_q^m(\Omega)}.$$

Lemma 2. *Let* $0 \le q \le \ell \le s_1$*. Let* $\phi \in \mathcal{L}^2\left((0,T);\mathcal{H}^1(\Omega)\cap\mathcal{H}^\ell(\Omega)\right)$ *and let* $\tilde{\phi}$ *be the corresponding* $\mathcal{L}^2$ *projection of* ϕ *into* $\mathcal{S}_h$*. If for some integer* $\jmath \ge 0$*,* $\left(\frac{\partial}{\partial t}\right)^\jmath \phi \in \mathcal{L}^2\left((0,T);\mathcal{H}^1(\Omega)\cap\mathcal{H}^\ell(\Omega)\right)$*, then* $\left(\frac{\partial}{\partial t}\right)^\jmath \tilde{\phi} \in \mathcal{L}^2\left((0,T);\mathcal{S}_h\right)$ *and*

$$\left|\left|\left(\frac{\partial}{\partial t}\right)^\jmath \left(\phi - \tilde{\phi}\right)\right|\right|_{\mathcal{L}^2((0,T);\mathcal{H}^q(\Omega))} \le K_0 h^{s-q} \left|\left|\left(\frac{\partial}{\partial t}\right)^\jmath \phi\right|\right|_{\mathcal{L}^2((0,T);\mathcal{H}^s(\Omega))},$$

for some constant K_0 *independent of* ϕ, q, h, ℓ*, where* $s = \min(\ell, s_1)$*.*

There will also be occasion to employ the following lemma whose proof can be found in Brenner and Scott (1994) in Corollary 4.8.9.

Lemma 3. *Let* $\phi \in \mathcal{W}_\infty^1(\Omega)$ *and let* $\tilde{\phi}$ *be the* $\mathcal{L}^2$ *projection of* ϕ *into* $\mathcal{S}_h$*. Then, the first-order spatial derivatives of* $\tilde{\phi}$ *are bounded above in* $\mathcal{L}^\infty\left((0,T);\mathcal{L}^\infty(\Omega)\right)$ *by a positive constant* K_0*.*

Finally, we let $K, K_i, (i = 0,1,2,..)$ and ϵ be generic constants not necessarily the same at every occurrence.

3. Characteristic shallow water equations

3.1. The characteristic form

The characteristic formulation of the SWE is based on manipulating the governing equations into a form in which the time derivative and the advective term are absorbed into a directional derivative.

Let $\boldsymbol{\tau}$ be a unit vector in the direction $(\boldsymbol{u}, 1)$ so that $\boldsymbol{\tau} = \frac{1}{\alpha}(\boldsymbol{u}, 1)$, with $\alpha = |\boldsymbol{\tau}| = \sqrt{|\boldsymbol{u}|^2 + 1}$. Then, define

$$\alpha \frac{\partial \phi}{\partial \boldsymbol{\tau}} \equiv \boldsymbol{u} \cdot \nabla \phi + \frac{\partial \phi}{\partial t},$$

as the directional derivative of ϕ in the direction $\boldsymbol{\tau}$ as similarly done in Douglas and Russell (1982) .

Thus, we can write SWE in *characteristic form* as

$$\alpha \frac{\partial H}{\partial \boldsymbol{\tau}} + H(\nabla \cdot \boldsymbol{u}) = 0, \tag{3}$$

$$\alpha \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\tau}} + g\nabla(H - h_\mathrm{b}) - \mu \Delta \boldsymbol{u} + \tau_\mathrm{b} \boldsymbol{u} + \boldsymbol{F} = 0. \tag{4}$$

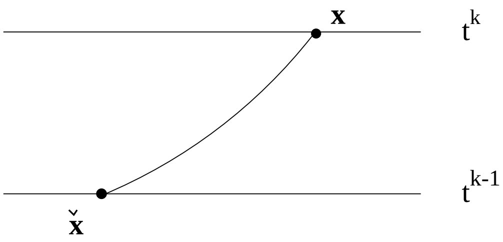

Fig. 2. Characteristic

Douglas and Russell (1982) observed theoretically that it is the much smaller norms of $\frac{\partial^2 H}{\partial \boldsymbol{\tau}^2}$ and $\frac{\partial^2 \boldsymbol{u}}{\partial \boldsymbol{\tau}^2}$ compared to the norms of $\frac{\partial^2 H}{\partial t^2}$ and $\frac{\partial^2 \boldsymbol{u}}{\partial t^2}$ (obtained in standard time-stepping procedures) that allow larger time-steps to be taken in advection-dominated flow.

The time-stepping procedure (along the characteristic lines determined by the method of characteristics) in combination with any spatial discretization has been referenced in the literature as the modified method of characteristics (MMOC). Specifically, the MMOC together with the Galerkin finite element method constitute the Characteristic-Galerkin (CG) method.

Parametrizing $\boldsymbol{x}$ with respect to t, the characteristic is the local solution to the initial-value problem

$$\left.\begin{aligned} \frac{d\boldsymbol{x}(t)}{dt} &= \boldsymbol{u}(\boldsymbol{x}(t),t), \qquad t \in (t^{k-1}, t^k), \\ \boldsymbol{x}(t^k) &= \boldsymbol{x}. \end{aligned}\right\} \tag{5}$$

The solution is computed by backtracing along the characteristic until $t = t^{k-1}$ is reached to determine the "foot" of the characteristic $\check{\boldsymbol{x}}$ (see Fig. 2).

We approximate the solution to this problem using Euler's method, let

$$\check{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}, t^k)\Delta t.$$

In the rest of the paper, let $\check{f} = f(\check{\boldsymbol{x}})$ and $f^k(\boldsymbol{x}) = f(\boldsymbol{x}, t^k)$.

3.2. Weak formulation

A weak form of (3)-(4) is

$$\left(\alpha \frac{\partial H}{\partial \boldsymbol{\tau}}, v\right) + (H(\nabla\boldsymbol{\cdot}\boldsymbol{u}), v) = 0, \quad \forall v \in \mathcal{H}^1(\Omega), \tag{6}$$

$$\begin{aligned} &\left(\alpha \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{\tau}}, \boldsymbol{w}\right) - (g(H - h_{\mathrm{b}}), \nabla\boldsymbol{\cdot}\boldsymbol{w}) + \mu\,(\nabla \boldsymbol{u}, \nabla \boldsymbol{w}) \\ &\quad + (\tau_{\mathrm{b}}\boldsymbol{u}, \boldsymbol{w}) + (\boldsymbol{F}, \boldsymbol{w}) = 0, \quad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{H}^1}(\Omega), \end{aligned} \tag{7}$$

where we assume $(H, \boldsymbol{u})$ satisfy the following initial conditions

$$H(\boldsymbol{x}, 0) = H_0(\boldsymbol{x}), \quad \boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0(\boldsymbol{x}). \tag{8}$$

3.3. Some assumptions

To avoid technical difficulties associated with implementing boundary conditions along characteristics we shall assume that the solution is Ω-periodic. Hereafter, we shall understand that each Sobolev space is a periodic Sobolev space on Ω and understand the meaning of the associated norms accordingly.

