INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS, VOL. 22, 835–849 (1996)

COLLOCATION AND UPWINDING FOR THERMAL FLOW IN PIPELINES: THE LINEARIZED CASE

PHILIP T. KEENAN

Texas Institute for Computational and Applied Mathematics, University of Texas at Austin, Austin TX 78745, U.S.A.

SUMMARY

Simulating thermal effects in pipeline flow involves solving a coupled non-linear system of first-order hyperbolic equations. The advection term has two large eigenvalues of opposite signs, corresponding to the propagation of high-speed sound waves, and one eigenvalue close to or even equal to zero, representing the much slower fluid flow velocity, which transports temperature. Standard collocation methods work well for isothermal flow in pipelines, but the stagnating eigenvalue causes difficulties when thermal effects are included. In a companion paper we formulate and analyse a new numerical method for the non-linear system which arises in thermal modelling. The new method applies to general coupled systems of non-linear first-order hyperbolic partial differential equations with one degenerate eigenvalue. In the present paper we focus on a linearized constant coefficient form of the thermal flow equations. This substantially simplifies presentation of the error analysis for the numerical scheme. We also include numerical results for the method applied to the fully non-linear system. Both the error analysis and the numerical experiments show that the difficulties that come from the application of standard collocation can be overcome by using upwinded piecewise constant functions for the degenerate component of the solution.

KEY WORDS: non-linear first-order hyperbolic system; collocation method; upwinding; thermal pipeline simulation

1. INTRODUCTION

Non-linear systems of first-order hyperbolic partial differential equations arise in the modelling of many natural and man-made phenomena. For instance, the pressure, velocity and temperature of a fluid in a one-dimensional pipeline can be described by such a system. However, accurate simulation of such systems by numerical computation can be difficult. Luskin[1] analysed a collocation method which can be successfully applied to isothermal flow in pipelines. It does not, however, apply in certain common cases when thermal effects are modelled, because the temperature equation introduces a degenerate eigenvalue corresponding to stagnating flow. Keenan[2,3] defined and analysed a new numerical method for coupled systems of non-linear first-order hyperbolic partial differential equations with one degenerate eigenvalue, which extended in a certain direction the collocation method described by Luskin. Both methods have direct application to the study of one-dimensional fluid flow through pipelines.

The present paper is intended as an introductory companion piece to References 2 and 3. A number of technical details obscure the error analysis presented in that work because it treats the general non-linear case. The present paper describes and analyses the method in the context of a linear, constant coefficient system of equations based on the thermal pipeline equations described in References 2 and 3. In this special case the error analysis simplifies considerably. For additional details on the application of the new method to the full non-linear system of pipeline simulation equations, as

CCC 0271–2091/96/090835–15

Received 25 May 1992
Revised 6 June 1995

well as experimental results and convergence proofs in the general case of non-linear coupled systems, see References 2 and 3.

The model problem is presented in Section 2. Next the new numerical method is defined in Section 3. A representative theorem describing asymptotic convergence of the numerical method is presented in Section 4. Finally the proof of the theorem is presented in Section 5.

2. THE MODEL PROBLEM

Consider the following system of two coupled first-order constant coefficient hyperbolic partial differential equations in one space dimension:

$$p_t = v_s p_x + aT = 0, \qquad T_t + v_f T_x + bp = 0, \tag{1}$$

where $p = p(x, t)$ and $T = T(x, t)$ are sought in the region $x \in [0, 1]$, $t \in (0, 1]$. Here v_s, v_f, a and b are constants,

$$v_s \gg v_f \geqslant 0,$$

and $p(x, 0)$, $T(x, 0)$, $p(0, t)$ and $T(0, t)$ are given.

The p component here is based on the pressure in the pipeline equations and T is based on temperature. Here p is advected at a high speed v_s compared with T, which flows only slowly. In fact, v_f can equal zero, in which case T is said to stagnate. It is known that standard collocation can produce bizarre behaviour in problems where stagnation can occur, such as in the thermal simulation of pipeline flow. For instance, if $v_f = 0$ and $b = 0$, a change in T at $x = 0$ would be instantly propagated as a saw-tooth wave down the entire pipe.

To keep the analysis of the model problem simple, p and T are here only coupled through lower-order terms. The actual thermal pipeline equations include additional coupling through x-derivative terms and in the coefficient functions. Moreover, v_f becomes an additional unknown, representing the fluid velocity. All these features complicate the definition and analysis of the numerical method in the general case; see References 2 and 3 for details. Also note that all the ideas in both papers generalize immediately to the case of systems of $n > 2$ equations for which all the eigenvalues but one are bounded uniformly away from zero.

3. THE NUMERICAL METHOD

3.1. Discrete notation for collocation in one space dimension

Let N be a positive integer and let $\Delta x = 1/N$. Let

$$x_j = j\Delta x, \quad j = 0, 1, \ldots, N.$$

Similarly let

$$x_{j+1/2} = (j + \tfrac{1}{2})\Delta x, \quad j = 0, 1, \ldots, N - 1.$$

The x_j are called the knots and the $x_{j+1/2}$ the midpoints. For any function $u(x)$ let

$$u_j = u(x_j), \qquad u_{j+1/2} = u(x_{j+1/2}), \qquad u_{j,c} = \tfrac{1}{2}(u_j + u_{j+1}).$$

For any functions $f(x)$ and $g(x)$ let

$$(f, g)_{L^2} = \int_0^1 f(x)g(x)\,dx,$$

$$(f, g)_{m^2} = \sum_{j=0}^{N-1} f_{j+1/2}g_{j+1/2}\Delta x, \qquad (f, g)_{l^2} = \sum_{j=1}^{N-1} f_j g_j \Delta x + \tfrac{1}{2}(f_0 g_0 + f_N g_N)\Delta x.$$

The first is the usual L^2 inner product; the latter two are discrete versions taken at the midpoints or at the knots. Also define the following L^2-like norms:

$$\|f\|_{L^2} = \sqrt{(f, f)_{L^2}}, \qquad |f|_{m^2} = \sqrt{(f, f)_{m^2}}, \qquad |f|_{l^2} = \sqrt{(f, f)_{l^2}}.$$

Next adopt the following notation for L^∞-like norms:

$$\|f\|_{L^\infty} = \max_{x\in[0,1]} |f(x)|, \qquad |f|_{m^\infty} = \max_{j\in\{0,\ldots,N-1\}} |f_{j+1/2}|, \qquad |f|_{l^\infty} = \max_{j\in\{0,\ldots,N\}} |f_j|.$$

Let M be a positive integer and let $\Delta t = 1/M$. Let

$$t^n = n\Delta t, \quad n = 0, 1, \ldots, M.$$

Similarly let

$$t^{n+\theta} = (n + \theta)\Delta t, \quad n = 0, 1, \ldots, M - 1,$$

for any $\theta \in [0, 1]$. For any function $u(t)$ use

$$u^n = u(t^n), \qquad u^{n+\theta} = u(t^{n+\theta}), \qquad u^{n,\theta} = \theta u^{n+1} + (1 - \theta)u^n.$$

For any function $f(x, t)$ let

$$|f|_{l^\infty(L^2)} = \max_{n\in\{0,\ldots,M\}} \|f(\cdot, t^n)\|_{L^2}.$$

In general define the composition of any pair of time and space norms in an analogous way.

For any function $u(x)$ define an x-difference by

$$\partial_x u_{j+1/2} = \frac{u_{j+1} - u_j}{\Delta x}$$

and for any function $u(t)$ and any $\theta \in (0, 1]$ use the time difference

$$\partial_t u^{n+\theta} = \frac{u^{n+1} - u^n}{\Delta t}.$$

Let

$$\text{Poly}^k = \{f(x): f \text{ is a polynomial in } x \text{ of degree at most } k\}.$$

Let

$$P_l^k = \{f(x): f|_{[x_j, x_{j+1}]} \in \text{Poly}^k \quad \text{and} \quad f \in C^l\}.$$

In particular P_0^1 is the class of continuous piecewise linear functions on the mesh defined by the x_j, while P_{-1}^0 is the class of piecewise constant functions discontinuous at the x_j.

3.1.1. Example 1: Standard Collocation. Consider the scalar, linear, first-order hyperbolic partial differential equation

$$u_t + au_x = f,$$

with $a(x, t) > a_0 > 0$ for all $x \in [0, 1]$ and $t \in [0, 1]$, with $u(x, 0) = u_0(x)$ and $u(0, t) = u_l(t)$ given. Standard collocation is a natural way to compute an approximate solution. One seeks a function $U(x, t)$ approximating $u(x, t)$ and defined as follows. At each t^n, $U(x, t^n) \in P_0^1$. Between time levels U is extended by linear interpolation in time, meaning $U(x, t^{n+\theta}) = \theta U(x, t^{n+1}) + (1 - \theta)U(x, t^n)$ for all x, n and $\theta \in [0, 1]$. One takes $U(x, 0)$ to be an approximation to $u_0(x)$, such as the interpolant. To compute U^{n+1} from U^n requires $N+1$ equations. One is given by the boundary condition $U^{n+1}(0) = u_l(t^{n+1})$. For the rest choose a fixed $\theta \in [\frac{1}{2}, 1]$ and require $U(x, t)$ to satisfy the differential equation at the N points $(x_{j+1/2}, t^{n+\theta})$, $j = 0, \ldots, N-1$. This yields the system

$$(U_t)_{j+1/2}^{n+\theta} + a_{j+1/2}^{n+\theta}(U_x)_{j+1/2}^{n+\theta} = f_{j+1/2}^{n+\theta}, \tag{2}$$

with $j = 0, \ldots, N-1$. Because U is piecewise linear in space and in time, this reduces to the system

$$\frac{U_{j+1/2}^{n+1} - U_{j+1/2}^{n}}{\Delta t} + a_{j+1/2}^{n+\theta} \frac{U_{j+1}^{n+\theta} - U_j^{n+\theta}}{\Delta x} = f_{j+1/2}^{n+\theta} \tag{3}$$

or, in the above notation for discrete derivatives,

$$\partial_t U_{j+1/2}^{n+\theta} + a_{j+1/2}^{n+\theta} \partial_x U_{j+1/2}^{n+\theta} = f_{j+1/2}^{n+\theta}. \tag{4}$$

The phrase 'with $j = 0, \ldots, N-1$' will henceforth be suppressed as implied by context.

Define the set of collocation points

$$\mathscr{CP} = \{(x_{j+1/2}, t^{n+\theta}) : j = 0, \ldots, N-1 \quad \text{and} \quad n = 0, \ldots, M-1\}.$$

In equations such as (4) in which all the subscripts are $j + \frac{1}{2}$ and all the superscripts are $n + \theta$, the subscripts and superscripts will henceforth be suppressed. To remind the reader of this convention, the phrase 'on $\mathscr{CP}$' will be appended to such equations. This convection will allow the use of subscripts for indexing component equations and variables in the case of systems of equations. With this convention the equations for standard collocation become

$$\partial_t U + a \partial_x U = f \quad \text{on } \mathscr{CP}. \tag{5}$$

3.1.2. Example 2: Upwinding. Suppose now that $a(x, t)$ in Example 1 is no longer bounded away from zero. For simplicity in this example, however, assume $a(x, t) \geqslant 0$ for all x and t. To avoid the instabilities which arise from the application of standard collocation in this case, one can use a technique known as 'upwinding'.