We need to list some additional assumptions. Let ℓ be a positive integer, $\ell \leq s_1$. Suppose that for $(\boldsymbol{x}, t) \in \bar{\Omega} \times (0, T]$,

A1. the solutions $(H, \boldsymbol{u})$ to (6)-(8) exist and are unique,

A2. $\exists$ positive constants H_* and H^* such that $H_* \leq H(\boldsymbol{x}, t) \leq H^*$,

A3. μ is a positive constant,

A4. $\boldsymbol{F}(\boldsymbol{x}, t)$ is bounded,

A5. $H_0(\boldsymbol{x}) \in \mathcal{H}^\ell(\Omega)$,

A6. $\boldsymbol{u}_0(\boldsymbol{x}) \in \boldsymbol{\mathcal{H}^\ell}(\Omega)$,

A7. $H(\boldsymbol{x}, t) \in \mathcal{H}^\ell(\Omega) \cap \mathcal{W}^1_\infty(\Omega), \;\; t \in (0, T)$,

A8. $\boldsymbol{u}(\boldsymbol{x}, t) \in \boldsymbol{\mathcal{H}^\ell}(\Omega) \cap \boldsymbol{\mathcal{W}^1_\infty}(\Omega), \;\; t \in (0, T)$.

A9. $\frac{\partial^2 H}{\partial \boldsymbol{\tau}^2}$ and $\frac{\partial^2 \boldsymbol{u}}{\partial \boldsymbol{\tau}^2}$ are in $\mathcal{L}^2\left((0, T); \mathcal{L}^2(\Omega)\right)$.

4. Characteristic-Galerkin finite element approximation

4.1. Defining the finite element approximations

Define the Characteristic-Galerkin approximations to $(H, \boldsymbol{u})$ to be the maps $H_h : J_{\Delta t} \to \mathcal{S}_h$, $\boldsymbol{u}_h : J_{\Delta t} \to \boldsymbol{\mathcal{S}_h}$. Let the approximate characteristic be denoted by

$$\hat{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{u}_h^k(\boldsymbol{x}) \Delta t.$$

Let $\hat{f} = f(\hat{\boldsymbol{x}})$. Then, Characteristic-Galerkin approximations $(H_h^k, \boldsymbol{u}_h^k)$ satisfy

$$\left(\frac{H_h^k - \hat{H}_h^{k-1}}{\Delta t}, v\right) + \left(H_h^k(\nabla \cdot \boldsymbol{u}_h^k), v\right) = 0, \quad \forall v \in \mathcal{S}_h, \; k \geq 1, \tag{9}$$

$$\begin{aligned}\left(\frac{\boldsymbol{u}_h^k - \hat{\boldsymbol{u}}_h^{k-1}}{\Delta t}, \boldsymbol{w}\right) - \left(g(H_h^k - h_{\mathrm{b}}), \nabla \cdot \boldsymbol{w}\right) + \mu \left(\nabla \boldsymbol{u}_h^k, \nabla \boldsymbol{w}\right) \\ + \left(\tau_{b_h}^k \boldsymbol{u}_h^k, \boldsymbol{w}\right) + \left(\boldsymbol{F}_h^k, \boldsymbol{w}\right) = 0, \quad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{S}_h}, \; k \geq 1,\end{aligned} \tag{10}$$

with initial conditions

$$
H_h^0 = \tilde{H}_0(\boldsymbol{x}), \qquad \boldsymbol{u}_h^0 = \tilde{\boldsymbol{u}}_0(\boldsymbol{x}),
$$
$$
\hat{H}_h^0 = \tilde{H}_0(\hat{\boldsymbol{x}}), \qquad \hat{\boldsymbol{u}}_h^0 = \tilde{\boldsymbol{u}}_0(\hat{\boldsymbol{x}}), \tag{11}
$$

where $\tilde{H}_0 \in \mathcal{S}_h$ and $\tilde{\boldsymbol{u}}_0 \in \boldsymbol{\mathcal{S}_h}$ are the $\mathcal{L}^2$ projections of H_0 and $\boldsymbol{u}_0$.

In the sections that follow, we will derive an a priori error estimate for the Characteristic-Galerkin method described here.

4.2. Boundedness assumptions

Given $\mathcal{L}^2$ projections $\tilde{H} \in \mathcal{S}_h$, and $\tilde{\boldsymbol{u}} \in \boldsymbol{\mathcal{S}_h}$ of H and $\boldsymbol{u}$, we denote the projection errors in elevation and velocity as

$$
\psi_H = (H_h - \tilde{H}) \quad \text{and} \quad \boldsymbol{\psi_u} = (\boldsymbol{u}_h - \tilde{\boldsymbol{u}}),
$$

respectively; and we also denote the approximation errors in elevation and velocity as

$$
\theta_H = (H - \tilde{H}) \quad \text{and} \quad \boldsymbol{\theta_u} = (\boldsymbol{u} - \tilde{\boldsymbol{u}}).
$$

In order to derive our estimate, we make some boundedness assumptions on the approximate solutions. We then show that for h and Δt sufficiently small, and for s_1 sufficiently large, we can remove the estimate's dependence on the assumed bound of the approximations, being dependent instead on a smaller bound of the $\mathcal{L}^2$ projection of the true solution. First, for any time t, let K^* satisfy

$$
|\tilde{H}| + |\tilde{\boldsymbol{u}}| + |\nabla \tilde{\boldsymbol{u}}| \le K^*.
$$

Such a constant exists by Lemmas 2 and 3. We then assume that for $k = 0, \ldots, N$ there exists positive constants $K_{**} \le \frac{H_*}{2}$ and $K^{**} \ge 2K^*$ such that

B1. $K_{**} \le H_h^k \le K^{**}$,

B2. $\left|\boldsymbol{u}_h^k\right| \le K^{**}$, $\left|\nabla \boldsymbol{u}_h^k\right| \le K^{**}$.

5. Characteristic equation and properties therein

Before bounding terms, we will need to show that the characteristic equations have certain properties.

The following theorems have been proven in Russell (1985) and Ewing et al. (1984).

Theorem 1. *Suppose* $\boldsymbol{u}^k \in \mathcal{W}^1_\infty(\Omega)$*. For* $t^k \in J_{\Delta t}, t^k \geq \Delta t$*, let* $\mathcal{F}_{\check{\boldsymbol{x}}}(\boldsymbol{x}) \equiv \boldsymbol{x} - \boldsymbol{u}^k(\boldsymbol{x})\Delta t$*. Then,* $\mathcal{F}_{\check{\boldsymbol{x}}}$ *is a differentiable homeomorphism for* Δt *sufficiently small.*

Theorem 2. *Let Assumption* **B2** *hold. For* $t^k \in J_{\Delta t}, t^k \geq \Delta t$*, let* $\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x}) \equiv \boldsymbol{x} - \boldsymbol{u}_h^k(\boldsymbol{x})\Delta t$*. Then,* $\mathcal{F}_{\hat{\boldsymbol{x}}}$ *is a differentiable homeomorphism for* Δt *sufficiently small.*

Now we can obtain the following generalization of a lemma found in Dawson et al. (1989).

Lemma 4. *Assume that* $\boldsymbol{u}_h^k(\boldsymbol{x})$ *has bounded first partial derivatives in space* $\forall k$*, (Assumption* **B2***). Then, for* Δt *sufficiently small, an arbitrary function* $f \in \mathcal{L}^2(\Omega)$ *satisfies*

$$\frac{1}{2\Delta t}\left[\left(\hat{f},\hat{f}\right) - (f,f)\right] \leq K_1||f||^2 + \epsilon||f||^2,$$

where,

$$K_1 = K_1\left(||\nabla\cdot\boldsymbol{u}_h||_{\mathcal{L}^\infty(\Omega)}\right).$$

Proof. Following closely the arguments of Lemma 3.1 in Dawson et al. (1989), let

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{u}_h^k(\boldsymbol{x})\Delta t \equiv \mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x}).$$

From the boundedness assumptions on the first-order spatial partial derivatives of $\boldsymbol{u}_h$, observe that the inverse of the Jacobian of this transformation satisfies

$$\left|J_{\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x})}{}^{-1}\right| = 1 + \nabla\cdot\boldsymbol{u}_h^k(\boldsymbol{x})\Delta t + O(\Delta t^2).$$

Therefore, given that the differentiable homeomorphism $\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x})$ maps the periodic Ω into itself, consider the following change of variables:

$$\begin{aligned}\left(\hat{f},\hat{f}\right) &= \int_\Omega f(\boldsymbol{y})f(\boldsymbol{y})\,d\boldsymbol{x} = \int_\Omega f(\boldsymbol{y})f(\boldsymbol{y})\left|J_{\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x})}{}^{-1}\right|\,d\boldsymbol{y}\\ &= \int_\Omega f(\boldsymbol{y})f(\boldsymbol{y})\left[1 + \nabla\cdot\boldsymbol{u}_h^k(\boldsymbol{x})\Delta t + O(\Delta t^2)\right]\,d\boldsymbol{y}.\end{aligned}$$

Now, subtracting (f,f) from $\left(\hat{f},\hat{f}\right)$ yields

$$\begin{aligned}&\frac{1}{2\Delta t}\left[\left(\hat{f},\hat{f}\right) - (f,f)\right]\\ &\quad= \frac{1}{2\Delta t}\left\{\int_\Omega f(\boldsymbol{y})f(\boldsymbol{y})\left[1 + \nabla\cdot\boldsymbol{u}_h^k(\boldsymbol{x})\Delta t + O(\Delta t^2)\right]\,d\boldsymbol{y}\right.\end{aligned}$$

$$
\begin{aligned}
&\quad - \int_\Omega f(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \Big\} \\
&= \frac{1}{2\Delta t} \left\{ \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y}) \left[1 + \nabla\cdot \boldsymbol{u}_h^k(\boldsymbol{x}) \Delta t + O(\Delta t^2) \right] \, d\boldsymbol{y} \right. \\
&\quad \left. - \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y})\, d\boldsymbol{y} \right\} \\
&= \frac{1}{2} \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y}) \left[\nabla\cdot \boldsymbol{u}_h^k(\boldsymbol{x}) + O(\Delta t) \right] \, d\boldsymbol{y} = W_1 + W_2,
\end{aligned}
$$

where the second equality comes from the fact that $\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x})$ is a differentiable homeomorphism onto itself.

In Dawson et al. (1989), term W_1 (with $\nabla\cdot\boldsymbol{u}(\boldsymbol{x})$ instead of $\nabla\cdot\boldsymbol{u}_h^k(\boldsymbol{x})$) was bounded by first adding and subtracting $\nabla\cdot\boldsymbol{u}(\boldsymbol{y})$ to get two terms W_{1a} and W_{1b}. The second term, W_{1b} was straightforward to bound using the assumption that $\nabla\cdot\boldsymbol{u}$ is bounded in $\mathcal{L}^\infty(\Omega)$. The first term W_{1a} was bounded using the Mean-Value Theorem on $\nabla\cdot\boldsymbol{u}$, assuming that $\nabla(\nabla\cdot\boldsymbol{u})$ exists and is bounded in $\mathcal{L}^\infty(\Omega)$. Here, we weaken these assumptions by not splitting W_1 into two terms and instead writing

$$
\begin{aligned}
W_1 &= \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y}) (\nabla\cdot \boldsymbol{u}_h^k(\boldsymbol{x}))\, d\boldsymbol{y} \\
&= \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y}) \left(\nabla\cdot \boldsymbol{u}_h^k \left(\mathcal{F}_{\hat{\boldsymbol{x}}}^{-1}(\boldsymbol{y}) \right) \right) \, d\boldsymbol{y} \\
&\le K_1 ||f||^2,
\end{aligned}
$$

where $K_1 = K_1 \left(||\nabla\cdot \boldsymbol{u}_h^k||_{\mathcal{L}^\infty(\Omega)} \right)$.

Now, note that for Δt sufficiently small,

$$
W_2 = \frac{O(\Delta t)}{2} \int_\Omega f(\boldsymbol{y}) f(\boldsymbol{y})\, d\boldsymbol{y} \le \epsilon ||f||^2.
$$

Therefore,

$$
\frac{1}{2\Delta t} \left[\left(\hat{f}, \hat{f} \right) - (f, f) \right] \le K_1 \left(\left|\left| \nabla\cdot \boldsymbol{u}_h^k \right|\right|_{\mathcal{L}^\infty(\Omega)} \right) ||f||^2 + \epsilon ||f||^2.
$$

We will also need to develop another technique based on the definition of the characteristic map, as done in Russell (1985), Ewing et al. (1984) and Dawson et al. (1989).