Again one seeks a function $U(x, t)$ approximating $u(x, t)$. Now however, at each t^n, $U(x, t^n) \in P_{-1}^0$. Between time levels U is again extended by linear interpolation in time, meaning $U(x, t^{n+\theta}) = \theta U(x, t^{n+1}) + (1 - \theta)U(x, t^n)$ for all x, n and $\theta \in (0, 1]$. Again one takes $U(x, 0)$ to be a suitable approximation to $u_0(x)$.

Requiring U to satisfy (2) no longer seems sensible, as the U_x term vanishes. Note that so far $U(x_j, t)$ is undefined for all j and t. To avoid losing information about the slope of U, one extends the definition of U by setting

$$U(x_j, t^n) = U(x_{j-1/2}, t^n),$$

for $j = 1, \ldots, N$, and

$$U(0, t^n) = u_l(t^n).$$

One continues to define U at intermediate times by linear interpolation in time. The spatial asymmetry in the definition of $U(x_j, t^n)$ is due to the assumed asymmetry in the sign of the coefficient $a(x, t)$. One

can interpret u as an entity being advected by a velocity a. U is defined by looking 'upwind' relative to this velocity, whence the name of the technique.

To compute U^{n+1} from U^n requires N equations. One chooses a fixed $\theta \in [\frac{1}{2}, 1]$ and requires $U(x, t)$ to *approximately* satisfy the differential equation at the N points $(x_{j+1/2}, t^{n+\theta}), j = 0, \ldots, N-1$. That is, rather than require (2), one instead requires (3). This equation is well defined because of the extended definition of U. As in Example 1, this equation can be written compactly as (4). Using the same convention of suppressing the subscripts and superscripts, collocation using upwinded piecewise constants can be written as

$$\partial_t U + a\partial_x U = f \quad \text{on } \mathscr{C}\mathscr{P},$$

just as in (5) for piecewise linears.

It turns out that upwinding amounts to adding extra numerical diffusion to the numerical method, which provides the stability missing from standard collocation. Had one 'downwinded' instead, the extra diffusion would have the wrong sign, making the method less stable rather than more.

Figure 1 illustrates the upwinding process in the case of velocity flowing to the right. The dots at (x_i, T_i) show the temperature value at each node. The full lines illustrate the piecewise constant representation of temperature between nodes. The broken line shows how a meaningful 'slope' can be defined for this discontinuous piecewise constant function, which is used as an approximation to the x-derivative. The broken line connects (x_i, T_i) with one or the other adjacent node value based on the local sign of the fluid velocity.

3.1.3. Component notation. In both the previous examples the convention of dropping spatial subscripts and temporal superscripts was employed. This convention will be continued throughout this paper to simplify the notation and to allow the use of subscripts denoting vector and matrix components.

In particular, if A is a matrix, one writes A_{ij} for the component of A in the ith row and jth column. The identity matrix is represented by the Kronecker delta symbol, defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

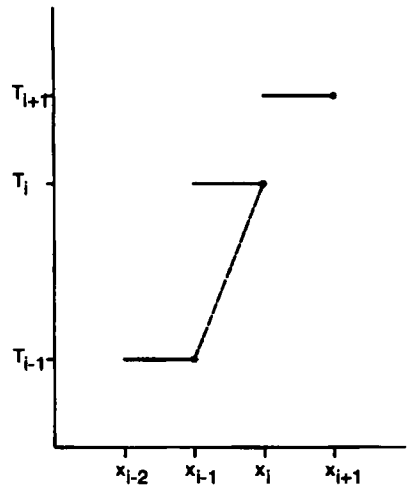

Figure 1. Upwinding illustration

The summation convention will be used throughout this paper; it implies summations over all repeated component indices. Thus if B is another matrix of compatible dimensions, one defines $A_{ij}B_{jk}$ by

$$A_{ij}B_{jk} = \sum_j A_{ij}B_{jk}.$$

3.2. Defining the new numerical method

Consider now the model problem (1) written in vector form

$$u_t + Au_x + Bu = 0, \tag{6}$$

with $u = (p, T)^T$. Here superscript T means transpose. The model problem was chosen to make the matrix A diagonal, which simplifies the analysis of the method. Notice that T_x only occurs in the T equation—in the general case the equations must be rewritten to make this happen.

The new numerical method combines standard collocation and upwinding as follows. One seeks a vector function $U(x, t)$ approximating $u(x, t)$. At each t^n, U_1 is to be in P_0^1, but $U_2 \in P_{-1}^0$. Moreover, the definition of U_2 is extended to the knots x_j by upwinding as in Example 2. In particular one defines

$$U_2(x_j, t^n) = U_2(x_{j-1/2}, t^n),$$

since $v_f \geqslant 0$. In the general case treated in References 2 and 3, the velocity is one of the unknowns and can change sign, which complicates the definition of upwinding.

When $x_{j-1/2}$ falls outside the domain [0, 1], one inserts the corresponding boundary condition instead. Note how the upwinding technique fits naturally with the specified boundary condition:

$$U_2(x_0, t^n) = T(0, t^n).$$

Each component U_k is defined to be piecewise linear in time between the t^n. $U(x, 0)$ is taken to be a suitable approximation to $u^0(x)$; it can be the interpolant. To incorporate the remaining boundary condition, one sets

$$U_1(x_0, t^n) = p(0, t^n).$$

The new numerical method determines U^{n+1} from U^n by requiring that U satisfy a certain linear system of equations at the collocation points $(x_{j+1/2}, t^{n+\theta})$, subject to the special interpretation of $U_{2,x}$ described in Example 2. Per the conventions previously described, this discrete linear system can be written as

$$\partial_t U + A\partial_x U + BU = 0, \tag{7}$$

where, as with all following discrete equations, the phrase 'on $\mathscr{C}\mathscr{P}$' is to be understood.