For a general function $f(\boldsymbol{x})$ defined over Ω, the expansion of $f(\check{\boldsymbol{x}})$ about $f(\hat{\boldsymbol{x}})$ using Taylor's theorem with integral remainder gives,

$$
f(\check{\boldsymbol{x}}) - f(\hat{\boldsymbol{x}}) = \int_{\hat{\boldsymbol{x}}}^{\check{\boldsymbol{x}}} \frac{\partial f}{\partial \boldsymbol{z}}(\boldsymbol{z})\, d\boldsymbol{z}
$$

where $\boldsymbol{z}$ is the unit vector in the direction $\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}$. Letting $\bar{z} \in [0,1]$ parametrize the segment from $\hat{\boldsymbol{x}}[\bar{z}=0]$ to $\check{\boldsymbol{x}}[\bar{z}=1]$, then

$$\begin{aligned} f(\check{\boldsymbol{x}}) - f(\hat{\boldsymbol{x}}) &= \left[\int_0^1 \frac{\partial f}{\partial \boldsymbol{z}}\left((1-\bar{z})\hat{\boldsymbol{x}} + \bar{z}\check{\boldsymbol{x}}\right)\, d\bar{z}\right](\check{\boldsymbol{x}} - \hat{\boldsymbol{x}}) \\ &\equiv \mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})\,(\check{\boldsymbol{x}} - \hat{\boldsymbol{x}})\,, \end{aligned}$$

where,

$$\begin{aligned} \mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}}) &= \int_0^1 \frac{\partial f}{\partial \boldsymbol{z}}\left((1-\bar{z})\hat{\boldsymbol{x}} + \bar{z}\check{\boldsymbol{x}}\right)\, d\bar{z} \\ &= \int_0^1 \frac{\partial f}{\partial \boldsymbol{z}}\left(Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})\right)\, d\bar{z}, \end{aligned}$$

with $Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x}) = (1-\bar{z})\hat{\boldsymbol{x}}(\boldsymbol{x}) + \bar{z}\check{\boldsymbol{x}}(\boldsymbol{x})$.

Lemma 5. *Let* $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. *Let* $\nabla f(\boldsymbol{x}) \in \mathcal{L}^p(\Omega)$, $(\check{\boldsymbol{x}} - \hat{\boldsymbol{x}}) \in \mathcal{L}^q(\Omega)$, *and* $g(\boldsymbol{x}) \in \mathcal{L}^r(\Omega)$, *Then,*

$$\begin{aligned} &\int_\Omega g(\boldsymbol{x})\left(f(\check{\boldsymbol{x}}) - f(\hat{\boldsymbol{x}})\right)\, d\boldsymbol{x} \\ &\quad= \int_\Omega \mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})\,(\check{\boldsymbol{x}} - \hat{\boldsymbol{x}})\, g(\boldsymbol{x})\, d\boldsymbol{x} \\ &\quad\le ||\mathcal{I}_f||_{\mathcal{L}^p(\Omega)} ||\check{\boldsymbol{x}} - \hat{\boldsymbol{x}}||_{\mathcal{L}^q(\Omega)} ||g||_{\mathcal{L}^r(\Omega)} \\ &\quad\le K||\nabla f||_{\mathcal{L}^p(\Omega)} ||\check{\boldsymbol{x}} - \hat{\boldsymbol{x}}||_{\mathcal{L}^q(\Omega)} ||g||_{\mathcal{L}^r(\Omega)}. \end{aligned}$$

Proof. Ewing et al. (1984) establish that $Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})$ is a differentiable homeomorphism (except they use an extrapolated approximate velocity instead of the approximate velocity proper) using arguments similar to those showing that $\mathcal{F}_{\hat{\boldsymbol{x}}}(\boldsymbol{x})$ is a differentiable homeomorphism. Then, we can establish the following results.

Case $p \in [1, \infty)$:

$$\begin{aligned} \int_\Omega |\mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})|^p\, d\boldsymbol{x} &= \int_\Omega \left| \int_0^1 \frac{\partial f}{\partial \boldsymbol{z}}\left(Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})\right)\, d\bar{z}\right|^p\, d\boldsymbol{x} \\ &\le \int_0^1 \int_\Omega \left|\frac{\partial f}{\partial \boldsymbol{z}}\left(Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})\right)\right|^p\, d\boldsymbol{x}\ \ d\bar{z}. \end{aligned}$$

Letting $\boldsymbol{y} = Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})$ and changing variables above, yields

$$\begin{aligned} \int_\Omega |\mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})|^p\, d\boldsymbol{x} &\le \int_0^1 \int_\Omega \left|\frac{\partial f}{\partial \boldsymbol{z}}(\boldsymbol{y})\right|^p \left|J_{Y_{\check{\boldsymbol{x}},\hat{\boldsymbol{x}}}(\boldsymbol{x})}{}^{-1}\right|\, d\boldsymbol{y}\ \ d\bar{z} \\ &\le K \int_0^1 \int_\Omega \left|\frac{\partial f}{\partial \boldsymbol{z}}(\boldsymbol{y})\right|^p\, d\boldsymbol{y}\ \ d\bar{z} \end{aligned}$$

Thus

$$\int_{\Omega} |\mathcal{I}_f(\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})|^p \ d\boldsymbol{x} \le K||\nabla f||^p_{\mathcal{L}^p(\Omega)}.$$

Case $p = \infty$ is straightforward. □

6. Error estimate

6.1. The error equations

Write down the error equations resulting from subtracting (6)-(8) from (9)-(11), respectively, as

$$\begin{aligned}
&\left(\frac{\psi_H^k - \hat{\psi}_H^{k-1}}{\Delta t}, v\right) \\
&= \left(\frac{\theta_H^k - \check{\theta}_H^{k-1}}{\Delta t}, v\right) - \left(\frac{\check{\tilde{H}}^{k-1} - \hat{\tilde{H}}^{k-1}}{\Delta t}, v\right) \\
&\quad + \left((H^k - H_h^k)\nabla\cdot\boldsymbol{u}_h^k, v\right) + \left(H^k\nabla\cdot(\boldsymbol{u}^k - \boldsymbol{u}_h^k), v\right) \\
&\quad - \left(\zeta^k, v\right), \quad \forall v \in \mathcal{S}_h,\ k \ge 1, \qquad (12)
\end{aligned}$$

and

$$\begin{aligned}
&\left(\frac{\boldsymbol{\psi}_{\boldsymbol{u}}^k - \hat{\psi}_{\boldsymbol{u}}^{k-1}}{\Delta t}, \boldsymbol{w}\right) + \mu\left(\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k, \nabla\boldsymbol{w}\right) + \left({\tau_{b_h}}^k\boldsymbol{\psi}_{\boldsymbol{u}}^k, \boldsymbol{w}\right) \\
&= \left(\frac{\boldsymbol{\theta}_{\boldsymbol{u}}^k - \check{\boldsymbol{\theta}}_{\boldsymbol{u}}^{k-1}}{\Delta t}, \boldsymbol{w}\right) - \left(\frac{\check{\tilde{\boldsymbol{u}}}^{k-1} - \hat{\tilde{\boldsymbol{u}}}^{k-1}}{\Delta t}, \boldsymbol{w}\right) \\
&\quad - \left(g(H^k - H_h^k), \nabla\cdot\boldsymbol{w}\right) + \mu\left(\nabla\boldsymbol{\theta}_{\boldsymbol{u}}^k, \nabla\boldsymbol{w}\right) \\
&\quad + \left({\tau_{\mathrm{b}}}^k\boldsymbol{\theta}_{\boldsymbol{u}}^k, \boldsymbol{w}\right) + \left(({\tau_{\mathrm{b}}}^k - {\tau_{b_h}}^k)\tilde{\boldsymbol{u}}^k, \boldsymbol{w}\right) \\
&\quad + \left(\boldsymbol{F}^k - \boldsymbol{F}_h^k, \boldsymbol{w}\right) + \left(\boldsymbol{\sigma}^k, \boldsymbol{w}\right), \quad \forall \boldsymbol{w} \in \boldsymbol{\mathcal{S}_h},\ k \ge 1, \qquad (13)
\end{aligned}$$

with

$$\psi_H^0 = \hat{\psi}_H^0 = 0, \quad \boldsymbol{\psi}_{\boldsymbol{u}}^0(\boldsymbol{x}) = \hat{\psi}_{\boldsymbol{u}}^0 = 0, \qquad (14)$$

and the truncation terms ζ^k and $\boldsymbol{\sigma}^k$ are defined as follows

$$\zeta^k = \alpha^k\frac{\partial H^k}{\partial\boldsymbol{\tau}} - \left(\frac{H(\boldsymbol{x}, t^k) - H(\check{\boldsymbol{x}}, t^{k-1})}{\Delta t}\right),$$

$$\boldsymbol{\sigma}^k = \alpha^k\frac{\partial\boldsymbol{u}^k}{\partial\boldsymbol{\tau}} - \left(\frac{\boldsymbol{u}(\boldsymbol{x}, t^k) - \boldsymbol{u}(\check{\boldsymbol{x}}, t^{k-1})}{\Delta t}\right).$$

6.2. Bounding the errors

Sum together (12) and (13) using test functions $v = \psi_H^k$ and $\boldsymbol{w} = \boldsymbol{\psi}_{\boldsymbol{u}}^k$. Now use the inequality $(a-b,a) \geq \frac{1}{2}(a^2-b^2)$ once with $a = \psi_H^k$ and $b = \hat{\psi}_H^{k-1}$ and a second time with $a = \boldsymbol{\psi}_{\boldsymbol{u}}^k$ and $b = \hat{\psi}_{\boldsymbol{u}}^{k-1}$. Use the definition of the $\mathcal{L}^2$ projection and finally, add and subtract the two terms $\left|\left|\psi_H^{k-1}\right|\right|^2$ and $\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|$ to the result to obtain

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(\left|\left|\psi_H^k\right|\right|^2 - \left|\left|\psi_H^{k-1}\right|\right|^2\right) + \frac{1}{2\Delta t}\left(\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2 - \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2\right) \\
&\quad +\mu\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2 + \left|\left|\sqrt{\tau_{b_h}^{\ k}}\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2 \\
&\leq \frac{1}{2\Delta t}\left(\left|\left|\hat{\psi}_H^{k-1}\right|\right|^2 - \left|\left|\psi_H^{k-1}\right|\right|^2\right) \\
&\quad + \left(\frac{\theta_H^k - \check{\theta}_H^{k-1}}{\Delta t}, \psi_H^k\right) - \left(\frac{\check{\tilde{H}}^{k-1} - \hat{\tilde{H}}^{k-1}}{\Delta t}, \psi_H^k\right) \\
&\quad + \left((\theta_H^k - \psi_H^k)\nabla\cdot\boldsymbol{u}_h^k, \psi_H^k\right) + \left(H^k\nabla\cdot(\boldsymbol{\theta}_{\boldsymbol{u}}^k - \boldsymbol{\psi}_{\boldsymbol{u}}^k), \psi_H^k\right) \\
&\quad - \left(\zeta^k, \psi_H^k\right) \\
&\quad + \frac{1}{2\Delta t}\left(\left|\left|\hat{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2 - \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2\right) \\
&\quad + \left(\frac{\boldsymbol{\theta}_{\boldsymbol{u}}^k - \check{\boldsymbol{\theta}}_{\boldsymbol{u}}^{k-1}}{\Delta t}, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) - \left(\frac{\check{\tilde{\boldsymbol{u}}}^{k-1} - \hat{\tilde{\boldsymbol{u}}}^{k-1}}{\Delta t}, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) \\
&\quad - \left(g(\theta_H^k - \psi_H^k), \nabla\cdot\boldsymbol{\psi}_{\boldsymbol{u}}^k\right) + \mu\left(\nabla\boldsymbol{\theta}_{\boldsymbol{u}}^k, \nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k\right) \\
&\quad + \left((\tau_{\mathrm{b}}{}^k - \tau_{b_h}^{\ k})\tilde{\boldsymbol{u}}^k, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) + \left(\boldsymbol{F}^k - \boldsymbol{F}_h^k, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) - \left(\boldsymbol{\sigma}^k, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) \\
&= T_1 + \cdots + T_6 + S_1 + \cdots + S_8. \qquad (15)
\end{aligned}$$

From Lemma 4, with $f = \psi_H^{k-1}$ in T_1 and with $f = \boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}$ in S_1, we immediately have that

$$T_1 \leq K_1\left|\left|\psi_H^{k-1}\right|\right|^2 + \epsilon\left|\left|\psi_H^{k-1}\right|\right|^2,$$
$$S_1 \leq K_1\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2 + \epsilon\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2.$$