This discrete linear system may be solved very efficiently using a straightforward modification of the standard algorithm for tridiagonal matrices.

4. THEORETICAL RESULTS

Assumption 1

Assume $\theta \in (\frac{1}{2}, 1]$ is a given constant. Assume there is a constant K_0 independent of Δx and Δt such that

$$\frac{1}{K_0} \leqslant \frac{\Delta x}{\Delta t} \leqslant K_0$$

as both Δx and Δt go to zero. Moreover, assume the system (1) with given initial and boundary data has a unique solution which is smooth for all $t \in [0, 1]$.

Theorem 1

Consider the model problem (1). Let Assumption 1 hold. Define a discrete solution U by (7). Then there is a constant C which depends on K_0 and on Sobolev norms for u but remains bounded even when $v_f = 0$, and which is otherwise independent of Δx and Δt, such that for Δx and Δt sufficiently small,

$$\|U - u\|_{l^\infty(L^2)} \leqslant C\Delta x.$$

This theorem is representative of those presented in References 2 and 3.

As described in References 2 and 3, the results in this paper generalize to non-linear first-order hyperbolic systems of any size where exactly one eigenvalue is not bounded uniformly away from zero.

4.1. Computational results

The new numerical method for the fully non-linear thermal pipeline equations was implemented by the author in C++ and was informally compared (during 1990) against then-state-of-the-art commercial codes in widespread use. These codes used *ad hoc* methods to incorporate temperature effects, which generally require very small time steps to maintain stability. Such time step limitations are poorly understood, since convergence analyses do not exist for these methods. Owing to the proprietary nature of commercial pipeline codes, a detailed comparison cannot be presented. However, the new method does seem to be able to use much longer time steps than the comparison methods and the analysis does not require any limitation on the time step. For instance, there is no CFL constraint as would occur in an explicit method. This is important in networks of pipelines of different lengths, where the CFL time step for the system would be limited by the smallest natural time step in the network.

Numerical experiments were conducted for the fully non-linear thermal pipeline equations,[2,3] in which the advection matrix does depend on temperature and the velocity can change sign. Two illustrations of actual computed solutions follow.

For the first example consider a 150 km insulated pipe with a 75 cm internal diameter, carrying gaseous methane. Initially everything is at rest, with a pressure of 8000 kPa and a temperature of 20 °C throughout the pipe. The ends of the pipe are then opened and the outlet pressure is dropped to 5500 kPa over 1 min. Over the next 8–16 h the flow evolves to a steady state. The solution after 12 h was computed using 10 km space intervals and 10 min time steps. Figure 2 illustrates the resulting pressure, velocity and temperature along the pipe. To fit all three variables on one graph, temperature is shown in degrees celsius, velocity in metres per second and pressure in megapascals.

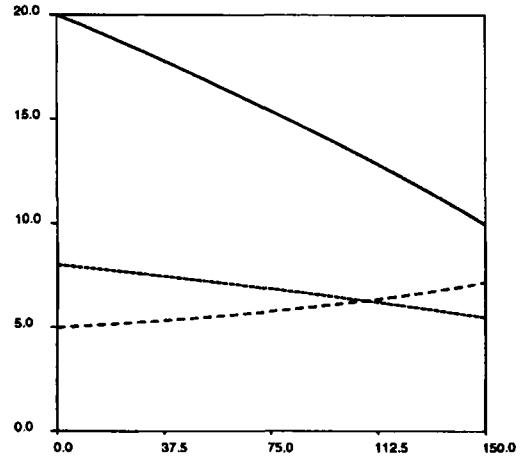

Figure 2. Steady state. Key: ——, T in deg. C; – – –, v in m/s; - - - - -, p in mPa

Table I. Approximate convergence rates

	Steady state		Small waves	
	L^2	L^∞	L^2	L^∞
Pressure	2·0	1·9	0·9	0·9
Velocity	1·1	1·2	0·9	0·9
Temperature	1·3	1·4	1·0	0·9

For the second example consider a 100 km insulated pipe with a 60 cm internal diameter, carrying liquid *n*-octane. Initially everything is at rest, with a pressure of 1400 kPa and a temperature of 20 °C throughout the pipe. Then a 10 s pulse of extra pressure is applied at $x = 0$. This creates a smooth travelling wave in pressure which propagates down the pipe at the sonic velocity of 1·6 km s^{-1}. The pressure wave has an amplitude equal to 10 per cent of the initial pressure, i.e. 140 kPa. As it travels, it excites identical-looking pulses in velocity and temperature. The solution during the first 45 s was computed using 1 km space intervals and $\frac{5}{8}$ s time steps. Figure 3 illustrates the resulting pressure wave at 15, 30 and 45 s. Notice the decay in the wave amplitude due both to friction and to numerical dissiption in the upwinding process. In both examples the friction factor was 0·014.

Table I indicates the convergence rates obtained for pressure, velocity and temperature in each of the two scenarios described above. In each case both the L^2 and L^∞ norms of the error were measured for each component, as compared with a reference solution computed on a much finer mesh. As Δx was decreased, Δt was decreased proportionately. For the steady state simulation the norm of the error 12 h after opening the valves was examined. For the small-amplitude wave case the norm of the error was examined after 15 s. In both cases $\theta = 0{\cdot}6$.