Using Cauchy-Schwarz, assumptions A2 and A3, and the inequality $ab \leq \epsilon a^2 + \frac{1}{4\epsilon}b^2$, the bounds on T_4, T_5, S_4 and S_5 are straightforward:

$$T_4 \leq K_1\left|\left|\theta_H{}^{k-1}\right|\right|^2 + K_1\left|\left|\psi_H^{k-1}\right|\right|^2 + K_1\left|\left|\psi_H^k\right|\right|^2,$$

$$T_5 \leq \epsilon\left|\left|\nabla \boldsymbol{\psi}_{\boldsymbol{u}}^{k-1}\right|\right|^2 + K\left|\left|\nabla\boldsymbol{\cdot}\boldsymbol{\theta}_{\boldsymbol{u}}^{k-1}\right|\right|^2 + K\left|\left|\psi_H^k\right|\right|^2,$$
$$S_4 \leq \epsilon\left|\left|\nabla \boldsymbol{\psi}_{\boldsymbol{u}}^{k}\right|\right|^2 + K\left|\left|\theta_H^k\right|\right|^2 + K\left|\left|\psi_H^k\right|\right|^2,$$
$$S_5 \leq \epsilon\left|\left|\nabla \boldsymbol{\psi}_{\boldsymbol{u}}^{k}\right|\right|^2 + K\left|\left|\nabla\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2.$$

Using the definition of the $\mathcal{L}^2$ projection, assumption A8, the Mean Value Theorem, and Cauchy-Schwarz, we get

$$\begin{aligned}
T_2 &= \frac{1}{\Delta t}\int_{\Omega}\left(\theta_H^k - \check{\theta}_H^{k-1}\right)\psi_H^k\, dx \\
&= \frac{1}{\Delta t}\int_{\Omega}\left({\theta_H}^{k-1} - \check{\theta}_H^{k-1}\right)\psi_H^k\, dx \\
&\leq \frac{1}{\Delta t}\left|\left|\nabla{\theta_H}^{k-1}\right|\right|\,||\boldsymbol{x} - \check{\boldsymbol{x}}||_{\mathcal{L}^\infty(\Omega)}\left|\left|\psi_H^k\right|\right| \\
&= \left|\left|\nabla{\theta_H}^{k-1}\right|\right|\,\left|\left|\boldsymbol{u}^k\right|\right|_{\mathcal{L}^\infty(\Omega)}\left|\left|\psi_H^k\right|\right| \\
&\leq K\left|\left|\nabla{\theta_H}^{k-1}\right|\right|^2 + K\left|\left|\psi_H^k\right|\right|^2.
\end{aligned}$$

Note, that the term $\left|\left|\nabla{\theta_H}^{k-1}\right|\right|$ accounts for the suboptimality of the error estimate we will derive. Douglas and Russell (1982) handle a similar term by bounding $\left(\frac{\theta_H^k - \check{\theta}_H^{k-1}}{\Delta t}\right)$ in the $\mathcal{H}^{-1}$ norm since then $\left|\left|\frac{\theta_H^k - \check{\theta}_H^{k-1}}{\Delta t}\right|\right|_{\mathcal{H}^{-1}(\Omega)} \approx \left|\left|{\theta_H}^{k-1}\right|\right|$. However, the test function must then be measured in the $\mathcal{H}^1$ norm. Since we won't have a term on the LHS of the error equations in which to hide this latter term, we do not find it useful to apply a duality argument.

Similarly,

$$\begin{aligned}
S_2 &= \frac{1}{\Delta t}\int_{\Omega}\left(\boldsymbol{\theta}_{\boldsymbol{u}}^k - \check{\boldsymbol{\theta}}_{\boldsymbol{u}}^{k-1}\right)\boldsymbol{\psi}_{\boldsymbol{u}}^k\, dx \\
&\leq K\left|\left|\nabla\boldsymbol{\theta}_{\boldsymbol{u}}^{k-1}\right|\right|^2 + K\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2.
\end{aligned}$$

The bounds on T_3 and S_3 can be determined using a parametrization argument made in Ewing et al. (1984) and in Russell (1985). Use Lemma 5 with $g = \psi_H^k$ and $f = \tilde{H}^{k-1}/\Delta t$ in T_3 and with $g = \boldsymbol{\psi}_{\boldsymbol{u}}^k$ and $f = \tilde{\boldsymbol{u}}^{k-1}/\Delta t$ in S_3. Recall from Lemma 2 that $\nabla\tilde{H}^{k-1} \in \mathcal{L}^\infty(\Omega)$ and $\nabla\tilde{\boldsymbol{u}}^{k-1} \in \mathcal{L}^\infty(\Omega)$. Finally, use the inequality $\left(ab \leq \epsilon a^2 + \frac{1}{4\epsilon}b^2\right)$ to obtain

$$T_3 = \int_{\Omega}\left(\frac{\check{\tilde{H}}^{k-1} - \hat{\tilde{H}}^{k-1}}{\Delta t}\right)\psi_H^k\, dx \leq K\left|\left|\psi_H^k\right|\right|\,\left|\left|\nabla\tilde{H}^{k-1}\right|\right|_{\mathcal{L}^\infty(\Omega)}\left|\left|\boldsymbol{u}^k - \boldsymbol{u}_h^k\right|\right|$$

$$\leq K\left|\left|\psi_H^k\right|\right|^2 + K\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2 + K\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2,$$

$$S_3 = \int_\Omega \left(\frac{\check{\tilde{\boldsymbol{u}}}^{k-1} - \hat{\tilde{\boldsymbol{u}}}^{k-1}}{\Delta t}\right)\boldsymbol{\psi}_{\boldsymbol{u}}^k\, dx \leq K\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right| \left|\left|\nabla\tilde{\boldsymbol{u}}^{k-1}\right|\right|_{\mathcal{L}^\infty(\Omega)}\left|\left|\boldsymbol{u}^k - \boldsymbol{u}_h^k\right|\right|$$
$$\leq 2K\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2 + K\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2.$$

In bounding T_6 and S_8, recall $\alpha = |\boldsymbol{\tau}|$ and $\left|\left|\alpha^4\right|\right|_{\mathcal{L}^\infty(\Omega)}$ is bounded by assumption A8. Now, following Russell (1985), we find

$$\left|\left|\zeta^k\right|\right|^2 \leq K\Delta t\int_{t^{k-1}}^{t^k}\left|\left|\frac{\partial^2 H}{\partial\boldsymbol{\tau}^2}\right|\right|^2 dt,$$
$$\left|\left|\boldsymbol{\sigma}^k\right|\right|^2 \leq K\Delta t\int_{t^{k-1}}^{t^k}\left|\left|\frac{\partial^2 \boldsymbol{u}}{\partial\boldsymbol{\tau}^2}\right|\right|^2 dt$$

to get

$$T_6 \leq K\Delta t\int_{t^{k-1}}^{t^k}\left|\left|\frac{\partial^2 H}{\partial\boldsymbol{\tau}^2}\right|\right|^2 dt + K\left|\left|\psi_H^k\right|\right|^2,$$
$$S_8 \leq K\Delta t\int_{t^{k-1}}^{t^k}\left|\left|\frac{\partial^2 \boldsymbol{u}}{\partial\boldsymbol{\tau}^2}\right|\right|^2 dt + K\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2.$$