Figure 4 is a log–log plot of the errors at Δx decreases. It shows the L^∞ norm of the error in pressure in each of the two scenarios described above. In each case the base 10 logarithm of the error is plotted against the base 10 logarithm of the number of spatial intervals. These same sample points were used in constructing Table I. The pressure is in pascals here. The graphs for velocity and temperature look very similar and so are not shown here.

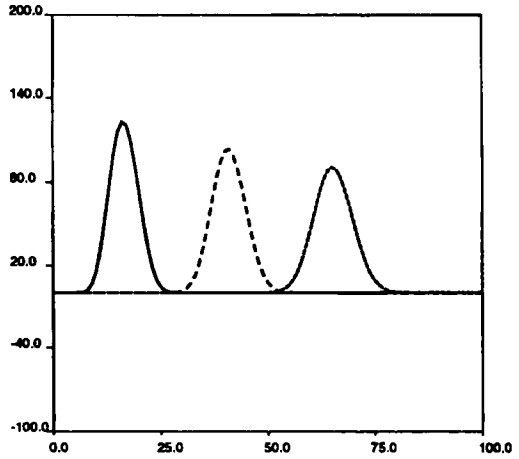

Figure 3. Small-amplitude wave. Key: pressure in kPa, $p0 = 1400$ kPa; ——, $p(15\text{ s}) - p0$; – – –, $p(30\text{ s}) - p0$; - - - - -, $p(45\text{ s}) - p0$

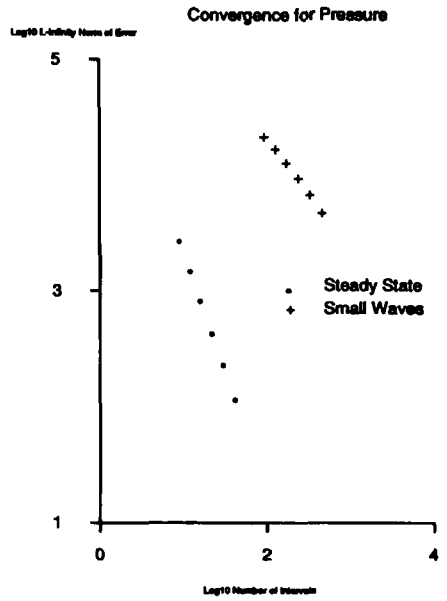

Figure 4. Convergence of error

The table and graph illustrate the empirical observation that even though the first-order nature of piecewise constants appears in the asymptotic convergence rates, in practice one may obtain close-to-second-order convergence rates. This is not too surprising, since the pressure and velocity approximations are second-order (for $\theta = \frac{1}{2}$) and isothermal pressure–velocity simulations give good results in many situations. Temperature effects generally occur on a much slower time scale than sonic effects, so one may expect the constant on the first-order error terms to be small relative to typical practical values of Δx. In fact, one sees that in the nearly steady state case, where the temperature varies only slowly, the convergence is indeed approximately second-order, at least for pressure, over the parameter range shown. Note that this range is more than sufficient for practical computations, since the errors involved are well below the error of measurement in a real pipeline. Note also that first-order effects dominate in the small-amplitude wave case, since here the temperature changes as sharply and rapidly as the pressure and velocity.

It is interesting to note the effect of temperature on the sonic speed. Using an isothermal model produces a substantial change in the sonic velocity. In the methane pipeline the sonic speed decreases from 415 m s^{-1} in the adiabatic case to 350 m s^{-1} in the isothermal case. In the octane pipeline it decreases from 1630 to 1312 m s^{-1}. This means that the pulses in Figure 3 would travel about 20 per cent slower if temperature effects were omitted, despite the fact that the overall temperature in that example is virtually constant.

5. ERROR ANALYSIS

5.1. Overview

The *a priori* error bound stated in the theorem is derived from an energy estimate based on applying the discrete scheme to $U - W$, where W is a discrete interpolant of u. The 'error equation' satisfied by $U - W$ is an inhomogeneous version of the discrete scheme itself, with truncation error terms on the right-hand side. In the general case the error equation is then diagonalized by changing variables,

following the earlier work of Thomeé[4] and Luskin,[1] though for the model problem it is already in diagonal form. Next an evolution inequality (28) is developed for certain norms of the error, using a discrete m^2 inner product of the diagonalized error equation with a certain test function. The test function is the sum of three terms representing the error, its time derivative and its space–time second derivative, each weighted with a carefully chosen power of Δx. The time derivative test function is the only one not used in Reference 1. In developing this evolution inequality, there are many terms to estimate; these are summarized in a tableau and bounded one by one. Finally the evolution inequality is used to derive the error bounds; in the linear case presented here, the usual Gronwall lemma suffices for this last step.

For a proof in the intermediate case of linear *variable* coefficient systems, see also Reference 5.

5.2. *The error equation*

Convention 1

In what follows, let C be a generic constant whose value in any particular equation depends on various Sobolev norms of u, on the constants in the model problem and on the constant K_0 of Assumption 1, but which is otherwise *independent* of the discretization parameters Δx, Δt and θ.

Throughout the proof assume that Assumption 1 holds. The number 2 occurs throughout owing to the special treatment of T in the model problem, but the proof generalizes to any size of system with exactly one degenerate eigenvalue.