To get the bound on S_6, we recall the bound obtained for the same term in Chippada et al(1999). Recalling Lemma 3 and assumptions B1 and B2, we obtain

$$S_6 \leq K\left[\left|\left|\psi_H^k\right|\right|^2 + \left|\left|\theta_H^k\right|\right|^2 + \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\right]$$
$$+K\left[\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2 + \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\right].$$

Finally, to get the bound on S_7, we again recall the bound obtained for the same term in Chippada et al(1999). From assumptions A2, A4 and B1, we obtain

$$S_7 \leq \left(f_{\mathrm{c}}\boldsymbol{k}\times\boldsymbol{\theta}_{\boldsymbol{u}}^k, \boldsymbol{\psi}_{\boldsymbol{u}}^k\right) + \left(\frac{\theta_H^k - \psi_H^k}{H^k H_h^k}, \boldsymbol{\tau}_{\mathrm{ws}}\boldsymbol{\psi}_{\boldsymbol{u}}^k\right)$$
$$\leq K\left(\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2 + \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\right) + K\left|\left|\theta_H^k\right|\right|^2 + \epsilon\left|\left|\psi_H^k\right|\right|^2.$$

Multiplying (15) by Δt, summing over k, $k = 1, \cdots, N$ using the bounds on $T_1, \cdots, S_8$, and collecting terms yields

$$\frac{1}{2}\left|\left|\psi_H^N\right|\right|^2 + \frac{1}{2}\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^N\right|\right|^2 + \sum_{k=1}^N\left|\left|\sqrt{\tau_{b_h}^{\,k}}\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t + \mu\sum_{k=1}^N\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}\right|\right|^2\Delta t$$

$$\begin{aligned}
&\leq \frac{1}{2}\left|\left|\psi_H^0\right|\right|^2 + \frac{1}{2}\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^0\right|\right|^2 + \epsilon\sum_{k=1}^{N}\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t + K\sum_{k=1}^{N}\left|\left|\psi_H^k\right|\right|^2\Delta t \\
&+K\sum_{k=1}^{N}\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t + K\sum_{k=1}^{N}\left|\left|\theta_H^k\right|\right|^2_{\mathcal{H}^1(\Omega)}\Delta t + K\sum_{k=1}^{N}\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2_{\mathcal{H}^1(\Omega)}\Delta t \\
&+K\Delta t^2\left|\left|\frac{\partial^2 H}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} + K\Delta t^2\left|\left|\frac{\partial^2 \boldsymbol{u}}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}. \qquad (16)
\end{aligned}$$

Hide $\epsilon\sum_{k=1}^{N}\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t$ on the left side of (16) and use the fact that $\psi_H^0 = 0, \boldsymbol{\psi}_{\boldsymbol{u}}^0 = 0$, to get

$$\begin{aligned}
&\frac{1}{2}\left|\left|\psi_H^N\right|\right|^2 + \frac{1}{2}\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^N\right|\right|^2 + \sum_{k=1}^{N}\left|\left|\sqrt{\tau_{b_h}^k}\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t + \frac{\mu}{2}\sum_{k=1}^{N}\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}\right|\right|^2\Delta t \\
&\leq K_2\sum_{k=1}^{N}\left|\left|\psi_H^k\right|\right|^2\Delta t + K_3\sum_{k=1}^{N}\left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t + K\sum_{k=1}^{N}\left|\left|\theta_H\right|\right|^2_{\mathcal{H}^1(\Omega)}\Delta t \\
&+K\sum_{k=1}^{N}\left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}\right|\right|^2_{\mathcal{H}^1(\Omega)}\Delta t \\
&+K\Delta t^2\left(\left|\left|\frac{\partial^2 H}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} + \left|\left|\frac{\partial^2 \boldsymbol{u}}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}\right). \qquad (17)
\end{aligned}$$

Finally, apply the discrete Gronwall's Lemma to obtain

$$\begin{aligned}
&\left|\left|\psi_H^N\right|\right|^2 + \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^N\right|\right|^2 + \sum_{k=1}^{N}\left[\left|\left|\sqrt{\tau_{b_h}}\boldsymbol{\psi}_{\boldsymbol{u}}\right|\right|^2 + \left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}\right|\right|^2\right]\Delta t \\
&\leq \bar{K}K\left[\sum_{k=1}^{N}\left[\left|\left|\theta_H^k\right|\right|^2_{\mathcal{H}^1(\Omega)} + \left|\left|\boldsymbol{\theta}_{\boldsymbol{u}}^k\right|\right|^2_{\mathcal{H}^1(\Omega)}\right]\Delta t\right. \\
&\left.+\Delta t^2\left(\left|\left|\frac{\partial^2 H}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} + \left|\left|\frac{\partial^2 \boldsymbol{u}}{\partial\boldsymbol{\tau}^2}\right|\right|^2_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}\right)\right], \qquad (18)
\end{aligned}$$

where $\bar{K} = \exp\left(\sum_{k=1}^{N}\left(\frac{K_4}{1-\Delta t K_4}\right)\Delta t\right)$, $K_4 = 2\max\{K_2, K_3\}$, and Δt is sufficiently small.