Consider the model problem in vector form (6), with the numerical method given by (7). Recall that U is piecewise linear in time, U_1^n is piecewise linear in space and U_2^n is discontinuous piecewise constant, upwinded by v_f. It will now be useful to introduce a discrete interpolant W of u. Such a function is defined in the same discrete space as U. Therefore $U - W$ is also in the discrete space and thus is easier to analyse than $U - u$. Define W by $W_1^n(x_j) = u_1^n(x_j)$ and $W_2^n(x_{j+1/2}) = u_2^n(x_{j+1/2})$, with W_2^n at the knots upwinded by v_f, so $W_2^n(x_j) = W_2^n(x_{j-1/2})$, for $j = 1, \ldots, N$.

Define the total error

$$\Psi = u - U,$$

the discrete error

$$\zeta = W - U$$

and the approximation error

$$e = u - W.$$

Under reasonable conditions it is standard to show that e is small; since $\Psi = e + \zeta$, it will suffice to show that ζ is small. In particular, the interpolant W satisfies the equation

$$\partial_t W + A\partial_x W + BW = TE, \tag{8}$$

where the *local truncation error* TE is given by

$$TE = (\partial_t W - u_t) + A(\partial_x W - u_x) + B(W - u).$$

By standard calculations involving Taylor series expansions, one can show that for some C independent of Δx, Δt and θ,

$$|TE^{n+\theta}|_{m^2} \leqslant C[\Delta x + \Delta t^2 + (\theta - \tfrac{1}{2})\Delta t], \quad \text{for all } n. \tag{9}$$

Similarly one may show

$$|(TE_1^{n+1+\theta} - TE_1^{n+\theta})|_{m^2} \leqslant C(\Delta x \Delta t + \Delta t^2). \tag{10}$$

Note that this equation does not apply to the second component of the truncation error, which in general is one order less accurate owing to the use of piecewise constants.

Subtracting (8) and (7) shows that the discrete error ζ satisfies

$$\begin{matrix} \boxed{A} & \boxed{B} & \boxed{D} & \boxed{C} \\ \partial_t \zeta + A\partial_x \zeta + B\zeta = TE, \end{matrix} \tag{11}$$

where each of the four terms is labelled with a letter for future convenience.

Thus the discrete error ζ satisfies an inhomogeneous version of the same equation satisfied by U.

In the general case the next step in the proof is to diagonalize the matrix A by changing variables. This requires introducing extra notation. Fortunately, in the model problem A is already diagonal. The nonlinear version of (11) is further complicated by the appearance of the usual 'shower of terms' from differentiating A and B.

Let

$$P = \begin{Bmatrix} 1 & 0 \\ 0 & 0 \end{Bmatrix}.$$

Now define the test function to be used in the energy analysis:

$$\begin{matrix} \boxed{1} & \boxed{2} & \boxed{3} \\ \varphi = \zeta + \alpha \Delta x P \partial_t \zeta + \beta \Delta x P A \partial_t \partial_x \zeta, \end{matrix} \tag{12}$$

where α and β are two unspecified but non-negative parameters which will be determined later and which will be independent of Δx and Δt. In the general case an extra factor appears in some terms of φ to handle the boundary conditions, but in the constant coefficient model problem it is not needed.

5.3. *The 12 product terms*

Now form the vector inner product of both sides of (11) with φ, producing another equation involving 12 product terms which must be considered individually. The following chart summarizes the situation:

	A	B	C	D
1	L	L	R	R
2	L	R	R	R
3	L	L	R	R

The five terms marked with an 'L' are primarily 'helping' or 'left-hand-side terms'; the other seven are right-hand-side terms. The analysis of each product term is conducted as follows. The product equation holds at every point of $\mathscr{GP}$; for each time level $t^{n+\theta}$ multiply by Δx and sum over all $x_{j+1/2}$, thus forming the discrete spatial midpoint-based m^2 norm.

For each right-hand-side term an upper bound will be given for the sum over the $x_{j+1/2}$. For the terms marked L, a positive lower bound will be given instead. The bounds may not be obvious at first, but they follow in straightforward ways from the properties of the objects involved, in particular from knowing that ζ is piecewise linear in time and either piecewise linear or piecewise constant in space.

For non-constant coefficients, additional lower-order terms would appear in the following bounds, as in the non-linear case. However, the constant coefficient analysis does include the analysis of all the highest-order terms.

Later some right-hand-side terms will be hidden by direct subtraction and Gronwall's inequality will be used to handle the rest.

Define

$$\hat{\mathscr{P}}_x = \{x_{j+1/2} : j = 0, \ldots, N-1\}.$$

For any function $f(t)$ let

$$(f^{n+\theta})^+ = f^{n+1}$$

and

$$(f^{n+\theta})^- = f^n.$$

First consider the three main 'helping terms' A–1, A–2 and B–3.

5.3.1. A–1. The steps leading to the bound are spelled out in some detail for this first left-hand-side term:

$$\partial_t \zeta \cdot \zeta \geqslant \tfrac{1}{2}\partial_t(\zeta^2) + (\theta - \tfrac{1}{2})\Delta t(\partial_t \zeta)^2. \tag{13}$$

The above equation holds at each point of $\mathscr{C}\mathscr{P}$, hence for each t^n,

$$\sum_{\hat{\mathscr{P}}_x} \partial_t \zeta \cdot \zeta \Delta x \geqslant \tfrac{1}{2}\partial_t |\zeta|_{m^2}^2 + (\theta - \tfrac{1}{2})\Delta t |\partial_t \zeta|_{m^2}^2.$$

Here C is a generic constant independent of Δx, Δt; as always, it can depend on norms of u.

5.3.2. A–2

$$\sum_{\hat{\mathscr{P}}_x} \partial_t \zeta \cdot \alpha \Delta x P \partial_t \zeta \Delta x \geqslant \alpha \Delta x |P \partial_t \zeta|_{m^2}^2. \tag{14}$$

5.3.3. B–3

$$\sum_{\hat{\mathscr{P}}_x} \partial_x(A\zeta) \cdot \beta \Delta x P \partial_t \partial_x(A\zeta) \Delta x \geqslant \beta \Delta x [\tfrac{1}{2}\partial_t |P \partial_x(A\zeta)|_{m^2}^2 + (\theta - \tfrac{1}{2})\Delta t |P \partial_t \partial_x(A\zeta)|_{m^2}^2]. \tag{15}$$

Next upper bounds for right-hand-side terms are derived, beginning with the easiest ones.

5.3.4. D–1

$$\left| \sum_{\hat{\mathscr{P}}_x} B\zeta \cdot \zeta \Delta x \right| \leqslant C |\zeta|_{m^2}^2. \tag{16}$$

5.3.5. B–2

$$\left| \sum_{\hat{\mathscr{P}}_x} \partial_x(A\zeta) \cdot \alpha \Delta x P \partial_t \zeta \Delta x \right| \leqslant \frac{\alpha \Delta x}{64} |P \partial_t \zeta|_{m^2}^2 + \alpha \Delta x C |P \partial_x(A\zeta)|_{m^2}^2. \tag{17}$$

5.3.6. D–2

$$\left|\sum_{\hat{\mathscr{P}}_x} B\zeta \cdot \alpha\Delta x P\partial_t\zeta\Delta x\right| \leqslant \frac{\alpha\Delta x}{64}|P\partial_t\zeta|_{m^2}^2 + \alpha\Delta x C|\zeta|_{m^2}^2. \tag{18}$$

5.3.7. C–1

$$\sum_{\hat{\mathscr{P}}_x} TE \cdot \zeta\Delta x \leqslant C(|\zeta|_{m^2}^2 + |TE|_{m^2}^2). \tag{19}$$

5.3.8. C–2

$$\sum_{\hat{\mathscr{P}}_x} TE \cdot \alpha\Delta x P\partial_t\zeta\Delta x \leqslant \frac{\alpha\Delta x}{64}|P\partial_t\zeta|_{m^2}^2 + \alpha\Delta x C|TE|_{m^2}^2. \tag{20}$$

5.3.9. B–1. In the constant coefficient case this term contains only helping terms, so a lower bound is derived:

$$\sum_{\hat{\mathscr{P}}_x} \partial_x(A\zeta) \cdot \zeta\Delta x \geqslant \tfrac{1}{2}[(v_s l_1 \zeta_1^2) + (v_f \zeta_2^2)]|_{x=0}^{x=1}. \tag{21}$$

5.3.10. A–3. Note that ζ_2 does not appear in this term, so only piecewise linear functions need be considered:

$$\beta\Delta x \sum_{\hat{\mathscr{P}}_x} \partial_t\zeta \cdot P\partial_x(A\partial_t\zeta)\Delta x \geqslant \frac{\beta\Delta x}{2}[v_s l_1(\partial_t\zeta_1)^2]]|_{x=0}^{x=1}. \tag{22}$$

The spatial boundary terms in (21) and (22) turn out to give non-negative helping terms. In the general non-linear case this requires a slightly more general test function with a parameter to be chosen sufficiently small relative to certain $O(1)$ constants depending only on u. This is based on a trick used by Luskin[1] and pioneered by Thomeé.[4] In the present case, however, it is clear by inspection. The boundary term in (21) is

$$\tfrac{1}{2}[-v_s\zeta_1^2(0) - v_f\zeta_2^2(0) + v_s\zeta_1^2(1) + v_f\zeta_2^2(1)].$$

Now $\zeta(0) = 0$ by choice of boundary conditions and the remaining terms are non-negative because of the sign of the velocities. Similar arguments apply to the $\partial_t\zeta$ terms from (22).

5.3.11. C–3. Use the following formula for summation by parts in time at $t = t^{n+\theta}$, in which, for clarity, time superscripts are not suppressed:

$$a^{n+\theta}\partial_t b^{n+\theta} = \frac{1}{\Delta t}a^{n+\theta}(b^{n+1} - b^{n-1}) = \frac{1}{\Delta t}(a^{n+\theta}b^{n+1} - a^{n-1+\theta}b^{n-1}) - \frac{1}{\Delta t}(a^{n+\theta} - a^{n-1+\theta})b^{n+1}. \tag{23}$$

Thus

$$\begin{aligned}\sum_{\hat{\mathscr{P}}_x} TE \cdot \beta\Delta x P\partial_t\partial_x(A\zeta)\Delta x = {} & \frac{\beta\Delta x}{\Delta t}\sum_{\hat{\mathscr{P}}_x}[TE^{n+\theta} \cdot P\partial_x(A\zeta^{n+1}) - TE^{n-1+\theta} \cdot P\partial_x(A\zeta^{n-1})]\Delta x \\ & - \beta\frac{\Delta x}{\Delta t}\sum_{\hat{\mathscr{P}}_x}(TE^{n+\theta} - TE^{n+\theta-1}) \cdot P\partial_x(A\zeta^{n+1})\Delta x.\end{aligned} \tag{24}$$

The first term will telescope when summed over n. One can bound the second sum by

$$C\beta\Delta x|P\partial_x(A\zeta^+)|^2_{m^2} + \beta\frac{\Delta x}{\Delta t^2}|TE^{n+1+\theta} - TE^{n+\theta}|^2_{m^2}.$$