Therefore,

$$\left|\left|\psi_H^N\right|\right|^2 + \left|\left|\boldsymbol{\psi}_{\boldsymbol{u}}^N\right|\right|^2 + \sum_{k=1}^{N}\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}\right|\right|^2\Delta t \leq K\left(h^{2(\ell-1)} + \Delta t^2\right).$$

To complete the proof, we use the same argument as in Chippada et al. (1998), namely, when $\Delta t = o(h), s_1 \geq 3, \ell > 2$, we obtain for $\phi = \{H, \boldsymbol{u}\}$ (with ϕ_h understood to mean the already defined H_h and $\boldsymbol{u}_h$, respectively):

$$\begin{aligned} \left|\phi_h^N\right| &\leq \left|\psi_\phi^N\right| + \left|\tilde{\phi}^N\right|, \\ &\leq K h^{-1}\left(h^{l-1} + \Delta t\right) + K^* \\ &< 2K^* \\ &\leq K^{**}, \end{aligned}$$

and

$$\begin{aligned} H_h^N &= H^N - (\theta_H^N - \psi_h^N) \\ &\geq H_* - K h^{-1}(h^{l-1} + \Delta t) \\ &> H_*/2 \\ &\geq K_{**}. \end{aligned}$$

Finally, balancing Δt and h, we find that if $\Delta t = o(h^2), s_1 \geq 4, \ell > 3$, then

$$\begin{aligned} \left|\nabla\cdot\boldsymbol{u}_h^N\right| &\leq \left|\nabla\boldsymbol{u}_h^N\right| \\ &\leq \left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^N\right| + \left|\nabla\tilde{\boldsymbol{u}}^N\right| \\ &\leq K\Delta t^{-1/2}h^{-1}\left(\sum_{k=0}^{N}\left|\left|\nabla\boldsymbol{\psi}_{\boldsymbol{u}}^k\right|\right|^2\Delta t\right)^{1/2} + K^* \\ &\leq K\Delta t^{-1/2}h^{-1}\left(h^{l-1} + \Delta t\right) + K^* < 2K^*. \end{aligned}$$

Thus, we have proved the following:

Theorem 3. *Let $s_1 \geq 2$ and $1 \leq \ell \leq s_1$. Let $(H(\cdot, t^k), \boldsymbol{u}(\cdot, t^k))$ be Ω-periodic solutions to (6)-(8) at time $t = t^k$. Let $(H_h^k, \boldsymbol{u}_h^k)$ be the Characteristic -Galerkin approximations to $(H, \boldsymbol{u})$. If assumptions A1-A9 hold, with reasonable assumptions on surface and body forces, and with Δt sufficiently small, then $\exists$ a constant $\bar{K} = \bar{K}(T, s_1, K_*, K^*, K_{**}, K^{**})$ such that*

$$\begin{aligned} &\left|\left|H(\boldsymbol{x}, t^N) - H_h^N\right|\right| + \left|\left|\boldsymbol{u}(\boldsymbol{x}, t^N) - \boldsymbol{u}_h^N\right|\right| \\ &\quad + \left(\sum_{k=1}^{N}\left|\left|\nabla\boldsymbol{u} - \nabla\boldsymbol{u}_h\right|\right|^2\Delta t\right)^{1/2} \leq \bar{K}\left(h^{\ell-1} + \Delta t\right). \end{aligned}$$

If $h, \Delta t$ are sufficiently small, in particular $\Delta t = o(h)$, $s_1 \geq 3$ and $l > 2$, then we can remove the boundedness assumptions on $\boldsymbol{u}_h$ and on H_h, but not on $\nabla\cdot\boldsymbol{u}_h$. Finally, for h sufficiently small, $\Delta t = o(h^2)$, $s_1 \geq 4$ and $l > 3$, then

$$\bar{K} = \bar{K}(T, s_1, K_*, K^*).$$

Remark 1. The scheme described above may be made semi-explicit, thus decoupling (6)-(7), by using an extrapolated velocity $Eu_h^k = 2u_h^{k-1} - u_h^{k-2}$ in determining the approximate characteristic, and in lagging the term $H(\nabla\cdot\boldsymbol{u}_h)$ in (9) by evaluating it at t^{k-1}. The estimates above carry through, at the expense of some additional time truncation error terms. In particular, we obtain an $O(\Delta t)$ term involving the time derivative of $H(\nabla\cdot\boldsymbol{u}_h)$.

References

1. Adams, R.A. (1978): Sobolev Spaces. Pure Appl. Math., vol. 65, Academic Press, New York
2. Bernardi, C., Pironneau, O.(1991): Numerical analysis for the shallow water equations. Communications in Partial Differential Equations **16**, 59–104
3. Brenner, S.C., Ridgway Scott, L.R. (1994): The Mathematical Theory of Finite Element Methods, Texts in Appl. Math., vol. 15, Springer-Verlag, New York
4. Chippada, S., Dawson, C.N., Martínez, M.L., Wheeler, M.F. (1998): Finite element approximations to the system of shallow water equations, Part I: Contiuous time a priori error estimates. SIAM Journal of Numerical Analysis **35**, 692–711
5. Chippada, S., Dawson, C.N., Martínez, M.L., Wheeler, M.F. (1999): Finite element approximations to the system of shallow water equations, Part II: Discrete time a priori error estimates. SIAM Journal of Numerical Analysis **36**, 226–250
6. Dawson, C.N., Russell, T.F., Wheeler, M.F. (1989): Some improved error estimates for the modified method of characteristics. SIAM Journal of Numerical Analysis **26**, 1487–1512
7. Ewing, R.E., Russell, T.F., Wheeler, M.F. (1984): Convergence analysis of an approximation of miscible displacement in porous media by mixed finite element methods and a modified method of characteristics. Computer Methods in Applied Mechanics and Engineering **47**, 73–92
8. Gent, P. (1993): The energetically consistent shallow water equations. Journal of the Atmospheric Sciences **50**, 1323–1325
9. Russell, T.F., Douglas, J. (1982): Numerical methods for convection-dominated diffusion problems based on combining the method of characterisitcs with finite element or finite difference procedures. SIAM Journal of Numerical Analysis **19**, 871–885
10. Luettich, R.A., Westerink, J.J., Scheffner, N.W. (1991): ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. Technical Report 1, Department of the Army, U.S. Army Corps of Engineers, Washington, D.C. 20314-1000
11. Lynch, D.R., Gray, W.G. (1979): A wave equation model for finite element tidal computations. Computers and Fluids **7**, 207–228
12. Russell, T.F. (1985): Time-stepping along characteristics with incomplete iteration for a Galerkin approximation of miscible displacement in porous media. SIAM Journal of Numerical Analysis **22**, 970–1013
13. Weiyan, T. (1992): Shallow Water Hydrodynamics, Elsevier Oceanography Series., vol.55 Elsevier, Amsterdam
14. Zienkiewicz, O.C., Ortiz, P. (1995a): An improved finite element model for shallow water problems. In Graham F. Carey, editor, Finite Element Modeling of Environmental Problems: Surface and Subsurface Flow and Transport, Chap. 4, pp. 61–84. John Wiley & Sons, Chichester
15. Zienkiewicz, O.C., Ortiz, P. (1995): A split-characteristic based finite element model for the shallow water equations. International Journal for Numerical Methods in Fluids **20**, 1061–1080