5.3.12. D–3. As in C–3, sum by parts:

$$\sum_{\hat{\mathscr{P}}_x} B\zeta \cdot \beta\Delta x P\partial_t\partial_x(A\zeta)\Delta x = \frac{\beta\Delta x}{\Delta t}\sum_{\hat{\mathscr{P}}_x}[B\zeta \cdot P\partial_x(A\zeta^+) - (B\zeta)^{n-1+\theta} \cdot P\partial_x(A\zeta^-)]\Delta x$$
$$- \beta\Delta x\sum_{\hat{\mathscr{P}}_x} BP(\zeta^{n+\theta} - \zeta^{n+\theta-1}) \cdot P\partial_x(A\zeta^+)\Delta x. \tag{25}$$

One can bound the second sum by

$$C\beta\Delta x|P\partial_x(A\zeta^+)|^2_{m^2} + \frac{\alpha\Delta x}{64}(|P\partial_t\zeta|^2_{m^2} + |P\partial_t\zeta^{n+\theta-1}|^2_{m^2}).$$

5.4. The evolution inequality

Now collect all 12 terms to form an evolution inequality. In the general non-linear case one must carefully formulate induction hypotheses in order to analyse this inequality. In the present case, however, the ordinary discrete Gronwall lemma will suffice.

In particular take α to be small based on some other order-one constants. Next take β small relative to α, again based on order-one constants, and finally require Δx and Δt to be sufficiently small with respect to these other constants.

A number of right-hand-side terms now can be directly subtracted off from left-hand-side terms. These are terms with a smaller multiplier on them, usually written as $\frac{1}{64}$ above. This results in the inequality

$$\partial_t|\zeta|^2_{m^2} + \Delta x|P\partial_t\zeta|^2_{m^2} + \Delta x\partial_t|AP\partial_x\zeta|^2_{m^2} + TS - \frac{\Delta x}{64}|P\partial_t\zeta^{n+\theta-1}|^2_{m^2}$$
$$\leqslant C(|\zeta^+|^2_{m^2} + |\zeta^-|^2_{m^2} + \Delta x|AP\partial_x\zeta^+|^2_{m^2} + \Delta x|AP\partial_x\zeta^-|^2_{m^2} + \Delta x^2), \tag{26}$$

where use was made of Assumption 1 and equations (9) and (10). Here TS stands for the telescoping terms in C–3 and D–3. Multiplying by Δt, summing on n and using the fact that the initial error ζ^0 is zero by construction, one obtains

$$|\zeta^N|^2_{m^2} + \Delta x|P\partial_t\zeta|^2_{m^2(m^2)} + \Delta x|AP\partial_x\zeta^N|^2_{m^2} \leqslant C(|\zeta|^2_{l^2(m^2)} + \Delta x|AP\partial_x\zeta|^2_{l^2(m^2)} + \Delta x^2). \tag{27}$$

Gronwall's lemma then implies

$$|\zeta|^2_{l^\infty(m^2)} + |AP\partial_x\zeta|^2_{l^\infty(m^2)} \leqslant C\Delta x^2. \tag{28}$$

Careful reading of the proof shows that one can prove a stronger theorem than claimed, in the constant coefficient case, but the intent here has been to present the theorem and proof of the non-linear case in a simpler context. See also Reference 5 for details of the theorems one can prove in the linear variable coefficient case.

6. CONCLUSIONS

Standard collocation methods work well for isothermal flow in pipelines, but the stagnating eigenvalue causes difficulties when thermal effects are included.

In the context of linear, constant coefficient systems we have presented and analysed a numerical method which extends standard collocation by using upwinded piecewise constant functions for the degenerate component of the solution. Both the error analysis and the numerical results indicate that the new method overcomes the difficulties that come from the application of standard collocation to problems with one degenerate eigenvalue. In particular the new method is not subject to any stability limitation on the time step, even when the degenerate eigenvalue is zero, whereas standard collocation develops erroneous oscillations in these situations. Moreover, informal comparisons (during 1990) against then-state-of-the-art commercial pipeline simulators in widespread use indicated that the new method did seem to be able to successfully use much longer time steps than the comparison methods.

In a companion paper[2,3] these results are extended to general non-linear systems of first-order hyperbolic equations with one degenerate eigenvalue, such as arise in the modelling of thermal effects in one- dimensional pipeline flow.

ACKNOWLEDGEMENTS

This research was supported in part by the Department of Energy, the State of Texas Governor's Energy Office and project grants from the National Science Foundation. The author was also supported in part by an NSF Postdoctoral Fellowship.

REFERENCES

1. M. Luskin, 'An approximation procedure for nonsymmetric, nonlinear hyperbolic systems with integral boundary conditions', *SIAM J. Numer. Anal.*, **16**, 145–164 (1979).
2. P. T. Keenan, 'Thermal simulation of pipeline flow', *SIAM J. Numer. Anal.*, **32**, 1225–1262 (1995).
3. P. T. Keenan, 'Thermal simulation of pipeline flow', *Ph.D. Thesis*, Department of Mathematics, University of Chicago, 1991.
4. V. Thomeé, 'A stable difference scheme for the mixed boundary problem for a hyperbolic, first order system in two dimensions', *J. Soc. Ind. Appl. Math.*, **10**, 229–245 (1962).
5. P. T. Keenan, 'An error estimate for a new scheme for the general variable coefficient linearized thermal pipeline equations', *Tech. Rep. 90–20*, Department of Computer Science, University of Chicago, 1990.
6. M. Luskin and T. Blake, 'The existence of a global weak solution to the nonlinear waterhammer problem', *Commun. Pure Appl. Math.*, **35**, 697–735 (1982).
7. E. B. Wylie and V. Streeter, *Fluid Transients*, McGraw Hill, New York, 1978.